Can the site-frequency spectrum distinguish exponential population growth from multiple-merger coalescents?

Bjarki Eldon[1,*], Matthias Birkner[2], Jochen Blath[1], Fabian Freund[3]

December 16, 2014

Author affiliations:

1. TU Berlin, Institut für Mathematik
10623 Berlin, Germany

2. JGU Mainz, Institut für Mathematik
55099 Mainz, Germany

3. University of Hohenheim, Institute of plant breeding, seed science, and population genetics
70599 Stuttgart, Germany

*: corresponding author

Running title: Multiple mergers and population growth

Keywords: coalescent, multiple mergers, population growth, approximate maximum likelihood test, approximate bayesian computation, site-frequency spectrum

corresponding author:
Bjarki Eldon
TU Berlin, Institut für Mathematik
Straße des 17. Juni 136
10623 Berlin, Germany
Email: eldon@math.tu-berlin.de
Phone: +49 30 314 25762
Fax: +49 30 314 21695

Abstract

The ability of the site-frequency spectrum (SFS) to reflect the particularities of gene genealogies exhibiting multiple mergers of ancestral lines as opposed to those obtained in the presence of population growth is our focus. An excess of singletons is a well-known characteristic of both population growth and multiple mergers. Other aspects of the SFS, in particular the weight of the right tail, are, however, affected in specific ways by the two model classes. Using an approximate likelihood method and minimum-distance statistics, our estimates of statistical power indicate that exponential and algebraic growth can indeed be distinguished from multiple merger coalescents, even for moderate sample size, if the number of segregating sites is high enough. A normalized version of the SFS is also used as a summary statistic in an approximate Bayesian computation (ABC) approach. The results give further positive evidence as to the general eligibility of the SFS to distinguish between the different histories.

Introduction

The site-frequency spectrum (SFS) at a given locus is one of the most important and popular statistics based on genetic data sampled from a natural population. In combination with the postulation of the assumptions of the infinitely-many sites mutation model (WATTERSON, 1975) and a suitable underlying coalescent framework, the SFS allows one to draw inference about evolutionary parameters, such as coalescent parameters associated with multiple-merger coalescents or population growth models.

The Kingman coalescent, developed by KINGMAN (1982a,b,c), HUDSON (1983a,b) and TAJIMA (1983), describing the random ancestral relations among DNA sequences drawn from natural populations, is a prominent and widely-used model from which one can make predictions about genetic diversity. Many quantities of interest, such as the expected values and covariances of the SFS associated with the Kingman coalescent, are easily computed thanks to results by FU (1995). The robustness of the Kingman coalescent is quite remarkable; indeed a large number of genealogy models can be shown to have the Kingman coalescent, or a variant thereof, as their limit process, cf. e.g. MÖHLE (1998). A large volume of work is thus devoted to inference methods based on the Kingman coalescent; see e.g. DONNELLY and TAVARÉ (1995), HUDSON (1990), NORDBORG (2001), HEIN *et al.* (2005) or WAKELEY (2007) for reviews.

However, many evolutionary histories can lead to significant deviations from the Kingman coalescent model. Such deviations can be detected using a variety of statistical tools, such as Tajima's D (TAJIMA, 1989a), Fu and Li's D (FU and LI, 1993) or Fay and Wu's H (FAY and WU, 2000), which are all functions of the SFS. However, they do not always allow to identify the actual evolutionary mechanisms leading to such deviations. Developing statistical tools that allow to distinguish between different evolutionary histories is, therefore, of fundamental importance.

The present work focuses on properties of the (folded and unfolded) SFS in the infinitely-many sites model for three population histories: (1) classical Kingman coalescent, (2) population growth, in particular exponential population growth, and (3) high fecundity coupled

with skewed offspring distributions (HFSOD), resulting in gene genealogies being described by so-called Lambda-coalescents (Sagitov, 1999; Pitman, 1999; Donnelly and Kurtz, 1999). Briefly, multiple merger coalescents may be more appropriate for organisms exhibiting HFSOD than the Kingman coalescent (cf. eg. Beckenbach, 1994; Árnason, 2004; Eldon and Wakeley, 2006; Sargsyan and Wakeley, 2008; Hedgecock and Pudovkin, 2011), see also a recent review by Tellier and Lemaire (2014).

Recent population growth as well as multiple-merger coalescents may lead to an excess of singletons in the SFS compared to the classical Kingman coalescent based SFS, which e.g. contributes to shifting Tajima (1989b)' D values to the negative. Indeed, Durrett and Schweinsberg (2005) prove that Tajima's D will be negative, at least for large sample size, under fairly general multiple-merger coalescents.

The associated genealogical trees are, however, qualitatively different. While moderate fluctuations in population size lead to a time-change of the Kingman coalescent (Kaj and Krone, 2003), multiple merger coalescents by definition change the topology of the genealogical tree. There is thus hope that each demographic effect leaves specific signatures in the resulting SFS, not only with respect to an excess of singletons, but for example also with respect to its right tail.

Indeed, one observes that the Kingman coalescent will not be a good match to genetic data containing a large fraction of singleton polymorphisms (relative to the total number of polymorphisms) due to a lack of free (coalescent) parameters, as opposed to multiple merger and population growth models, which both can predict an excess of singletons. Encouragingly, multiple merger and growth models exhibit noticeable differences in the bulk of the site-frequency spectrum, in particular in the lumped tail (see Figure 1; see also (Figs. 4, S2 Neher and Hallatschek, 2013)). In Figure 1, the normalised expected spectrum $\varphi_i^{(n,\Pi)}$ (see Eq. (2)) for a given coalescent Π, ie. the expected spectrum scaled by the expected total number of segregating sites, is compared for different multiple-merger coalescents (`B` – Beta-coalescents; Schweinsberg, 2003), or (`D` – Dirac coalescents; Eldon and Wakeley, 2006), and exponential (`E`) and algebraic (`A`) growth models leading to time-changed King-

man coalescents, for sample size (number of leaves) n as shown. Details for these coalescent models are recalled at the beginning of the Supporting Information. The first five classes (representing relative length of external branches, two-leaf branches, etc.) are shown, with classes from six onwards collected together (labelled as '5+'). In Figure 1, the relative external branch lengths were matched between the different coalescent processes. Even though the relative external branch lengths, and by implication the number of singletons relative to the total number of segregating sites, can be matched between the different processes, the collapsed tail (group '5+' in Figure 1) differs noticeably between the multiple-merger coalescents and the growth models. One also observes that the parameters have been chosen to match $\varphi_1^{(n,\Pi)}$ for $\Pi \in \{\mathtt{A}, \mathtt{D}, \mathtt{E}\}$ with $\varphi_1^{(n,\mathtt{B})}$ when $\alpha = 1$, where α is the coalescent parameter associated with B. Thus $\varphi_1^{(n,\mathtt{B})}$ is maximized for the given n (since $\alpha \in [1, 2]$), but $\varphi_1^{(n,\mathtt{D})}$, $\varphi_1^{(n,\mathtt{E})}$, and $\varphi_1^{(n,\mathtt{A})}$ can all increase by increasing the relevant parameters (ψ, β, or γ).

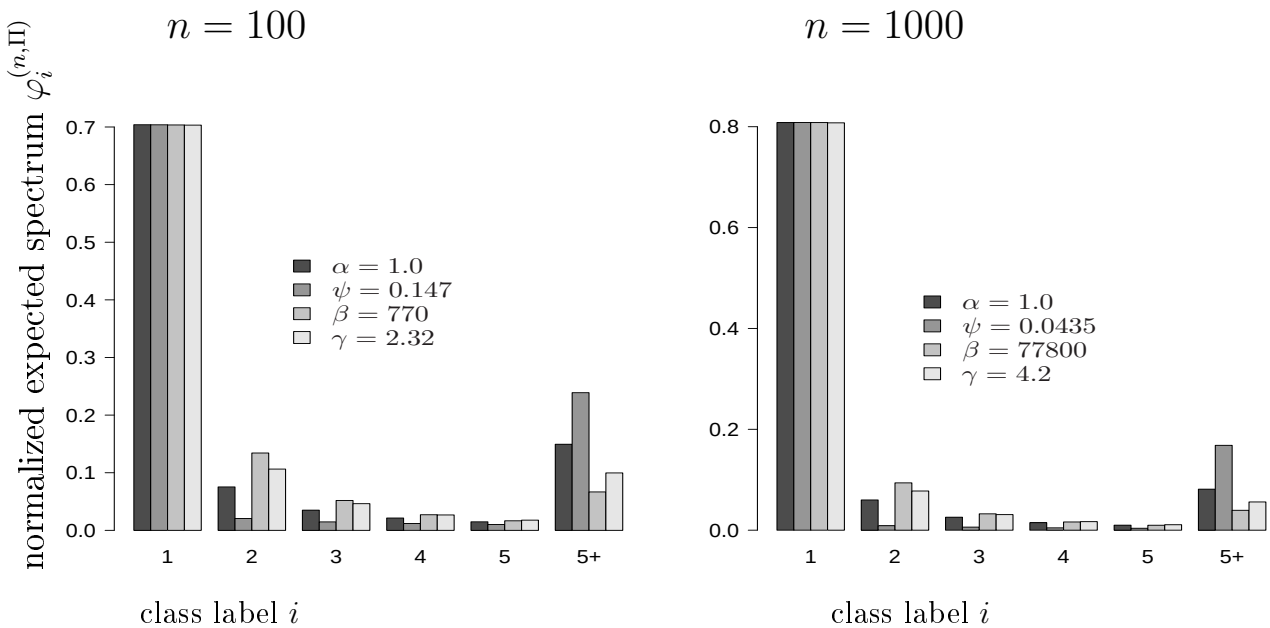
Matching the relative external branch lengths $\varphi_1^{(n,\Pi)}$ (see Eq. (2)) and observing how the rest of the normalised expected spectrum behaves, as illustrated in Figure 1, gives hope that multiple merger processes may be distinguished from (at least) particular population growth models with adequate statistical power. In the limit of large n, for the Kingman coalescent, $\varphi_1^{(n),\mathtt{K}} = O(1/\log(n))$.

Inference methods for distinguishing population growth from the usual Kingman coalescent have been extensively studied, see e.g. Tajima (1989a), Slatkin and Hudson (1991), Rogers and Harpending (1992), Kaj and Krone (2003) and Sano and Tachida (2005). Simulation-based work include Ramírez-Soriano *et al.* (2008), who consider the statistical power of several tests under population size increase and decrease, and the impact of recombination. Ramos-Onsins and Rozas (2002) consider the statistical power of statistics based on the site-frequency spectrum to distinguish deterministic population growth from the Kingman coalescent. On the theoretical side Myers *et al.* (2008), Bhaskar and Song (2014) and Kim *et al.* (2014) consider principal questions of identifiability of demographic histories. In particular, Bhaskar and Song (2014) show theoretically that complete knowledge of the SFS for large sample sizes carries enough information to fully recover demographic history

under mild assumptions on the possible fluctations of the demography.

Detecting multiple merger coalescents in populations deviating from the Kingman coalescent assumptions is a relatively new direction of research. Indeed, deriving inference methods based on multiple merger coalescents has only just begun (Eldon and Wakeley, 2006; Birkner and Blath, 2008; Eldon, 2011; Birkner *et al.*, 2011, 2013a,b; Steinrücken *et al.*, 2013; Koskela *et al.*, 2013; Rödelsperger *et al.*, 2014). In particular, Birkner *et al.* (2013b) obtain recursions for the expected site-frequency spectrum associated with Lambda-coalescents. In the present work we address the issue of distinguishing multiple merger coalescents from exponential population growth, by proposing statistical tests based on the (normalized) SFS, estimating statistical power for interval hypotheses via simulation. Since we can only work with approximate likelihood functions, and our methods, in particular the so-called 'fixed-S-method', can be sensitive to an (unknown) true coalescent mutation rate $\theta/2$, we complement our analysis by an approximate Bayesian computation approach (ABC; Rubin, 1984; Tavaré *et al.*, 1997; Pritchard *et al.*, 1999; Cucala and Marin, 2013; Baragatti and Pudlo, 2014).

Figure 1: Matching $\varphi_1^{(n),\Pi}$ (see Eq. (2)) for the different coalescent processes $\Pi \in \{\mathtt{A}, \mathtt{B}, \mathtt{D}, \mathtt{E}\}$, with number of leaves n as shown. Expected values were computed exactly. The processes and their associated parameters are algebraic growth ($\mathtt{A}$, γ); Beta$(2-\alpha, \alpha)$-coalescent ($\mathtt{B}$, α); Dirac coalescent ($\mathtt{D}$, ψ); exponential growth ($\mathtt{E}$, β). The values with label '5+' represent the collapsed tail $\sum_{i>5} \varphi_i^{(n,\Pi)}$.



Theory and Methods

Basic properties of the site-frequency spectrum

Consider a sample of n DNA sequences taken at a given genetic locus and assume that we can distinguish between derived (new mutations) and ancestral states. For $n \in \mathbb{N}$ let $[n] := \{1, \ldots, n\}$. We denote by $\xi_i^{(n)}$ the total number of sites at which the mutant base appears $i \in [n-1]$ times. Then,

$$\underline{\xi}^{(n)} := \left(\xi_1^{(n)}, \ldots, \xi_{n-1}^{(n)}\right)$$

is referred to as the *unfolded* site-frequency spectrum based on the n DNA sequences. If mutant and wild-type cannot be distinguished, one often considers the *folded* spectrum $\underline{\eta}^{(n)} := \left(\eta_1^{(n)}, \ldots, \eta_{\lfloor n/2 \rfloor}^{(n)}\right)$, where ancestral and derived states are not distinguished, and hence

$$\eta_i^{(n)} := \frac{\xi_i^{(n)} + \xi_{n-i}^{(n)}}{1 + \delta_{i,n-i}}, \quad 1 \leq i \leq \lfloor n/2 \rfloor,$$

(Fu, 1995), where $\delta_{i,j} = 1$ if $i = j$, and is zero otherwise. In this study, we will mostly be concerned with the unfolded site-frequency spectrum. Define $\zeta_i^{(n)} := \xi_i^{(n)}/|\xi^{(n)}|$ where $|\xi^{(n)}| := \xi_1^{(n)} + \cdots + \xi_{n-1}^{(n)}$ denotes the total number of segregating sites. Thus, $\underline{\zeta}^{(n)} = \left(\zeta_1^{(n)}, \ldots, \zeta_{n-1}^{(n)}\right)$ is the 'normalized' unfolded SFS, with the convention that $\underline{\zeta}^{(n)} = 0$ in the trivial case of complete absence of segregating sites ($|\xi^{(n)}| = 0$).

In order to compute expected values, variances and covariances of the SFS, an explicit underlying probabilistic model is needed. In the following we assume that the genealogy of a sample can be described by a coalescent process, more precisely by either (a timechange of) the Kingman coalescent or a multiple-merger coalescent. In addition, the infinitely-many-sites mutation model (Watterson, 1975) is assumed, and mutations are modeled by a Poisson-process on the coalescent branches with rate $\theta/2$. With this parametrization, the expected number of segregating sites in a sample of size 2, and hence the expected number of pairwise differences in a sample from the population, equals θ.

Closed-form expressions for the expected values and (co)variances of $\underline{\xi}^{(n)}$ have been determined in Fu (1995) when associated with the Kingman coalescent. One can represent the expected values of $\underline{\xi}^{(n)}$ in a unified way using the results of Griffiths and Tavaré (1998), Kaj and Krone (2003) and Birkner *et al.* (2013b), that allow to treat the expected values (and covariances) of the SFS for all coalescent models in question.

Let $\Pi^{(n)} = (\Pi_t^{(n)}, t \geq 0)$ be a (partition-valued exchangeable) coalescent process started from n leaves (partition blocks) corresponding to the random genealogy of a sample of size n. By discussing 'leaves' rather than DNA sequences we are emphasizing our viewpoint of the genealogy as a random graph, where the leaves are a particular kind of a vertex. Our interest is in the topology of the genealogy and how it is reflected in the associated site-frequency spectrum.

If the initial number of leaves is not specified, we simply speak of Π. One may think of Π as the Kingman coalescent, but the point is that the following result will stay true also for externally time-changed Kingman coalescents as well as multiple merger coalescents (a.k.a. 'Lambda'- or 'Xi'-coalescents in the mathematical literature), and even externally time-changed multiple merger coalescents.

