Annual Review of Ecology, Evolution, and Systematics

Using Genomic Data to Infer Historic Population Dynamics of Nonmodel Organisms

Annabel C. Beichman,[1] Emilia Huerta-Sanchez,[2,3] and Kirk E. Lohmueller[1,4]

[1]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, California 90095, USA; email: klohmueller@ucla.edu

[2]Department of Molecular and Cell Biology, University of California, Merced, California 95343, USA

[3]Current affiliation: Department of Ecology and Evolutionary Biology, Brown University, Providence, Rhode Island 02912, USA

[4]Interdepartmental Program in Bioinformatics and Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, California 90095, USA

Annu. Rev. Ecol. Evol. Syst. 2018. 49:433–56

First published as a Review in Advance on August 16, 2018

The *Annual Review of Ecology, Evolution, and Systematics* is online at ecolsys.annualreviews.org

https://doi.org/10.1146/annurev-ecolsys-110617-062431

Keywords

coalescent, demographic inference, nonmodel organisms, statistical inference, whole-genome sequence data

Abstract

Genome sequence data are now being routinely obtained from many nonmodel organisms. These data contain a wealth of information about the demographic history of the populations from which they originate. Many sophisticated statistical inference procedures have been developed to infer the demographic history of populations from this type of genomic data. In this review, we discuss the different statistical methods available for inference of demography, providing an overview of the underlying theory and logic behind each approach. We also discuss the types of data required and the pros and cons of each method. We then discuss how these methods have been applied to a variety of nonmodel organisms. We conclude by presenting some recommendations for researchers looking to use genomic data to infer demographic history.

1. WHY GENOMIC DATA?

The focus of this review is on inferring demographic history from genomic data. By genomic data, we mean genetic variation data from multiple loci sampled from throughout the autosomal genome. This could take the form of whole-genome sequence (WGS) data or thousands of loci sampled throughout the genome using methods such as restriction site–associated DNA sequencing (RAD-Seq), sequence capture, or RNA-Seq. With the advent of next-generation sequencing technology, it is becoming routine to collect this type of data from nonmodel species.

In this review, demographic inference refers to finding a particular model describing population size changes as well as population split and mixture events over time. Genomic data have important practical benefits for inference of demographic history. To gain insights from genomic data, we need models that describe genetic variation, such as the coalescent (see **Supplemental Note 1** and **Supplemental Figure 1**). The coalescent is a probabilistic model that underlies how neutral genetic variation data can be generated under a particular demographic model. Said another way, a given demographic model can produce many possible genealogies. These distinct genealogies will lead to different patterns of genetic variation. Each position of the genome has a particular genealogy describing its history (see **Supplemental Note 1** for further detail on the coalescent). Sites near to each other that lack historical recombination will share the exact same genealogy (**Figure 1**). Loci far apart from each other or on different chromosomes essentially have independent genealogies, each drawn from the set of possible genealogies from the particular demographic model. Sequence data from a single locus (e.g., a mitochondrial genome or a single short segment of autosomal sequence lacking recombination) provide one independent realization of the evolutionary process. Because this is only one of the essentially infinite number of possible genealogies for the underlying demography, inferences based on a single locus have considerable uncertainty. Increasing the numbers of individuals sequenced will not solve this problem. Additional individuals will still be a part of this same genealogy. The only way to reduce the uncertainty of the demographic model is to increase the number of genealogies sampled. Sampling many loci throughout the genome yields a set of nearly independent genealogies that contain a wealth of information for demographic inference.

Supplemental Material >

2. METHODS FOR DEMOGRAPHIC INFERENCE

2.1. Methods Based on the Site Frequency Spectrum

The site frequency spectrum (SFS) refers to the number of single-nucleotide polymorphisms (SNPs) at particular frequencies in a sample of individuals (**Figure 1**). For the SFS shown in **Figure 1**, for example, there are five singletons, or sites where only one chromosome in the sample carries the derived (i.e., mutant) allele. The SFS can be constructed from a single genomic region, the entire genome, or a particular category of sites (e.g., only those SNPs that are in noncoding portions of the genome). Importantly, the SFS treats all SNPs in the data set as independent of one another, ignoring the correlation structure [i.e., linkage disequilibrium (LD)] among markers (**Figure 1**).

In almost all cases, resequencing data are preferred to create the SFS, as computing an SFS from SNP genotype data can be fraught with difficulties relating to how SNPs were ascertained (Clark et al. 2005). As the SFS is a function of the sample size, sequencing data from multiple individuals is required. While there is no specific minimum number of individuals necessary, larger numbers of individuals (even in the hundreds or thousands) will increase the ability to infer recent demographic events (within the last hundred generations). Smaller sample sizes, say, at least

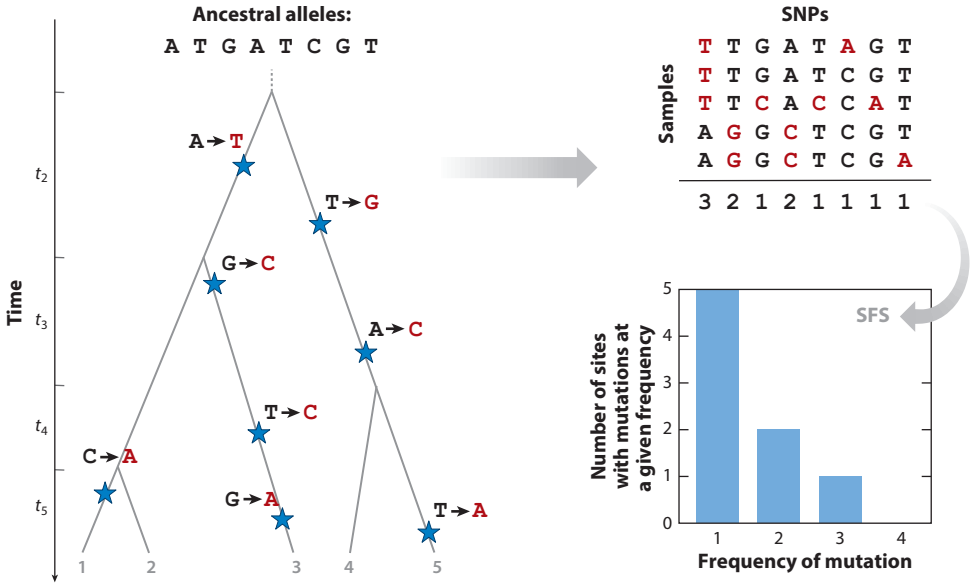

Figure 1

The locations of mutations on the coalescent genealogy (*left*) give rise to patterns of genetic variation data (*top right*). The SFS depicts the mutational patterns seen in the genetic variation data. Abbreviations: SFS, site frequency spectrum; SNPs, single nucleotide polymorphisms.

10 individuals, will enable inference of more ancient processes. Since SNPs are treated as unlinked, the SFS can be computed from sequencing of short, random fragments, such as those resulting from RAD-Seq approaches (see below).

The SFS is a useful summary statistic because it is influenced by the history of the population (**Figure 2**). Coalescent theory (see **Supplemental Note 1**) provides some guidance for how demography impacts the SFS. Different demographic scenarios change the shape and branch lengths of the underlying genealogies, which consequently change the SFS. The theory of the expected SFS in a randomly mating population of constant size was developed in the 1990s (Fu 1995) and later extended to include population size changes (Griffiths & Tavaré 1998, Polanski & Kimmel 2003, Polanski et al. 1998).

Supplemental Material >

Population growth, going forward in time, means that the current population size is larger than the ancient population size (**Figure 2**). The probability of lineages coalescing in a particular generation is lower in a larger population than in a smaller population (see **Supplemental Note 1**). As such, the population growth scenario predicts that genealogies should have long external branches, corresponding to the large population size where the probability of coalescing in a particular generation is lower, and shorter internal ones, where the per generation probability of coalescing is higher (**Figure 2**). Mutations randomly added to these genealogies would likely be singletons, as mutations have a higher probability of falling on external branches, because most of the genealogy is composed of external branches. Thus, population growth results in the SFS being skewed toward a greater proportion of low-frequency SNPs and singletons (**Figure 2**). A population contraction is a decrease in population size going forward in time. It generates a pattern opposite to that of population growth. Under a contraction, the probability of coalescing in a particular generation is higher in the current population. Population contractions result in a lower proportion of low-frequency variants compared to a population of constant size.

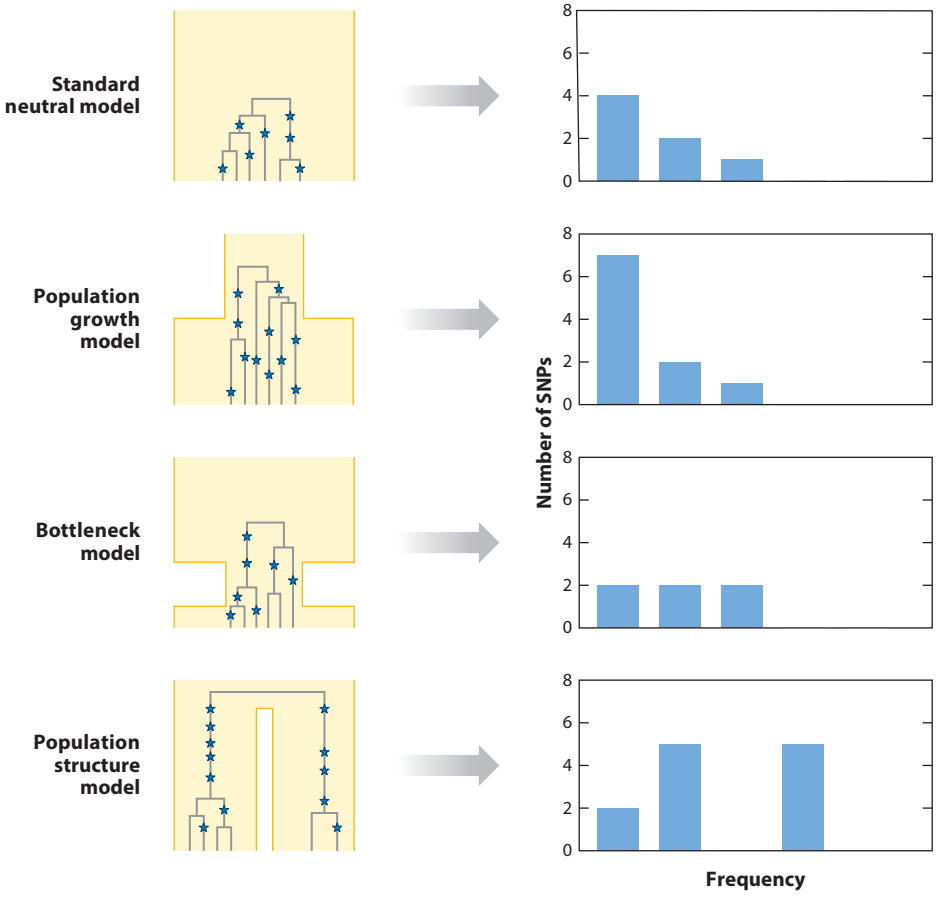

Figure 2

Population history influences the shape of genealogies and the SFS. The yellow shaded areas on the left denote the history of each population. These demographic histories give rise to the genealogies shown within each model. Blue stars denote mutations that occur on the genealogies. The histograms denote the SFS for each model that is generated from the mutational pattern that occurred on the genealogy. Abbreviations: SFS, site frequency spectrum; SNPs, single nucleotide polymorphisms.

