Applications of Bayesian Skyline Plots and Approximate Bayesian Computation for Human Demography

Fernando A. Villanea, Andrew Kitchen, Brian M. Kemp

Human Biology, Volume 91, Number 4, Fall 2019, pp. 279-296 (Article)

Published by Wayne State University Press

REVIEW

Applications of Bayesian Skyline Plots and Approximate Bayesian Computation for Human Demography

Fernando A. Villanea,[1] Andrew Kitchen,[2] and Brian M. Kemp[3]*

ABSTRACT

Bayesian methods have been adopted by anthropologists for their utility in resolving complex questions about human history based on genetic data. The main advantages of Bayesian methods include simple model comparison, presenting results as a summary of probability distributions, and the explicit inclusion of prior information into analyses. In the field of anthropological genetics, for example, implementing Bayesian skyline plots and approximate Bayesian computation is becoming ubiquitous as means to analyze genetic data for the purpose of demographic or historic inference. Correspondingly, there is a critical need for better understanding of the underlying assumptions, proper applications, and limitations of these two methods by the larger anthropological community. Here we review Bayesian skyline plots and approximate Bayesian computation as applied to human demography and provide examples of the application of these methods to anthropological research questions. We also review the two core components of Bayesian demographic analysis: the coalescent and Bayesian inference. Our goal is to describe their basic mechanics in an attempt to demystify them.

Understanding the demographic history of populations is one of the central research interests of anthropological geneticists. Some of the most exciting anthropological genetic studies have demographic history questions at their core, such as understanding the structure of modern human populations (e.g., Ramachandran et al. 2005; Wang et al. 2007), the routes taken by our human ancestors as they migrated across the planet (e.g., Macaulay et al. 2005), major population migrations such as the Bantu expansion (e.g., Berniell-Lee et al. 2009), migrations associated with the spread of Indo-European languages (e.g., Ammerman and Cavalli-Sforza 1984), and understanding the origins of modern populations, such as Native Americans (e.g., Bonatto and Salzano 1997a, 1997b; Tamm et al. 2007). The importance of demography to anthropology has driven the implementation of dedicated analytical methods, including Bayesian skyline plots (BSPs) and approximate Bayesian computation (ABC).

BSP and ABC methods are built around a simple mathematical model: the coalescent (or *n*-coalescent; Kingman 1982a). The coalescent explores the history of a population by creating a gene genealogy (or genealogical tree) representing the relationship between individuals, using methods similar to the study of phylogenetics (Wakeley

[1]Center for Computational Molecular Biology, Brown University, Providence, Rhode Island, USA.
[2]Department of Anthropology, The University of Iowa, Iowa City, Iowa, USA.
[3]Department of Anthropology, University of Oklahoma, Norman, Oklahoma, USA.

*Correspondence to: Fernando A. Villanea, Brown University, 164 Angell Street, Providence, RI 02914 USA. E-mail: fervillanea@gmail.com.

KEY WORDS: BAYESIAN INFERENCE, DEMOGRAPHY, POPULATION GENETICS, ANTHROPOLOGICAL GENETICS.

Human Biology, Fall 2019, v. 91, no. 4, pp. 279–296. doi: 10.13110/humanbiology.91.4.04.

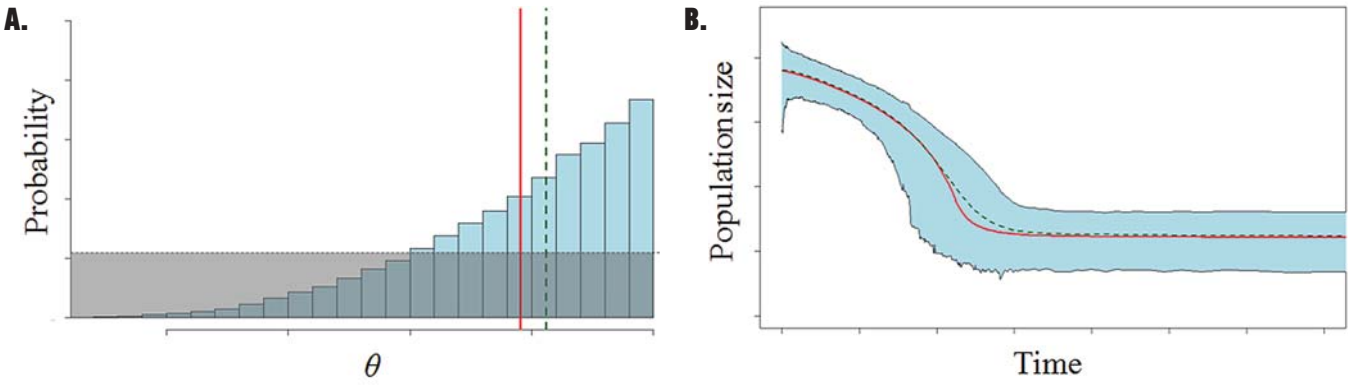

FIGURE 1. A, Graphic representation of prior and posterior distributions for a Bayesian analysis. The area under the bars represents the posterior probability density function of parameter *θ*; the shaded area represents the prior probability density function of parameter *θ*. The prior density in this case is flat or uniform (horizontal line) across the entire interval. Point estimators such as the mean (solid line) or median (dashed line) are used to generate a posterior point value for *θ*. B, Graphic representation of a BSP. The posterior density function for the value of a parameter, in this case, effective population size, is represented over time. The shaded area represents the 95% credibility interval; the lines represent the posterior point estimators mean (solid line) and median (dashed line).

2009). The history of a population can be represented by a single genealogical tree, but different hypotheses of how a population has evolved are represented as a set of many possible trees. The goal of Bayesian demographic analyses is to determine the adequacy of each genealogical tree to represent the true biological history of a population, as best supported by the empirical data.

To appreciate how competing models of evolution are discriminated in any demographic study, it is important to understand how probability is assigned to a genealogical tree, the probability that one tree represents true biological history. This probability is referred to as the *likelihood* or *likelihood function* and can be calculated through various statistical approaches (Box 1). Regardless of the selected statistical approach, likelihood calculations are difficult and become computationally prohibitive as evolutionary histories depend on more and more unknown parameters (Bertorelle et al. 2010). The second theoretical component of Bayesian demographic analyses is thus the statistics approach behind it, known as *Bayesian inference*.

Bayesian inference expands on the likelihood framework by adopting the use of prior probabilities through Bayes's theorem (Box 2). Bayesian inference allows for user-modified priors, which simplify the estimation of probability calculations, at least enough to be viable under current computational resources. In addition, the Bayesian philosophy is not concerned with identifying the single best demographic model but instead summarizes the adequacy of competing models as a distribution of probabilities (Konigsberg and Frankenberg 2013), which is a more realistic approach to modeling natural processes, as often our models are only approximations of reality (Figure 1A). This summary distribution has the added benefit of functioning as an intuitive way to incorporate uncertainty into the presentation of results. The popularity of the Bayesian approach has resulted in the development of software capable of powerful yet simple analyses (Table 1). Most of these software packages are distributed freely but nevertheless require specialized expertise to use effectively.

The goal of this review is to familiarize the anthropological geneticist audience with two commonly used methods in Bayesian demographic analysis: (1) Bayesian skyline plots (BSPs), a popular demographic model of variable population size trajectories over time, and (2) approximate Bayesian computation (ABC), a promising method for exploring diverse evolutionary scenarios employing simulated data. Because the theoretical foundations of both the neutral coalescent and Bayesian inference are critical to meet these objectives, they are reviewed here as well. These methods have become increasingly prevalent in anthropological literature; therefore, a working understanding of their theoretical fundamentals is of value for both maintaining the transparency of and disseminating results. We conclude by presenting some case studies of how these methods can be used to address anthropological questions about demographic changes in the human past.

BOX 1. Frequentist and Bayesian Statistics

Both frequentist and Bayesian statistics share a large body of underlying theory, developed concurrently, to solve similar problems. Likewise, both schools of statistics have converged on similar terminology and compatible representations of results, as they easily supplement each other (Puga et al. 2015b). However, some of their underlying elements remain distinct, and these features are particularly important to understand to apply both approaches effectively.

To formalize the distinction mathematically let there be a population of data N, from which we have the sample n, which is but a portion of the entire population. For any variable of interest in the data, the frequency histogram of all possible values is called the *population distribution*. The population N possesses immutable characteristics, or parameters, such as a mean μ (Krzywinski and Altman 2013b). If $n = N$, then the population mean μ is known; otherwise, it must be described in terms of probability. This is where the differences in philosophy commence.

Frequentist Probability

Frequentist probability, also known as *physical* or *objective probability*, is associated with repeatable processes that occur at a given rate (i.e., occurring at some frequency during a long set of trials). The probability of an outcome is then estimated as a measure of the relative frequency of the occurrence of that outcome from a lengthy number of trials. For example, the probability of rolling a 20 on a 20-sided die [$P(20)$] is approximately given by the frequency of times a 20 is rolled (n_{20}) over the total number of trials (n_t). At this point, it should be noted that relative frequency is a poor approximation of the "true" frequency when the number of trials is low, but as the number of trials approaches infinity the relative frequency becomes exactly the true frequency (i.e., the law of large numbers). Similarly, when frequentist statistics are used to estimate an aspect of the total population of data N, for example, the mean, μ, using a sample n, the mean of the sample, $\bar{x}$, is a bad approximation for μ when the sample size is small and becomes better as the sample size increases. Finally, when $n = N$, then $\bar{x} = \mu$. That is, when all samples have been observed in the population, the mean is known.