Given n and a coalescent model Π, let $(Y_t^{(n)})_{t\geq 0}$ be the block counting process of the underlying coalescent $\Pi^{(n)}$ started from n lineages, i.e. $Y_t^{(n)}$ gives the number of ancestral lines (blocks) present/active at (backwards) time t. For $2 \leq k < n$, let $T_k^{(n)}$ be the random amount of time that $(Y_t^{(n)})_{t\geq 0}$ spends in state k. Given a coalescent $\Pi^{(n)}$ started from n (unlabelled) lineages, denote by $p^{(n),\Pi}[k,i]$ the probability that, *conditional* on the event that $Y_t^{(n)} = k$ for some time point t, a given one of the k blocks subtends exactly $i \in [n-1]$ leaves. A general representation of $\mathbb{E}^{\Pi,\theta}\left[\xi_i^{(n)}\right]$ is then

$$\mathbb{E}^{\Pi,\theta}\left[\xi_i^{(n)}\right] = \frac{\theta}{2}\sum_{k=2}^{n-i+1} p^{(n),\Pi}[k,i]\cdot k \cdot \mathbb{E}^{\Pi}\left[T_k^{(n)}\right], \quad i \in [n-1]. \tag{1}$$

The normalized expected SFS, $\varphi_i^{(n,\Pi)}$, for $i \in [n-1]$, is defined as

$$\varphi_i^{(n,\Pi)} = \frac{\sum_{k=2}^{n-i+1} p^{(n),\Pi}[k,i] \cdot k \cdot \mathbb{E}^{\Pi}\left[T_k^{(n)}\right]}{\sum_{\ell=2}^{n} \ell\, \mathbb{E}^{\Pi}\left[T_\ell^{(n)}\right]}. \tag{2}$$

where the denominator in (2) is the expected total tree length when starting from n leaves. One can interpret the quantity $\varphi_i^{(n,\Pi)}$ as the probability that a mutation, under the infinitely many sites assumption and the coalescent model Π, with known ancestral types, appears i times in a sample of size n. Importantly, $\varphi_i^{(n,\Pi)}$ is not a function of the mutation rate, unlike $\mathbb{E}^{\Pi,\theta}\left[\xi_i^{(n)}\right]$. One can also view $\varphi_i^{(n,\Pi)}$ as a first-order approximation of the expected value $\mathbb{E}^{\Pi,\theta}\left[\zeta_i^{(n)}\right]$ of the normalised SFS.

As examples for Π we will consider the classical Kingman coalescent (`K`), exponential (`E`) and algebraic (`A`) growth models, and the Beta$(2-\alpha,\alpha)$ (`B`) and the Dirac (`D`) multiple-merger coalescents, as recalled in the Supporting Information. Simulations suggest that $\varphi_i^{(n,\mathtt{B})}$ is a good approximation of $\mathbb{E}^{\mathtt{B},\theta}\left[\zeta_i^{(n)}\right]$ when α is not too close to 1, and n and θ are not too small (Birkner *et al.*, 2013b). Similar conclusions hold in the case of exponential and algebraic growth (results not shown).

One can use the recursive formulas obtained by Birkner *et al.* (2013b) to compute $\mathbb{E}^{\Pi,\theta}\left[\xi_i^{(n)}\right]$ associated with Lambda-coalescents. To compute $\varphi_i^{(n,\Pi)}$ associated with growth models we use the results of Polanski and Kimmel (2003), whose recursions are recalled in the Supporting Information.

A comparison of the observed $\zeta_i^{(n)}$ (instead of $\xi_i^{(n)}$) to an expected value $\mathbb{E}^{\Pi,\theta}\left[\zeta_i^{(n)}\right]$ – obtained under a particular coalescent model Π – enables one to do inference without having to jointly estimate the mutation rate θ using e.g. a minimum-distance statistic. Indeed, it appears that under any coalescent model Π, $\mathbb{E}^{\Pi,\theta}\left[\zeta_i^{(n)}\right]$ is almost constant as a function of the mutation rate θ (unless θ is very small); we provide some evidence for this in the Supporting Information, see (S12). Unfortunately, there seems to be no explicit way of representing $\mathbb{E}^{\Pi,\theta}\left[\zeta_i^{(n)}\right]$ as a simple function of the associated coalescent parameters and sample size n. As mentioned above, one may instead work with $\varphi_i^{(n,\Pi)}$.

Timescales, segregating sites and mutation rates

The choice of a multiple merger coalescent model (resp. demographic history) Π and its underlying parameters strongly affects classical estimates for the coalescent mutation rate $\theta/2$ (i.e. the Poisson rate at which mutations appear on coalescent branches). Assume w.l.o.g. for all multiple merger coalescents in question that the underlying coalescent measure Λ is always a probability measure: This normalisation fixes the coalescent time unit as the expected time to the most recent common ancestor of two individuals sampled uniformly from the population.

Given an observed number of segregating sites S in a sample of size n, a common estimate $\hat{\theta}^{\Pi}$ of the scaled mutation rate θ associated with coalescent model Π is the Watterson estimate

$$\hat{\theta}^{\Pi} := \frac{2S}{\mathbb{E}^{\Pi}\left[B^{(n)}\right]}, \tag{3}$$

where $\mathbb{E}^{\Pi}\left[B^{(n)}\right]$ is the expectation of the total tree length $B^{(n)}$ of an (n-) coalescent model Π. One can of course also estimate θ as a (different) linear combination of the site-frequency spectrum [(cf. Achaz, 2009) in the case of the Kingman coalescent]. Using the recursions for $\mathbb{E}^{\Pi,\theta}\left[\xi_i^{(n)}\right]$ obtained by Birkner *et al.* (2013b), one can also estimate θ, using either (3) or a linear combination of the expected SFS, in case of a Lambda-coalescent.

Given an estimate $\hat{\theta}^{\Pi}$, and knowledge of the mutation/substitution rate $\hat{\mu}$ per year at the locus under consideration, allows one to find a real-time embedding, of the coalescent history via the approximate identity

$$\text{coal. time unit} \times \frac{\hat{\theta}^{\Pi}}{2} \approx \text{year} \times \hat{\mu}. \tag{4}$$

cf. (Steinrücken *et al.*, 2013, Section 4.2), which of course depends on Π.

If one has additional information on the specific reproductive mechanisms of an approximating population model, this can even enable one to estimate the model census population size. For example, given a Cannings population model (Cannings, 1974, 1975) of fixed size N, let c_N be the probability that two gene copies, drawn uniformly at random and with-

out replacement from a population of size N, derive from a common parental gene copy in the previous generation. While for the usual haploid Wright-Fisher model $c_N = 1/N$, in a class of population models studied by SCHWEINSBERG (2003) leading to the Beta$(2-\alpha, \alpha)$-coalescent, c_N is proportional to $1/N^{\alpha-1}$, for $\alpha \in (1, 2]$ (but note that the proportionality constant depends on finer details of the particular model). By a limit theorem for Cannings models of MÖHLE and SAGITOV (2001), one coalescent time unit corresponds to approximately $1/c_N$ generations in the original model with population size N. Thus the mutation rate $\tilde{\mu}$ at the locus under consideration per individual per generation must be scaled with $1/c_N$ (as noted e.g. in ELDON and WAKELEY (2006)), and the relation between $\tilde{\mu}$, the coalescent mutation rate $\theta^{\Pi}/2$ and c_N is then given by the (approximate) identity

$$c_N \approx \frac{2\tilde{\mu}}{\theta^{\Pi}}. \tag{5}$$

In particular, if the Cannings model class (and thus c_N as a function of N) is known, N can be estimated via (5).

In this context, it is important to note that different population models on very different timescales can still have the Kingman coalescent as their ancestral limit process; two examples are the Wright-Fisher ($O\left(c_N^{-1}\right) = N$) and the Moran ($O\left(c_N^{-1}\right) = N^2$) models. This is certainly also the case for multiple merger coalescents. In particular, c_N is a priori *not* a function of the limiting coalescent model (this appears to be a rather frequent misperception).

Distinguishing real- and coalescent-timescales is important, because non-linear scaling may otherwise easily lead to confusion: For example, the expected total tree length $\mathbb{E}^{\mathsf{B}}[B^{(n)}]$ (measured in coalescent time units) *decreases* as a function of $\alpha \in (1, 2]$, while the corresponding quantity (measured in real-time generations) $\mathbb{E}^{\mathsf{B}}[B^{(n)}]/c_N$ *increases* in $\alpha \in (1, 2]$ (cf. Figure S1 in the supporting information).

The time-scaling applied to a classical Wright-Fisher model with *fluctuating* population size [as in KAJ and KRONE (2003)] in order to obtain a (time-changed) Kingman coalescent is recalled in the Supporting Information, see in particular Equation (S2). Again, the estimate (3) of θ depends on the growth model and growth parameter.

Approximate likelihood ratio tests for the SFS

Our aim is to construct a statistical test to distinguish between the model classes $\mathtt{E}$, $\mathtt{A}$, $\mathtt{D}$ and $\mathtt{B}$ (which intersect exactly in the Kingman coalescent $\mathtt{K}$). In order to distinguish, say, $\mathtt{E}$ from $\mathtt{B}$, based on an observed site-frequency spectrum $\underline{\xi}^{(n)}$ with sample size n and segregating sites $S = |\underline{\xi}^{(n)}|$, a natural approach is to construct a likelihood-ratio test.

Recall that we think of our observed spectrum as a realization of a coalescent tree with n leaves obtained from a coalescent model Π, with mutations distributed on the tree according to an independent Poisson($\theta/2$)-process. For each model Π from classes $\{\mathtt{E}, \mathtt{A}, \mathtt{D}, \mathtt{B}\}$, the coalescent will be uniquely determined by a single coalescent parameter $\beta \in [0, \infty)$ (for $\mathbb{E}$), $\gamma \in [0, \infty)$ (for $\mathtt{A}$), $\psi \in [0, 1]$ (for $\mathtt{D}$) and $\alpha \in [1, 2]$. (for $\mathtt{B}$)[1].

Suppose our null-hypothesis H_0 is presence of recent exponential population growth ($\mathtt{E}$) with (unknown) parameter $\beta \in [0, \infty)$, and we wish to test it against the alternative H_1 hypothesis of a multiple merger coalescent, say, the Beta$(2-\alpha, \alpha)$-coalescent ($\mathtt{B}$) for (unknown) $\alpha \in [1, 2]$, where $\beta = 0$ and $\alpha = 2$ correspond to the Kingman coalescent. In this framework, the coalescent mutation rate θ is not directly observable, but plays the role of a nuisance parameter. In particular, it is the interplay of the coalescent model Π and the mutation rate parameter θ that governs the law of the observed number of mutations, cf. the discussion in the previous section. To take θ explicitly into account, one could test

$$H_0 \,:\, (\Pi, \theta) \in \Theta^{\mathtt{E}} := \big\{(\beta, \theta) : \beta \in [0, \infty), \theta \in (0, \infty)\big\}$$

(exponential growth) against

$$H_1 \,:\, (\Pi, \theta) \in \Theta^{\mathtt{B}} := \big\{(\alpha, \theta) : \alpha \in [1, 2], \theta \in (0, \infty)\big\}$$

if the Beta-coalescent family is the alternative (by slight abuse of notation, we identify the coalescent model Π with the corresponding coalescent parameter β (resp. α) in each

[1] The beta-coalescent is also well-defined for $\alpha \in (0, 2]$, but we restrict to a smaller parameter range corresponding to the population model in Schweinsberg (2003).

model class when appropriate). The underlying parameter ranges are two-dimensional; and although an explicit likelihood-ratio test based on methods described in SIMONSEN *et al.* (1995) can be constructed, it will likely pose computational challenges.

Instead, given an observed number of segregating sites $S = s$, we simplify our framework by employing the 'fixed-s-method' discussed e.g. in DEPAULIS and VEUILLE (1998) and RAMOS-ONSINS and ROZAS (2002). Here we treat the observed number of segregating sites as a *fixed parameter* $s \in \mathbb{N}$, not as (observation of a) random variable S. We will thus obtain the empirical distributional quantities of our test by Monte-Carlo simulations, placing uniformly at random s mutations along the branches of the simulated tree.

The fixed-s-method is different from generating samples for a given θ by conditioning on $S = s$, yet the fixed-s-method usually leads to reasonable tests when the true θ is close to the Watterson estimate $\hat{\theta}(\Pi, s)$ based on s [WALL and HUDSON (2001)]. However, it can lead to substantial deviations from the condional distribution if θ is extreme (see e.g. MARKOVTSOVA *et al.* (2001), who show that for a test in a related framework, the probability of rejection can be substantially different from 5%). Regarding this caveat, DEPAULIS *et al.* (2001) take a Bayesian viewpoint and show that those values of θ that lead to unreliable tests are highly unlikely given s. We address this issue by using rejection sampling to check the robustness of the fixed-s-method against varying θ (see Supporting Information). Our analysis is complemented with an ABC approach using the normalized frequency spectrum $\underline{\zeta}^{(n)}$, which should be insensitive to the actual value of θ as long as θ is not too small, cf. (S12).

By fixing $S = s$ and treating it as a parameter of our test, we may consider the new pair of hypotheses

$$H_0^s \,:\, \Pi \in \Theta_s^{\mathtt{E}} := \big\{\beta : \beta \in [0, \infty)\big\}$$

and

$$H_1^s \,:\, \Pi \in \Theta_s^{\mathtt{B}} := \big\{\alpha : \alpha \in [1, 2]\big\}.$$

Define a likelihood-function $L(\Pi, \underline{k}^{(n)}, s)$ for the observed frequency spectrum $\underline{k}^{(n)} =$

$\left(k_1^{(n)}, \ldots, k_{n-1}^{(n)}\right)$ with fixed $|\underline{k}^{(n)}| = s$ under the coalescent model $\Pi \in \Theta_s^{\mathtt{E}}$ (resp. $\Pi \in \Theta_s^{\mathtt{B}}$), by

$$L(\Pi, \underline{k}^{(n)}, s) = \mathbb{P}^{\Pi,s}\big\{\xi_i^{(n)} = k_i^{(n)}, i \in [n-1]\big\} = \mathbb{E}^{\Pi}\left[\frac{s!}{k_1^{(n)}! \cdots k_{n-1}^{(n)}!} \prod_{i=1}^{n-1} \left(\frac{B_i^{(n)}}{B^{(n)}}\right)^{k_i^{(n)}}\right], \tag{6}$$

where the $B_i^{(n)}$ are the random length of branches subtending $i \in [n-1]$ leaves and $B^{(n)}$ is the total branch length of the coalescent under Π. The fixed-s-paradigm thus leads to a mixture of multinomial distributions, where the parameters are given by the respective relative branch lengths. The hope is that the location of the maximum of $L(\Pi, \underline{k}^{(n)}, s)$ is typically not far from the location of the corresponding coordinate of the maximizer in the full two-dimensional 'explicit-θ'-model, in which one can additionally maximize over all $\theta \in [0, \infty)$.

Now we can construct a 'likelihood-ratio' test based on $L(\Pi, \underline{k}^{(n)}, s)$ via the likelihood-ratio function

$$\varrho_{(\mathtt{E},\mathtt{B};s)}(\underline{\xi}^{(n)}) := \frac{\sup\left\{L(\Pi, \underline{k}^{(n)}, s),\ \Pi \in \Theta_s^{\mathtt{E}}\right\}}{\sup\left\{L(\Pi, \underline{k}^{(n)}, s),\ \Pi \in \Theta_s^{\mathtt{B}}\right\}}. \tag{7}$$

Given a significance level $a \in (0,1)$ (say, $a = 0.05$), let $\varrho^*_{(\mathtt{E},\mathtt{B};s)}(a)$ be the a-quantile of $\varrho_{(\mathtt{E},\mathtt{B};s)}(\underline{\xi}^{(n)})$ under $\mathtt{E}$, chosen as the largest values so that

$$\sup_{\Pi \in \Theta_s^{\mathtt{E}}} \mathbb{P}^{\Pi,s}\big\{\varrho_{(\mathtt{E},\mathtt{B};s)}(\underline{\xi}^{(n)}) \le \varrho^*_{(\mathtt{E},\mathtt{B};s)}(a)\big\} \le a. \tag{8}$$

The decision rule that constitutes the 'fixed-s-likelihood-ratio test', given s and sample size n, is

$$\text{reject } H_0^s \quad \Longleftrightarrow \quad \varrho_{(\mathtt{E},\mathtt{B};s)}(\underline{\xi}^{(n)}) \le \varrho^*_{(\mathtt{E},\mathtt{B};s)}(a).$$

This formulation is free of the nuisance parameter θ. To assess the justification of the 'fixed-s-assumption', we investigate how close this is to the corresponding quantiles for different values of θ, including the Watterson estimator

$$\hat{\theta} = \hat{\theta}(\Pi; s) = \frac{2s}{\mathbb{E}^{\Pi}[B^{(n)}]} \tag{9}$$

cf. (3), for selected choices of Π. The agreement appears reasonably good and seems to increase with sample size, cf. the supplementary material.