Population bottlenecks, or periods of reduced population size, followed by an expansion can have any number of effects on the genealogies and thus on the SFS, depending on the parameters of the bottleneck. If the bottleneck is very severe, then all lineages will coalesce during the bottleneck, generating genealogies and SFSs that look like those of population growth. If the bottleneck is less severe, then some lineages will not coalesce in the bottleneck and will make their way back to the larger ancestral population (**Figure 2**). Once there, they will take longer to coalesce in this larger population. As such, the genealogies will have longer internal branches and the SFS will have fewer low-frequency variants than an SFS from a constant-sized population.

Population structure also affects the SFS (**Figure 2**). If an equal number of lineages is sampled from the two different populations, the migration rate is low, and the populations split from each other long ago, then it is likely that lineages within each subpopulation will coalesce with each other prior to coalescing with lineages from the other subpopulation. These genealogies will

have long internal branches. Mutations falling on these branches will be carried exclusively by individuals in that subpopulation. In the extreme case, these mutations will be fixed differences between the two populations. If data from the two populations are combined when making the SFS, this type of population structure will result in an excess of intermediate-frequency SNPs in the SFS. If more lineages are sampled from one subpopulation than the other, the bin of the SFS corresponding to the number of descendants of the long internal branch will be higher (**Figure 2**). Higher migration rates or a more recent split will result in a less pronounced skew in the SFS.

Because demographic history can impact the SFS, it is a useful summary statistic to infer demographic parameters. Typically, researchers will tabulate the empirical SFS from their resequencing data. Then, coalescent theory, either in the form of analytical calculations (Bhaskar et al. 2015, Keinan et al. 2007, Marth et al. 2004) or simulations (Adams & Hudson 2004, Excoffier et al. 2013, Nielsen 2000), is used to generate the predicted SFS for a particular demographic model. Note that here a demographic model consists of both the type of model (e.g., growth model, bottleneck, multipopulation split model) and the particular parameter values (e.g., migration rate, population size, time of size change). Once the SFS predicted by the demographic model has been generated, the fit of the predicted SFS to the empirical SFS can be assessed, typically in a likelihood framework.

There are two ways that the SFS can be used for statistical inference. The first approach uses the proportions of SNPs at different frequencies in the sample that follows a multinomial distribution (Adams & Hudson 2004, Nielsen 2000), yielding a multinomial likelihood function. Alternatively, inference can be done using the number of SNPs at different frequencies in the data, yielding a Poisson likelihood function with the rate parameter coming from the demographic model (Hartl et al. 1994). The Poisson likelihood approach has the advantage of using more information from the data but requires information about mutation rates. Using either approach, researchers can search over the space of demographic parameters to find the parameter values that generate SFSs that most closely match the observed SFS, and thereby maximize the likelihood function. If SNPs are not independent of each other, the same approach can be used, but the likelihoods will be composite likelihoods rather than true likelihoods. Composite likelihood estimators have similar properties to likelihood estimators and are consistent (Wiuf 2006), but the asymptotic confidence intervals are too small. Instead, appropriate confidence intervals are found through bootstrapping (Gutenkunst et al. 2009, Keinan et al. 2007) or approaches that seek to correct the likelihood function for LD (Coffman et al. 2016).

A number of software packages implement the methods to infer demographic parameters from the SFS. The fastsimcoal2 software (Excoffier et al. 2013) uses coalescent simulations to generate the predicted SFS for demographic parameters and can consider both single- and multipopulation models. For inference using more than one population, this method can use the multidimensional SFS, essentially a two-dimensional histogram, to infer split times and migration rates. Because coalescent simulations are used, it can be time consuming to run for large numbers of populations or samples. Further, many simulation replicates must be run to obtain reliable estimates of the likelihood. Another software package for demographic inference that uses the SFS with coalescent models is FastNeutrino (Bhaskar et al. 2015), which analytically calculates the SFS for a population size change model. Thus, FastNeutrino is faster, especially for large samples, but currently it can only consider a single population.

Beyond using the coalescent to obtain the predicted SFS for a given demographic model, other approaches model genetic drift forward in time and use diffusion equations to approximate the binomial distribution of allele frequency change (Kimura 1964). Inference can be done in the likelihood framework as discussed above. One popular program that uses diffusion theory to generate the SFS from a demographic model is δaδi (Gutenkunst et al. 2009), which can work

with single populations or jointly model up to three populations simultaneously. Another program, MultiPop, uses polynomials to solve the forward diffusion approximation and can yield more exact results under certain models (Lukic & Hey 2012).

2.2. Approximate Bayesian Computation with Summary Statistics

Another computational approach used for demographic inference is approximate Bayesian computation (ABC) (Beaumont et al. 2002, Fu & Li 1997, Marjoram et al. 2003, Pritchard et al. 1999, Tavaré et al. 1997). The idea behind this approach is to infer demographic parameters from multiple summary statistics, such as the average number of pairwise differences between sequences, F_{ST}, or LD decay. The summary statistics are computed on the real data set and from an ensemble of coalescent simulations under a prior distribution of parameter values (**Figure 3**). Demographic parameter values from the prior distribution are then accepted if they generate summary statistics that are close to the values seen in the empirical data, thus yielding the posterior parameter distribution. Several excellent review articles on ABC methods in population genetics have been published (Beaumont 2010, Bertorelle et al. 2010, Csilléry et al. 2010, Marjoram 2013, Sunnåker et al. 2013).

There are several advantages to ABC approaches. Firstly, because coalescent simulations are used to generate summary statistics from demographic models, very complex and realistic models can be explored. Secondly, ABC methods are fairly general and can be applied to different platforms. Previous studies have used SNP genotype data (Wollstein et al. 2010), resequencing data from short fragments in multiple individuals (Fagundes et al. 2007), or WGS data on a small number of individuals (Robinson et al. 2016). Ancient DNA can also be used in conjunction with modern DNA in ABC inference frameworks (see **Supplemental Note 2** and **Supplemental Figure 2**). Of course, different summary statistics will need to be employed depending on the data type and the particular parameters that can be estimated.

Supplemental Material >

The ABC approach begins by selecting a particular demographic model of arbitrary complexity, from which one wishes to infer parameters (**Figure 3**). A prior distribution is specified for each parameter of interest. Often, uniform priors are chosen. For parameters that tend to vary over many orders of magnitude, like migration rates and selection coefficients, log-normal distributions can be chosen. When in doubt, multiple prior distributions can be considered to assess the robustness of the results. Coalescent simulations are run based on parameters drawn from the priors, and simulations should match the specifics of the empirical data used in the study, including the numbers of individuals sequenced, the number of loci, and other ascertainment conditions. The resulting summary statistics (multiple statistics are allowed) from the simulated data set are compared to those from the empirical data, and if close enough, the parameter values used to simulate the data are retained and contribute to the posterior distribution. If, however, the simulated summary statistics are not close in value to those of the empirical data, the parameters are rejected. This process is repeated until enough replicates (at least a few thousand) are retained. The distribution of the accepted parameter values will represent an approximate posterior distribution. This approach, as outlined above, is a simple rejection sampling approach that has been widely used in population genetics and other fields (Marjoram 2013, Tavaré et al. 1997).

A major methodological advance in ABC is that rather than implementing the algorithm described above, parameter sets can be more heavily weighted if the resulting simulation more closely matches the summaries of real data (Beaumont et al. 2002). In this framework, one can accept parameter values that are not as close a match to the empirical data, increasing the efficiency of the ABC. This framework has been implemented in local linear regression models (Beaumont et al. 2002) as well as other nonlinear regression approaches (Blum & François 2010).

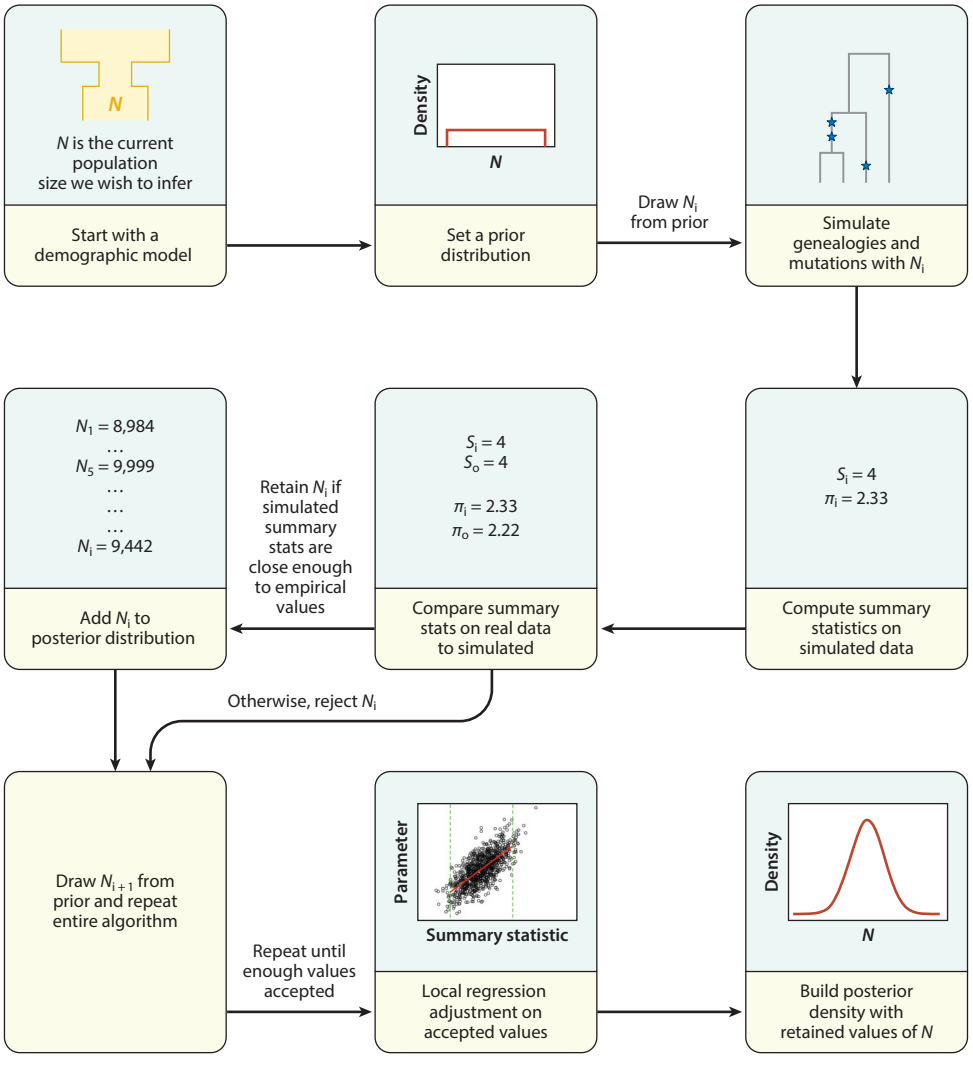

Figure 3

Workflow for an approximate Bayesian computation (ABC) approach to demographic inference. Here, we wish to infer N, the current population size in the bottleneck model. The empirical data for this example consist of a region of the genome where $S_o = 4$ and $\pi_o = 2.22$, and summary statistics calculated from simulation replicate i are labeled S_i and π_i.