Interval estimation is commonly used to calculate unknown parameters of the population N, as an alternative to providing a single estimator value. A *confidence interval* (CI) is a range of values that should contain the true value of the parameter for a given relative frequency, or *confidence level* (this is different from a *credibility interval*, as explained below). For example, given a confidence level of 95%, we are 95% confident that the CI contains the true value of the parameter. This means that, if we were to repeat an experiment multiple times and construct the corresponding CIs, we would expect 95% of those CIs to contain the true value of μ (Krzywinski and Altman 2013a).

The central ideas of frequentist probability are commonly applied in hypothesis testing in the form of the familiar p-value. To connect the frequentist philosophy with hypothesis testing, it is important to formalize the null hypothesis in terms of a distribution of expected observations. To generate a null hypothesis or null distribution, we need a control or reference, and we have to assume that all the random fluctuations inherent in measuring that control or reference can be characterized. If this is possible, we can construct the null distribution, which has a mean μ corresponding to the value of the reference, and variance determined by the inherent random fluctuations (Krzywinski and Altman 2013c). The purpose of a statistical test is to determine how a new observation compares to this distribution and, in particular, to determine how far removed it is from the mean μ. The significance of the difference between the observation and μ is determined by first calculating the proportion of the null distribution that is equal to or more extreme than the observation (this is the p-value) and then comparing that to a proportion of most extreme values that are a priori defined as outliers (this is the α-value). The value of α is used to calculate maximum and minimum threshold values beyond which the observation is considered significantly different from μ.

A p-value is the probability of observing a value equal to or more extreme than our cutoff value α, assuming that the null hypothesis is true (Krzywinski and Altman 2013c). Thus, when a p-value is found to be less than a standard α of 0.05 (i.e., $p < 0.05$), it is an observation that falls in the most extreme 5% of all observations relative to the mean μ. Now we can connect the concepts of hypothesis testing and CIs, as the confidence level is the complement to its significance level, or α. For example, a 95% confidence interval represents all estimates of μ that would not be considered significant at the $\alpha \leq 0.05$ level.

Bayesian Probability

Bayesian probability, in contrast, is conditional probability. A conditional probability measures the chance of an outcome given another outcome, as explained by Bayes's theorem (see Box 2). Returning to the example of population N and its mean μ, under Bayesian theory both the sample n and its mean $\bar{x}$ are treated more fluidly and described probabilistically. A very important distinction is that Bayesian statistics considers the sample n a *realized* sample; it treats the sample as the only source of data, D, while frequentist statistics is concerned with repeated sampling.

Another key distinction is that frequentist statistics treats the true mean μ of population N as fixed but also unknown and approximates its value through the mean $\bar{x}$ (except in the rare instances when $n = N$). In Bayesian statistics, it is possible to estimate the true mean μ by associating a conditional probability to it, that is, μ given the data, or $P(\mu|D)$. This is referred to as the *posterior probability* of parameter μ given some known data D. The likelihood of a parameter is proportional to the probability density function, which can be used to determine the probability of μ having a value in any given interval, called a *credibility interval*. A credible interval is the region with the highest posterior density. The credibility interval represents a range of values that contain the true parameter of the population for a given probability level. For example, a 95% highest-posterior-density region contains 95% of the area under the posterior density curve, representing all posterior densities that are equal to or greater than a given value. Because the posterior density has been "normalized" so that it integrates to 1.0, the 95% highest posterior density is 0.95. A 95% credibility interval thus consists of values that include the true value of μ with 95% probability (Casella 2008).

Because in Bayesian statistics μ can take a range of values, each with an associated conditional probability, it is typical to represent this distribution with a summary statistic and credibility intervals. The mean value of the probability distribution is typically reported for symmetrical probability distributions, whereas in cases with an asymmetric distribution the median or modal (i.e., highest probability) value are often more characteristic of the distribution and should be used.

BOX 2. Bayesian Inference

Bayesian inference is a statistical method based on Bayes's theorem, in which the probability of a hypothesis is updated based on prior evidence and a is model created to explain the data (Konigsberg and Frankenberg 2013). In Bayesian inference, probability is treated as *conditional probability*, the probability of an outcome given another outcome (Casella 2008; Puga et al. 2015b). At the core of Bayesian inference is Bayes's theorem (Puga et al. 2015a), in which the probability of a model M given the data D is described by $P(M|D)$, calculated as follows:

$$P(M\,|\,D) = \frac{P(M\,|\,D)\cdot P(M)}{P(D)}$$

Here, $P(D|M)$ is referred to as the *likelihood*, and it describes the compatibility of the data, given a model (specifically, it is the probability of the model M producing the data D). $P(M)$ is the probability of the model M before the data D are observed, also known as the *prior probability* or simply a *prior*. A prior represents the degree of belief in the values that a parameter can take, and it modifies the likelihood to produce the probability of a model given the data $P(M|D)$. The $P(M|D)$ is referred to as the *posterior probability*. Finally, $P(D)$ represents the probability of the data. For discrete cases, it is the sum of $P(D|M)\cdot P(M)$ across the different models. For continuous cases, it is the integral of the product across M. Critically, when posterior probabilities are calculated using the same data, $P(D)$ takes the same value in all independent calculations (because the empirical data are the same for all); it is therefore a fixed scalar of $P(M|D)$ and is often ignored:

$$P(M\,|\,D) \propto P(D\,|\,M)\cdot P(M)$$

Model testing in Bayesian frameworks is relatively straightforward and is usually performed using Bayes factors. A *Bayes factor* is the ratio of the prior odds of two hypotheses (i.e., the odds of model M_1 over model M_2) to the posterior odds of the hypotheses (Kass and Raftery 1995). The Bayes factor K is thus the ratio of the two marginal likelihoods of the models integrated across all model parameters:

$$K = \frac{P(D\,|\,M_1)}{P(D\,|\,M_2)}$$

Conveniently, Kass and Raftery (1995) provide a scale to discriminate between models based on the value of the ratio K (borrowed from Jeffreys 1998). Notably, though this scale has assumed some authority in the field of Bayesian inference, it is itself a suggestion when interpreting the importance of Bayes factor values (as are schemes regarding the significance of p-values). It is also important to remember when calculating Bayes factors that most coalescent or phylogenetic software packages report probabilities and likelihoods in $\log_e$ units, because likelihoods of phylogenies and genealogies can be exceedingly small.

Table 1. Popular Software for Bayesian Analysis Applied to Population Demography

Program	Main Function	Reference	Website
abc for R	Generates simulated population sequence data and provides analysis for ABC	Csilléry et al. 2012	https://cran.r-project.org/web/packages/abc/index.html
BaySICS	Performs ABC analyses by means of coalescent simulations	Sandoval-Castellanos et al. 2014	https://omictools.com/baysics-tool
BEAST	Bayesian estimation of population parameters by tree sampling	Drummond and Rambaut 2007	http://beast.bio.ed.ac.uk/
BEAST 2	Bayesian estimation of population parameters by tree sampling	Bouckaert et al. 2019	https://www.beast2.org/
phylodyn (for R)	Bayesian estimation of population parameters by tree sampling	Palacios et al. 2019	https://github.com/JuliaPalacios/phylodyn
DIYABC	Generates simulated population sequence data and provides analysis for ABC	Cornuet et al. 2014	http://www1.montpellier.inra.fr/CBGP/diyabc/
IM, IMA, IMA2	Migration between populations by Bayesian allele assignment	Hey and Nielsen 2007	https://bio.cst.temple.edu/~hey/software
LAMARC	Bayesian estimation of population parameters	Kuhner 2006	http://evolution.genetics.washington.edu/lamarc/index.html
MIGRATE, MIGRATE-N	Migration between populations (Bayesian and ML-based)	Beerli 2006	http://popgen.sc.fsu.edu/Migrate/Download.html
MrBayes	Generates phylogenetic trees for multiple species	Ronquist and Huelsenbeck 2003	http://mrbayes.sourceforge.net/
ms, msprime	Generates simulated coalescent trees	Kelleher et al. 2016	https://msprime.readthedocs.io/en/stable/#
Serial SimCoal	Generates simulated heterochronous population sequence data for ABC	Anderson et al. 2005	http://web.stanford.edu/group/hadlylab/ssc/index.html
SIMCOAL, fastsimcoal	Generates simulated population sequence data for ABC	Excoffier et al. 2000, 2013; Excoffier and Foll 2011	http://cmpg.unibe.ch/software/simcoal/, http://cmpg.unibe.ch/software/fastsimcoal2/
SLiM	Forward simulation of populations	Haller and Messer 2019	https://messerlab.org/slim/
Structure	Population structuring by Bayesian allele assignment	Pritchard et al. 2000; Falush et al. 2003, 2007; Hubisz et al. 2009	http://web.stanford.edu/group/pritchardlab/structure.html

ABC, approximate Bayesian computation; ML, maximum likelihood.