The corresponding power function of the test, that is, the probability to reject a false null-hypothesis, is given by

$$G_{(\mathtt{E},\mathtt{B};s)}(\Pi) = \mathbb{P}^{\Pi}\{\varrho_{(\mathtt{E},\mathtt{B};s)}(\underline{\xi}^{(n)}) \leq \varrho^{*}_{(\mathtt{E},\mathtt{B};s)}(a,S)\}, \quad \Pi \in \Theta^{\mathtt{B}}_{s}. \tag{10}$$

The likelihood (6) cannot be represented as a simple formula involving the coalescent parameters; one can approximate (6) via a Monte Carlo approach but this is computationally expensive. An approximation is

$$L(\Pi, \underline{k}^{(n)}, s) \approx \frac{s!}{k_1^{(n)}! \cdots k_{n-1}^{(n)}!} \prod_{i=1}^{n-1} \left(\varphi_i^{(n,\Pi)}\right)^{k_i^{(n)}} \tag{11}$$

where we replaced the random quantities $B_i^{(n)}/B^{(n)}$ in (6) by $\varphi_i^{(n,\Pi)} = \mathbb{E}^{\Pi}[B_i^{(n)}]/\mathbb{E}^{\Pi}[B^{(n)}]$ (2).

Interestingly, an approximate maximum likelihood method based on (11) is equivalent to the following approach: Consider a family of (approximate) likelihood functions

$$\widetilde{L}(\Pi, \underline{\xi}^{(n)}, s) = \prod_{i=1}^{n-1} e^{-\frac{\hat{\theta}(\Pi,s)}{2}\mathbb{E}^{\Pi}[B^{(n)}]\varphi_i^{(n,\Pi)}} \frac{\left(\frac{\hat{\theta}(\Pi,s)}{2}\mathbb{E}^{\Pi}[B^{(n)}]\varphi_i^{(n,\Pi)}\right)^{\xi_i^{(n)}}}{\xi_i^{(n)}!} \tag{12}$$

where $\hat{\theta}(\Pi, s) = 2s/\mathbb{E}^{\Pi}[B^{(n)}]$ is the Watterson estimator for the mutation rate under a Π-coalescent with n leaves when $S = s$ segregating sites are observed, recall (3). In (12), $\widetilde{L}$ is well defined even if $|\underline{\xi}^{(n)}| \neq s$.

The rationale behind (12) is simple: It pretends that the classes are approximately independent and Poisson-distributed (this is of course not literally true but encouraged by the fact that the off-diagonal entries of the covariance matrix of $\underline{\xi}^{(n)}$ are small compared to the diagonal terms, see Birkner *et al.* (2013b); our approximation is different from Sawyer and Hartl (1992)'s Poisson random field which applies to unlinked sites).

For fixed s, we can view $(\Pi, \hat{\theta}(\Pi, s))$ as parametrizing a one-dimensional curve in the full

two-dimensional space $H_0 \cup H_1$ defined by the requirement that $\mathbb{E}^{\Pi,\hat{\theta}(\Pi,s)}[S] = s$.

The two approximate maximum likelihood approaches based on (11) and on (12) are equivalent; indeed

$$\widetilde{L}(\Pi, \underline{k}^{(n)}, s) = \prod_{i=1}^{n-1} e^{-s\varphi_i^{\Pi,(n)}} \frac{(s\varphi_i^{\Pi,(n)})^{k_i^{(n)}}}{k_i^{(n)}!} = e^{-s} L(\Pi, \underline{k}^{(n)}, s), \tag{13}$$

since the $\varphi_i^{(n,\Pi)}$ sum to 1. Hence, both likelihood functions differ only by the fixed prefactor e^{-s}, so that they attain their maximum at the same position.

Thus, now we consider the statistic

$$\widetilde{\varrho}_{(\mathtt{E},\mathtt{B})}(\underline{\xi}^{(n)}) := \frac{\sup\left\{\widetilde{L}(\Pi, \underline{k}^{(n)}, |\underline{k}^{(n)}|),\, \Pi \in \Theta^{\mathtt{E}}\right\}}{\sup\left\{\widetilde{L}(\Pi, \underline{k}^{(n)}, |\underline{k}^{(n)}|),\, \Pi \in \Theta^{\mathtt{B}}\right\}} \tag{14}$$

(where $\Theta^{\mathtt{E}}$ and $\Theta^{\mathtt{B}}$ refer to the projection of H_0, respectively H_1, on the coalescent parameter). For a given value of s, we can then (by simulations using the fixed-s-approach) determine approximate quantiles $\widetilde{\varrho}^*_{(\mathtt{E},\mathtt{B};s)}(a)$ associated with a significance level a as in (8), and base our test on the criterion $\widetilde{\varrho}_{(\mathtt{E},\mathtt{B})}(\underline{\xi}^{(n)}) \leq \widetilde{\varrho}^*_{(\mathtt{E},\mathtt{B};s)}(a)$. Similary, the (approximate) power function $\widetilde{G}_{(\mathtt{E},\mathtt{B};s)}$ can be estimated using simulations.

An alternative approach to the (approximate) likelihood-based tests would be rejection rules based on minimal distance statistics:

$$\varrho^{(d)}_{(\mathtt{E},\mathtt{B})}(\underline{\xi}^{(n)}) := \frac{\inf\{d(\underline{\varphi}^{(n,\Pi)}, \underline{\xi}^{(n)}), \Pi \in \Theta^{\mathtt{E}}_s\}}{\inf\{d(\underline{\varphi}^{(n,\Pi)}, \underline{\xi}^{(n)}), \Pi \in \Theta^{\mathtt{B}}_s\}}, \tag{15}$$

for some suitable distance measure d (e.g. the ℓ_p distance with $p = 2$) with corresponding power function $G^{(d)}_{(\mathtt{E},\mathtt{B};s)}$. We will not discuss the theoretical justification of this method. However, we will use the ℓ_2-distance between normalized expected spectra under various coalescent models to produce three-dimensional 'heatmaps' that give some intuitive insight on how a pair of different models out of $\{\mathtt{E}, \mathtt{B}, \mathtt{A}, \mathtt{D}\}$ relates to each other depending on the underlying pair of coalescent parameters (cf. Results).

We conclude this section by a remark on *lumping.* One often observes $k_i^{(n)} = 0$ for most i greater than some (small) number m in observed data, in particular for large n. It thus seems natural to consider (approximate) likelihood functions for 'lumped spectra' (e.g. collapsing all entries in classes to the right of some number m into one class m^+), as we have done eg. in Figure 1. Another natural type of lumping may be to collect together classes so that $\sum_i \varphi_i^{(n)} \geq x$ for some $x \in (0, 1/2]$. This may not always be feasible, though, if the individual $\varphi_i^{(n)}$ quickly become quite small, and we will refrain from going into a more detailed theoretical discussion of optimal lumpings. However, we will see in our subsequent ABC analysis, that adequate lumping can improve the reliability of our model-selection procedure.

Approximate Bayes factors and model selection

In view of the approach and notation of the previous section, an analogous way of model selection could be based on a *Bayes factor* of the form

$$\varrho^{\mathcal{B}}_{(\mathtt{B},\mathtt{E})}(\underline{\zeta}^{(n)}) := \frac{\int_{\Theta_s^{\mathtt{B}}} \widetilde{L}(\Pi, \underline{\zeta}^{(n)}; s)\, d\pi_{\mathtt{B}}(\alpha)}{\int_{\Theta_s^{\mathtt{E}}} \widetilde{L}(\Pi, \underline{\zeta}^{(n)}; s)\, d\pi_{\mathtt{E}}(\beta)}, \tag{16}$$

(and similar for all other combinations of classes $\mathtt{A}, \mathtt{D}, \mathtt{E}, \mathtt{B}$), given a pair of priors $\pi_{\mathtt{B}}, \pi_{\mathtt{E}}$ on $\Theta_s^{\mathtt{B}}, \Theta_s^{\mathtt{E}}$. While the approach will work in principle also for the two-dimensional prior ranges $\Theta^{\mathtt{B}}$ and $\Theta^{\mathtt{E}}$, we will present the (approximate) Bayesian methods with one-dimensional prior ranges (where $S = s$ is treated as a fixed parameter, motivated from the 'fixed-s-approach'), so that they complement our previous methods.

Our simulations will be obtained using the rationale behind (12), that is, after simulating a tree according to a given coalescent parameter, say α from $\pi_{\mathtt{B}}$, mutations are placed on the tree according to a Poisson process with mutation rate $\hat{\theta}(\Pi; s)$ estimated using (3). However, in (16), we use the normalized site frequency spectrum $\underline{\zeta}^{(n)}$ as observed statistics, since it should be more insensitive to the 'true' coalescent mutation rate θ [potentially deviating from $\hat{\theta}(\Pi, s)$], as argued in the corresponding section in the Supporting Information, and thus yield

more robust results. $\tilde{L}$ in Eq. 16 thus denotes the likelihood function of the observed nSFS under the chosen coalescent model with mutation parameter given by Watterson's estimator based on s. Since we estimate the mutation rate based on s, the information loss of using the nSFS instead of the SFS should be only slight. We also experimented with ABC based on simulations using the fixed-s-method, i.e. distributing a fixed number of mutations uniformly on the simulated tree, and generally found higher misclassification probabilities (results not shown).

To overcome the problem of exact computation of $\widetilde{L}(\Pi, \underline{\zeta}^{(n)}; s)$, which appears infeasible in practice, we employ approximate Bayesian methods (see e.g. Beaumont (2010)) based on the ℓ_2-distance between observed and simulated nSFS.

Bayes factors based on further (lumped) distances d and/or the folded nSFS may of course also be considered. In line with classical Bayes factor philosophy (cf. e.g. Kass and Raftery (1995)), one interprets an observed value of $\varrho^{\mathcal{B}}_{(\mathtt{B},\mathtt{E})} \gg 1$ as evidence in favor of $\Theta^{\mathtt{B}}_s$ over $\Theta^{\mathtt{E}}_s$.

For the ABC analyses, we consider as before on exponential growth (E), algebraic growth (A), and Beta- (B), and Dirac- (D) coalescents. Given sample size n and number of segregating sites s, again the coalescent model classes can be parametrised by a single parameter each, which are the exponential growth rate $\beta \in [0, \infty)$, algebraic growth rate $\gamma \in [0, \infty)$; Beta coalescent parameter $\alpha \in [1, 2]$, and mass point location $\psi \in (0, 1]$ for the Dirac coalescent.

For convenience, we employ a simple rejection-based ABC scheme to approximate the Bayes factor for the model (class) comparison given an observed nSFS (resp. folded and/or lumped versions, which can be treated analogously). First, select a number of models (out of $\{\mathtt{E}, \mathtt{B}, \mathtt{A}, \mathtt{D}\}$) that should be compared, say E and B, and choose the corresponding prior distributions on the correspoding coalescent parameter ranges. To simulate say n_{r} independent samples of the nSFS from each model, say from E, independently generate n_{r} coalescent parameters from the prior $\pi_{\mathtt{E}}$, and a corresponding coalescent tree Π for each generated coalescent parameter. Distribute independently Poisson mutations with parameter $\hat{\theta}(\Pi; s)/2$ on each such tree and record the corresponding normalized site frequency spectra. This should

be done independently for all models.

Then fix a tolerance level $x \in (0,1)$ and count the number of simulations $N_{\mathtt{E}}$, $N_{\mathtt{B}}$ from each model that are among the $100 \cdot x\,\%$ best fits with respect to the ℓ_2 distance to the observed nSFS $\underline{\zeta}^{(n)}$ (the 'accepted' simulations). Here, we use an additional scaling by dividing each class (resp. lumped class) in the nSFS by the median (if non-zero) within this class observed in all simulations as implemented in the `R` package `abc` (Csilléry *et al.*, 2012). The Bayes factor for model `E` vs. `B` can then be approximated by

$$\varrho^{\mathcal{B}}_{(\mathtt{E},\mathtt{B})}(\underline{\zeta}^{(n)}) \approx \frac{N_{\mathtt{E}}}{N_{\mathtt{B}}}.$$

To assess how well this ABC approach allows one to distinguish, say, `E` from `B` (or, more generally, simultaneously among $\{\mathtt{E},\mathtt{B},\mathtt{A},\mathtt{D}\}$), we use two approaches from the `R` package `abc`. Both are based on leave-one-out cross-validation. More precisely, we pick n_{cv} simulations at random from each model, treat them as the observed value of the nSFS and then run the ABC approach with the same parameters and simulations as above. For each cross-validation sample, say $\underline{\zeta}^{(n)}_{\mathtt{E}}(i)$, $i \in [n_{\mathrm{cv}}]$ from model `E`, we record the counts of accepted simulations $N_{\mathtt{A}}(\underline{\zeta}^{(n)}_{\mathtt{E}}(i)), N_{\mathtt{B}}(\underline{\zeta}^{(n)}_{\mathtt{E}}(i))$ and $N_{\mathtt{D}}(\underline{\zeta}^{(n)}_{\mathtt{E}}(i))$ from the model classes $\mathtt{A},\mathtt{B},\mathtt{D}$, `E` (recall that the chosen cross-validation sample is left out). As measures for the distinction ability of this approach, we record for each model class, borrowing notation from Stoehr *et al.* (2014),

- the (estimated) mean posterior probabilities π for model `B` given the observed nSFS under the true model `E`, say,

  $$\mathbb{E}^{\mathtt{E}}\left[\pi(\mathtt{B}|\underline{\zeta}^{(n)})\right] \approx \frac{1}{n_{\mathrm{cv}}}\sum_{i=1}^{n_{\mathrm{cv}}} \frac{N_{\mathtt{B}}\big(\underline{\zeta}^{(n)}_{\mathtt{E}}(i)\big) + N_{\mathtt{E}}\big(\underline{\zeta}^{(n)}_{\mathtt{E}}(i)\big)}{N_{\mathrm{a}}}$$

  where $N_{\mathrm{a}} = N_{\mathtt{A}}(\underline{\zeta}^{(n)}_{\mathtt{E}}(i)) + N_{\mathtt{B}}(\underline{\zeta}^{(n)}_{\mathtt{E}}(i)) + N_{\mathtt{E}}(\underline{\zeta}^{(n)}_{\mathtt{E}}(i)) + N_{\mathtt{D}}(\underline{\zeta}^{(n)}_{\mathtt{E}}(i))$ is the number of accepted simulations;

- the (estimated) mean misclassification probabilities

$$\mathbb{E}^{\mathtt{E}}\left[\pi\left(\min_{\mathtt{Y}\neq\mathtt{B}} \varrho^{\mathcal{B}}_{(\mathtt{B},\mathtt{Y})} \geq 1 \,|\, \underline{\zeta}^{(n)}\right)\right] \approx \frac{1}{n_{\mathrm{cv}}} \sum_{i=1}^{n_{\mathrm{cv}}} 1_{\left\{ N_{\mathtt{B}}\left(\underline{\zeta}^{(n)}_{\mathtt{E}}(i)\right) \geq N_{\mathtt{Y}}\left(\underline{\zeta}^{(n)}_{\mathtt{E}}(i)\right) \;\forall\; \mathtt{Y}\neq\mathtt{B}\right\}}$$

for $\mathtt{Y} \in \{\mathtt{A}, \mathtt{E}, \mathtt{D}\}$. To ease notation we will from now on omit n in the formulae.

In practice, we need to efficiently generate samples of the nSFS under the different models which can be achieved by backward-in-time coalescent simulations. For the exponential growth models (`E`), we use Hudson's ms (Hudson, 2002) as implemented in the `R` (R Core Team, 2012) package phyclust (Chen, 2011). For algebraic growth models, the Beta-coalescents (`B`) and Dirac coalescents (`D`), we use custom `R` and `C` scripts to generate samples of the nSFS, available at `http://page.math.tu-berlin.de/~eldon/programs.html`. To conduct the actual ABC analysis including cross-validation techniques, we employed the `R` package abc (Csilléry *et al.* (2012)). Additionally, since we use Watterson's estimator to set the mutation rate within each model, we compute the mean total length of each coalescent model as described in Supporting Information.