Newer approaches involve using deep learning to match the nonlinear relationship between combinations of summary statistics and parameter values (Sheehan & Song 2016).

ABC methods have several disadvantages. Firstly, the methods are highly computationally intensive, often involving billions of coalescent simulations, which may limit the parameter space explored and can lead to incorrect inference. However, the simplest ABC approaches are easily parallelizable and computational power is continuing to increase, potentially ameliorating this concern. Secondly, ABC approaches may not be efficient if the prior distributions are wide or if the type of model being considered is unlikely to generate the observed data. Under these conditions, a high proportion of parameter values will be rejected, although the local regression

approach discussed above will mitigate some of this concern. A more serious drawback is that the success of ABC inference may depend on the choice of summary statistics. If the summary statistics are not capturing enough information from the data or the relevant information for the parameters of interest, then the posterior distributions will appear similar to the prior distribution, suggesting that the data are noninformative. Deciding which summaries are relevant is more of an art, although some intuition can be gained from understanding the type of models being analyzed. For example, if one is estimating the migration rate and split times between populations, a statistic related to population differentiation, like F_{ST}, is more informative. Recent work has provided more systematic approaches to picking the optimal set of summary statistics (Joyce & Marjoram 2008, Jung & Marjoram 2011, Nunes & Balding 2010).

A number of software packages implementing ABC approaches have been developed. Popular packages include DIY-ABC (Cornuet et al. 2008, 2014), popABC (Lopes et al. 2009), and ABC-toolbox (Wegmann et al. 2010). These packages are similar to each other in that they can be used to analyze a variety of types of data, and they rely on other software to perform the coalescent simulations. Researchers interested in applying ABC approaches should investigate these methods to see whether one approach may have certain advantages for their specific question and data.

In addition to ABC methods, approximate likelihood approaches have been devised by other researchers (Weiss & Von Haeseler 1998). These approaches compute the likelihood of the model parameters, given the summary statistics (i.e., probability of the summary statistics under the model), which is possible when the distribution of the summary statistics is approximately known. For example, Plagnol & Wall (2006) assumed that a set of summary statistics (Tajima's D; Fu and Li's D; F_{ST}) was multivariate normal and Lohmueller et al. (2009) assumed that the number of haplotypes and the count of the most common haplotype in windows across the genome follow a multinomial distribution.

2.3. Identity by State/Identity by Descent

Summaries of allele frequencies, such as the SFS, assume that sites are independent and the associated inference methods do not account for linkage between sites. To model linkage, there are inference methods based on haplotype patterns: the arrangement of specific alleles occurring in the same chromosome within a given genetic segment. Given a collection of individual DNA sequences, similarity between the haplotypes in a given region of the genome may be due to identity by descent (IBD) or identity by state (IBS). IBD corresponds to having shared sequence content due to shared ancestry (the samples are descendants from a recent common ancestor and no recombination events have occurred to break up that tract of DNA), and IBS corresponds to the sequences being identical in their composition alone, whether or not they have shared ancestry. IBD does not always imply IBS, as new mutations may have occurred on the given genetic segment, so it is not required to have identical sequence composition despite shared ancestry. Conversely, IBS does not always imply IBD. Calling IBD tracts requires an additional inference step, which is not trivial, and several tools (e.g., GERMLINE, diCal-IBD, HaploScore, Relate-Admix, Refined IBD, IBDseq; Browning & Browning 2013a,b; Durand et al. 2014; Gusev et al. 2009; Moltke & Albrechtsen 2014; Tataru et al. 2014) have been developed to identify IBD segments in the genome [see Ramstetter et al. (2017) for comparative accuracy between several IBD calling methods]. By contrast, IBS is computed directly from the observed data, making it sensitive to several factors affecting data quality, such as phasing errors, sequencing errors, and missing data.

2.3.1. Identity by descent. IBD is informative about the demography of the population because population-level parameters directly influence IBD (Browning & Browning 2013a, 2015; Gusev

et al. 2012). The effective population size, N_e, under the neutral model is inversely related to the time to the most recent common ancestor (TMRCA) between any randomly chosen pair of individuals in the population. Smaller population sizes would result in more recent TMRCAs. A population bottleneck would cause many loci to coalesce during the bottleneck (a high coalescence rate corresponding to a low population size), while periods of large population size would result in a low coalescence rate. The TMRCA can be estimated by knowing the level of IBD sharing between two individuals, as it gives a direct readout of ancestry. Since shorter IBD segments correspond to longer TMRCAs and longer IBD segments correspond to shorter TMRCAs, differences in the ancestral (old) population sizes have an effect on the overall level of sharing of short IBD segments between individuals, whereas more recent population size changes affect the level of sharing of long IBD segments (Gusev et al. 2012). Other demographic features, like increased probability of mating within a particular community, will increase IBD over baseline levels. Population expansions will lead to a faster exponential drop-off in the distribution of lengths of shared IBD segments, due to an increasing effective population size. However, IBD sharing can be higher than expected despite overall population expansion if there is population substructure (demes) with high levels of intrapopulation migration. In essence, the high level of sharing within the smaller demes (smaller effective population size) is propagated to neighboring demes to increase overall sharing.

Geographic proximity and population substructure are more likely to result in higher IBD. For example, Ralph & Coop (2013) studied IBD sharing within and between European human populations to look for trends in geography and found that individuals from neighboring populations share considerable numbers of common ancestors (2–12) within the last 1,500 years, showing the high level of relatedness among Europeans.

To use IBD patterns for demographic inference, one first needs to define the tracts from the genetic variation data. This can be done using dense genotype data or WGS data. Current methods to call IBD tracts tend to perform better at detecting longer IBD tracks (>2 cM) and shorter tracts may be more likely to be called in error (Chiang et al. 2016). IBD tracks provide valuable information about common genetic ancestry and recent evolutionary events because IBD tracks are broken up over time due to recombination.

There are methods to exploit long-range IBD information shared between individuals and infer recent demographic events using large sample sizes. For example, Palamara et al. (2012) examined the relationship between the distribution of shared IBD segment lengths and demographic history and showed that they can infer demographic events occurring as recently as tens of generations ago. Intuitively, more recent common ancestry is associated with a smaller population size and would result in a higher fraction of longer IBD segments between individuals. They inferred two expansion events in an Ashkenazi Jewish population interspersed with a founder event.

2.3.2. Identity by state. Unlike IBD, the related IBS measure is computed directly from the observed data, making it easier to calculate but more prone to sequencing or phasing errors. In one application, Harris & Nielsen (2013) inferred complex demographic events such as divergence times, admixture times, and population size changes from the distribution of the lengths of IBS haplotypes between pairs of chromosomes. One advantage of using IBS instead of IBD is that IBS is directly observable, while IBD needs to be inferred, and IBS can infer both ancient and recent demographic events. The IBS-based method works on phased WGS data and has been successfully applied to human data sets from the 1000 Genomes Project, Native American genomes, and sequence data from polar bears (Harris & Nielsen 2013, Liu et al. 2014, Raghavan et al. 2015).

To demonstrate the value of demographic inference from IBS distributions, Harris & Nielsen (2013) use coalescent simulations (and their approximation formulas) to show that the IBS

distribution is able to accurately determine admixture times between two populations, with more recent admixture leading to an increase in longer shared IBS tracts, as expected. The advantage of IBS over the SFS is that IBS incorporates information from linkage between sites, and unlike most methods based on the hidden Markov model [e.g., pairwise sequentially Markovian coalescent (PSMC), discussed below], it can fit more complex demographic scenarios and use larger samples sizes.

2.4. Sequentially Markovian Coalescent Methods

The distribution of hetero- and homozygous sites across the genome contains information about the set of genealogies that gave rise to the genome. Loci containing a large proportion of homozygous sites are inferred to have recent coalescent times, while regions with many heterozygous sites are inferred to have more ancient coalescence times. The resulting distribution of coalescent times inferred across even a single individual's genome can provide insight into coalescence rates through time, and more indirectly, past demographic events.

The PSMC (Li & Durbin 2011) leverages this information to infer changes in N_e through time from a single genome. PSMC requires WGS data from at least one individual. Sequencing reads must be mapped to a reference genome and filtered for high-confidence sites. PSMC uses the presence of SNPs across 100-bp windows of the genome to infer the distribution of coalescent times using a hidden Markov model (with the TMRCAs as the hidden states, recombination events as the transitions, and mutations as the emissions; see **Supplemental Figure 3*a***), and ultimately, the coalescence rate over a given number of time segments. The inverse of the coalescence rate can be rescaled by the user's choice of mutation rate and generation time, to yield a trajectory of N_e over time (see **Supplemental Figure 3*b***).

Supplemental Material >

An extension of the method, the multiple sequentially Markovian coalescent (MSMC) (Schiffels & Durbin 2014), uses the improved sequentially Markov coalescent (SMC′) algorithm (Marjoram & Wall 2006), which enables more accurate recombination rate modeling, to infer the coalescence rate from the most recent time to coalescence between any two haplotypes across multiple genomes (up to eight phased haplotypes). The goal of using multiple genomes is to better infer recent demography. By including more genomes, there are more recent coalescent events that provide more information about recent population sizes.