The Coalescent

In 1982, John Kingman derived a mathematical model describing the process of lineages merging—or coalescing—back in time, eventually reaching a common ancestor (Kingman 1982a, 1982b), which has since been called the "Kingman coalescent," or "*n*-coalescent" (Box 3). The coalescent can be understood as a representation of the historical relationship between related individuals, conceptualized as a genealogical tree, that models the effects of genetic drift looking backward in time (Rice 2004). From a biological perspective, the coalescent is concerned with tracing copies of genetic elements back in time to a single ancestral copy, the most recent common ancestor (MRCA) of that stretch of DNA. The coalescent is a powerful approach: it allows inference of unknown biological processes from a small sample of individuals, as long as they share a common history. Because understanding how the coalescent generates demographic information is paramount to conceptualizing how to relate its conclusions to biological questions, we summarize classic coalescent theory here (for in-depth summaries, see Rice 2004; Wakeley 2009).

The coalescent parallels the Wright-Fisher population model, which provides an accessible mathematical foundation (Hudson 1983; Tajima 1983). Under the Wright-Fisher model, a population of size N is assumed to be finite and constant over time. As a consequence of random, neutral fluctuations in reproductive success, some lineages will be lost. In the Wright-Fisher model this is known as genetic drift, but from the perspective of lineages merging back in time, this process produces the coalescent. Importantly, we need only consider the direct ancestors of extant individuals in a population when considering the coalescent process going backward in time. This makes coalescent analysis computationally very efficient relative to forward-in-time methods, which must track a greater number of individuals. This is one reason coalescent theory is a major theoretical framework for modern population genetics. Moreover, it is flexible in accommodating various biological models, providing the ability to explore diverse evolutionary processes under a single theoretical umbrella (Wakeley 2009; see Box 3).

The basic model for the coalescent is a sample of two alleles whose lineages coalesce back in time into their MRCA. Of primary interest in this process is estimating how many generations back in time are required until the lineages coalesce. This time is determined by a coalescent rate, which is inversely proportional to the size of the population, N, from which the alleles were drawn. For two alleles drawn from a population of diploid organisms, the probability of coalescing in the previous generation is

$$P_C = 1/2N,$$

and the complementary probability of not coalescing is

$$P_{NC} = 1 - \frac{1}{2N}.$$

(The rates for haploid organisms or genomes, as for example with mtDNA, are $1/N$ and $1 - 1/N$). To estimate $P_{C,t}$, the probability of coalescing at time t, one calculates the probability of not coalescing at all times in previous generations, $t - 1$, and then multiplies that by the probability of coalescing in the next generation, generation t:

$$P_{C,t} = \left(1 - \frac{1}{2N}\right)^{t-1} \cdot \frac{1}{2N}$$

(Rice 2004). This equation forms the basis of coalescent theory and defines how genealogies and waiting times between coalescent events provide information about population sizes. (For an expanded explanation of coalescent theory, see Box 3.)

The basic coalescent is derived from an idealized population, such as a Wright-Fisher population, with several important properties. These include no intralocus recombination, no selection, a single population (i.e., no population substructure or migration), and a constant population size. Genealogies that deviate from expectations indicate populations that do not conform to the assumptions of a Wright-Fisher ideal population. Coalescent theory allows those deviations to be quantified. Importantly, the basic coalescent model can be extended by relaxing the assumptions of the Wright-Fisher model. Coalescent models now exist that handle fluctuations in population size, complex population structures (including migration and metapopulation models), recombination, and natural selection. Some applications of these modifications to the basic coalescent include estimates of inbreeding between modern

BOX 3. The *n*-Coalescent

In the coalescent, copies of genetic elements are traced back in time to form a genealogy of the elements that describes their ancestor and descendent relationships. In this genealogy, each time two elements have a common ancestor, their lineages join to form one ancestral lineage. This event is called a *coalescent event*, in which two descendent lineages coalesce into one ancestral lineage in a process moving from the present into the past. The patterning of coalescent events in time provides information about the past dynamics of a population, such as fluctuations in population size or migration.

Under the Kingman *n*-coalescent model (Kingman 1982a, 1982b), a sample of size *n* is taken from a population *N* and the genealogy of these *n* individuals is traced as they coalesce back in time, ultimately reaching their most recent common ancestor (MRCA; see Figure 3.1) This process involves $n - 1$ coalescent events (i.e., events connecting two descendent lineages) going backward in time. Any two lineages may coalesce, and they do so at a per generation rate, one inversely proportional to the size of the population, $1/2N$ for diploids ($1/N$ for haploids), adjusted by the possible pairs of lineages [$k(k - 1)/2$ for k lineages]. Intuitively, large populations contain many distantly related individuals and thus have low rates of coalescence, while small populations contain closely related individuals and thus have high rates of coalescence.

The coalescent events define a branching tree of relationships between the *n* individuals called a *genealogy* or *coalescent tree*. Associated with the genealogy are $n - 1$ time intervals T_i between coalescent events. These intervals represent the waiting time for subsequent coalescent events, and their duration varies inversely with the rate of coalescence (i.e., the higher the rate, the shorter the waiting time). Collectively, these intervals sum to the time of most recent common ancestry, T_{MRCA}, which is how long in the past the MRCA of the sample existed. Figure 3.1 shows a representative genealogy for a sample size of $n = 6$, with waiting times indicated.

Kingman (1982a, 1982b) demonstrated that as *N* tends to very large values, the coalescent intervals T_i are independent and exponentially distributed, and coalescence can be modeled as a Poisson process with rate $k(k - 1)/4N$ for diploid genes. This indicates that the first coalescent event before the present should occur relatively recently and that the last two lineages should take the longest time to coalesce. Because the T_i intervals are independent and exponentially distributed (though not identically distributed), calculating T_{MRCA}, one of the most interesting parameters of a population, is straightforward and relatively easy. A classic result of coalescent theory is that T_2 (the interval directly prior to the MRCA, during which there are only two lineages) is on average $2N$ generations for diploid populations ($1N$ generations for haploids). The expected T_{MRCA} for a large sample of diploids is $4N$ generations ($2N$ for haploids). Notice that the time scale referred to here is in discrete generations (although coalescent time itself is continuous) and that time moves backward into the past. This is because the coalescent is a continuous approximation of a discrete descendent-ancestor process that occurs by tracing ancestry from the individuals in the sample to the generation of their common ancestor.

Another key feature of the coalescent is that every possible coalescent tree is equally probable, as any two individuals have an equal probability of coalescing at every event. That every tree is a possible genealogy poses a challenge for relatively large data sets. For example, given 10 individuals, there are more than 2 million possible unrooted trees; this number rises to more than 2.5 billion when considering rooted trees with ordered coalescent events (i.e., coalescent trees) (Edwards 1970; Wakeley 2009). Therefore, the analysis of realistically sized data sets requires a way to focus on the genealogies that have the highest likelihoods (i.e., contribute the most information about the demographic process). Genetic data may be used to identify these most probable genealogies. It is here that techniques borrowed from phylogenetic analysis can be employed to determine the set of genealogies most consistent with the data [and thus with the largest $P(D|M)$, or likelihood]. The genealogies with the greatest likelihood are those that contribute the most to inferences about the underlying demographic process; thus, focusing on these genealogies is one way to sort through the huge number of possible genealogies.

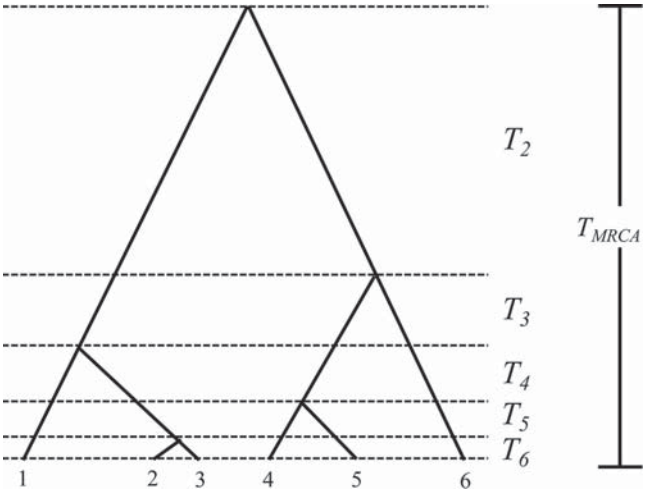

FIGURE 3.1. Graphical representation of the coalescent process for $n = 6$ individuals. Time is considered to move backward as lineages coalesce, commencing at the present time (bottom). Each interval of T_n represents the time in generations between coalescent events, adding to T_{MRCA}.

Several key assumptions must be made when considering the coalescent, because they directly affect the shape of genealogies by distorting them from their expected distributions. This is critical, as the coalescent is powerful as a method for inferring aspects of a population from the shape of the genealogical process, so violations of these assumptions will bias inferences of the demographic history of a population. Some of these assumptions include the neutrality of genetic elements under study, lack of population subdivision, no migration, constant population size over time, and the absence of recombination within the genetic locus under study (Wakeley 2009). Extensions of the basic coalescent can include these features in the coalescent framework. The ability of the basic coalescent model to be extended to allow for a variety of demographic and evolutionary scenarios is a strength of coalescent theory.

human and Neanderthal populations (Serre et al. 2004), migration rates between Asian and Native American populations (Ray et al. 2010), and natural selection in modern human populations with complex evolutionary histories (Akey et al. 2004). To summarize, the key insight from coalescent theory is that all biological processes affecting a population can be studied under one theoretical umbrella, with a single standardized unit: waiting times between coalescent events.