Results

Power estimates of approximate likelihood-ratio tests

To assess the sensitivity of our approximate likelihood ratio test associated with the likelihood ratio function (7), we estimate its power $\widetilde{G}_{(\mathtt{E},\mathtt{B};s)}$ from the analogue of (10) based on the approximate likelihood from (12) as a function of α (Figure 2**A**) with $H_0^s = \Theta_s^{\mathtt{E}}$ and $H_1^s = \Theta_s^{\mathtt{B}}$; and estimate $\widetilde{G}_{(\mathtt{B},\mathtt{E};s)}$ as a function of β with $H_0^s = \Theta_s^{\mathtt{B}}$ and $H_1^s = \Theta_s^{\mathtt{E}}$ (Figure 2**B**).

As shown in Figure 2, reasonably high power is obtained to reject $\Theta_s^{\mathtt{E}}$ for $n = 500$; and even for smaller sample size $n = 100$, but the power also depends, as one would expect, on the size of the test. As a sidenote, we remark that the power estimates $\widetilde{G}_{(\mathtt{E},\mathtt{B};s)}$, as a function of α, are right at the size of each corresponding test when $\alpha = 2$ (the Kingman case), as required.

The mtDNA-genome analysis of Carr and Marshall (2008), who scanned whole mitochondrial genomes (15,655 bp) of the highly fecund Atlantic cod (*Gadus morhua*), prompted us to briefly investigate the power (Figures S3, S4) with number of segregating sites $s = 300$. This is nearly the total number of polymorphisms (298) observed among the 32 mtDNA genomes sampled by Carr and Marshall (2008). Our results show that while we may not quite have enough power when $n = 30$ and $s = 300$, (Figure S4**A**), we would be in good shape for $n = 100$ (Figures S3, S4**B**). It would be very interesting to analyse such a sample once available, since it appears to be an open debate whether Beta-coalescents should be favoured over classical models (including recent population growth) in HFSOD populations (cf. eg. Steinrücken *et al.* (2013)).

Another quite striking observation is that the power of our test is apparently non-monotone as a function of β when $H_0^s = \Theta_s^{\mathtt{B}}$, in particular for smaller Type I error. We will present a possible heuristic explanation for this in the discussion section.

Rather high power in general is obtained when comparing $\Theta_s^{\mathtt{E}}$ and $\Theta_s^{\mathtt{D}}$ associated with the Dirac coalescent (Figure S2) for $(n, s) = (200, 50)$.

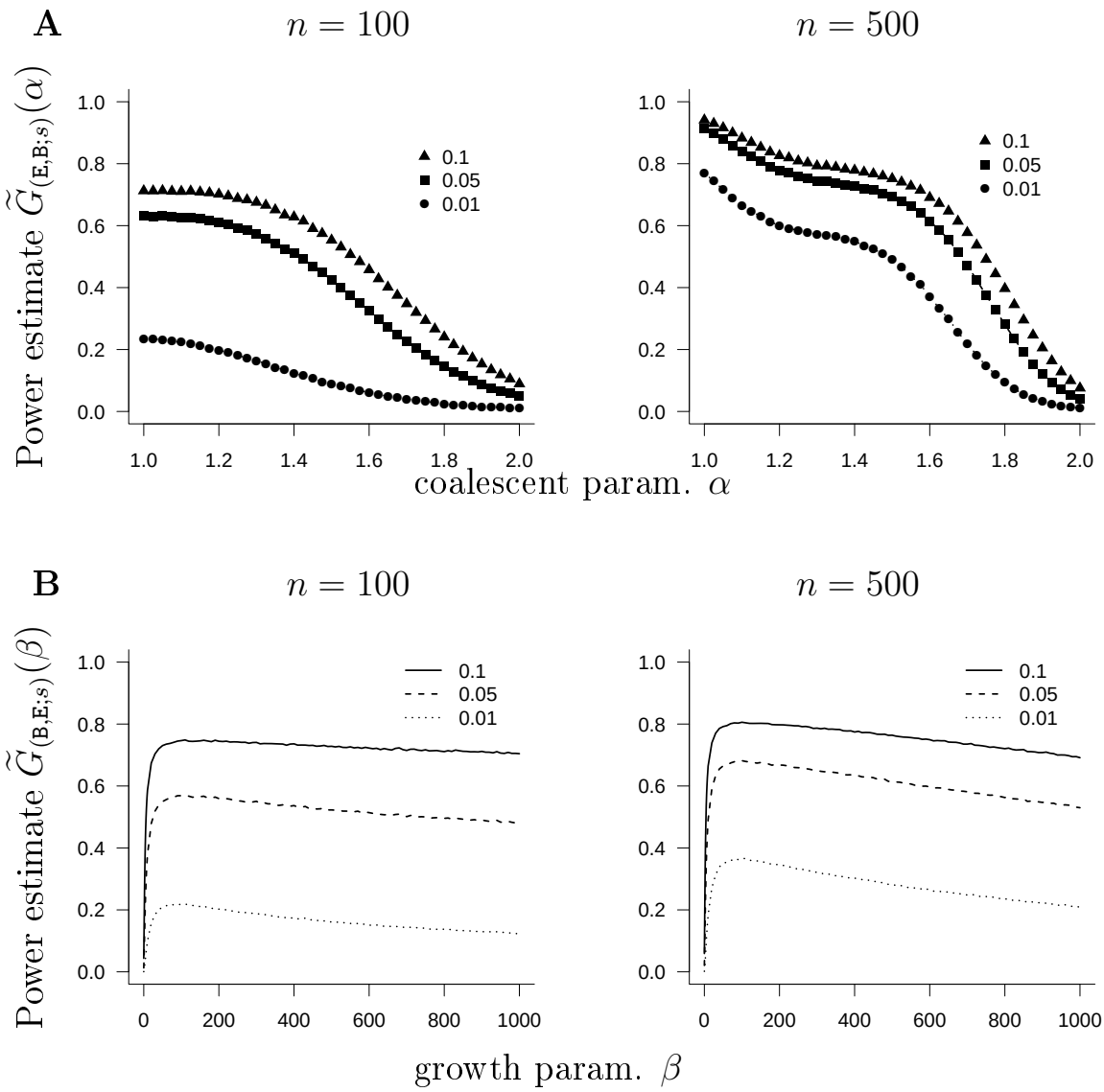
For further combinations, we refer to the supporting material, where $\Theta_s^{\mathtt{A}}$, associated with

algebraic growth, is compared with $\Theta_s^{\mathtt{B}}$ in Figure S6, and with $\Theta_s^{\mathtt{D}}$ in Figure S5. The power functions $\widetilde{G}_{(\mathtt{A},\mathtt{D};s)}(\psi)$ (Figure S5) are decidedly non-monotone, as is $\widetilde{G}_{(\mathtt{A},\mathtt{B};s)}(\alpha)$ (Figure S6**A**).

We conclude with a short remark on the sensitivity of our results on *lumping* of classes in the observed spectrum. Indeed, our power estimates suggest that keeping at least the first five classes of the SFS intact, and collecting the rest into one other class, has little effect on the power of the test (results not shown). Keeping only the singleton ($\xi_1^{(n)}$) class intact, and collecting all the rest into one class, however, significantly diminishes power (results not shown).

Availability of source-code: `C` (cf. Kernighan and Ritchie, 1988) code written for estimating the power of our tests, where use was made of the GNU Scientific Library (Galassi *et al.*, 2013), is available at
`http://page.math.tu-berlin.de/~eldon/programs.html`.

Figure 2: **A**: Estimate of $\widetilde{G}_{(\mathtt{E},\mathtt{B};s)}$ from (10) based on the approximate likelihood from (12) as a function of α (no lumping) with number of leaves n as shown, and $s = 50$. **B**: Estimate of $\widetilde{G}_{(\mathtt{B},\mathtt{E};s)}$ from (10) based on the approximate likelihood from (12) as a function of β (no lumping) with number of leaves n as shown, and $s = 50$. The symbols denote the size of the test as shown in the legend. The interval hypotheses are discretized to $\Theta_s^{\mathtt{E}} = \{\beta : \beta \in \{0, 1, 2, \ldots, 10, 20, \ldots, 1000\}\}$ and $\Theta_s^{\mathtt{B}} = \{\alpha : \alpha \in \{1, 1.025, \ldots, 2\}\}$. In **A**, the Beta$(2-\alpha, \alpha)$-coalescent is the alternative; in **B** exponential growth is the alternative.



Mean-squared distance landscapes for the normalized expected SFS under different growth and coalescent models

Given the potential ability to distinguish between growth- and multiple-merger coalescent models invites the following question: How does the distance between $\varphi_i^{(n),\Pi_1}$ and $\varphi_i^{(n),\Pi_2}$ behave as a function of the underlying coalescent and growth parameters? Is it possible to visibly identify a one-dimensional curve given by coalescent parameter pairs corresponding to (Π_1, Π_2) along which minimal distance is achieved? Figures 3 and 4 are a brief effort to understand the relation between the normalised expected SFS for the models in question by graphing the ℓ_2-distance $\mathrm{d}_2^{(n)}(\mathtt{X}, \mathtt{Y}) = \left(\sum_{i=1}^{n-1}\left(\varphi_i^{(n,\mathtt{X})} - \varphi_i^{(n,\mathtt{Y})}\right)^2\right)^{1/2}$ as a function of the coalescent and growth parameters associated with `X` and `Y`. In Figure 3 `E` is compared to `B` and `D`. In Figure 4, `A` is compared to `B` and `D`. In Figures 3 and 4, the upper panels show the distance as the respective growth parameter ranges from 0 to 1000, while the lower panels 'zoom in' on the range from 0 to 10.

Figure 3 indicates the presence of a region, essentially a curve in the two-dimensional (α, β) parameter space, along which the lowest ℓ_2 distance is reached. However, one should be aware that this curve shifts in space when the sample size n is increased (data not shown).

Figure 3 suggests that we should have good power to distinguish between algebraic growth and Beta-coalescents. However, this seems not to be the case for distinguishing algebraic growth from Dirac-coalescents: Extreme growth (large γ) seems to produce an almost star-shape like genealogy - consequently the distance to a Dirac coalescent with ψ close to 1 becomes very small (recall that $\psi = 1$ exactly corresponds to the star-shaped coalescent).

Figure 3: The ℓ_2-distance $d_2^{(n)}(\mathtt{E},\mathtt{X})$ for $\mathtt{X} \in \{\mathtt{B},\mathtt{D}\}$ of the normalized expected spectra $\varphi_i^{(n,\mathtt{E})}$ (see Eq. (2)) and $\varphi_i^{(n,\Pi)}$ as a function of α ($\mathtt{X} = \mathtt{B}$) (resp. ψ; $\mathtt{X} = \mathtt{D}$) and β $\mathtt{E}$ for number of leaves $n = 100$. Expected values were computed exactly. The gridpoints are $\alpha \in \{1, 1.025, \ldots, 2\}$, $\psi \in \{0.01, 0.02, \ldots, 0.1, 0.15, 0.2, \ldots, 0.95\}$, and for β as shown.

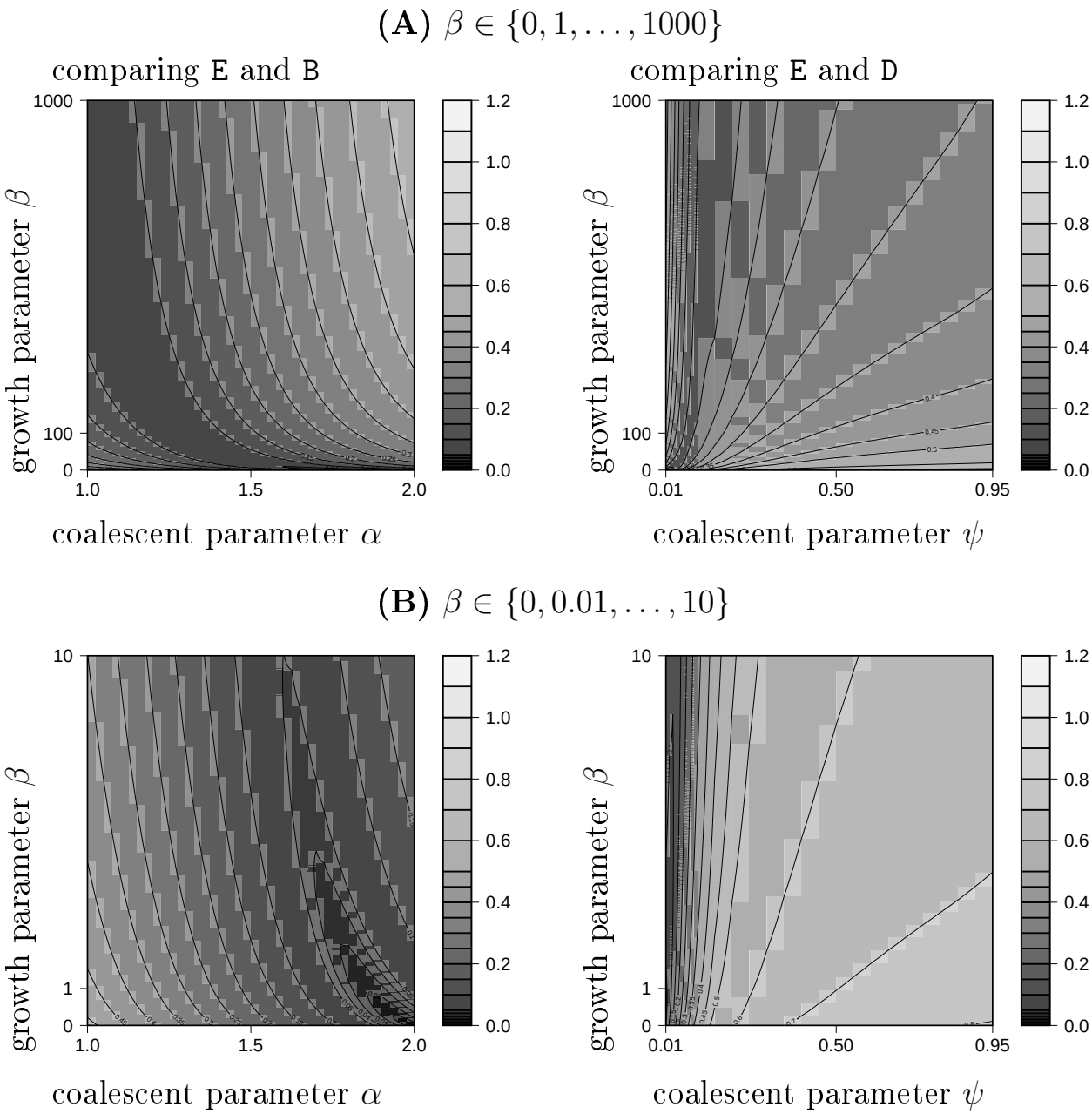
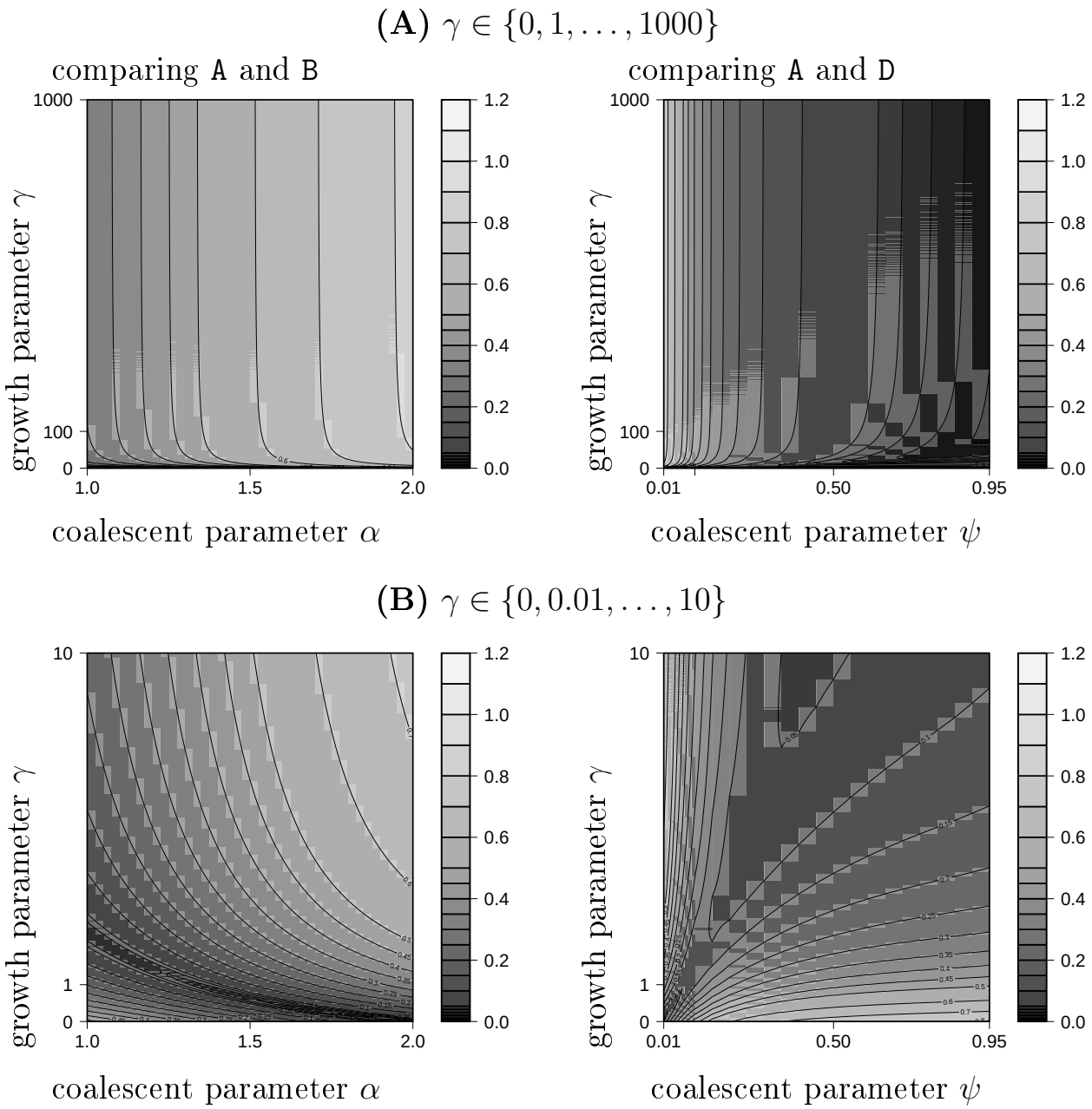