PSMC and MSMC are staples of de novo genome papers and a reasonable option for an initial query into the history of coalescence rates for a population. However, caution should be used when interpreting the resulting demographic trajectories. Changes in gene flow can yield trajectories that mimic changes in population size, and PSMC/MSMC demographic histories do not always predict other summaries of the data, such as the SFS (Beichman et al. 2017, Chikhi et al. 2018, Mazet et al. 2016, Orozco-terWengel 2016).

Newer methods continue to leverage the distribution of coalescence rates, while incorporating other summaries of genetic variation. Minimal assumption genomic inference of coalescence (MAGIC) (Weissman & Hallatschek 2017) uses a range of window sizes across the genome to infer the single-locus distribution of coalescent times from any number of genomes, without explicitly having to model recombination. MAGIC enables users to use simulations to test the resulting models against empirical summaries of the coalescent process. This allows the coalescent time distribution to be used as a summary of the data rather than a literal record of demographic history. Another inference method, called SMC^{++} (Terhorst et al. 2017), incorporates the SFS based on hundreds of genomes with the distribution of coalescent times from a single genome. This enables information from many genomes to be incorporated without adding a major computational burden. While MAGIC and SMC^{++} can be run on a single genome, their real promise is in the

incorporation of hundreds of WGSs into demographic inference. While numerous WGSs may not yet be a reality for many nonmodel organisms, these approaches will become more accessible over time.

2.5. Generalized Phylogenetic Coalescent Sampler

Unlike the previously discussed approaches, a program called the generalized phylogenetic coalescent sampler (G-PhoCS) uses the full sequence data from multiple individuals for demographic inference (Gronau et al. 2011). G-PhoCS is typically run on many independent, short genomic fragments scattered throughout the genome, which could be generated from whole-genome, sequence capture, or RAD-Seq approaches. It is assumed that there is no recombination within each fragment. Several individuals are required for inference. G-PhoCS gains its ability to infer parameters from the mutational patterns at many independent loci across the genome. Because it uses the maximum amount of information possible from the available data, this approach is ideal from a statistical inference perspective.

The G-PhoCS method is a significant extension of previous full-likelihood programs for genomic data (Burgess & Yang 2008, Rannala & Yang 2003). G-PhoCS performs inference under a multipopulation demographic model whose parameters consist of population split time(s), relative population sizes, and migration bands (i.e., the migration rates between populations at specific time points). Genealogies are generated under particular demographic models, and then, conditional on that genealogy, the probability of the sequence data for each fragment is computed. Importantly, most genealogies have a very low probability of generating the observed sequence data. Thus, G-PhoCS does not sample genealogies at random but instead uses Metropolis–Hastings sampling in a Markov chain Monte Carlo approach to preferentially sample genealogies that are likely to be compatible with the data, increasing efficiency of the inference. G-PhoCS provides posterior distributions of the demographic parameters of interest.

An advantage of G-PhoCS is that it makes use of the full information from the data. Additionally, it is designed for inference under complex demographic models, such as multipopulation models with migration. The major disadvantage of G-PhoCS is that it is computationally demanding, often requiring weeks to run. Additionally, it is less able to infer recent population size changes compared to other approaches, which may be important, depending on the goal of the inference.

3. DEMOGRAPHIC INFERENCE IN NONMODEL SPECIES

3.1. Methods Based on the Site Frequency Spectrum

The SFS based on whole-genome resequencing has been used in numerous studies to infer the demographic history of nonhuman organisms, from model species (Garud et al. 2015), to wild populations (Zhan et al. 2014, Zhao et al. 2013), to domesticated animals and crops (Lam et al. 2010, Lin et al. 2014, Qi et al. 2013, Shapiro et al. 2013, Wang et al. 2016).

Douglas et al. (2015) used whole-genome sequencing and the SFS to infer the demographic history of several species in the plant genus *Capsella* and used the resulting demographic history as the basis for inferring the distribution of fitness effects and the impacts of polyploidy on selection. *Capsella bursa-pastoris* (shepherd's purse) is a tetraploid species. Douglas et al. (2015) used whole-genome sequencing of the *C. bursa-pastoris* genome and comparison with its diploid relatives to determine that it has a hybrid origin from the ancestors of two modern diploid *Capsella* lineages with more restricted ranges. The authors then sequenced 9–13 additional genomes from each of

the three species and built a joint SFS. The fastsimcoal2 software (Excoffier et al. 2013) was used to infer demographic parameters for four different models of speciation, which included population size changes and different amounts of gene flow. They determined that a model of exponential growth with no gene flow was the best fitting model and the two diploid *Capsella* lineages had an ancient divergence ~900 kya, with the hybrid *C. bursa-pastoris* occurring much more recently (100–300 kya).

RAD-Seq and other genotyping-by-sequencing (GBS) approaches allow researchers to affordably sequence many individuals in a population by generating sequences of short segments of DNA from across the genome (Andrews et al. 2016). While these segments may not be useful for studies of selection or gene function, they are a good way to build an SFS based on many individuals for use in demographic inference (Egger et al. 2017, Filatov et al. 2016, Harvey et al. 2016, Kautt et al. 2016, Lanier et al. 2015, Le Moan et al. 2016, Lipshutz et al. 2017, Sovic et al. 2016, Tine et al. 2014, Trucchi et al. 2014). Sequence capture can also be used to generate enough data to build an SFS without having to resequence whole genomes (Crawford et al. 2017, Harvey et al. 2016, Laurent et al. 2016, Singh et al. 2013).

Laurent et al. (2016) were interested in differential coloration between two species of White Sands lizards (*Sceloporus cowlesi* and *Aspidoscelis inornata*). In both species, lizards that live on pale-colored sand at White Sands National Monument, New Mexico, have an adaptive pale coloration, while those living elsewhere across the range of the two species live on dark sands and have dark coloration. To explore the demographic history of these populations and detect a signature of selection underlying this phenotype, the authors sampled 24 individuals from each species across three populations (one white sand and two dark sand sites). They used a custom sequence capture approach to sequence the region around the candidate coloration gene *Mc1r* in addition to hundreds of loci scattered throughout the genome. For each species, they generated a joint SFS between the populations and inferred demographic parameters for six models using both fastsimcoal2 (Excoffier et al. 2013) and δaδi (Gutenkunst et al. 2009). In both species, they found recent divergence between the light soil and dark soil populations, which tallies with the geological history of the region (the light soil is the result of recent gypsum deposits 7,000 ya), with different light-soil colonization times between the species.

RNA-Seq is another popular way to generate large amounts of data inexpensively. However, RNA-Seq also has numerous pitfalls, including issues with sample preservation and quality, the ability to call SNPs (Han et al. 2015), and a potential for linked selection to confound demographic inference (Gazave et al. 2014, McCoy et al. 2014, Schrider et al. 2016). Nevertheless, RNA-Seq data have been used to generate the SFS and infer demographic history in a number of species (Combosch et al. 2017, Edwards et al. 2016, McCoy et al. 2014, Stuglik & Babik 2016). McCoy et al. (2014) tested the viability of a joint SFS based on SNPs derived from RNA-Seq data for demographic inference in δaδi (Gutenkunst et al. 2009) by comparing the recorded history of the introduced Gillette's checkerspot butterfly (*Euphydryas gillettii*) population to their inferred demographic parameters. The population parameters they estimated were concordant with the known history of the founding of the population, although they still recommend RAD-Seq as the preferable option for demographic studies since RAD-Seq is less likely to be confounded by linked selection in genic regions.

3.2. Approximate Bayesian Computation

ABC approaches were originally used on limited loci: a handful of nuclear loci, mitochondrial DNA, or microsatellites (Fagundes et al. 2007, Pritchard et al. 1999, Thornton & Andolfatto

2006). Now, whole-genome sequencing yields genome-wide summary statistics that can be used for the inference of demographic parameters in an ABC framework (Boitard et al. 2016, Nater et al. 2017, Prado-Martinez et al. 2013, J.A. Robinson et al. 2016, J.D. Robinson et al. 2014). Nater et al. (2017) sequenced 37 whole genomes to identify a new orangutan species (*Pongo tapanuliensis*) in Sumatra, south of Lake Toba. They used the ABC approach of Beaumont et al. (2002) to infer demographic parameters using numerous summary statistics, including Watterson's estimate of the population mutation rate (θ), nucleotide diversity (π), Tajima's *D*, average sequence divergence, and others, calculated within and between populations from putatively neutral loci throughout the genome. The authors further explored their results using MSMC (Schiffels & Durbin 2014) and G-PhoCS (Gronau et al. 2011). The resulting best-fit model had three deep lineages of orangutan and indicated that the fragile population (<800 individuals) south of Lake Toba is a distinct orangutan lineage, descended from an ancestral population that was the source for the northern Sumatra and Borneo orangutan species.

Sequence capture and GBS lend themselves to ABC inference as summary statistics and can be generated for a reduced portion of the genome, making them useful approaches for surveying natural and domesticated populations lacking in large-scale sequencing resources (Combosch et al. 2017, Cornille et al. 2016, Shafer et al. 2015). Shafer et al. (2015) explored the utility of GBS data for ABC demographic inference using extensive simulations to develop a set of best-practice recommendations. They implemented their best practices to infer the demographic history of high Arctic and central Arctic Atlantic walrus (*Odobenus rosmarus rosmarus*) populations based on GBS sequencing of 30 walrus individuals. They inferred divergence between the walrus populations around the time of the last glacial maximum, with subsequent population declines and asymmetric migration of individuals moving from the high Arctic population to the central Arctic. The population decline they detected was not found by a previous microsatellite study (Andersen et al. 2009), indicating an improvement using GBS data, although they also acknowledged that ABC is underpowered to detect bottlenecks, especially bottlenecks followed by recovery.

RNA-Seq data can also be used to generate the summary statistics necessary for ABC demographic inference. Nabholz et al. (2014) carried out RNA-Seq on 9 domesticated (*Oryza glaberrima*) and 10 wild African rice (*Oryza barthii*) plants and aligned them to the Asian rice (*Oryza sativa*) reference genome. They used a simple two-population demographic model to estimate the timing and magnitude of the African rice domestication bottleneck. Using the number of SNPs, Tajima's *D*, and F_{ST} as summary statistics, they found that the domestication bottleneck in African rice was severe, resulting in a reduction of 98% of the ancestral population, and twice as strong as that estimated in maize by Wright et al. (2005).

3.3. Identity by State/Identity by Descent Tracts

IBS/IBD tract methods have not been as widely employed in demographic inference of nonmodel species compared to other methods since they require many WGSs or dense genotype data. However, in a notable use of the approach, Liu et al. (2014) used IBS tract information to infer the demographic history of polar bears (*Ursus maritimus*) and brown bears (*Ursus arctos*). They generated high-coverage genome sequences from 18 polar bears and 10 brown bears, plus an additional 61 low-coverage polar bear genomic sequences. They then used the IBS tract approach of Harris & Nielsen (2013) to infer demographic parameters, including levels of gene flow, for both species. They found the divergence of polar bears to be more recent than that inferred by other studies, only 343–479 kya, which they credit to inferring a complex demographic history with divergence

rather than a basic isolation–migration model that did not incorporate fluctuations in population size. They demonstrated that a simple model cannot recapitulate IBS tract patterns in the data, whereas their best-fit model accurately predicts the distribution of IBS tracts in both species.