Bayesian Inference

The coalescent does not necessitate the use of Bayesian inference; however, the structure and logic of the Bayesian philosophy comport well with coalescent analysis (see Box 2). For example, population parameters can be estimated using either maximum likelihood (ML), another commonly used statistical approach, or Bayesian inference. As a toy example, for the parameter θ (the population mutation rate, equal to four times the effective population size, N_e, multiplied by the mutation rate, $4N_e\mu$) an ML estimate would be determined by finding the value of θ that maximizes the probability of the data [i.e., the likelihood function, $P(D|\theta)$]. This single value of θ would comprise the ML estimate. A Bayesian estimate of θ also involves the likelihood function $P(D|\theta)$, but this probability distribution is modified by a distribution of expected values of θ [i.e., the prior distribution, $P(\theta)$]. The two composite probability distributions, when multiplied and "normalized" so that the product integrates to 1.0, produce the probability distribution of θ from the data (i.e., the posterior probability, $P(\theta|D)$; see Box 2). The goal of Bayesian inference is consideration of the complete posterior probability distribution of θ, which is quite unlike the goal of ML of finding the point estimation for the best-fitted θ. The Bayesian method involves updating the model [the prior probability of θ, $P(\theta)$] after observing some data [the likelihood function, $P(D|\theta)$] and producing a continuous distribution of θ values with their associated probability density. As an added benefit, providing a distribution of probabilities explicitly defines uncertainty around any single point estimate. Konigsberg and Frankenberg (2013) provide an excellent review of Bayesian inference applied generally to anthropology outside of anthropological genetics.

Three aspects of Bayesian inference prove advantageous. First, the ability to select a prior distribution reflects belief that some values of the parameter of interest are more probable than others, which is reasonable given the progressive nature of science. Thus, a posterior distribution from a previous study may serve as the prior for a new study. Second, the posterior distribution is used to test the reasonableness of the prior distribution, which allows for model improvement by recursively adjusting the priors. Third, model comparison is simple in Bayesian inference: different priors can be compared and evaluated using the Bayes factor test, which provides the ratio of the probability of two models having generated the observed data, thus allowing for the selection of the best-fitting model (Box 2). As an additional benefit over ML, Bayesian model comparisons can occur between two wholly unrelated (nonnested) models (Kass and Raftery 1995).

Managing results as density functions in Bayesian inference is of particular utility for coalescent analysis. It is a reasonable assumption that for any population of individuals there will be a multitude of probable genealogical arrangements and patterns of ancestry. The use of a distribution of possible genealogies accounts for uncertainty by integrating across all probable parameter values. In a Bayesian coalescent analysis, the likelihood function is the likelihood of the genealogy given the genetic data. Prior probability distributions are placed on all model parameters, such as mutation rates, amino acid substitutions (e.g., nucleotide base frequencies and transition-transversion ratios), and population sizes. The posterior probability is a collection of the likelihoods of all genealogies modified by the product of all the priors.

Demographic Inference (Bayesian Skyline Plots)

One aspect of demography that has been of particular interest is ascertaining historical changes in population size. The coalescent, as originally conceived, assumes a population does not change in size. We know that for many populations and species this is not true, and could not be. When a

BOX 4. Markov Chain Monte Carlo Sampling

Markov chain Monte Carlo (MCMC) is a sampling algorithm that has enabled the use of modern Bayesian demographic methods by allowing for an approximation of the likelihood equations, which are more practical to estimate computationally. The typical MCMC algorithm commences by sampling a genealogy, testing its fit to the data, and then proposing a new genealogy by making a random, but small, change in the genealogical tree topology. The chain accepts this new genealogy given a probability calculated from the ratio of Bayesian posteriors between the new genealogy $P(\bar{x})$ and old genealogy $P(x)$. If the ratio is higher [if $P(\bar{x})/P(x) \geq 1$], the "chain" adopts the new genealogy and abandons the previous one. Even if the new posterior probability is lower [if $P(\bar{x})/P(x) \leq 1$], the chain may accept the new genealogy with probability proportional to the ratio (≤1). This is critical to ensure that the posterior distribution is composed of a mix of diverse genealogies, proportional to their likelihood, in true Bayesian form. If, for example, only $P(\bar{x})/P(x) \geq 1$ steps are considered, such that only higher-probability genealogies are collected (as an optimization algorithm might do), the posterior distribution will be incomplete, because it will not sample lower-probability genealogies.

MCMC samples step by step, starting at a random genealogy and generally moving toward better-fitting ones. This is far from a perfect process for two main reasons. First, an individual chain may get "stuck," moving between closely related genealogies, each not considered truly independent for the purposes of parameter estimation. For this reason, MCMC analysis records only a fraction of the total visited trees; for example, only every 1,000th tree may contribute to the posterior distribution, assuming that the interim will permit the chain to explore unrelated (but not truly independent) genealogies. Second, each chain starts from a random point, which may not be a very likely tree. For this reason, MCMC analysis often ignores the first few thousand steps (called burn-in), only including later sampled genealogies as part of the distribution, assuming that the chain will move to a likelier group of trees after this interval. To alleviate these limitations, a typical analysis will run thousands of chains, each contributing to the sampling of genealogies. The final MCMC sampling is distributed closely to the true posterior distribution (one composed of all possible genealogies), allowing for the estimation of the posterior density or posterior probabilities of interest at a scale that is computationally practical.

population undergoes a change in size, the coalescent is forced to accommodate this as a change in the waiting time between coalescent events. Thus, unexpected or unusual wait times can be reinterpreted as changes in effective population size. Many extensions of the coalescent have included populations that grow or shrink with deterministic demographic functions (Griffiths and Tavaré 1994; Donnelly and Tavaré 1995; Wilson and Balding 1998; Beaumont 1999) or sudden shifts between demographic trajectories (e.g., Shapiro et al. 2004). These extensions of the basic coalescent models are used to reveal the past dynamics of a population history.

Perhaps the most popular demographic model in use today is the Bayesian skyline plot (BSP), which allows the effective population size to change in a piecewise fashion at coalescent events (Ho and Shapiro 2011; Drummond et al. 2012). The predecessors to the BSP, classic (Pybus et al. 2000) and generalized (Strimmer and Pybus 2001) skyline plots, estimate changes in effective population size over time based on a single genealogy, similar to ML inference. The BSP improved on these models by estimating changes in effective population size from a distribution of genealogies in a Bayesian fashion, using Markov chain Monte Carlo (MCMC) algorithms (Box 4). This both integrates over the uncertainty in genealogies and allows for the calculation of credibility intervals (Box 1) for effective population size (Heled and Drummond 2008). Another advantage of a BSP, perhaps one that is just as pivotal for its popularity among researchers, is the ability to render its output in a visually pleasing graphical format (Figure 1B). This is the familiar plot found at the center of many studies of historical human population dynamics (e.g., Kitchen et al. 2008; Mulligan et al. 2008; Atkinson et al. 2009).

Under a BSP demographic model, effective population size is allowed to change an arbitrary number of times. The BSP model assumes that effective population size remains constant between change points but can instantaneously change at coalescent events (Drummond et al. 2005). The change points are determined by grouping neighboring coalescent events such that each group is associated with a single constant population size that persists across all coalescent events, with changes occurring at the transition from one group to another. The minimum number of groups is 1, which reduces the BSP to a constant population demographic function, and the maximum is $n - 1$

for a sample of n individuals, which provides for as many changes in effective population size (N_e) as there are coalescent events. The number of groups is fixed a priori, and though for most data sets use of an intermediate number of groups (e.g., 5 or 10) does not affect the analysis, an excessive number of groups will inhibit efficient MCMC performance. On the other hand, too few groups will not capture complex population histories. In these extreme cases, it is necessary to evaluate manually how the number of groups affects the fit of the model, which is often a slow and lengthy process.

It should be noted that choosing a BSP demographic model is itself a prior, in the sense that the user is allowing for piece-wise changes in population size. More stringent demographic models exist, such as constant population size, exponential growth, or logistic growth. These models can even be combined manually and tested against each other (Pybus and Rambaut 2002). Simpler models often fit the data better than models allowing for "free" population size change, such as a BSP. However, testing each and every demographic model can be extremely time-consuming, with no guarantee that an adequately fitting model will be found (Drummond et al. 2005). The BSP provides an easy alternative, in which user input is reduced to deciding on the number of groups of coalescent events. It is noteworthy that, because a BSP is in itself a prior, it is recommended that the fit of the empirical data to the model be tested by performing a Bayes factor model comparison (Box 2), with at least a few other demographic priors, such as the constant population model and the exponential model. This comparative step is important to ensure that biological conclusions are driven by the data and not by model selection.