Figure 4: The ℓ_2-distance $d_2^{(n)}(\mathtt{A},\mathtt{X})$ for $\mathtt{X} \in \{\mathtt{B},\mathtt{D}\}$ of the normalized expected spectra $\varphi_i^{(n,\mathtt{A})}$ (see Eq. (2)) and $\varphi_i^{(n,\mathtt{X}}$ as a function of α ($\mathtt{X} = \mathtt{B}$) (resp. ψ; $\mathtt{X} = \mathtt{D}$) and γ ($\mathtt{A}$) for number of leaves $n = 100$. Expected values were computed exactly. The gridpoints are $\psi \in \{0.01, 0.02, \ldots, 0.1, 0.15, 0.2, \ldots, 0.95\}$, $\alpha \in \{1, 1.025, \ldots, 2\}$; for γ as shown.



Mean misclassification & posterior probabilities for the ABC approach

In this section, we analyse in how far an ABC approach using the nSFS (resp. the folded nSFS, abbreviated as nfSFS) and their lumped variants as summary statistic support our claim that one can distinguish between exponential growth $\mathtt{E}$ and the Beta$(2-\alpha, \alpha)$-coalescent$\mathtt{B}$ as well as between $\mathtt{E}$, $\mathtt{B}$ and Dirac coalescents $\mathtt{D}$ or between $\mathtt{B}$, $\mathtt{D}$ and algebraic growth $\mathtt{A}$. The distinction ability of the ABC model comparison is assessed based on the simulation procedure and notation described in the methods section. Priors were uniform over the full range of coalescent parameters in models $\mathtt{B}, \mathtt{D}$, and uniform until a maximal cutoff for $\mathtt{E}$ (on $[0, \beta_{\max}]$) and $\mathtt{A}$ (on $[0, \gamma_{\max}]$). We discretized the parameter range for the growth models by using increments of 1 or 10 for exponential growth (the first used in all multiple-model comparisons, the latter in the pairwise comparisons between $\mathtt{E}$ and $\mathtt{B}$) and increments of 1 for algebraic growth. If not specified otherwise, we used $\beta_{\max} = \gamma_{\max} = 1000$. We fixed a sample size $n = 200$. The number of replications was set to $n_{\mathrm{r}} = 2 \times 10^5$. See Tables 1-3 and S4-S8 for the estimates of posterior probabilities and misclassification probabilities (some with one replication) with various degrees of lumping and various parameter settings for $n = 200$. For an example with higher sample size $n = 1278$, see Table S9.

Table 1: Approximations of the mean posterior probabilities and misclassification probabilities for the ABC model comparison between models $\mathtt{E}$ and $\mathtt{B}$ for tolerance $x = 0.01$, $s = 15$ segregating sites; and using either the nSFS or the nfSFS as summary statistics. n_{cv} denotes the number of cross-validations; 'lump' indicates which mutation classes are lumped into one class.

fold	lump	n_{cv}	$\mathbb{E}^{\mathtt{B}}\left[\pi(\mathtt{E}\|\underline{\zeta})\right]$	$\mathbb{E}^{\mathtt{E}}\left[\pi(\mathtt{B}\|\underline{\zeta})\right]$	$\mathbb{E}^{\mathtt{B}}\left[\pi(\varrho^{\mathcal{B}}_{(\mathtt{E},\mathtt{B})} \geq 1\|\underline{\zeta})\right]$	$\mathbb{E}^{\mathtt{E}}\left[\pi(\varrho^{\mathcal{B}}_{(\mathtt{E},\mathtt{B})} \leq 1\|\underline{\zeta})\right]$
no	10+	24000	0.30	0.25	0.26	0.13
no	50+	12000	0.32	0.29	0.26	0.17
no	100+	1200	0.33	0.29	0.28	0.17
yes	10+	24000	0.32	0.25	0.28	0.12
yes	50+	12000	0.34	0.29	0.28	0.16
yes	no	12000	0.34	0.29	0.29	0.16

The estimated error probabilities range from moderate to low values. Mean posterior probabilities $\mathbb{E}^{\mathtt{B}}\left[\pi(\mathtt{E}|\underline{\zeta})\right] \approx 30\%$ indicate a correct classification probability $\mathbb{E}^{\mathtt{B}}\left[\pi(\mathtt{B}|\underline{\zeta})\right] \approx 70\%$, which shows that our method has good distinguishing ability. As expected, lower tolerance generally leads to smaller errors, as do larger mutation rates, while using the folded nSFS increases them. Appropriate lumping seems to decrease the error probabilities in many occasions, see e.g. Table 1 where a positive effect for strong lumping is observed for $s = 15$ segregating sites, whereas for $s = 75$ in Table S4, moderate lumping seems to be more appropriate (both tables show the comparison of models B and E). Not surprisingly exponential growth rates closer to zero are harder to distinguish from the Beta$(2-\alpha, \alpha)$-coalescent models than higher growth rates (see Tables S4 and S6). The ABC model comparison distinguishes especially well, even for $s = 15$, between exponential growth from Dirac coalescents and also algebraic growth from Beta$(2 - \alpha, \alpha)$-coalescent(Tables 3, S8, 2, S7). It should be noted that for relatively low number of segregating sites ($s = 15$), some comparisons (algebraic growth with Dirac coalescents, Beta coalescents with Dirac coalescents) can lead to common misclassification, but this effect vanishes for larger s. For $s = 75$, Dirac coalescents can be distinguished relatively well from Beta$(2 - \alpha, \alpha)$-coalescent (Tables 3, S8, 2, S7).

Table 2: Approximations of the mean posterior probabilities for the ABC model comparison between models E, B, D for tolerance $x = 0.005$, sample size $n = 200$ and $s = 15$ or $s = 75$. The nfSFS was used as summary statistics.

s	lump	n_{cv}	$\mathbb{E}^{\mathtt{B}}\left[\pi(\mathtt{E}\|\underline{\zeta})\right]$	$\mathbb{E}^{\mathtt{B}}\left[\pi(\mathtt{D}\|\underline{\zeta})\right]$	$\mathbb{E}^{\mathtt{E}}\left[\pi(\mathtt{B}\|\underline{\zeta})\right]$	$\mathbb{E}^{\mathtt{E}}\left[\pi(\mathtt{D}\|\underline{\zeta})\right]$	$\mathbb{E}^{\mathtt{D}}\left[\pi(\mathtt{B}\|\underline{\zeta})\right]$	$\mathbb{E}^{\mathtt{D}}\left[\pi(\mathtt{E}\|\underline{\zeta})\right]$
15	10+	24000	0.29	0.12	0.24	0.02	0.55	0.03
15	50+	12000	0.39	0.10	0.22	0.02	0.54	0.05
15	no	12000	0.42	0.10	0.21	0.02	0.55	0.07
75	10+	24000	0.22	0.09	0.12	0.00	0.11	0.01
75	50+	12000	0.26	0.10	0.12	0.00	0.11	0.01

Table 3: Approximations of the mean posterior probabilities for the ABC model comparison between models A, B, D for tolerance $x = 0.005$, sample size $n = 200$ and $s = 15$ or $s = 75$. The nfSFS was used as summary statistics.

s	lump	n_{cv}	$\mathbb{E}^{\mathtt{B}}\left[\pi(\mathtt{A}\|\underline{\zeta})\right]$	$\mathbb{E}^{\mathtt{B}}\left[\pi(\mathtt{D}\|\underline{\zeta})\right]$	$\mathbb{E}^{\mathtt{A}}\left[\pi(\mathtt{B}\|\underline{\zeta})\right]$	$\mathbb{E}^{\mathtt{A}}\left[\pi(\mathtt{D}\|\underline{\zeta})\right]$	$\mathbb{E}^{\mathtt{D}}\left[\pi(\mathtt{B}\|\underline{\zeta})\right]$	$\mathbb{E}^{\mathtt{D}}\left[\pi(\mathtt{A}\|\underline{\zeta})\right]$
15	10+	24000	0.02	0.12	0.01	0.03	0.16	0.55
15	50+	12000	0.02	0.11	0.01	0.03	0.24	0.57
75	10+	24000	0.01	0.09	0.01	0.05	0.10	0.29
75	50+	12000	0.01	0.08	0.01	0.05	0.15	0.32

Discussion

The development of methods to distinguish between different (time-changed) coalescent scenarios for the underlying genealogy of a population on the basis of observed data is an important task, in particular since the choice of an underlying coalescent model affects the estimated coalescent mutation rate $\hat{\theta}$ via (3) and a potential real-time embedding of the genealogy based on (5). The identification of an appropriate coalescent model may also give hints about the underlying reproductive mechanisms present in a population. By way of example, multiple-merger coalescents may indicate presence of HFSOD in the population.

While inference methods for distinguishing population growth from the usual Kingman coalescent have been extensively studied, see e.g. Tajima (1989a), Slatkin and Hudson (1991), Rogers and Harpending (1992), Kaj and Krone (2003) and Sano and Tachida (2005), and sophisticated theoretical results on the question of identifiability of demographic histories have been obtained (cf. eg. Myers *et al.* (2008), Bhaskar and Song (2014) and Kim *et al.* (2014)), none of these studies have addressed multiple merger coalescents. In fact, only few results, eg. on the statistical properties of the SFS in multiple merger coalescents, see Birkner *et al.* (2013b), are available.

For the particular case of distinguishing multiple merger coalescents from population growth scenarios, this decision problem is complicated by the fact that the patterns of genetic variation produced by the two demographic effects, and summarized in the SFS, are expected to be similar: Both lead to an excess of singletons compared to a classical Kingman-coalescent based genealogy. However, while it is usually possible to match the predicted number of singletons with the observed number in various special cases for both models, the bulk and tail of the spectrum will typically differ, cf. Figure 1 for some examples.

The present paper is thus aimed at exploiting and quantifying these differences. However, for feasibility, we had to restrict both the scope and the employed methods of our analysis. The first (restrictive) decision in the design of our analysis was the selection of certain subfamilies of Lambda-coalescents and demographic growth scenarios that we deemed suitable for investigation. The reason for restricting to sub-classes of Lambda-coalescents is that

the full class of multiple merger coalescents is in one-to-one relation with the uncountable and non-parametric set of finite measures Λ on $[0, 1]$, which drastically complicates statistical questions, while most of these coalescents do not appear to have a clear biological motivation in terms of a natural underlying population model. Considering the whole Lambda-coalescent class would also raise theoretical questions related to the unique identifyability of multiple merger coalescents on the basis of the SFS, related Myers *et al.* (2008) and Bhaskar and Song (2014), which is mathematically challenging and outside the scope of this work.

Hence, in case of the multiple merger coalescents, we restricted our attention to the class of Beta-coalescents (`B`) and the Dirac-coalescents (`D`). These classes appeared particularly interesting to us, since they both interpolate in a parametric way between the boundary points of Kingman-coalescent (`K`) via the Bolthausen-Sznitman coalescent (`B` at $\alpha = 1$) to the 'star-shaped-coalescent' (`B`, $\alpha = 0$; `D`, $\psi = 1$) - where the whole genealogy collapses to a single line in a single large merger event - among the multiple merger coalescents. Beta-coalescents have been studied frequently in the literature (see, eg., Birkner *et al.* (2005), Bertoin and Le Gall (2003), Hallatschek and Neher (2013), Steinrücken *et al.* (2013), Birkner *et al.* (2013b)) and are related to a population model with high fecundity and skewed offspring distribution (HFSOD), Schweinsberg (2003). Dirac-coalescents have been chosen for their simplicity from a mathematical standpoint and have also been investigated in Eldon and Wakeley (2006). The parameter of the Dirac coalescent has a clear interpretation as the fraction of the population that is replaced in each single HFSOD reproductive event.

For similar reasons, we restricted demographic scenarios to two basic parametric growth models. Exponential growth (`E`) is certainly a natural model in the presence of a super-critical branching population model without geographic or resource restrictions. Our second choice, algebraic growth model (`A`) appears maybe less natural, but can reflect situations in which there are spatial- or resource limitations and has been analysed in the mathematical literature, eg. Schweinsberg (2010). We refrain from more complicated scenarios, such as models with different epochs of exponential growth, and recent models including super-exponential growth (Reppell *et al.*, 2014), which indeed could be investigated with similar

methods.

Regarding statistical methodology, one could construct likelihood-based tests on the full two-dimensional parameter spaces for Π and θ, given by $\Theta^{\mathtt{E}}$ and $\Theta^{\mathtt{B}}$ as outlined in the methods section, however, this will likely yield considerable computational challenges. Instead, we opted to employ approximate likelihood-methods based on the 'fixed-s-method' as eg. done by Ramos-Onsins and Rozas (2002), reducing our test to a one-dimensional situation, where (14) does not depend on θ at all.

Based on this method, we derive an approximate likelihood-ratio test based on a Poissonisation of the SFS via (12) for interval hypotheses, including large ranges of parameters such as the growth parameter β in model (E) or the coalescent parameter α in model (B). By considering the power of our test, a key result in this setup is that, even for moderate sample size, (B) and (E) can be distinguished reasonably well for substantial parts of the parameter space of α and β.

A well-known criticism of this method is its sensitivity upon the true yet unknown coalescent mutation rate θ, cf. Markovtsova *et al.* (2001). We checked by rejection sampling (cf. the supplementary material), conditioning on $S = s$, that for various fixed values of (Π, θ) the rejection probability in our proposed test would be reasonably close to the 'true' rejection probability as long as the 'true' θ is close enough to the Watterson estimate $2s/\mathbb{E}^{\Pi}[B^{(n)}]$, in line with similar observations (for different test statistics) made by Wall and Hudson (2001).

Given additional information about the exact coalescent and growth parameters could lead one to test the point hypotheses (eg. (B) with fixed α vs. (E) with fixed β). Indeed, in this case, higher power can be achieved, even for relatively small numbers of segregating sites ($s = 20$), as expected (data not shown).

Distance plots (eg. Figure 3) over two-dimensional parameter ranges indicate a one-dimensional curve along which the minimal distance is reached. Note that both approaches (maximum-(approximate)-likelihood and minimum ℓ_2-distance) could be linked if asymptotic normality of our estimators could be established - this is a theoretical question for future

work.

Finally, we consider decision rules for the normalized spectrum $\zeta_i^{(n)}$ associated with models $\mathtt{A}, \mathtt{B}, \mathtt{E}, \mathtt{D}$ based on a simple rejection-based ABC analysis. More sophisticated techniques are available (see BEAUMONT (2010) for an overview) that may improve the prediction accuracy. Empirical misclassification probabilities show, for a reasonable sample size of $n = 200$, at least moderate success in distinguishing between the 4 model classes even for as few as $s = 15$ segregating sites. Note though that depending on the model class comparison to perform, reasonable error probabilities may only be achieved at higher mutation rates/a higher number s. This indicates that the genealogies produced by the different model classes (at least for suitable sample sizes) are different enough to be distinguished, but that mutation rates have to be high enough so that these differences are mirrored in the SFS.