3.4. Pairwise Sequentially Markovian Coalescent

PSMC and MSMC have been widely used for the study of nonhuman species (Albert et al. 2013; Bosse et al. 2014; Fitak et al. 2016; Freedman et al. 2014; Groenen et al. 2012; Holliday et al. 2016; Hung et al. 2014; Ibarra-Laclette et al. 2013; Murray et al. 2017; Nadachowska-Brzyska et al. 2013, 2015, 2016; Prado-Martinez et al. 2013; Wang et al. 2016; Zhao et al. 2013). The relatively low barrier to entry (a single genome) allows demographic inference in many species for which other methods would be impossible, especially in extinct species with only a few historical or ancient genome sequences available (Hung et al. 2014, Murray et al. 2017, Palkopoulou et al. 2015, Prüfer et al. 2014, Schubert et al. 2014). Palkopoulou et al. (2015) sequenced two woolly mammoth (*Mammuthus primigenius*) genomes, one from ∼4 kya, the other from ∼45 kya, to relatively high coverage of 17× and 11×, respectively. They carried out PSMC analyses on both mammoth genomes and found that both individuals showed a concordant demographic history in the ancient past, with a steep decline in the more recent past in the trajectory of the less ancient sample (representing the demographic history after the death of the older mammoth individual).

3.5. Generalized Phylogenetic Coalescent Sampler

G-PhoCS has been used alone and in conjunction with other methods to infer demographic history from whole genomes (Carbone et al. 2014, Freedman et al. 2014, Hung et al. 2014, McManus et al. 2015, Wang et al. 2016) and RAD-Seq data (Harrington et al. 2018). Harrington et al. (2018) used both SFS-based inference in fastsimcoal2 and G-PhoCS on RAD-Seq data from 35 red diamond rattlesnake (*Crolatus ruber*) samples along the coast of Baja California. They found evidence for a phylogeographic break between populations that is concordant with existing subspecies taxonomy and morphological differences. To explain this break, they tested multiple models of divergence and admixture between populations using fastsimcoal2, and then compared the parameters estimated from the SFS to those estimated in G-PhoCS under an isolation–migration model. They found the best support for a model of isolation followed by secondary contact during the Pleistocene, with G-PhoCS and fastsimcoal2 differing somewhat on the timing of divergence and ancestral population size.

4. RECOMMENDATIONS

4.1. Workflow and Method Choice

Here, we provide some guidelines to decide which approach to demographic inference is most appropriate for the research question being asked. However, we do not address the question of selecting the appropriate demographic scenario to be input into the inference procedure. Choosing the appropriate demographic scenario (e.g., bottleneck, growth, population substructure, and combinations thereof) may not be trivial, since it depends on the biology and ecology of the organism, research question, and type of data that will be generated. Performing qualitative analysis for data exploration, such as principal component analysis (PCA) (Price et al. 2006) or the use of Admixture (Alexander et al. 2009) or TreeMix (Pickrell & Pritchard 2012), can provide useful intuition about population structure or admixture and therefore help narrow the choices

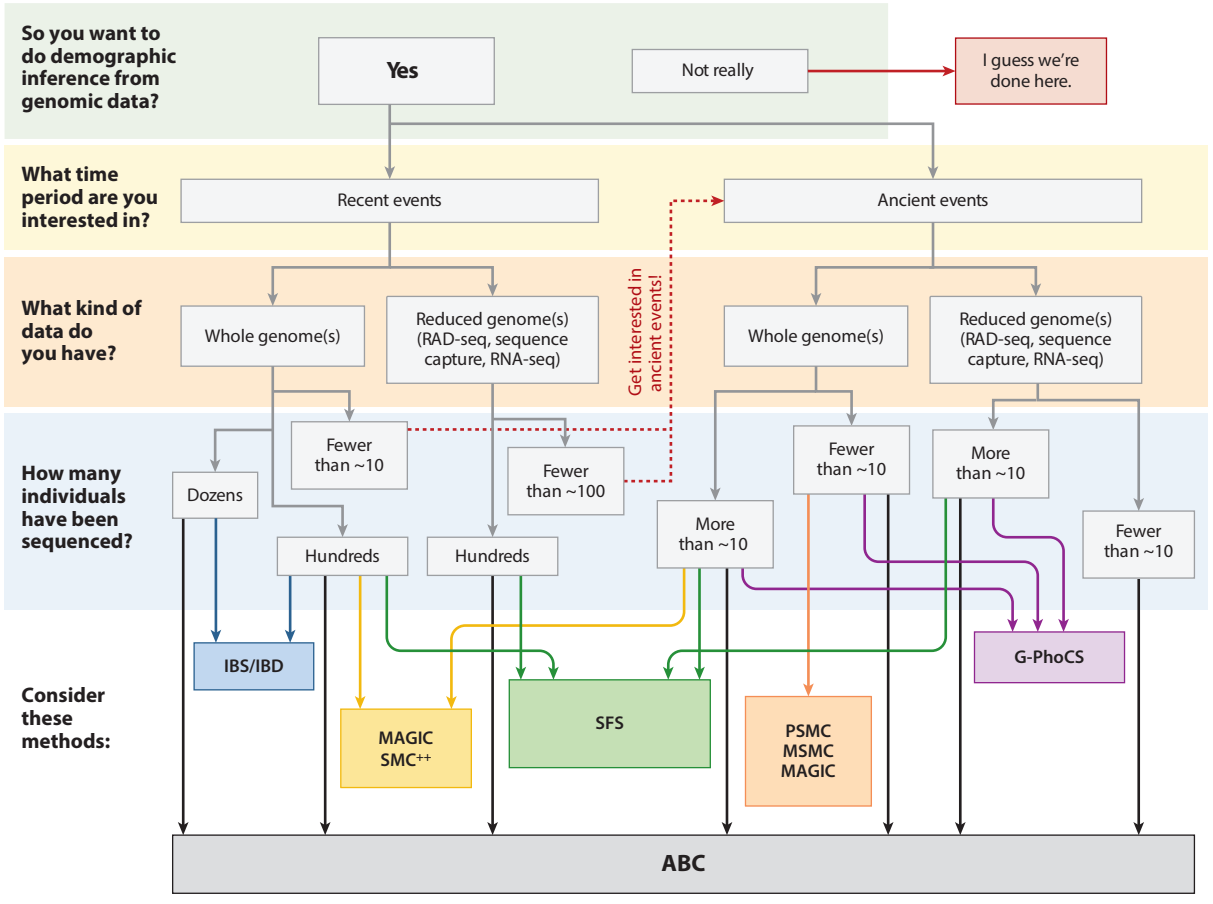

Figure 4

Decision tree to determine which demographic inference methods are appropriate for a given biological question and data. The choice of method depends on the time period the researcher is interested in (e.g., recent events occurring within the last ~100–1,000 generations of the population; ancient events happening in deeper time), the type of sequence data, and the sample size. Note that sample sizes are rough recommendations and not hard cutoffs. Abbreviations: ABC, approximate Bayesian computational inference methods; G-PhoCS, generalized phylogenetic coalescent sampler; IBD, identity by descent methods; IBS, identity by state methods; MAGIC, minimal assumption genomic inference of coalescence; P/MSMC, pairwise/multiple sequentially Markovian coalescent methods; RAD-Seq, restriction site associated DNA sequencing; SFS, site frequency spectrum–based inference methods; SMC^{++}, a sequentially Markovian coalescent method that also uses the site frequency spectrum.

of demographic models. However, while these are informative approaches to visualizing the data, they do not provide estimates of population demography and are sensitive to different models that can give rise to similar patterns in a PCA, STRUCTURE, or Admixture plot (DeGiorgio & Rosenberg 2013, Kalinowski 2011, Lawson et al. 2017, McVean 2009).

We summarize our guidelines for choosing methods of demographic inference in the decision tree shown in **Figure 4**. The tree first asks the researcher to consider what timeframe of events they are interested in, either recent events within the last several hundred generations of a population (such as human-driven bottlenecks that are often the focus of conservation geneticists) or more ancient history (such as ancient speciation or domestication events, response to climactic events,

or pre-exploitation dynamics). The choice of method, then, depends on the type of genomic data and how many individuals have been sequenced, but note that the sample sizes given in the decision tree are very loose guidelines; for instance, an SFS approach could still be used with only eight individuals. While not pictured in the decision tree, within each method, the particular software to use depends on further considerations, such as the computational resources, number of populations under consideration, and necessity for extensive model testing. We have listed several popular software implementations for each approach when discussing each method (summarized in **Supplemental Table 1**), but as new software packages are constantly being developed, this list is not exhaustive. Importantly, **Figure 4** can also be used to guide researchers on the type of data they should collect for their particular research questions. Below, we make several additional recommendations for demographic inference in nonmodel species.

Supplemental Material >

4.2. Data Quality

Figure 4 assumes that the researcher's sequence data are of sufficient coverage and quality to be used for demographic inference and, for most methods, that putatively neutral loci can be identified. The data quality requirements vary across methods.

The data requirements for SFS-based inference are flexible, and if the number of individuals is large, the coverage can be low, as we only need accurate population allele frequency estimates, not individual genotypes. Software tools such as analysis of next generation sequencing data (ANGSD) infer the SFS and account for the uncertainty due to several factors (e.g., difference in sequencing coverage, sequencing error rates). Using simulations, Nielsen et al. (2012) showed that even with 5 MB of data, 10 diploid individuals, and 1× coverage, their method leads to accurate inferences of the SFS with the method implemented in ANGSD. However, given higher levels of uncertainty in real data, we would recommend higher levels of coverage (>3×) for low numbers of samples (∼10 individuals), if possible. Note that some demographic scenarios such as rapid recent population growth (<200 generations in the past) create a signature of an excess of rare alleles in the SFS, and therefore large sample sizes (>1,000 individuals) would be required to detect these rare variants. The SFS inference framework also assumes independence of sites, which will cause difficulties for demographic inference if the variant data are all confined to genomic regions in high LD.

Nadachowska-Brzyska et al. (2016) found that genome-wide coverage, per site filtering, and the amount of missing data can all alter the shape of PSMC trajectories and therefore recommend requiring genome coverage of greater than 18× and less than 25% missing data to carry out PSMC inference, with per site filtering of more than 10 reads per site.