The fact that in the basic BSP the number of transitions in effective population size is determined a priori, with no governing principle, can be problematic. First, a poor choice in the number of groups may lead to large credibility intervals or even prevent convergence of the analysis on an accurate estimate, reducing any confidence in the results (Heled and Drummond 2008). Second, having a few large transitions in effective population size between groups (in the fashion of steps) is an artificial representation of the historical reality of a natural population, in which transitions are expected to be gradual. To address this problem, extensions on the BSP have been made to specifically remove the necessity of a strong prior determining the number of transitions over time. The extended Bayesian skyline plot (EBSP) is a modification of the BSP in which the genetic data are referenced at each coalescent event to estimate a new effective population size, by the same means of variable selection as would apply to any other parameter (Heled and Drummond 2008). In an EBSP, transitions in effective population size are not reported for each coalescent event; instead, a pass-fail test is calculated from the likelihood function at each transition, for which the effective population will either remain the same as the previous interval or will change to reflect the newly calculated parameter. The decision to change or not change the population size is predicated on probability rather than being deterministic. Thus, the EBSP keeps with the tenants of Bayesian philosophy, where changes in effective population size are reported as a distribution with associated probabilities.

By their nature, Bayesian analyses revolve around model improvement as a means to approximate natural processes, even when data are uninformative or incomplete. Under these circumstances, a poorly resolved likelihood function will still yield parameter estimation, which will be driven by the prior distribution rather than the data, resulting in incorrect biological inference (Konigsberg and Frankenberg 2013). For this reason, it is important to test the statistical power of any analysis. A commonly used metric of statistical power is the effective sample size (ESS), which estimates the average number of independent (noncorrelated) data points in the posterior distribution of sampled genealogies, ensuring that the MCMC chain has sampled a diverse mix of genealogies. While there is no hard limit on how large an appropriate ESS should be, values under 200 are not recommended (Kuhner 2009). Software packages such as TRACER (Drummond and Rambaut 2007) provide the tools for ESS estimation and other simple qualitative estimations of analysis appropriateness, including comparisons of the outputs of multiple independent replicate analyses, providing a visual check of the convergence of posterior distributions on similar values across the runs. Finally, it is important here to caution that different software packages use different functions

for ESS calculation, which may lead to different values of ESS.

Approximate Bayesian Computation

As whole-genome and whole-exome sequencing becomes more common, and sampling of previously underrepresented human populations is under way, our ability to answer questions about human origins is progressively becoming limited by computational power rather than data availability. The advantage of Bayesian inference over traditional model resolution lies in the incorporation of shortcuts into the calculations of complex likelihoods. For example, MCMC integration (Box 4) rarifies the sample of possible genealogical trees, necessitating a simplified likelihood equation. However, as data become more complete, questions of human history become more detailed, and models become more realistic, the resolution of likelihood functions becomes computationally intractable (Beaumont 2010; Bertorelle et al. 2010). As a compromise, advanced Bayesian methods can provide further shortcuts, at the cost of increased mathematical and statistical understanding on the part of the researchers, reviewers, and audience.

Approximate Bayesian computation (ABC), a model rejection approach, has become widely adopted for its ability to discriminate among complex models of demographic history (Chan et al. 2006; Ramakrishnan and Hadly 2009; Csilléry et al. 2010). ABC bypasses solving likelihood functions by instead simulating data based on prior hypotheses of population evolution. ABC then compares the output of simulations to the empirical data, assigning each hypothesis a probability and generating a distribution of parameters and probabilities akin to a Bayesian posterior distribution (Beaumont 2010). ABC is further simplified by limiting data comparison to summary statistics (see below), which represent useful characteristics of the data in simplified form (Figure 2).

In applying ABC for demography, complete genealogies are simulated computationally to produce sets of sequences, in most cases using the coalescent as a foundation. These simulations are constrained by prior input from the user. Programs such as SimCoal, Serial SimCoal, fastsimcoal, ABC for R, and DIYABC (see Table 1), as well as BaySICS (Sandoval-Castellanos et al. 2014), facilitate the simulation of sequences and provide tools for model-fitting analysis as well. Other programs such as msprime and SLiM (see Table 1) can be used to generate simulations, although they require external means for model comparison.

Summary Statistics

Summary statistics aim to distill the largest amount of information into the simplest possible form (Csilléry et al. 2010). Changes in population demography are expected to generate specific patterns of nucleotide variation, reflected in summary statistics calculated from a given genealogy. Hence, past demographic events can be inferred from these summary statistics (Chan et al. 2006; Ramakrishnan and Hadly 2009). The cornerstone of ABC lies in selecting simulated genealogical trees that fit closely to summary statistics of the empirical data, reflecting the true demographic history by approximation.

The selection of summary statistics for use in an ABC analysis is not trivial, as each statistic is more or less susceptible to various evolutionary processes (Miró-Herrans and Mulligan 2012). For example, F_{ST} values are useful for estimating migration between populations but are not informative when migration is absent from a model. Conversely, Tajima's D (Tajima 1989) and its component summary statistics, segregating sites (S), and the average pairwise differences (Π), all provide insight into recent demographic events such as bottlenecks or population expansions but do not capture such effects as migration. During model comparison, a first impulse might be to include any and all information possible, including summary statistics that have no impact on the analysis. This line of thought, however, is incorrect. The addition of summary statistics increases the complexity of the analysis multiplicatively (Bertorelle et al. 2010). The addition of more information decreases the accuracy of model-fitting methods, making discrimination between models much more difficult.

Unfortunately, no general principles govern the proper number of summary statistics for a given ABC analysis (Bertorelle et al. 2010). While it is common wisdom to limit the number of statistics to two or three (Chan et al. 2006; Csilléry et al. 2010), each analysis should be optimized

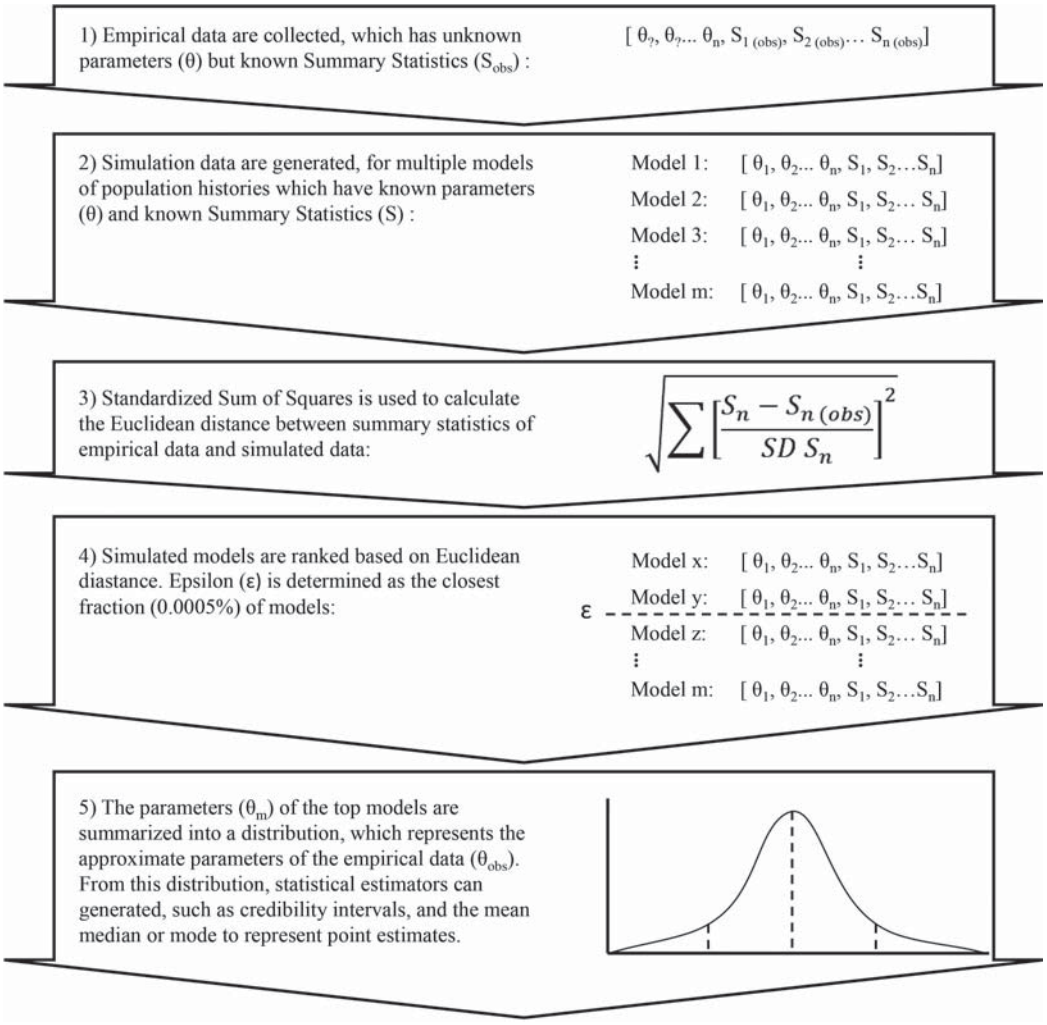

FIGURE 2. Basic steps of an approximate Bayesian computation (ABC) analysis.

individually to find balance between complexity and accuracy (see "Power Analysis to Optimize ABC," below).