In practice, our results could be used to design studies that allow to distinguish between different conjectured scenarios with suitable power. For example, in marine species, such as Atlantic cod (cf. eg. BIRKNER *et al.*, 2013b) and Pacific oysters (SARGSYAN and WAKELEY, 2008), it has been suggested that certain multiple merger coalescents could be more appropriate to describe underlying genealogies, and a reproductive mechansism (HFSOD) for population models has been proposed. We have put this to a test by performing an ABC model comparison between our four model classes for the Atlantic cod data of ÁRNASON (2004). The model comparison clearly rejected both Dirac coalescents and algebraic growth as potential models.

While our ABC analysis indicates exponential growth is slightly favored over the Beta$(2-\alpha, \alpha)$-coalescent, evidence is not really strong enough to rule out the latter model class. This may indicate that the SFS information of the ÁRNASON (2004) data does not have enough polymorphic sites to distinguish between $\mathtt{E}$ and $\mathtt{B}$ clearly enough, our posterior predictive checks revealed that likely neither model class explains the data completely. However, rejection of the $\mathtt{D}$ and $\mathtt{A}$ model classes suggest that models that predict star-shaped genealogies don't fit the data well. ÁRNASON (2004) uses a maximum likelihood estimation method (KUHNER *et al.*, 1998) and standard tests of neutrality (TAJIMA, 1989b; FU, 1997) to rule

out exponential population growth.

At this point we would like to point out that while our methods and results are exemplified in certain special coalescent and growth models, they could be modified to cover different frameworks.

Before ending the discussion we wish to comment on a few interesting side-issues that appeared during the analysis.

Non-monotonicity of the power function. At first glance, the observed non-monotonicity of the power function in the exponential growth parameter β when comparing with certain multiple merger coalescents (cf. e.g. Figures 2) may appear strange. However, the following example may suggest a heuristic way to understand such behaviour in a relatively simple special case. Suppose one wants to distinguish between an exponential growth model and a multiple merger coalescent with a substantial Kingman component and a small weight on large multiple mergers (eg. as in the Dirac-coalescent with ψ close to 1). That means that most of the time, the multiple merger coalescent will behave like a Kingman-coalescent (producing frequent binary mergers), but with small rate, comprehensive multiple mergers may occur. Certainly, when the growth parameter β is small, the exponential growth model will yield a pattern of variability close to a Kingman coalescent, and hence the power of a test to distinguish between both will be small if the Kingman component has weight close to 1. As β increases, the power to distinguish from a Kingman coalescent will increase, in line with intuition. However, as β becomes very large, lineages will coalesce after *very* short time in the exponential growth model. Such a scenario is certainly different from a Kingman coalescent, but could produce patterns of variability closer to a multiple merger coalescent with a drastic merger after very short time. Seeing such a merger in the very recent past has some cost (according to the weight Dirac component near 1), but appears more likely than observing large amounts of Kingman-like mergers within a unnaturally short time interval, thus leading to a relative decrease in power of associated test. This last effect is nicely illustrated by the upper-right scenario in Figure 4 in the case of large (algebraic) growth parameter.

Effect of 'lumping'. It is intriguing to see that using the complete nSFS as summary statistics in the ABC approach can yield higher errors than using intermediate (resp. strong) lumpings of the nSFS. A possible explanation is as follows. Consider the approximate likelihood function (12). Assume that the distribution of the SFS is approximately composed of independent Poisson distributions with parameter $\frac{\theta}{2}\mathbb{E}^{\Pi}\left[B_i^{(n)}\right]$ for $i \in [n-1]$. For a Poisson-distributed random variable X with parameter κ, we have $\frac{\sqrt{\mathrm{Var}(X)}}{\mathbb{E}(X)} = \frac{1}{\sqrt{\kappa}}$, thus showing that smaller Poisson parameters yield a higher amount of variation relative to their expected value. Hence, classes in the SFS with small underlying branch lengths (which tend to be in the right tail of the SFS) and/or a low mutation rate show relatively more variation compared to their contribution to the total number of mutations than those with longer branches or if the mutation rate is higher. Lumping such classes together, under (12), yields again a Poisson-distributed lumped class, but with Poisson parameter being the sum of parameters from the classes lumped together. Thus, the variation within this class relative to its contribution to the total number of mutations is reduced by lumping. If different coalescent models show different mean behaviour of (lumped) classes, lumping reduces noise and thus increases the chance to correctly identify the underlying model. Naturally, this effect is weakened by higher mutation rates and/or higher sample size n (e.g., consider the limit results for the SFS in Berestycki *et al.* (2013) and Kersting and Stanciu (2013)).

Thus, using an appropriate weighing of the variables in the nSFS (resp. SFS) should improve the power to distinguish between model classes. It would also be a worthwile future study to see whether a one-dimensional summary of the SFS similar to Tajima's D or Fay and Wu's H, as described in Achaz (2009), could yield similar or even higher power to distinguish between the model classes than the complete (possibly reweighted) nSFS.

Acknowledgements: J. Blath and B. Eldon were supported by Deutsche Forschungsgemeinschaft (DFG) grant BL 1105/3-1, and M. Birkner by DFG grant BI 1058/2-1, as parts of SPP Priority Programme 1590. F. Freund thanks Luca Ferretti and Guillaume Achaz (SMILE, Collège de France, Paris) for discussions about the site-frequency spectrum.

We would like to thank two anonymous referees whose insightful comments and constructive criticism helped to improve the presentation.

References

Abramowitz, M., and I. A. Stegun, editors, 1964 *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. Number 55 in Applied Mathematics Series. National Bureau of Standards, Washington, D.C.

Achaz, G., 2009 Frequency spectrum neutrality tests: one for all and all for one. Genetics **183**: 249–258.

Árnason, E., 2004 Mitochondrial cytochrome b variation in the high-fecundity Atlantic cod: trans-Atlantic clines and shallow gene genealogy. Genetics **166**: 1871–1885.

Baragatti, M., and P. Pudlo, 2014 An overview on approximate bayesian computation. ESAIM **44**: 291–299.

Beaumont, M. A., 2010 Approximate bayesian computation in evolution and ecology. Annual Review of Ecology, Evolution, and Systematics **41**: 379–406.

Beckenbach, A. T., 1994 Mitochondrial haplotype frequencies in oysters: neutral alternatives to selection models. In B. Golding, editor, *Non-Neutral Evolution*. Chapman & Hall, New York, 188–198.

Berestycki, J., N. Berestycki, and V. Limic, 2013 A sampling formulae for Lambda-coalescents. To appear (arXiv:1201.6512).

Berestycki, N., 2009 Recent progress in coalescent theory. Ensaios Mathématicos **16**: 1–193.

Bertoin, J., and J.-F. Le Gall, 2003 Stochastic flows associated to coalescent processes. Probab. Theory Related Fields **126**: 261–288.

Bhaskar, A., and Y. Song, 2014 Descartes' rule of signs and the identifiability of population demographic models from genomic variation data. Ann Statist **42**: 2469–2493.

Birkner, M., and J. Blath, 2008 Computing likelihoods for coalescents with multiple collisions in the infinitely many sites model. J Math Biol **57**: 435–465.

Birkner, M., J. Blath, M. Capaldo, A. Etheridge, M. Möhle, *et al.*, 2005 Alpha-stable branching and beta-coalescents. Electron. J. Probab. **10**: no. 9, 303–325 (electronic).

Birkner, M., J. Blath, and B. Eldon, 2013a An ancestral recombination graph for diploid populations with skewed offspring distribution. Genetics **193**: 255–290.

Birkner, M., J. Blath, and B. Eldon, 2013b Statistical properties of the site-frequency spectrum associated with lambda-coalescents. Genetics **195**: 1037–1053.

Birkner, M., J. Blath, and M. Steinrücken, 2011 Importance sampling for lambda-coalescents in the infinitely many sites model. Theor Popul Biol **79**: 155–173.

Cannings, C., 1974 The latent roots of certain Markov chains arising in genetics: a new approach. I. Haploid models. Adv Appl Prob **6**: 260–290.

Cannings, C., 1975 The latent roots of certain Markov chains arising in genetics: a new approach. II. Further haploid models. Adv Appl Prob **7**: 264–282.

Carr, S. M., and H. D. Marshall, 2008 Intraspecific phylogeographic genomics from multiple complete mtDNA genomics in Atlantic cod (*Gadus morhua*): origins of 'codmother', transatlantic vicariance, and midglacial population expansion. Genetics **180**: 381–389.

Chen, W.-C., 2011 *Overlapping Codon Model, Phylogenetic Clustering, and Alternative Partial Expectation Conditional Maximization Algorithm.* Ph.D. thesis, Iowa State University, Ames, Iowa. Http://gradworks.umi.com/34/73/3473002.html.

Csilléry, K., O. François, and M. G. B. Blum, 2012 ABC: an R package for approximate bayesian computation (ABC). Methods in Ecology and Evolution **3**: 475–479.

Cucala, L., and J. Marin, 2013 Bayesian inference on a mixture model with spatial dependence. J Comp Graph Stats **22**: 584–597.

Depaulis, F., S. Mousset, and M. Veuille, 2001 Haplotype tests using coalescent simulations conditional on the number of segregating sites. Molecular biology and evolution **18**: 1136–1138.

Depaulis, F., and M. Veuille, 1998 Neutrality tests based on the distribution of haplotypes under an infinite-site model. Molecular Biology and Evolution **15**: 1788.

Donnelly, P., and T. G. Kurtz, 1999 Particle representations for measure-valued population models. Ann Probab **27**: 166–205.

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

Durrett, R., and J. Schweinsberg, 2005 A coalescent model for the effect of advantageous mutations on the genealogy of a population. Stoch Proc Appl **115**: 1628–1657.

Eldon, B., 2011 Estimation of parameters in large offspring number models and ratios of coalescence times. Theor Popul Biol **80**: 16–28.

Eldon, B., and J. Wakeley, 2006 Coalescent processes when the distribution of offspring number among individuals is highly skewed. Genetics **172**: 2621–2633.

Fay, J. C., and C.-I. Wu, 2000 Hitchhiking under positive darwinian selection. Genetics **155**: 1405–1413.

Fu, Y. X., 1995 Statistical properties of segregating sites. Theor Popul Biol **48**: 172–197.

Fu, Y. X., 1997 Statistical tests of neutrality against population growth, hitchhiking, and background selection. Genetics **147**: 915–925.

Fu, Y. X., and W. H. Li, 1993 Statistical tests of neutrality of mutations. Genetics **133**: 693–709.

Galassi, M., J. Davies, J. Theiler, B. Gough, G. Jungman, *et al.*, 2013 *GNU Scientific Library Reference Manual*, third edition. ISBN 0954612078.

Griffiths, R. C., and S. Tavaré, 1998 The age of a mutation in a general coalescent tree. Comm Statistic Stoch Models **14**: 273–295.

Hallatschek, O., and R.-H. Neher, 2013 Genealogies of rapidly adapting populations. PNAS **110(2)**: 437–442.

Hedgecock, D., and A. I. Pudovkin, 2011 Sweepstakes reproductive success in highly fecund marine fish and shellfish: a review and commentary. Bull Mar Sci **87**: 971–1002.

Hein, J., M. H. Schierup, and C. Wiuf, 2005 *Gene genealogies, variation and evolution*. Oxford University Press, Oxford, UK.

Hudson, R. R., 1983a Properties of a neutral allele model with intragenic recombination. Theor Popul Biol **23**: 183–201.

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

Hudson, R. R., 1990 Gene genealogies and the coalescent process. In D. J. Futuyma and J. Antonovics, editors, *Oxford surveys in evolutionary biology*, volume 7. Oxford University Press, Oxford, 1–44.

Hudson, R. R., 2002 Generating samples under a wright-fisher neutral model. Bioinformatics **18**: 337–338.

Jeffreys, H., 1961 *Theory of Probability*. Oxford University Press, Oxford, UK, 3rd edition.

Kaj, I., and S. Krone, 2003 The coalescent process in a population with stochastically varying size. J Appl Probab **40**: 33–48.

Kass, R. E., and A. E. Raftery, 1995 Bayes factors. Journal of the American Statistical Association **90**: 773–795.

Kernighan, B. W., and D. M. Ritchie, 1988 *The C programming language*. Prentice Hall, Englewood Cliffs, New Jersey, second edition.

Kersting, G., and I. Stanciu, 2013 The internal branch lengths of the Kingman coalescent. To appear (arXiv:1303.4562).

Kim, J., E. Mossel, M. Z. Rácz, and N. Ross, 2014 Can one hear the shape of a population history? To appear (arXiv preprint arXiv:1402.2424) .

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

Kingman, J. F. C., 1982b Exchangeability and the evolution of large populations. In G. Koch and F. Spizzichino, editors, *Exchangeability in probability and statistics*. North-Holland, Amsterdam, 97–112.

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

Koskela, J., P. Jenkins, and D. Spanò, 2013 Computational inference beyond Kingman's coalescent. submitted .

Kuhner, M. K., J. Yamato, and J. Felsenstein, 1998 Maximum likelihood estimation of population growth rates based on the coalescent. Genetics **149**: 429–434.

Markovtsova, L., P. Marjoram, and S. Tavaré, 2001 On a test of Depaulis and Veuille. Molecular biology and evolution **18**: 1132–1133.

Milgram, M. S., 1985 The generalized integro-exponential function. Math Comp **44**: 443–458.

Möhle, M., 1998 Robustness results for the coalescent. J Appl Probab **35**: 438–447.

Möhle, M., and S. Sagitov, 2001 Classification of coalescent processes for haploid exchangeable coalescent processes. Ann Probab **29**: 1547–1562.

Myers, S., C. Fefferman, and N. Patterson, 2008 Can one learn history from the allelic spectrum? Theor Popul Biol **73**: 342–348.

Neher, R. A., and O. Hallatschek, 2013 Genealogies of rapidly adapting populations. Proceedings of the National Academy of Sciences **110**: 437–442.

Nordborg, M., 2001 Coalescent theory. In D. J. Balding, M. J. Bishop and C. Cannings, editors, *Handbook of statistical genetics*, chapter 25. John Wiley & Sons, Chichester, UK, 2nd edition, 179–212.

Pitman, J., 1999 Coalescents with multiple collisions. Ann Probab **27**: 1870–1902.

Polanski, A., A. Bobrowski, and M. Kimmel, 2003 A note on distribution of times to coalescence, under time-dependent population size. Theor Popul Biol **63**: 33–40.

Polanski, A., and M. Kimmel, 2003 New explicit expressions for relative frequencies of single-nucleotide polymorphisms with application to statistical inference on population growth. Genetics **165**: 427–436.

Pritchard, J. K., M. T. Seielstad, A. Perez-Lezaun, and M. W. Feldman, 1999 Population growth of human Y chromosomes: a study of Y chromosome microsatellites. Mol Biol Evol **16**: 1791–1798.

R Core Team, 2012 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Ramírez-Soriano, A., S. E. Ramos-Onsins, J. Rozas, F. Calafell, and A. Navarro, 2008 Statistical power analysis of neutrality tests under demographic expansions, contractions and bottlenecks with recombination. Genetics **179**: 555–567.

Ramos-Onsins, S. E., and J. Rozas, 2002 Statistical properties of new neutrality tests against population growth. Mol Biol Evol **19**: 2092–2100.

Reppell, M., M. Boehnke, and S. Zöllner, 2014 The impact of accelerating faster than exponential population growth on genetic variation. Genetics **196**: 819–828.

Rödelsperger, C., R. A. Neher, A. M. Weller, G. Eberhardt, H. Witte, *et al.*, 2014 Characterization of genetic diversity in the nematode pristionchus pacificus from population-scale resequencing data. Genetics **196**: 1153–1165.

Rogers, A. R., and H. C. Harpending, 1992 Population growth makes waves in the distribution of pairwise genetic differences. Mol Biol Evol **9**: 552–569.

Rubin, D. B., 1984 Bayesian justifiable and relevant frequency calculations for the applied statistician. Ann Stats **12**: 1151–1172.

Sagitov, S., 1999 The general coalescent with asynchronous mergers of ancestral lines. J Appl Probab **36**: 1116–1125.

Sano, A., and H. Tachida, 2005 Gene genealogy and properties of test statistics of neutrality under population growth. Genetics **169**: 1687–1697.

Sargsyan, O., and J. Wakeley, 2008 A coalescent process with simultaneous multiple mergers for approximating the gene genealogies of many marine organisms. Theor Popul Biol **74**: 104–114.