4.3. Natural Selection

All of the methods for demographic inference discussed herein assume that the loci being considered are neutrally evolving. However, some loci in the genome are also affected by natural selection. Selection can occur in a number of different ways (i.e., positive, purifying, balancing) and each type can affect both the functional mutations themselves and linked neutral variation. There is considerable evidence that neutral sites in the *Drosophila* and human genomes located close to genes and in regions of low recombination have been affected by selection (Enard et al. 2014, Hernandez et al. 2011, Lohmueller et al. 2011, Sella et al. 2009). While the effects of selection in many nonmodel species remain to be studied, it is likely that they have had at least some impact on patterns of genetic variation.

The presence of selection can adversely impact demographic inference. For example, purifying selection will result in a reduction in genetic variation combined with a skew toward low-frequency

variants. Such patterns can be interpreted as evidence of population bottlenecks and/or growth. In a rigorous simulation study, Schrider et al. (2016) examined the effect of selective sweeps on different methods of demographic inference. They found that sweeps can give false evidence of population size changes when using ABC and PSMC methods and that specific demographic parameter estimates may also be biased by sweeps. Further, they found that different inference methods are more sensitive to sweeps than others. Methods based on the SFS seemed to be slightly more robust than ABC or PSMC approaches, potentially because they use the average SFS across the genome rather than the full distribution of a statistic across the genome, which could contain outlier loci affected by selection. In another recent study, Murray et al. (2017) suggested that the passenger pigeon has experienced intense linked selection across the genome and that this resulted in PSMC plots not reflecting population size changes over time.

In order to mitigate the effects of selection, one should focus on neutrally evolving loci as much as possible. The effects of selection on demographic inference are likely to be most problematic in compact genomes, where it is difficult to pick sites far from genes. Avoiding sites located in or near exons is a good first step. Second, if recombination rate information is available, then sites in regions of high recombination should be used. Schrider et al. (2016) also recommend not using the variances in summary statistics in ABC approaches as variance may be inflated by linked selection. Lastly, it is our hope that methods that allow joint inference of demography and selection will become available in the future, which will enable more accurate inferences of both processes.

4.4. Limitations of Model Selection

The methods for demographic inference discussed herein excel at estimating the parameters of a particular demographic model from the data. However, these approaches are less suited to determining the particular type of demographic model (e.g., bottleneck versus structured population). In fact, assuming different types of demographic models can give vastly different parameter estimates and biological interpretations. For example, ABC approaches may falsely infer the presence of a bottleneck when fitting a population size change model if the true demography of the population involved population structure (Chikhi et al. 2010, Nielsen & Beaumont 2009), although newer approaches can explicitly test for this (Peter et al. 2010). Further, in the PSMC type framework, it has been shown that population structure can give false evidence of population size changes (Mazet et al. 2016). Unfortunately, there is no obvious way to determine the optimal type of demographic model for every situation. Prior biological information and exploratory data analyses offer some ideas of which models to try. Ideally, researchers should try fitting different types of plausible models and conduct parameter inference under these models. Further, as discussed below, it is imperative that researchers assess model fit, in particular, using features of the data not utilized in the inference process (Beichman et al. 2017, Gutenkunst et al. 2009, Thornton & Andolfatto 2006).

4.5. Combining Multiple Approaches

In the decision tree depicted in **Figure 4**, nearly every scenario yields multiple methods that could be tried, with ABC as a catchall that can be used in every situation. Since methods can be affected differently by unmodeled complexities of demography, we recommended comparing the results from multiple methods. While it can be a challenge to carry out more than one kind of demographic inference in nonhuman populations, many studies have employed multiple approaches to strengthen results (Carbone et al. 2014, Cornille et al. 2016, Freedman et al. 2014, Harrington et al. 2018, Hung et al. 2014, Liu et al. 2014, MacLeod et al. 2013, McManus et al.

2015, Prado-Martinez et al. 2013, Prates et al. 2016, Schubert et al. 2014, Shafer et al. 2015, Wang et al. 2016, Zhan et al. 2014).

Even if multiple demographic inference methods cannot be used, we recommend examining the fit of a demographic model to multiple summaries of the data whenever possible (see the approach of Beichman et al. 2017). Since features such as LD and the SFS can be affected differently by demographic events, combining the information contained by these different summaries of the genome has the potential to greatly improve demographic inference. New methods that combine several features of the data are highly promising (Boitard et al. 2016, Bunnefeld et al. 2015, Terhorst et al. 2017, Weissman & Hallatschek 2017), although few have been applied to nonhuman species as of yet.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

A.C.B. was supported by a National Science Foundation (NSF) Graduate Research Fellowship. E.H.S. was supported by NSF grant NSF-DEB-1557151. K.E.L. was supported by National Institutes of Health grant R35GM119856. The authors thank Dr. Kelsey Witt Dillon, Jesse Garcia, Bernard Kim, Chris Kyriazis, Dr. Clare Marsden, Jazlyn Mooney, Tanya Phung, Dr. Bradley Shaffer, and Dr. Ying Zhen for helpful comments on the manuscript.

LITERATURE CITED

Adams AM, Hudson RR. 2004. Maximum-likelihood estimation of demographic parameters using the frequency spectrum of unlinked single-nucleotide polymorphisms. *Genetics* 168(3):1699–712

Albert VA, Barbazuk WB, Der JP, Leebens-Mack J, Ma H, et al. 2013. The *Amborella* genome and the evolution of flowering plants. *Science* 342(6165):1241089

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19(9):1655–64

Andersen LW, Born EW, Doidge DW, Gjertz I, Wiig Ø, Waples RS. 2009. Genetic signals of historic and recent migration between sub-populations of Atlantic walrus *Odobenus rosmarus rosmarus* west and east of Greenland. *Endangered Species Res.* 9(3):197–211

Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA. 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. *Nat. Rev. Genet.* 17(2):81

Beaumont MA. 2010. Approximate Bayesian computation in evolution and ecology. *Annu. Rev. Ecol. Evol. Syst.* 41(1):379–406

Beaumont MA, Zhang W, Balding DJ. 2002. Approximate Bayesian computation in population genetics. *Genetics* 162(4):2025–35

Beichman AC, Phung TN, Lohmueller KE. 2017. Comparison of single genome and allele frequency data reveals discordant demographic histories. *G3* 7(11):3605–20

Bertorelle G, Benazzo A, Mona S. 2010. ABC as a flexible framework to estimate demography over space and time: some cons, many pros. *Mol. Ecol.* 19(13):2609–25

Bhaskar A, Wang YXR, Song YS. 2015. Efficient inference of population size histories and locus-specific mutation rates from large-sample genomic variation data. *Genome Res.* 25(2):268–79

Blum MGB, François O. 2010. Non-linear regression models for approximate Bayesian computation. *Stat. Comput.* 20(1):63–73

Boitard S, Rodríguez W, Jay F, Mona S, Austerlitz F. 2016. Inferring population size history from large samples of genome-wide molecular data—an approximate Bayesian computation approach. *PLOS Genet.* 12(3):e1005877

Bosse M, Megens H-J, Madsen O, Frantz LA, Paudel Y, et al. 2014. Untangling the hybrid nature of modern pig genomes: a mosaic derived from biogeographically distinct and highly divergent *Sus scrofa* populations. *Mol. Ecol.* 23(16):4089–102

Browning BL, Browning SR. 2013a. Detecting identity by descent and estimating genotype error rates in sequence data. *Am. J. Hum. Genet.* 93(5):840–51

Browning BL, Browning SR. 2013b. Improving the accuracy and efficiency of identity-by-descent detection in population data. *Genetics* 194(2):459–71

Browning SR, Browning BL. 2015. Accurate non-parametric estimation of recent effective population size from segments of identity by descent. *Am. J. Hum. Genet.* 97(3):404–18

Bunnefeld L, Frantz LA, Lohse K. 2015. Inferring bottlenecks from genome-wide samples of short sequence blocks. *Genetics* 201(3):1157–69

Burgess R, Yang Z. 2008. Estimation of hominoid ancestral population sizes under Bayesian coalescent models incorporating mutation rate variation and sequencing errors. *Mol. Biol. Evol.* 25(9):1979–94

Carbone L, Harris RA, Gnerre S, Veeramah KR, Lorente-Galdos B, et al. 2014. Gibbon genome and the fast karyotype evolution of small apes. *Nature* 513(7517):195

Chiang CWK, Ralph P, Novembre J. 2016. Conflation of short identity-by-descent segments bias their inferred length distribution. *G3* 6(5):1287–96

Chikhi L, Rodríguez W, Grusea S, Santos P, Boitard S, Mazet O. 2018. The IICR (inverse instantaneous coalescence rate) as a summary of genomic diversity: insights into demographic inference and model choice. *Heredity* 120(1):13

Chikhi L, Sousa VC, Luisi P, Goossens B, Beaumont MA. 2010. The confounding effects of population structure, genetic diversity and the sampling scheme on the detection and quantification of population size changes. *Genetics* 186(3):983–95

Clark AG, Hubisz MJ, Bustamante CD, Williamson SH, Nielsen R. 2005. Ascertainment bias in studies of human genome-wide polymorphism. *Genome Res.* 15(11):1496–1502

Coffman AJ, Hsieh PH, Gravel S, Gutenkunst RN. 2016. Computationally efficient composite likelihood statistics for demographic inference. *Mol. Biol. Evol.* 33(2):591–93

Combosch DJ, Lemer S, Ward PD, Landman NH, Giribet G. 2017. Genomic signatures of evolution in *Nautilus*—an endangered living fossil. *Mol. Ecol.* 26(21):5923–38

Cornille A, Salcedo A, Kryvokhyzha D, Glémin S, Holm K, et al. 2016. Genomic signature of successful colonization of Eurasia by the allopolyploid shepherd's purse (*Capsella bursa-pastoris*). *Mol. Ecol.* 25(2):616–29

Cornuet J-M, Pudlo P, Veyssier J, Dehne-Garcia A, Gautier M, 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(8):1187–89

Cornuet J-M, Santos F, Beaumont MA, Robert CP, Marin J-M, et al. 2008. Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. *Bioinformatics* 24(23):2713–19

Crawford JE, Alves JM, Palmer WJ, Day JP, Sylla M, et al. 2017. Population genomics reveals that an anthropophilic population of *Aedes aegypti* mosquitoes in West Africa recently gave rise to American and Asian populations of this major disease vector. *BMC Biol.* 15:16

Csilléry K, Blum MGB, Gaggiotti OE, François O. 2010. Approximate Bayesian computation (ABC) in practice. *Trends Ecol. Evol.* 25(7):410–18