A final word of caution: if the selected summary statistics lacks sufficiency, the resulting posterior introduces further approximation. A summary statistic is *sufficient* for a parameter if it provides just as much information to estimate the parameter as would the full data set (Csilléry et al. 2010). In practice, most summary statistics are not sufficient, so it is important to keep in mind that as the number of simulations considered is reduced (see next section), the resulting posterior may not necessarily approach the truth.

Model Fitting Algorithm

Once the number and identity of appropriate summary statistics have been selected, simulations are ranked based on their closeness to the summary statistics derived from the empirical data. The most commonly used method for estimating the distance between summary statistics is a simple standardized Euclidean error,

$$\sqrt{\sum\left[\frac{S_n - S_{n(\mathrm{obs})}}{SD\,S_n}\right]^2},$$

where the distance between each simulated data set and the empirical data is the normalized sum of squares of the summary statistics from a simulation (S_n) and a corresponding summary statistic drawn from the empirical data ($S_{n(\mathrm{obs})}$), divided by the standard deviation of the summary statistics from a simulation ($SD\,S_n$). Once a distance is associated with each simulated data set, distances can be easily sorted from smallest to largest. Notice that typically it is not a single simulation that is selected

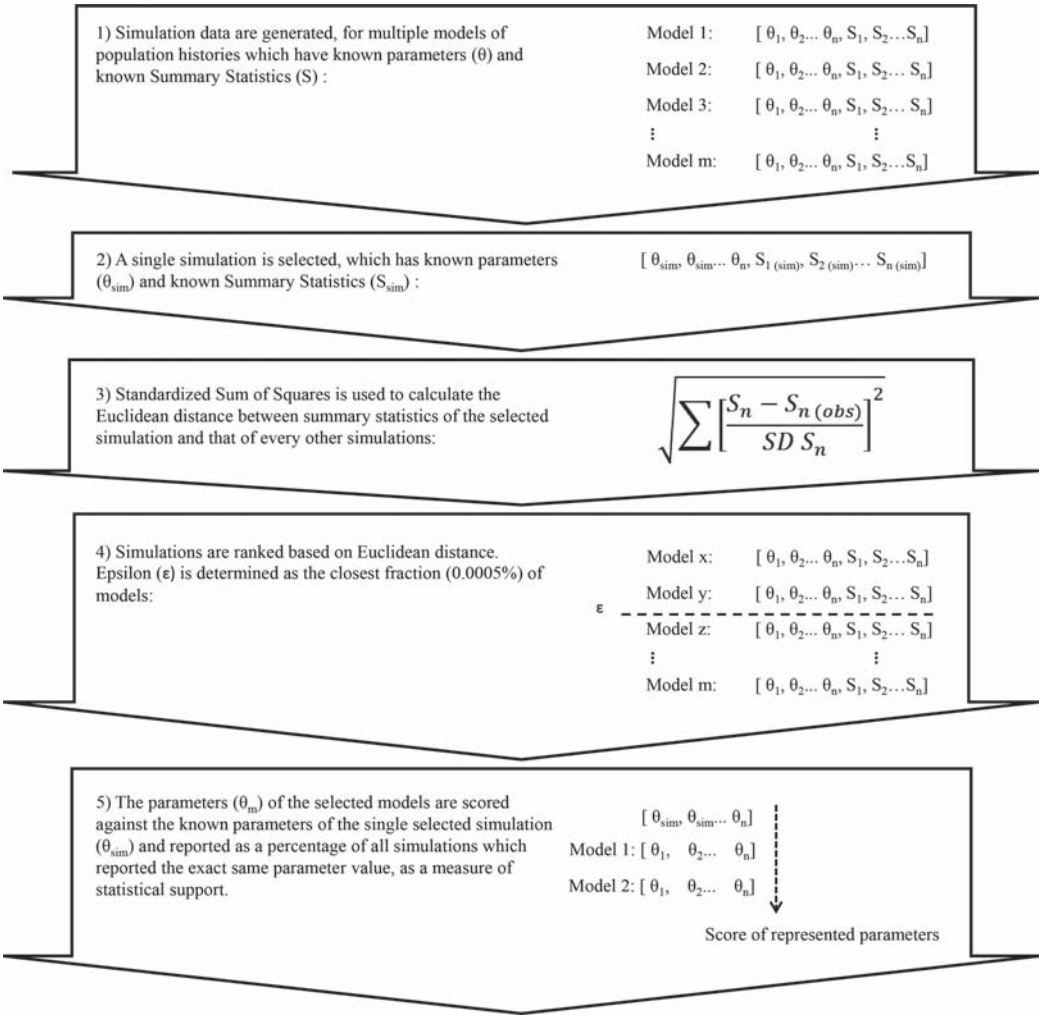

FIGURE 3. Basic steps to complete a power analysis for approximate Bayesian computation (ABC) as recommended in this study.

but a fraction of the total simulations, from which a distribution of parameters can be generated (e.g., the 1,000 closest simulations), in which case the cutoff value of accepted simulations is referred to as epsilon (ε).

The value for ε is user defined, and determining an appropriate cutoff value can be difficult, as once again no strict rules govern how it should be set. In most cases, ε is chosen independently of the distance values in the simulations themselves; most choose to select a static ε value prior to analysis, for example, the top 0.001% of all simulations. It is informally acceptable to estimate parameter values from such a small fraction, as typically the number of simulations tends to rank on a scale of 10^6 or larger. It is important to note that, if the chosen ε is very small, then a large number of simulations are discarded; in contrast, if ε is too large, the posterior distribution will be poorly characterized. Of course, generating an appropriate number of simulations can be computationally intensive as well, in which case other methods are available for parameter estimations, such as approximate approximate Bayesian computation (AABC; see Buzbas and Rosenberg 2015) and, more recently, various machine learning methods, beyond the scope of this review.

Parameter Estimation

The estimation of parameters through ABC presents the same advantages as Bayesian inference. By generating a distribution of parameters calculated from the selected simulations, each parameter value is assumed to occur in the distribution proportional to the likelihood of it occurring in the natural population. Note that the likelihood is not actually computed but instead approximated, lending to the name of this method: "approximate"

Bayesian computation. The collection of all parameter values and their occurrences then represents an approximation of the posterior probability distribution for that parameter (Beaumont 2010). By this understanding, parameter values can be calculated from the distribution of simulated parameters itself, which can be reported as an interval estimator (95% credibility interval is the most common), or even as a point estimate by using a mean, a median, or more appropriately, a mode, given most distributions will be asymmetric.

Power Analysis to Optimize ABC

Bayesian analyses often require extensive manipulation by the user, in the form of prior selection and experimental design. Similarly, the accuracy of ABC can be influenced by the choice of models studied, the number of simulations generated, the choice of summary statistics, and the choice of ε. Because of the intricacies of experimental design, it is important to generate the means to maximize the success of an analysis. Here we recommend a simple method to use the simulations generated prior to model fitting as a form of statistical power analysis for evaluating posterior distributions. The advantage of testing for statistical power prior to running an analysis on empirical data lies in the ability to modify the experimental design, either by expanding the number of simulations or by adjusting the summary statistics to maximize success.

Recall that empirical data has two qualities of interest: parameters, which are unknown and to be estimated, and summary statistics, which are known. Simulated data, conversely, possesses both known parameters and known summary statistics. Therefore, it is possible to treat individually simulated data as test data, which can be run through the model-fitting algorithm to produce an estimated parameter that can be scored against the known parameters of the test data, and thus ascertain the accuracy of the analysis (Figure 3). In addition, because computational resources can be limiting, prior testing can determine if a sufficient number of simulations have been generated and, more important, if expending resources generating further simulations is likely to increase statistical power.

Finally, it is important to observe the limits of ABC inference. In practical terms, genetic data may be of limited information, sampled from too few individuals, or exhibiting too few polymorphisms to converge on an acceptable answer. From a theoretical perspective, ABC will always provide a solution and thus always provide a distribution of parameter values. However, it is entirely possible that these parameter values do not reflect nature, simply because a more appropriate model was never considered. It is also possible for multiple demographic models to result in similar genealogies, providing no clear answers.

Examples of BSPs and ABC Applied to Anthropological Questions

Bayesian coalescent methods are becoming a common approach in anthropological genetics because they can be intuitive to apply and computationally viable. Bayesian coalescent analyses can be used efficiently in exploratory studies when effective population size is treated as a free parameter, and can also relate changes in effective population size with chronological dates of archaeological or historical importance. The ability to date demographic processes in chronological time allows biological events to be placed firmly in contexts of ecological and geological relevance.

Another advantage of the coalescent framework is the capacity to readily integrate data from individuals sampled from multiple points in time, including ancient DNA (aDNA) data. Bayesian coalescent analysis can readily formulate genealogies that include individuals sampled from multiple points in time. These *heterochronous data sets* are defined as genetic data from not just individuals belonging to the tips of a genealogy but individuals who may be the direct ancestors for the tips of a genealogy (Anderson et al. 2005; Drummond and Rambaut 2007; Ramakrishnan and Hadly 2009). Heterochronous data sets include both sequences collected from various generations of the same population (e.g., in short-generation species such as viruses) and sequences recovered from long-dead organisms (aDNA).