Sawyer, S. A., and D. L. Hartl, 1992 Population genetics of polymorphism and divergence. Genetics **132**: 1161–1176.

Schweinsberg, J., 2003 Coalescent processes obtained from supercritical Galton-Watson processes. Stoch Proc Appl **106**: 107–139.

Schweinsberg, J., 2010 The number of small blocks in exchangeable random partitions. ALEA Lat. Am. J. Probab. Math. Stat. **7**: 217–242.

Simonsen, K. L., G. A. Churchill, and C. F. Aquadro, 1995 Properties of statistical tests of neutrality for dna polymorphism data. Genetics **141**: 413–429.

Slatkin, M., and R. R. Hudson, 1991 Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. Genetics **129**: 555–562.

Steinrücken, M., M. Birkner, and J. Blath, 2013 Analysis of DNA sequence variation within marine species using beta-coalescents. Theor Popul Biol **87**: 15–24.

Stoehr, J., P. Pudlo, and L. Cucala, 2014 Geometric summary statistics for ABC model choice between hidden Gibbs random fields. arXiv:1402.1380 .

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

Tajima, F., 1989a The effect of change in population size on DNA polymorphism. Genetics **123**: 597–601.

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

Tavaré, S., D. J. Balding, R. C. Griffiths, and P. Donnelly, 1997 Inferring coalescence times from DNA sequence data. Genetics **145**: 505–518.

Tellier, A., and C. Lemaire, 2014 Coalescence 2.0: a multiple branching of recent theoretical developments and their applications. Mol Ecol **23**: 2637–2652.

Wakeley, J., 2007 *Coalescent theory*. Roberts & Co, Greenwood Village.

Wall, J. D., and R. R. Hudson, 2001 Coalescent simulations and statistical tests of neutrality. Molecular biology and evolution **18**: 1134–1135.

Watterson, G. A., 1975 On the number of segregating sites in genetical models without recombination. Theor Pop Biol **7**: 1539–1546.

Supporting Information

Can the site-frequency spectrum distinguish exponential population growth from multiple-merger coalescents?

B. Eldon, M. Birkner, J. Blath, F. Freund

Multiple merger-coalescents and the model classes K, B and D

Recall that a multiple merger- or Lambda-coalescent, formally introduced by PITMAN (1999); SAGITOV (1999), and DONNELLY and KURTZ (1999), is a partition-valued exchangeable coalescent process determined by a finite measure Λ on $[0,1]$ which governs the dynamics of the process: If there are currently b blocks in the partition (i.e. b active ancestral lineages), k out of them merge at rate

$$\lambda_{b,k} = \int_{[0,1]} x^{k-2}(1-x)^{b-k}\Lambda(dx), \quad k = 2,..,b. \tag{S1}$$

For an overview of the theory see e.g. BERESTYCKI (2009) or, with a biological perspective, TELLIER and LEMAIRE (2014). When Λ is associated with the beta-distribution with parameters $2-\alpha$ and α for $1 \leq \alpha < 2$ (SCHWEINSBERG, 2003), these rates can be given explicitly by

$$\lambda_{b,k} = \frac{B(k-\alpha, b-k+\alpha)}{B(2-\alpha,\alpha)},$$

where $B(\cdot,\cdot)$ is the classical Beta-function. Such coalescents will be called Beta-coalescents, and constitute the model class B.

When Λ is associated with the Dirac coalescent (ELDON and WAKELEY, 2006), that is, $\Lambda(dx) = \delta_{\{\psi\}}(dx)$, for $\psi \in [0,1]$, we are in class D. Here, for $\psi \in (0,1]$, the rates are given by

$$\lambda_{b,k} = \frac{\psi^k(1-\psi)^{b-k}}{\psi^2}.$$

Both classes intersect in the Kingman coalescent (model K), which corresponds to $\alpha = 2$ and $\psi = 0$, and of course has coalescence rates

$$\lambda_{b,k} = \begin{cases} 1 & \text{if } k = 2, \\ 0 & \text{else,} \end{cases}$$

ie. only binary mergers are allowed. The Beta- and the Dirac coalescent each introduce a *coalescent* parameter (α, ψ), which can be estimated from genetic data (ELDON, 2011;

Birkner *et al.*, 2013b; Birkner and Blath, 2008; Steinrücken *et al.*, 2013).

Population models leading to coalescent classes K, B and D

It is well-known that the classical Wright-Fisher and the Moran model have scaling limits whose genealogy is described by a Kingman coalescent. For the more general Lambda-coalescents, Möhle and Sagitov (2001) give a full classification of all Cannings models that lead to any given Lambda-coalescent. The relevant time-scaling is determined by c_N, the probability that in a population of size N, two distinct ancestral lineages merge in the previous generation. It is important to keep in mind that many different population models can lead to the same limiting coalescent, and also that the timescale, determined by c_N, may vary between different models having the same limit. For the Kingman coalescent, the classical Wright Fisher model converges on the time-scale $c_N = 1/N$, whereas for the Moran model, it is of order $1/N^2$.

A popular model that leads to the Beta$(2-\alpha, \alpha)$-coalescent has been introduced by Schweinsberg (2003). For this model, the relevant time-scale is of order $1/N^{\alpha-1}$. Here, single individuals can produce positive fractions of the next generation in a single reproductive event (an instance of 'HFSOD') that can be related to stable branching processes, cf. Birkner *et al.* (2005). The size of the reproductive event is random and governed by the Beta-distribution. For details we refer to Schweinsberg (2003), and for a discussion of its biological relevance eg. to Steinrücken *et al.* (2013).

The Dirac coalescent has been investigated in Eldon and Wakeley (2006). It has a particularly simple interpretation: Given the coalescent parameter $\psi \in (0, 1]$, in each 'substantial' reproductive event, a fraction of $100 \cdot \psi\%$ of the generation die and are replaced by the offspring of a single parent (there can be other, 'non-substantial' reproductive events which, though potentially frequent, become invisible in the limit). This is an extreme case of HFSOD, and biologically it seems difficult to justify why the fraction ψ should always be the same. However, it is mathematically simple and interpolates between the Kingman coalescent $\psi = 0$ and the star-shaped coalescent $\psi = 1$, thus we included it in our study.

For details see ELDON and WAKELEY (2006).

Population with varying population size and the classes E and A

In KAJ and KRONE (2003), a time-changed n-coalescent under a general model of variable population size is derived. More precisely, the authors consider a haploid Wright-Fisher model with population size N at generation $r = 0$ and consider a population size process $M_N(r), r \in \mathbb{Z}$ of the form $M_N(r) = NX_N(r), r \in \mathbb{Z}$, that is, $X_N(r)$ describes the 'relative population size' at generation r. Under the assumption that $X_N(\lfloor Nt \rfloor), t \in \mathbb{R}$ converges to something non-degenerate (ie. bounded away from 0 and ∞), they get the well-known limiting result that a time-changed Kingman coalescent describes the genealogy, where the infinitesimal coalescence rates are given by $1/\nu(s)$, with

$$\nu(s) = \lim_{N\to\infty} X_N(\lfloor Ns \rfloor). \tag{S2}$$

Our exponential growth model E corresponds to a Kingman-coalescent with exponentially growing coalescence rates $\nu(s) = e^{\beta s}$, for $\beta \geq 0$, and can be obtained from a a growth rate of β/N per generation in the pre-limiting model, ie. $N_k = N(1 + \beta/N)^k$. Indeed,

$$\nu(t) = \lim_{N\to\infty} X_N(\lfloor Nt \rfloor) = \lim_{N\to\infty} \left(1 + \frac{\beta}{N}\right)^{Nt} = e^{\beta t}.$$

Thus, the size Nt generations ago is approximately $Ne^{-\beta t}$.

The model class A is given by Kingman-coalescents with algebraically growing coalescence rates, ie. $\nu(s) = s^\gamma$, for $\gamma \geq 0$. Note that if $\gamma = 0$ or $\beta = 0$, we recover the Kingman coalescent and are back in class K.

A population model for algebraic growth was considered in (SCHWEINSBERG, 2010, Section 1.4): Fix a population size N at the present generation 0, and for notational convenience also for generation -1 (this short period of constant population size will become irrelevant after time-rescaling). For a fixed growth parameter $\gamma > 0$, the population size at the k-th generation before the present (for $k \in \mathbb{N}$) is assumed to be $\lceil Nk^{-\gamma} \rceil$. Measuring time in units

of size $N^{\frac{1}{1+\gamma}}$ yields the limiting infinitesimal coalescence rate

$$\nu(t) = \lim_{N\to\infty} N^{\frac{1}{1+\gamma}} c_N(t,\gamma) = \lim_{N\to\infty} N^{\frac{1}{1+\gamma}} \frac{(N^{\frac{1}{1+\gamma}} t)^\gamma}{N} = t^\gamma,$$

where $c_N(t,\gamma)$ is the probability that two individuals in generation $N^{\frac{1}{1+\gamma}} t$ choose the same ancestor (uniformly out of the $N(N^{\frac{1}{1+\gamma}} t)^{-\gamma}$ individuals alive in that generation). Consider the time-change (for the scaling limit as $N \to \infty$)

$$T_t := \frac{t^{\gamma+1}}{\gamma+1} = \int_0^t s^\gamma \, ds.$$

Then, the genealogy of the algebraic growth model at previous generation t equals in law the state of a classical Kingman coalescent at time T_t. See Schweinsberg (2010) for details.

The expected SFS under variable population size

The effect of fluctuations in population size on the SFS has been investigated in various articles, see eg. Griffiths and Tavaré (1998), who derive an analog of (1), and Kaj and Krone (2003) who link the Wright-Fisher approximation (with fluctuating population size) with the limiting genealogy.

Recursions for the expected values and covariances of the site-frequency spectrum associated with moderate fluctuations in population size will now be briefly discussed. We will in particular consider numerically tractable recursions for the model classes `E` and `A`, based on work by Polanski *et al.* (2003) and Polanski and Kimmel (2003).

Consider a time-inhomogeneous Kingman coalescent, started in n lineages, where each pair of lines present at time $t \geq 0$ merges at a rate $\nu(t)$. Then, the expected frequency spectrum $\mathbb{E}[\xi_i^{(n),\nu}], i \in [n-1]$, is again of the form (1), and the time-change ν enters only in the distribution of the $T_k^{(n)} = T_k^{(n),\nu}$, $2 \leq k \leq n$, that is, the distribution of the lengths of the time intervals of the block-counting process $Y_t^{(n),\nu}$ during which there are exactly k lineages.

To evaluate $\mathbb{E}[\xi_i^{(n),\nu}]$ one needs information about $\mathbb{E}[T_k^{(n),\nu}]$. Define

$$S_j^{(n),\nu} := T_n^{(n),\nu} + T_{n-1}^{(n),\nu} + \cdots + T_j^{(n),\nu}, \quad j = n, \ldots, 2 \tag{S3}$$

to be the time at which the block counting process $Y^{(n),\nu}$ jumps from j to $j-1$ lineages (with the convention $S_{n+1}^{(n),\nu} := 0$). Abbreviate, for $t \geq 0$ and $j \in 2, \ldots, n$,

$$F(t) := \int_0^t \nu(u)\, du \quad \text{and} \quad a_j^{(\vartheta)} := \int_0^\infty e^{-\binom{j}{2}F(s)}\, ds, \tag{S4}$$

assuming that the first integral in (S4) is finite. It is possible to compute the marginal density of $S_m^{(n),\nu}$ using the well-known fact that the density of a convolution of exponentials with different rates can be written as a linear combination of exponential densities,

$$\mathbb{E}\big[S_m^{(n),\nu}\big] = \sum_{j=m}^{n} c_m^{(j,n)} a_j^{(\vartheta)}\,, \tag{S5}$$

where

$$c_m^{(j,n)} := \prod_{\substack{m \leq i \leq n \\ i \neq j}} \frac{\binom{i}{2}}{\binom{i}{2} - \binom{j}{2}} = (-1)^{j-m} \frac{(2j-1)m}{j(j-1)} \frac{\binom{n}{j}\binom{j+m-2}{j}\binom{j}{m}}{\binom{n+j-1}{j}}, \tag{S6}$$

(put $c_m^{(j,n)} = 0$ for $j < m$).

Polanski and Kimmel (2003) obtain numerically stable and efficient recursions to compute $\mathbb{E}^\Pi\left[B_i^{(n)}\right]$ associated with any time-changed Kingman coalescent Π as follows. For ϑ denoting the growth parameter associated with process Π,

$$\mathbb{E}^\Pi\left[B_i^{(n)}\right] = \sum_{j=2}^{n} W_{i,j}^{(n)} a_j^{(\vartheta)} \tag{S7}$$

where the constants $W_{i,j}^{(n)}$ can be computed recursively (Polanski and Kimmel, 2003);

$$\begin{aligned} W_{i,2}^{(n)} &= \frac{6}{n+1}, \\ W_{i,3}^{(n)} &= \frac{30(n-2i)}{(n+1)(n+2)}, \\ W_{i,j+2}^{(n)} &= \frac{(3+2j)(n-2i)}{j(n+j+1)} W_{i,j+1}^{(n)} - \frac{(1+j)(3+2j)(n-j)}{j(2j-1)(n+j+1)} W_{i,j}^{(n)}. \end{aligned} \tag{S8}$$

We now specify the main ingredient $a_j^{(\vartheta)}$ (depending on $F(t), t \geq 0$ and hence $\nu(t), t \geq 0$) explicitly for two important special cases:

a) Exponential growth. In the case of an exponentially growing population with growth parameter β, that is, $\nu(t) = e^{\beta t}$, we have

$$a_j^{(\beta)} = \frac{1}{\beta} \exp\left(\beta^{-1}\tbinom{j}{2}\right) E_1\left(\beta^{-1}\tbinom{j}{2}\right), \tag{S9}$$

where

$$E_1(t) := \int_t^\infty \frac{e^{-x}}{x}\, dx = \int_1^\infty \frac{e^{-tx}}{x}\, dx \tag{S10}$$

is an exponential integral function, c.f. e.g. (Abramowitz and Stegun, 1964, 5.1.1). One can use numerical integration schemes to compute $E_1(t)$ for smaller values of t (eg. $t < 50$). For larger values of t, one can use the approximation

$$E_1(t) = t^{-1} e^{-t} \sum_{k=0}^{K-1} k!(-t)^{-k}$$

(Milgram, 1985), which has error of order $O\left(K! t^{-K}\right)$.

b) Algebraic ('power law') growth. In the case of algebraic growth of the form

$\nu(t) = t^\gamma$ for some $\gamma > 0$, we have

$$a_j^{(\gamma)} = \frac{\Gamma\big(1/(\gamma+1)\big)}{(1+\gamma)^{\gamma/(\gamma+1)}} \binom{j}{2}^{-1/(\gamma+1)}. \tag{S11}$$

Based on Equation (23) in Fu (1995), it is also possible to compute the *variance* and the *covariances* of the SFS based on expressions for $\mathbb{E}_\nu[T_k^{(n),\nu} T_l^{(n),\nu}]$, $2 \le k, l \le n$, which in turn can be obtained from

$$\mathbb{E}_\nu[T_k^{(n),\nu} T_l^{(n),\nu}] = \mathbb{E}_\nu[S_k^{(n),\nu} S_l^{(n),\nu}] - \mathbb{E}_\nu[S_{k-1}^{(n),\nu} S_l^{(n),\nu}] - \mathbb{E}_\nu[S_k^{(n),\nu} S_{l-1}^{(n),\nu}] + \mathbb{E}_\nu[S_{k-1}^{(n),\nu} S_{l-1}^{(n),\nu}],$$

noting that, in the above notation,

$$\mathbb{E}\big[(S_m^{(n),\nu})^2\big] = \int_0^\infty s_m^2 \sum_{j=m}^n c_m^{(j,n)} \nu(s_m) \binom{j}{2} e^{-\binom{j}{2} F(s_m)}\, ds_m,$$

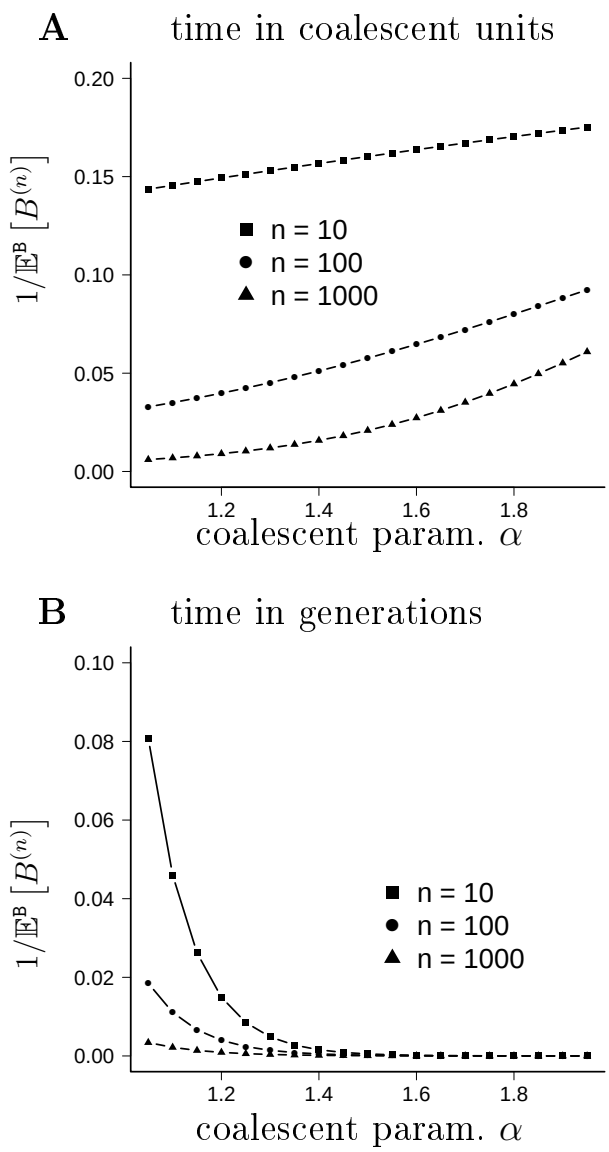
and

$$\mathbb{E}[S_m^{(n),\nu} S_k^{(n),\nu}] = \mathbb{E}\big[\mathbb{E}[S_m^{(n),\nu} | S_k^{(n),\nu}] S_k^{(n),\nu}\big],$$

where $\mathbb{E}[S_m^{(n),\nu} | S_k^{(n),\nu} = s_k]$ can be computed (it is the expectation under a regular conditional probability) as in (S5) replacing ν by $\tilde{\nu}(\cdot) := \nu(\cdot + s_k)$, $c_m^{(j,n)}$ by $\tilde{c}_m^{(j)} := c_m^{(j,k)}$ and F by $\tilde{F}(\cdot) = F(s_k + \cdot) - F(s_k)$.