DeGiorgio M, Rosenberg NA. 2013. Geographic sampling scheme as a determinant of the major axis of genetic variation in principal components analysis. *Mol. Biol. Evol.* 30(2):480–88

Douglas GM, Gos G, Steige KA, Salcedo A, Holm K, et al. 2015. Hybrid origins and the earliest stages of diploidization in the highly successful recent polyploid *Capsella bursa-pastoris*. *PNAS* 112(9):2806–11

Durand EY, Eriksson N, McLean CY. 2014. Reducing pervasive false-positive identical-by-descent segments detected by large-scale pedigree analysis. *Mol. Biol. Evol.* 31(8):2212–22

Edwards T, Tollis M, Hsieh P, Gutenkunst RN, Liu Z, et al. 2016. Assessing models of speciation under different biogeographic scenarios; an empirical study using multi-locus and RNA-seq analyses. *Ecol. Evol.* 6(2):379–96

Egger B, Roesti M, Böhne A, Roth O, Salzburger W. 2017. Demography and genome divergence of lake and stream populations of an East African cichlid fish. *Mol. Ecol.* 26(19):5016–30

Enard D, Messer PW, Petrov DA. 2014. Genome-wide signals of positive selection in human evolution. *Genome Res.* 24(6):885–95

Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M. 2013. Robust demographic inference from genomic and SNP data. *PLOS Genet.* 9(10):e1003905

Fagundes NJR, Ray N, Beaumont M, Neuenschwander S, Salzano FM, et al. 2007. Statistical evaluation of alternative models of human evolution. *PNAS* 104(45):17614–19

Filatov DA, Osborne OG, Papadopulos AS. 2016. Demographic history of speciation in a *Senecio* altitudinal hybrid zone on Mt. Etna. *Mol. Ecol.* 25(11):2467–81

Fitak RR, Mohandesan E, Corander J, Burger PA. 2016. The de novo genome assembly and annotation of a female domestic dromedary of North African origin. *Mol. Ecol. Resour.* 16(1):314–24

Freedman AH, Gronau I, Schweizer RM, Ortega-Del Vecchyo D, Han E, et al. 2014. Genome sequencing highlights the dynamic early history of dogs. *PLOS Genet.* 10(1):e1004016

Fu YX. 1995. Statistical properties of segregating sites. *Theor. Popul. Biol.* 48(2):172–97

Fu YX, Li WH. 1997. Estimating the age of the common ancestor of a sample of DNA sequences. *Mol. Biol. Evol.* 14(2):195–99

Garud NR, Messer PW, Buzbas EO, Petrov DA. 2015. Recent selective sweeps in North American *Drosophila melanogaster* show signatures of soft sweeps. *PLOS Genet.* 11(2):e1005004

Gazave E, Ma L, Chang D, Coventry A, Gao F, et al. 2014. Neutral genomic regions refine models of recent rapid human population growth. *PNAS* 111(2):757–62

Griffiths RC, Tavaré S. 1998. The age of a mutation in a general coalescent tree. *Stoch. Models* 14(1–2):273–95

Groenen MA, Archibald AL, Uenishi H, Tuggle CK, Takeuchi Y, et al. 2012. Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* 491(7424):393–98

Gronau I, Hubisz MJ, Gulko B, Danko CG, Siepel A. 2011. Bayesian inference of ancient human demography from individual genome sequences. *Nat. Genet.* 43(10):1031–34

Gusev A, Lowe JK, Stoffel M, Daly MJ, Altshuler D, et al. 2009. Whole population, genome-wide mapping of hidden relatedness. *Genome Res.* 19(2):318–26

Gusev A, Palamara PF, Aponte G, Zhuang Z, Darvasi A, et al. 2012. The architecture of long-range haplotypes shared within and across populations. *Mol. Biol. Evol.* 29(2):473–86

Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLOS Genet.* 5(10):e1000695

Han Y, Gao S, Muegge K, Zhang W, Zhou B. 2015. Advanced applications of RNA sequencing and challenges. *Bioinform. Biol. Insights* 9(Suppl. 1):29–46

Harrington SM, Hollingsworth BD, Higham TE, Reeder TW. 2018. Pleistocene climatic fluctuations drive isolation and secondary contact in the red diamond rattlesnake (*Crotalus ruber*) in Baja California. *J. Biogeogr.* 45(1):64–75

Harris K, Nielsen R. 2013. Inferring demographic history from a spectrum of shared haplotype lengths. *PLOS Genet.* 9(6):e1003521

Hartl DL, Moriyama EN, Sawyer SA. 1994. Selection intensity for codon bias. *Genetics* 138(1):227–34

Harvey MG, Smith BT, Glenn TC, Faircloth BC, Brumfield RT. 2016. Sequence capture versus restriction site associated DNA sequencing for shallow systematics. *Syst. Biol.* 65(5):910–24

Hernandez RD, Kelley JL, Elyashiv E, Melton SC, Auton A, et al. 2011. Classic selective sweeps were rare in recent human evolution. *Science* 331(6019):920–24

Holliday JA, Zhou L, Bawa R, Zhang M, Oubida RW. 2016. Evidence for extensive parallelism but divergent genomic architecture of adaptation along altitudinal and latitudinal gradients in *Populus trichocarpa*. *New Phytol.* 209(3):1240–51

Hung C-M, Shaner P-JL, Zink RM, Liu W-C, Chu T-C, et al. 2014. Drastic population fluctuations explain the rapid extinction of the passenger pigeon. *PNAS* 111(29):10636–41

Ibarra-Laclette E, Lyons E, Hernández-Guzmán G, Pérez-Torres CA, Carretero-Paulet L, et al. 2013. Architecture and evolution of a minute plant genome. *Nature* 498(7452):94–98

Joyce P, Marjoram P. 2008. Approximately sufficient statistics and Bayesian computation. *Stat. Appl. Genet. Mol. Biol.* 7(1). **https://doi.org/10.2202/1544-6115.1389**

Jung H, Marjoram P. 2011. Choice of summary statistic weights in approximate Bayesian computation. *Stat. Appl. Genet. Mol. Biol.* 10(1). **https://doi.org/10.2202/1544-6115.1586**

Kalinowski ST. 2011. The computer program STRUCTURE does not reliably identify the main genetic clusters within species: simulations and implications for human population structure. *Heredity* 106(4):625–32

Kautt AF, Machado-Schiaffino G, Torres-Dowdall J, Meyer A. 2016. Incipient sympatric speciation in Midas cichlid fish from the youngest and one of the smallest crater lakes in Nicaragua due to differential use of the benthic and limnetic habitats? *Ecol. Evol.* 6(15):5342–57

Keinan A, Mullikin JC, Patterson N, Reich D. 2007. Measurement of the human allele frequency spectrum demonstrates greater genetic drift in East Asians than in Europeans. *Nat. Genet.* 39(10):1251–55

Kimura M. 1964. Diffusion models in population genetics. *J. Appl. Probability* 1:177–232

Lam H-M, Xu X, Liu X, Chen W, Yang G, et al. 2010. Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nat. Genet.* 42(12):1053–59

Lanier HC, Massatti R, He Q, Olson LE, Knowles LL. 2015. Colonization from divergent ancestors: glaciation signatures on contemporary patterns of genomic variation in collared pikas (*Ochotona collaris*). *Mol. Ecol.* 24(14):3688–705

Laurent S, Pfeifer SP, Settles ML, Hunter SS, Hardwick KM, et al. 2016. The population genomics of rapid adaptation: disentangling signatures of selection and demography in White Sands lizards. *Mol. Ecol.* 25(1):306–23

Lawson D, van Dorp L, Falush D. 2017. A tutorial on how (not) to over-interpret STRUCTURE/ADMIXTURE bar plots. bioRxiv 066431. **https://doi.org/10.1101/066431**

Le Moan A, Gagnaire P-A, Bonhomme F. 2016. Parallel genetic divergence among coastal-marine ecotype pairs of European anchovy explained by differential introgression after secondary contact. *Mol. Ecol.* 25(13):3187–202

Li H, Durbin R. 2011. Inference of human population history from individual whole-genome sequences. *Nature* 475(7357):493–96

Lin T, Zhu G, Zhang J, Xu X, Yu Q, et al. 2014. Genomic analyses provide insights into the history of tomato breeding. *Nat. Genet.* 46(11):1220–26

Lipshutz SE, Overcast IA, Hickerson MJ, Brumfield RT, Derryberry EP. 2017. Behavioural response to song and genetic divergence in two subspecies of white-crowned sparrows (*Zonotrichia leucophrys*). *Mol. Ecol.* 26(11):3011–27

Liu S, Lorenzen ED, Fumagalli M, Li B, Harris K, et al. 2014. Population genomics reveal recent speciation and rapid evolutionary adaptation in polar bears. *Cell* 157(4):785–94

Lohmueller KE, Albrechtsen A, Li Y, Kim SY, Korneliussen T, et al. 2011. Natural selection affects multiple aspects of genetic variation at putatively neutral sites across the human genome. *PLOS Genet.* 7(10):e1002326

Lohmueller KE, Bustamante CD, Clark AG. 2009. Methods for human demographic inference using haplotype patterns from genomewide single-nucleotide polymorphism data. *Genetics* 182(1):217–31

Lopes JS, Balding D, Beaumont MA. 2009. PopABC: a program to infer historical demographic parameters. *Bioinformatics* 25(20):2747–49

Lukic S, Hey J. 2012. Demographic inference using spectral methods on SNP data, with an analysis of the human out-of-Africa expansion. *Genetics* 192(2):619–39

MacLeod IM, Larkin DM, Lewin HA, Hayes BJ, Goddard ME. 2013. Inferring demography from runs of homozygosity in whole-genome sequence, with correction for sequence errors. *Mol. Biol. Evol.* 30(9):2209–23

Marjoram P. 2013. Approximation Bayesian computation. *OA Genet.* 1(3):853

Marjoram P, Molitor J, Plagnol V, Tavaré S. 2003. Markov chain Monte Carlo without likelihoods. *PNAS* 100(26):15324–28