For example, Shapiro et al. (2004) used coalescent methods to estimate the timing of the decline of the now-extinct Beringian steppe bison (*Bison priscus*) using aDNA mitochondrial sequence data. The previous leading hypothesis posited that human presence was directly responsible

for the extinction of the Beringian steppe bison (the overkill hypothesis; see Grayson and Meltzer 2003). Shapiro et al.'s results date the beginning of the Beringian steppe bison population decline to 15,000 years earlier than the known presence of humans in Beringia, directly contradicting the expectation of the overkill hypothesis.

A later update to the study of the decline of the steppe bison by Lorenzen et al. (2011) illustrated how responses to climate change and human resource use during the last 50,000 years were species specific. Their study included mitochondrial aDNA from three other species of extinct megafauna: musk ox (*Ovibos moschatus*), woolly rhinoceros (*Coelodonta antiquitatis*), and wild horse (*Equus ferus*). This study integrated genetic, climatic, fossil, and human prehistory data to generate correlations between the presence of megafauna and humans and the shift in geographic ranges for both. To evaluate such long-term ecological interactions, they used a combination of BSP analysis and ABC, as a method for discrimination between various complex demographic models. The most influential finding of this study was that the decline in genetic diversity of the musk ox and woolly rhinoceros predated human presence and that climate change alone is a better explanation for their specific species declines. For the wild horse and steppe bison, a combination of climate change and human presence best explains their eventual extinction. The conclusions of Lorenzen et al. (2011) call for the consideration of species-specific responses to long-term climate change and anthropogenic stressors to infer causes of contemporary species declines.

Bayesian coalescent inference has also been central to the study of human population origins, particularly when archaeological or paleontological evidence is scarce or inconclusive. There is considerable interest is using coalescent methods to date known population expansions, including the expansions following the human migrations out of Africa. For example, BSPs made from 224 complete human mtDNA sequences indicate separate population expansions in South Asia 52,000 years ago, North and Central Asia 49,000 years ago, Europe 42,000 years ago, and the Middle East and North Africa 40,000 years ago (Atkinson et al. 2008). Later, Atkinson et al. (2009) provided evidence for a population expansion within Africa ca. 61,000–86,000 years ago, right before the human expansion out of Africa.

Similarly, Gignoux et al. (2011) used BSPs made from 425 mitochondrial coding region sequences to propose Holocene population expansions following the implementation of agriculture. To separate the signal of a population expansion from noise, they partitioned their mitochondrial sequences into lineages of hunter-gatherer or agricultural origin and generated independent BSPs from both. The authors found population expansions only in the agricultural lineages and dated them to 7,700 years ago in Europe, 4,700 years ago in Southeast Asia, and 4,600 years ago in sub-Saharan Africa.

Another area of interest has been determining the demographic history of the founder population from which contemporary Native Americans descend. Such questions have been approached with a combination of Bayesian tools, including early work by Hey (2005) using the Bayesian allele assignment program IM to ascertain the effective population size of the American founding population (later formalized in Hey and Nielsen 2007) and various later estimations (Fagundes et al. 2008a, 2008b). Of note, BSPs indicate the timing and trajectory of a large population expansion between 12,000 and 16,000 years ago, following human entry into the American continent (Kitchen et al. 2008; Mulligan et al. 2008).

A recent area of interest in human genetics is using the pattern of Neanderthal ancestry found in the genomes of living people to ascertain information about the admixture process between Neanderthals and anatomically modern humans. Originally, the breakdown of Neanderthal genome segments in modern human genomes indicated a time frame for admixture of 50,000–60,000 years ago, which is consistent with a single admixture event, prior to the diversification of East Asian and European lineages. However, Vernot and Akey (2015) used ABC on simulated neutral genome sequence data to reject a simple model of admixture. Instead, the authors advocate for admixture occurring multiple times: the first pulse of Neandertal gene flow into the population ancestral to East Asians and Europeans, and additional pulses after both populations had diverged.

Bayesian methods can even be used to estimate when humans first began wearing clothing. Toups et al. (2011) approached questions on the origin of

modern clothing by estimating the timing of the MRCA to the modern clothing lice species, which was estimated to at least 83,000 years and up to 170,000 years ago, implicating that clothing was developed by anatomically modern humans before the human species left Africa. Since clothes leave behind very little archaeological evidence, addressing this question from the straightforward means of looking for physical evidence had been difficult.

Concluding Remarks

> Essentially, all models are wrong, but some are useful.
>
> —George E. P. Box and Norman R. Draper, *Empirical Model-Building and Response Surfaces* (1987)

In the study of human demography, Bayesian coalescent methods offer a solution by placing part of that analytical burden on the user. BSP and ABC are powerful because they rely on preferential prior model selection, which can provide valuable insight by describing natural processes probabilistically. However, because Bayesian methods require intensive user input and because assigning priors can be arbitrary, constant trial and error is required. Equally dangerous is the interpretation of information-poor data sets, which can still provide parameter estimates, albeit bad ones. These limitations, far from damning Bayesian inference, instead highlight the importance of performing and interpreting these types of analyses from a place of understanding.

As scientists, we must pay close attention to the repeatability of results, check the convergence of posterior distributions on similar values, observe tests of statistical power such as ESS for Bayesian tree sampling analysis, and conduct statistical power analyses for ABC, steps that are paramount in providing informative advances for the community. Likewise, for reviewers and readers, a working understanding of these methods is fundamental for the advancement of the discipline, both for peer review and for voicing criticism.

ACKNOWLEDGMENTS

We thank Dr. Erkan Buzbas for consulting on the applications and limitations of approximate Bayesian computation, Katherine Brunson for helping us better understand questions of anthropological interests, and Nathan Layman for enlightening conversations about the philosophy of statistical inference. We also thank both reviewers, who added important commentary on the finer points of statistical inference.

Received 18 December 2019; accepted for publication 27 January 2020.

LITERATURE CITED

Akey, J. M., M. A. Eberle, M. J. Rieder et al. 2004. Population history and natural selection shape patterns of genetic variation in 132 genes. *PLoS Biol.* 2:1,591–1,599.

Ammerman, A. J., and L. L. Cavalli-Sforza. 1984. *The Neolithic Transition and the Genetics of Populations in Europe*. Princeton, NJ: Princeton University Press.

Anderson, C. N. K., U. Ramakrishnan, Y. L. Chan et al. 2005. Serial SimCoal: A population genetics model for data from multiple populations and points in time. *Bioinformatics* 21:1,733–1,734.

Atkinson, Q. D., R. D. Gray, and A. J. Drummond. 2008. mtDNA variation predicts population size in humans and reveals a major southern Asian chapter in human prehistory. *Mol. Biol. Evol.* 25:468–474.

Atkinson, Q. D., R. D. Gray, and A. J. Drummond. 2009. Bayesian coalescent inference of major human mitochondrial DNA haplogroup expansions in Africa. *Proc. R. Soc. Lond. B Biol. Sci.* 276:367–373.

Beaumont, M. A. 1999. Detecting population expansion and decline using microsatellites. *Genetics* 153:2,013–2,029.

Beaumont, M. A. 2010. Approximate Bayesian computation in evolution and ecology. *Annu. Rev. Ecol. Evol. Syst.* 41:379–406.

Beerli, P. 2006. Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. *Bioinformatics* 21:341–345.

Berniell-Lee, G., F. Calafell, E. Bosch et al. 2009. Genetic and demographic implications of the Bantu expansion: Insights from human paternal lineages. *Mol. Biol. Evol.* 26:1,581–1,589.

Bertorelle, G., A. Benazzo, and S. Mona. 2010. ABC as a flexible framework to estimate demography over space and time: Some cons, many pros. *Mol. Ecol.* 19:2,609–2,625.

Bonatto, S. L., and F. M. Salzano. 1997a. Diversity and age of the four major mtDNA haplogroups, and their implications for the peopling of the New World. *Am. J. Hum. Genet.* 61:1,413–1,423.

Bonatto, S. L., and F. M. Salzano. 1997b. A single and early

migration for the peopling of the Americas supported by mitochondrial DNA sequence data. *Proc. Natl. Acad. Sci. U. S. A.* 94:1,866–1,871.

Bouckaert, R., T. G. Vaughan, J. Barido-Sottani et al. 2019. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol*. 15:e1006650.

Buzbas, E. O., and N. A. Rosenberg. 2015. AABC: Approximate approximate Bayesian computation for inference in population-genetic models. *Theor. Popul. Biol.* 99:31–42.

Casella, G. 2008. Bayesians and frequentists: Models, assumptions, and inference. http://archived.stat.ufl.edu/casella/Talks/BayesRefresher.pdf.

Chan, Y. L., C. N. K. Anderson, and E. A. Hadly. 2006. Bayesian estimation of the timing and severity of a population bottleneck from ancient DNA. *PLoS Genet.* 2:e59.

Cornuet, J.-M., P. Pudlo, J. Veyssier et al. 2014. DIYABC v2.0: A software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics* 30:1,187–1,189.

Csilléry, K., M. Blum, O. Gaggiotti et al. 2010. Approximate Bayesian computation (ABC) in practice. *Trends Ecol. Evol.* 25:410–418.

Csilléry, K., O. François, and M. G. B. Blum. 2012. abc: An R package for approximate Bayesian computation (ABC). *Methods Ecol. Evol.* 3:475–479.

Donnelly, P., and S. Tavaré. 1995. Coalescents and genealogical structure under neutrality. *Annu. Rev. Genet.* 29:401–421.

Drummond, A., and A. Rambaut. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7:1–8.

Drummond, A. J., A. Rambaut, B. Shapiro et al. 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol. Biol. Evol.* 22:1,185–1,192.

Drummond, A. J., M. A. Suchard, D. Xie et al. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29:1,969–1,973.

Edwards, A. W. F. 1970. Estimation of the branch points of a branching diffusion process. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 32:155–174.

Excoffier, L., I. Dupanloup, E. Huerta-Sánchez et al. 2013. Robust demographic inference from genomic and SNP data. *PLoS Genet.* 9:1–17.

Excoffier, L., and M. Foll. 2011. fastsimcoal: A continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics* 27:1,332–1,334.

Excoffier, L., J. Novembre, and S. Schneider. 2000. SIMCOAL: A general coalescent program for the simulation of molecular data in interconnected populations with arbitrary demography. *J. Hered.* 91:506–509.

Fagundes, N. J. R., R. Kanitz, and S. L. Bonatto. 2008a. A reevaluation of the Native American mtDNA genome diversity and its bearing on the models of early colonization of Beringia. *PLoS One* 3:1–5.

Fagundes, N. J. R., R. Kanitz, R. Eckert et al. 2008b. Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. *Am. J. Hum. Genet.* 82:583–592.

Falush, D., M. Stephens, and J. K. Pritchard. 2003. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* 164:1567–1587.

Falush, D., M. Stephens, and J. K. Pritchard. 2007. Inference of population structure using multilocus genotype data: Dominant markers and null alleles. *Mol. Ecol. Notes* 7:574–578.

Gignoux, C. R., B. M. Henn, and J. L. Mountain. 2011. Rapid, global demographic expansions after the origins of agriculture. *Proc. Natl. Acad. Sci. U. S. A.* 108:6,044–6,049.

Grayson, D. K., and D. J. Meltzer. 2003. A requiem for North American overkill. *J. Archaeol. Sci.* 30:585–593.

Griffiths, R. C., and S. Tavaré. 1994. Sampling theory for neutral alleles in a varying environment. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 344:403–410.

Haller, B. C., and P. W. Messer. 2019. SLiM 3: Forward genetic simulations beyond the Wright-Fisher model. *Mol. Biol. Evol*. 36: 632–637.

Heled, J., and A. J. Drummond. 2008. Bayesian inference of population size history from multiple loci. *BMC Evol. Biol*. 8:1–15.

Hey, J. 2005. On the number of New World founders: A population genetic portrait of the peopling of the Americas. *PLoS Biol.* 3:e193.

Hey, J., and R. Nielsen. 2007. Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proc. Natl. Acad. Sci. U. S. A.* 104:2,785–2,790.

Ho, S. Y. W., and B. Shapiro. 2011. Skyline-plot methods for estimating demographic history from nucleotide sequences. *Mol. Ecol. Resour.* 11:423–434.

Hubisz, M. J., D. Falush, M. Stephens et al. 2009. Inferring weak population structure with the assistance of sample group information. *Mol. Ecol. Re*. 9:1,322–1,332.

Hudson, R. R. 1983. Testing the constant-rate neutral allele model with protein sequence data. *Evolution* 37:203–217.

Jeffreys, H. 1998. *The Theory of Probability.* Oxford: Oxford

University Press.

Kass, R. E., and A. E. Raftery. 1995. Bayes factors. *J. Am. Stat. Assoc.* 90:773–795.

Kelleher, J., A. M. Etheridge, and G. McVean. 2016. Efficient coalescent simulation and genealogical analysis for large sample sizes. *PLoS Comput. Biol.* 12:1–22.

Kingman, J. F. C. 1982a. The coalescent. *Stoch. Process. Their Appl.* 13:235–248.

Kingman, J. F. C. 1982b. On the genealogy of large populations. *J. Appl. Probab.* 19:27–43.

Kitchen, A., M. M. Miyamoto, and C. J. Mulligan. 2008. A three-stage colonization model for the peopling of the Americas. *PLoS One* 3:1–7.

Konigsberg, L. W., and S. R. Frankenberg. 2013. Bayes in biological anthropology. *Am. J. Phys. Anthropol.* 152:153–184.

Krzywinski, M., and N. Altman. 2013a. Points of significance: Error bars. *Nat. Methods* 10:921–922.

Krzywinski, M., and N. Altman. 2013b. Points of significance: Importance of being uncertain. *Nat. Methods* 10:809–810.

Krzywinski, M., and N. Altman. 2013c. Points of significance: Significance, *P* values and *t*-tests. *Nat. Methods* 10:1,041–1,042.

Kuhner, M. K. 2006. LAMARC 2.0: Maximum likelihood and Bayesian estimation of population parameters. *Bioinformatics* 22:768–770.

Kuhner, M. K. 2009. Coalescent genealogy samplers: Windows into population history. *Trends Ecol. Evol.* 24:86–93.

Lorenzen, E. D., D. Nogues-Bravo, L. Orlando et al. 2011. Species-specific responses of Late Quaternary megafauna to climate and humans. *Nature* 479:359–364.

Macaulay, V., C. Hill, A. Achilli et al. 2005. Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* 308:1,034–1,036.

Miró-Herrans, A. T., and C. J. Mulligan. 2012. Human demographic processes and genetic variation as revealed by mtDNA simulations. *Mol. Biol. Evol.* 30:244–252.

Mulligan, C. J., A. Kitchen, and M. M. Miyamoto. 2008. Updated three-stage model for the peopling of the Americas. *PLoS One* 3:1–4.

Palacios, J. A., A. Véber, L. Cappello et al. 2019. Bayesian estimation of population size changes by sampling Tajima's trees. *Genetics* 213:967–986.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945–959.

Puga, J. L., M. Krzywinski, and N. Altman. 2015a. Points of significance: Bayes' theorem. *Nat. Methods* 12:277–278.

Puga, J. L., M. Krzywinski, and N. Altman. 2015b. Points of significance: Bayesian statistics. *Nat. Methods* 12:377–378.

Pybus, O. G., and A. Rambaut. 2002. GENIE: Estimating demographic history from molecular phylogenies. *Bioinformatics* 18:1,404–1,405.

Pybus, O. G., A. Rambaut, and P. H. Harvey. 2000. An integrated framework for the inference of viral population history from reconstructed genealogies. *Genetics* 155:1,429–1,437.

Ramachandran, S., O. Deshpande, C. C. Roseman et al. 2005. Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proc. Natl. Acad. Sci. U. S. A.* 102:15,942–15,947.

Ramakrishnan, U. M. A., and E. A. Hadly. 2009. Using phylochronology to reveal cryptic population histories: Review and synthesis of 29 ancient DNA studies. *Mol. Ecol.* 18:1,310–1,330.

Ray, N., D. Wegmann, N. J. R. Fagundes et al. 2010. A statistical evaluation of models for the initial settlement of the American continent emphasizes the importance of gene flow with Asia. *Mol. Biol. Evol.* 27:337–345.

Rice, S. H. 2004. *Evolutionary Theory: Mathematical and Conceptual Foundations.* Sunderland, MA: Sinauer.

Ronquist, F., and J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1,572–1,574.

Sandoval-Castellanos, E., E. Palkopoulou, and L. Dalen. 2014. Back to BaySICS: A user-friendly program for Bayesian statistical inference from coalescent simulations. *PLoS One* 9:1–9.

Serre, D., A. Langaney, M. Chech et al. 2004. No evidence of Neandertal mtDNA contribution to early modern humans. *PLoS Biol.* 2:e57.

Shapiro, B., A. J. Drummond, A. Rambaut et al. 2004. Rise and fall of the Beringian Steppe bison. *Science* 306:1,561–1,565.

Strimmer, K., and O. G. Pybus. 2001. Exploring the demographic history of DNA sequences using the generalized skyline plot. *Mol. Biol. Evol.* 18:2,298–2,305.

Tajima, F. 1983. Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105:437–460.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595.

Tamm, E., T. Kivisild, M. Reidla et al. 2007. Beringian standstill and spread of Native American founders. *PLoS One* 9:1–6.

Toups, M. A., A. Kitchen, J. E. Light et al. 2011. Origin of clothing lice indicates early clothing use by anatomically

modern humans in Africa. *Mol. Biol. Evol.* 28:29–32.

Vernot, B., and J. M. Akey. 2015. Complex history of admixture between modern humans and Neandertals. *Am. J. Hum. Genet.* 96:448–453.

Wakeley, J. 2009. *Coalescent Theory: An Introduction.* Greenwood Village, CO: Roberts and Company.

Wang, S., C. M. Lewis, M. Jakobsson et al. 2007. Genetic variation and population structure in Native Americans. *PLoS Genet.* 3:e185.

Wilson, I. J., and D. J. Balding. 1998. Genealogical inference from microsatellite data. *Genetics* 150:499–510.