Marjoram P, Wall JD. 2006. Fast "coalescent" simulation. *BMC Genet.* 7:16
Marth GT, Czabarka E, Murvai J, Sherry ST. 2004. The allele frequency spectrum in genome-wide human variation data reveals signals of differential demographic history in three large world populations. *Genetics* 166(1):351–72
Mazet O, Rodriguez W, Grusea S, Boitard S, Chikhi L. 2016. On the importance of being structured: instantaneous coalescence rates and human evolution—lessons for ancestral population size inference? *Heredity* 116(4):362–71
McCoy RC, Garud NR, Kelley JL, Boggs CL, Petrov DA. 2014. Genomic inference accurately predicts the timing and severity of a recent bottleneck in a non-model insect population. *Mol. Ecol.* 23(1):136–50
McManus KF, Kelley JL, Song S, Veeramah KR, Woerner AE, et al. 2015. Inference of gorilla demographic and selective history from whole-genome sequence data. *Mol. Biol. Evol.* 32(3):600–12
McVean G. 2009. A genealogical interpretation of principal components analysis. *PLOS Genet.* 5(10):e1000686
Moltke I, Albrechtsen A. 2014. RelateAdmix: a software tool for estimating relatedness between admixed individuals. *Bioinformatics* 30(7):1027–28
Murray GGR, Soares AER, Novak BJ, Schaefer NK, Cahill JA, et al. 2017. Natural selection shaped the rise and fall of passenger pigeon genomic diversity. *Science* 358(6365):951–54
Nabholz B, Sarah G, Sabot F, Ruiz M, Adam H, et al. 2014. Transcriptome population genomics reveals severe bottleneck and domestication cost in the African rice (*Oryza glaberrima*). *Mol. Ecol.* 23(9):2210–27
Nadachowska-Brzyska K, Burri R, Olason PI, Kawakami T, Smeds L, Ellegren H. 2013. Demographic divergence history of pied flycatcher and collared flycatcher inferred from whole-genome re-sequencing data. *PLOS Genet.* 9(11):e1003942
Nadachowska-Brzyska K, Burri R, Smeds L, Ellegren H. 2016. PSMC analysis of effective population sizes in molecular ecology and its application to black-and-white *Ficedula* flycatchers. *Mol. Ecol.* 25(5):1058–72
Nadachowska-Brzyska K, Li C, Smeds L, Zhang G, Ellegren H. 2015. Temporal dynamics of avian populations during Pleistocene revealed by whole-genome sequences. *Curr. Biol.* 25(10):1375–80
Nater A, Mattle-Greminger MP, Nurcahyo A, Nowak MG, de Manuel M, et al. 2017. Morphometric, behavioral, and genomic evidence for a new orangutan species. *Curr. Biol.* 27(22):3487–3498.e10
Nielsen R. 2000. Estimation of population parameters and recombination rates from single nucleotide polymorphisms. *Genetics* 154(2):931–42
Nielsen R, Beaumont MA. 2009. Statistical inferences in phylogeography. *Mol. Ecol.* 18(6):1034–47
Nielsen R, Korneliussen T, Albrechtsen A, Li Y, Wang J. 2012. SNP calling, genotype calling, and sample allele frequency estimation from new-generation sequencing data. *PLOS ONE* 7(7):e37558
Nunes MA, Balding DJ. 2010. On optimal selection of summary statistics for approximate Bayesian computation. *Stat. Appl. Genet. Mol. Biol.* 9(1). **https://doi.org/10.2202/1544-6115.1576**
Orozco-terWengel P. 2016. The devil is in the details: the effect of population structure on demographic inference. *Heredity* 116(4):349
Palamara PF, Lencz T, Darvasi A, Pe'er I. 2012. Length distributions of identity by descent reveal fine-scale demographic history. *Am. J. Hum. Genet.* 91(5):809–22
Palkopoulou E, Mallick S, Skoglund P, Enk J, Rohland N, et al. 2015. Complete genomes reveal signatures of demographic and genetic declines in the woolly mammoth. *Curr. Biol.* 25(10):1395–400
Peter BM, Wegmann D, Excoffier L. 2010. Distinguishing between population bottleneck and population subdivision by a Bayesian model choice procedure. *Mol. Ecol.* 19(21):4648–60
Pickrell JK, Pritchard JK. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. *PLOS Genet.* 8(11):e1002967
Plagnol V, Wall JD. 2006. Possible ancestral structure in human populations. *PLOS Genet.* 2(7):e105
Polanski A, Kimmel M. 2003. New explicit expressions for relative frequencies of single-nucleotide polymorphisms with application to statistical inference on population growth. *Genetics* 165(1):427–36
Polanski A, Kimmel M, Chakraborty R. 1998. Application of a time-dependent coalescence process for inferring the history of population size changes from DNA sequence data. *PNAS* 95(10):5456–61
Prado-Martinez J, Sudmant PH, Kidd JM, Li H, Kelley JL, et al. 2013. Great ape genetic diversity and population history. *Nature* 499(7459):471–75
Prates I, Xue AT, Brown JL, Alvarado-Serrano DF, Rodrigues MT, et al. 2016. Inferring responses to climate dynamics from historical demography in neotropical forest lizards. *PNAS* 113(29):7978–85

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 38(8):904
Pritchard JK, Seielstad MT, Perez-Lezaun A, Feldman MW. 1999. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Mol. Biol. Evol.* 16(12):1791–98
Prüfer K, Racimo F, Patterson N, Jay F, Sankararaman S, et al. 2014. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* 505(7481):43
Qi J, Liu X, Shen D, Miao H, Xie B, et al. 2013. A genomic variation map provides insights into the genetic basis of cucumber domestication and diversity. *Nat. Genet.* 45(12):1510–15
Raghavan M, Steinrücken M, Harris K, Schiffels S, Rasmussen S, et al. 2015. Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science* 349(6250):aab3884
Ralph P, Coop G. 2013. The geography of recent genetic ancestry across Europe. *PLOS Biol.* 11(5):e1001555
Ramstetter MD, Dyer TD, Lehman DM, Curran JE, Duggirala R, et al. 2017. Benchmarking relatedness inference methods with genome-wide data from thousands of relatives. *Genetics* 207(1):75–82
Rannala B, Yang Z. 2003. Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics* 164(4):1645–56
Robinson JA, Ortega-Del Vecchyo D, Fan Z, Kim BY, vonHoldt BM, et al. 2016. Genomic flatlining in the endangered island fox. *Curr. Biol.* 26(9):1183–89
Robinson JD, Bunnefeld L, Hearn J, Stone GN, Hickerson MJ. 2014. ABC inference of multi-population divergence with admixture from unphased population genomic data. *Mol. Ecol.* 23(18):4458–71
Schiffels S, Durbin R. 2014. Inferring human population size and separation history from multiple genome sequences. *Nat. Genet.* 46(8):919–25
Schrider DR, Shanku AG, Kern AD. 2016. Effects of linked selective sweeps on demographic inference and model selection. *Genetics* 204(3):1207–23
Schubert M, Jónsson H, Chang D, Der Sarkissian C, Ermini L, et al. 2014. Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *PNAS* 111(52):E5661–69
Sella G, Petrov DA, Przeworski M, Andolfatto P. 2009. Pervasive natural selection in the *Drosophila* genome? *PLOS Genet.* 5(6):e1000495
Shafer A, Gattepaille LM, Stewart RE, Wolf JB. 2015. Demographic inferences using short-read genomic data in an approximate Bayesian computation framework: in silico evaluation of power, biases and proof of concept in Atlantic walrus. *Mol. Ecol.* 24(2):328–45
Shapiro MD, Kronenberg Z, Li C, Domyan ET, Pan H, et al. 2013. Genomic diversity and evolution of the head crest in the rock pigeon. *Science* 339(6123):1063–67
Sheehan S, Song YS. 2016. Deep learning for population genetic inference. *PLOS Comput. Biol.* 12(3):e1004845
Singh ND, Jensen JD, Clark AG, Aquadro CF. 2013. Inferences of demography and selection in an African population of *Drosophila melanogaster*. *Genetics* 193(1):215–28
Sovic MG, Carstens BC, Gibbs HL. 2016. Genetic diversity in migratory bats: results from RADseq data for three tree bat species at an Ohio windfarm. *PeerJ* 4:e1647
Stuglik MT, Babik W. 2016. Genomic heterogeneity of historical gene flow between two species of newts inferred from transcriptome data. *Ecol. Evol.* 6(13):4513–25
Sunnåker M, Busetto AG, Numminen E, Corander J, Foll M, Dessimoz C. 2013. Approximate Bayesian computation. *PLOS Comput. Biol.* 9(1):e1002803
Tataru P, Nirody JA, Song YS. 2014. diCal-IBD: demography-aware inference of identity-by-descent tracts in unrelated individuals. *Bioinformatics* 30(23):3430–31
Tavaré S, Balding DJ, Griffiths RC, Donnelly P. 1997. Inferring coalescence times from DNA sequence data. *Genetics* 145(2):505–18
Terhorst J, Kamm JA, Song YS. 2017. Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nat. Genet.* 49(2):303–9
Thornton K, Andolfatto P. 2006. Approximate Bayesian inference reveals evidence for a recent, severe bottleneck in a Netherlands population of *Drosophila melanogaster*. *Genetics* 172(3):1607–19
Tine M, Kuhl H, Gagnaire P-A, Louro B, Desmarais E, et al. 2014. European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nat. Commun.* 5:5770
Trucchi E, Gratton P, Whittington JD, Cristofari R, Maho YL, et al. 2014. King penguin demography since the last glaciation inferred from genome-wide data. *Proc. R. Soc. B.* 281(1787):20140528

Wang G-D, Zhai W, Yang H-C, Wang L, Zhong L, et al. 2016. Out of southern East Asia: the natural history of domestic dogs across the world. *Cell Res.* 26(1):21–33

Wegmann D, Leuenberger C, Neuenschwander S, Excoffier L. 2010. ABCtoolbox: a versatile toolkit for approximate Bayesian computations. *BMC Bioinform.* 11:116

Weiss G, von Haeseler A. 1998. Inference of population history using a likelihood approach. *Genetics* 149(3):1539–46

Weissman DB, Hallatschek O. 2017. Minimal-assumption inference from population-genomic data. *eLife* 6:e24836

Wiuf C. 2006. Consistency of estimators of population scaled parameters using composite likelihood. *J. Math. Biol.* 53(5):821–41

Wollstein A, Lao O, Becker C, Brauer S, Trent RJ, et al. 2010. Demographic history of Oceania inferred from genome-wide data. *Curr. Biol.* 20(22):1983–92

Wright SI, Bi IV, Schroeder SG, Yamasaki M, Doebley JF, et al. 2005. The effects of artificial selection on the maize genome. *Science* 308(5726):1310–14

Zhan S, Zhang W, Niitepold K, Hsu J, Haeger JF, et al. 2014. The genetics of monarch butterfly migration and warning colouration. *Nature* 514(7522):317–21

Zhao S, Zheng P, Dong S, Zhan X, Wu Q, et al. 2013. Whole-genome sequencing of giant pandas provides insights into demographic history and local adaptation. *Nat. Genet.* 45(1):67–71

Annual Review of Ecology, Evolution, and Systematics

Volume 49, 2018

Contents

Indexes

Errata

An online log of corrections to *Annual Review of Ecology, Evolution, and Systematics* articles may be found at http://www.annualreviews.org/errata/ecolsys