The estimate $1/\mathbb{E}^{\mathsf{B}}\left[B^{(n)}\right]$ as a function of α

Figure S1: Graphs of $1/\mathbb{E}^{\mathsf{B}}\left[B^{(n)}\right]$, the estimated value of $\theta/2$ per observed mutation when using the Watterson estimator (3) as a function of α (**A**), compare with (3); and the estimated value of μ per observed mutation (**B**), using (5) together with (3), and assuming the timescale $c_N = N^{1-\alpha}$. The number of leaves n are as shown. In **B**, time is converted into generations by multiplying $\mathbb{E}^{\mathsf{B}}\left[B^{(n)}\right]$ with $N^{\alpha-1}$, when $N = 10^5$.



Approximate insensitivity of the expected normalized SFS w.r.t. θ

In this section, we argue that for a random genealogical tree $\mathcal{T}$ with n leaves whose law is governed by a given coalescent mechanism Π, the expected nSFS $\mathbb{E}^{\Pi,\theta}\left[\left(\zeta_1^{(n)},\ldots,\zeta_{n-1}^{(n)}\right)\right]$ when the coalescent mutation rate is $\theta > 0$ is approximately constant as a function of θ. This is useful because it in a sense allows to "factor out" (i.e, ignore) the mutation rate parameter from a test problem when comparing different Π's. This also means that – at least when the observed number $|\xi^{(n)}|$ of segregating sites is reasonably large – the exact observed value $|\xi^{(n)}|$ does not add much additional information for tests based on the SFS.

Indeed, we can compute

$$\begin{aligned}
\mathbb{E}^{\Pi,\theta}\left[\zeta_i^{(n)}\right] &= \mathbb{E}^{\Pi,\theta}\left[\zeta_i^{(n)} 1_{\{|\xi^{(n)}|>0\}}\right] = \mathbb{E}^{\Pi,\theta}\left[\mathbb{E}^{\Pi,\theta}\left[\zeta_i^{(n)} 1_{\{|\xi^{(n)}|>0\}} \,\middle|\, \mathcal{T}\right]\right] \\
&= \mathbb{E}^{\Pi,\theta}\left[\mathbb{P}^{\Pi,\theta}(|\xi^{(n)}|>0\,|\,\mathcal{T})\frac{\mathbb{E}^{\Pi,\theta}\left[\zeta_i^{(n)} 1_{\{|\xi^{(n)}|>0\}} \,\middle|\, \mathcal{T}\right]}{\mathbb{P}^{\Pi,\theta}(|\xi^{(n)}|>0\,|\,\mathcal{T})}\right] \\
&= \mathbb{E}^{\Pi}\left[\left(1-e^{-\frac{\theta}{2}\sum_{i=1}^{n-1}B_i^{(n)}}\right)\frac{\frac{\theta}{2}\cdot B_i^{(n)}}{\frac{\theta}{2}\sum_{i=1}^{n}B_i^{(n)}}\right] \\
&= \mathbb{E}^{\Pi}\left[\frac{B_i^{(n)}}{\sum_{i=1}^{n}B_i^{(n)}}\right] - \mathbb{E}^{\Pi}\left[e^{-\frac{\theta}{2}\sum_{i=1}^{n-1}B_i^{(n)}}\frac{B_i^{(n)}}{\sum_{i=1}^{n}B_i^{(n)}}\right]. \qquad \text{(S12)}
\end{aligned}$$

Here, $B_i^{(n)}$ denotes the total length of all branches in $\mathcal{T}$ which subtend i leaves for $i = 1,\ldots,n-1$ and in the third line we used Lemma S1.1 below together with the fact that given $\mathcal{T}$ and θ, $\xi_i^{(n)}$, $i = 1,\ldots,n-1$ are independent and each $\xi_i^{(n)}$ is Poisson distributed with mean $\frac{\theta}{2}B_i^{(n)}$. Note that the first term in (S12) is independent of θ and the "correction" term is small unless θ is very small or $L_n := \sum_{i=1}^{n-1} B_i^{(n)}$, the total length of $\mathcal{T}$, is small under Π with substantial probability. Note that for each of the coalescent processes we consider in this investigation, it does hold that $L_n \to \infty$ as $n \to \infty$. Simulations also indicate that the distribution (not only the mean) of $\zeta_i^{(n)}$ does not depend much on θ (data not shown).

Lemma S1.1. *Let X_1, X_2 be independent Poisson-distributed variables with parameters a*

and b. Then,

$$\mathbb{E}\left[\frac{X_1}{X_1+X_2}\,\middle|\,(X_1+X_2)>0\right]=\frac{a}{a+b}.$$

Proof. X_1+X_2 as a sum of independent Poisson distributed random variables is again Poisson distributed with parameter $a+b$. We have

$$P(X_1=k,X_2=m-k|(X_1+X_2)>0)=\frac{P(X_1=k)P(X_2=m-k)}{P(X_1+X_2>0)}=\frac{a^k b^{m-k}}{k!(m-k)!}\frac{e^{-(a+b)}}{1-e^{-(a+b)}}$$

for $k\in\mathbb{N}_0$, $m\in\mathbb{N}$ with $k\leq m$. We compute

$$\begin{aligned}
&\mathbb{E}\left[\frac{X_1}{X_1+X_2}\,\middle|\,(X_1+X_2)>0\right]=\sum_{m=1}^{\infty}\sum_{k=0}^{m}\frac{k}{m}\frac{a^k b^{m-k}}{k!(m-k)!}\frac{e^{-(a+b)}}{1-e^{-(a+b)}}\\
=&\frac{e^{-(a+b)}}{1-e^{-(a+b)}}\sum_{m=1}^{\infty}\frac{a}{m(m-1)!}\sum_{k=1}^{m}\frac{(m-1)!}{(k-1)!((m-1)-(k-1))!}a^{k-1}b^{(m-1)-(k-1)}\\
=&\frac{e^{-(a+b)}}{1-e^{-(a+b)}}\frac{a}{a+b}\sum_{m=1}^{\infty}\frac{(a+b)^m}{m!}=\frac{e^{-(a+b)}}{1-e^{-(a+b)}}(e^{(a+b)}-1)\frac{a}{a+b}=\frac{a}{a+b}
\end{aligned}$$

□

Insensitivity of the fixed-s-method w.r.t. θ

To check the the robustness of our fixed-s-method against varying θ under rejection sampling (cf. e.g. Markovtsova *et al.* (2001), Wall and Hudson (2001)), we applied the following exact rejection sampling approach to simulate a coalescent tree conditional on a given number of observed segregating sites s. As input, the algorithm takes sample size n, number of segregating sites s, a coalescent model Π, and mutation rate θ, and returns a realisation of $\underline{\xi}^{(n)}$ with $|\underline{\xi}^{(n)}| = s$.

Rejection sampling algorithm :

(i) generate a coalescent tree according to Π, read off branch lengths $B_i^{(n)}$,

(ii) draw a total number of mutations S as realization of a Poisson random variable with parameter $(\theta/2)\sum_i B_i^{(n)}$,

(iii) if $S = s$ the required fixed number of segregating sites, keep the $B_i^{(n)}$, otherwise discard and draw again,

(iv) throw uniformly s mutations on the tree, $B_i^{(n)}$, so that the probability of a mutation falling into class i is $B_i^{(n)}/(\sum_i B_i^{(n)})$.

We then computed (approximately via rejection-sampling) the size of a conditional distribution based test if one employs quantiles of the fixed-s-method derived from (8). Of course, the hope is that both are reasonably close to each other, and this seems to hold relatively well if θ is close to the Watterson estimate $\hat{\theta}(\Pi, s)$ (9). In particular, the results (Tables (S1)–(S3)) show that the method is particularly robust against varying θ when exponential growth is taken as null model.

Table S1: Checking size of test given fixed-s quantiles associated with size $x\%$ and $\alpha \in \{1, 1.5\}$ with Beta$(2-\alpha, \alpha)$-coalescent as null model, and exponential growth as alternative, using rejection sampling with mutation rate θ as shown. Sample size $n = 100$, segregating sites $s = 50$. The estimate $(\theta_W(\alpha))$ is obtained from (9). All estimates from 10^5 iterates.

α	$x\%$	θ $(\theta_W(\alpha))$	size of test
1.0	10%	3.082453 $(\theta_W(1))$	0.10
		2.0	0.13
		3.0	0.10
		5.0	0.07
		7.0	0.06
	5%	3.082453 $(\theta_W(1))$	0.05
		2.0	0.07
		5.0	0.03
		7.0	0.02
	1%	3.082453 $(\theta_W(1))$	0.01
		2.0	0.02
		5.0	0.01
		7.0	0.002
1.5	10%	5.7638 $(\theta_W(1.5))$	0.11
		3.0	0.03
		5.0	0.09
		7.0	0.11
		10.0	0.13
	5%	5.7638 $(\theta_W(1.5))$	0.05
		3.0	0.01
		5.0	0.04
		7.0	0.06
		10.0	0.07
	1%	5.7638 $(\theta_W(1.5))$	0.01
		3.0	0.001
		5.0	0.01
		7.0	0.01
		10.0	0.02

Table S2: Checking size of test given fixed-s quantiles associated with size $x\%$ as shown using rejection sampling with mutation rate θ as shown for exponential growth as null model, and Beta$(2-\alpha, \alpha)$-coalescent as alternative. Sample size $n = 50$, segregating sites $s = 25$. The estimate $(\theta_W(\beta))$ is obtained from (9). All estimates from 10^5 iterates.

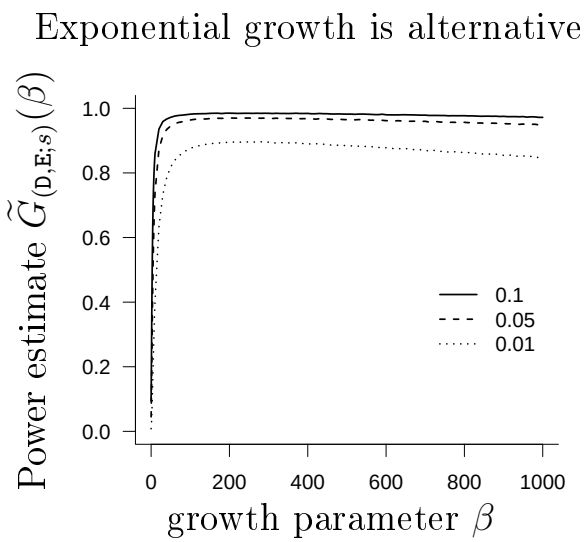
β	$x\%$	θ $(\theta_W(\beta))$	test size
1	10%	7.895425 $\theta_W(1)$	0.10
		5.0	0.10
		7.0	0.10
		9.0	0.10
		11.0	0.10
	5%	7.895425 $\theta_W(1)$	0.05
		5.0	0.05
		7.0	0.05
		9.0	0.05
		11.0	0.05
	1%	7.895425 $\theta_W(1)$	0.01
		5.0	0.01
		7.0	0.01
		9.0	0.01
		11.0	0.01
10	10%	16.33632 $\theta_W(10)$	0.10
		12.0	0.13
		14.0	0.12
		18.0	0.10
		20.0	0.10
	5%	16.33632 $\theta_W(10)$	0.05
		12.0	0.05
		14.0	0.05
		18.0	0.05
		20.0	0.05
	1%	16.33632 $\theta_W(10)$	0.01
		12.0	0.01
		14.0	0.01
		18.0	0.01
		20.0	0.01

Table S3: Checking size of test given fixed-s quantiles associated with size $x\%$ as shown using rejection sampling with mutation rate θ as shown for exponential growth as null model ($\beta = 1000$), and Beta$(2-\alpha, \alpha)$-coalescent as alternative. The estimate ($\theta_W(\beta)$) is obtained from (9). Sample size $n = 50$, segregating sites $s = 25$. All estimates from 10^5 iterates.

β	$x\%$	θ ($\theta_W(\beta)$)	test size
1000	10%	263.1798 $\theta_W(10^3)$	0.10
		259	0.10
		261	0.10
		265	0.10
		267	0.10
	5%	263.1798 $\theta_W(10^3)$	0.05
		259	0.05
		261	0.05
		265	0.05
		267	0.05
	1%	263.1798 $\theta_W(10^3)$	0.01
		259	0.01
		261	0.01
		265	0.01
		267	0.01

Estimation of power for $\Theta_0 = \Theta_s^{\mathtt{D}}$, $\Theta_1 = \Theta_s^{\mathtt{E}}$

Figure S2: Estimate of $\widetilde{G}_{(\mathtt{D},\mathtt{E};s)}$ from (10) based on the approximate likelihood (12) as a function of ψ (no lumping) with number of leaves $n = 100$ and $s = 50$. The line types denote the size of the test as shown in the legend. The interval hypotheses are discretized to $\Theta_s^{\mathtt{E}} = \{\beta : \beta \in \{0, 1, 2, \ldots, 10, 20, \ldots, 1000\}\}$ and $\Theta_s^{\mathtt{D}} = \{\psi : \psi \in \{0, 0.01, 0.02, \ldots, 0.1, 0.15, 0.2, \ldots, 0.95\}\}$. Reverting the hypotheses yield very similar results (not shown).



Estimation of power for $s = 300$

Figure S3: Estimate $\widetilde{G}_{(\mathtt{B},\mathtt{E};s)}(\beta)$ of power as a function of β for (**A**) $\beta \in \{0, 10, \ldots, 1000\}$; (**B**) $\beta \in \{0, 1, 2, \ldots, 9, 10, 20, \ldots, 1000\}$ when the Beta$(2-\alpha, \alpha)$-coalescent is the null hypothesis, and the test statistic is $\sup\{\tilde{\ell}(\Pi, \underline{\xi}^{(n)}, s), \Pi \in \Theta_s^{\mathtt{B}}\} - \sup\{\tilde{\ell}(\Pi, \underline{\xi}^{(n)}, s), \Pi \in \Theta_s^{\mathtt{E}}\}$ (7) , with $\tilde{\ell}(\Pi, \underline{\xi}^{(n)}, s)$ the log of the Poisson likelihood function (12) (no lumping). Values at $\beta = 0$ correspond to the Kingman coalescent. A total of 10^6 replicates for both quantiles and power estimates.

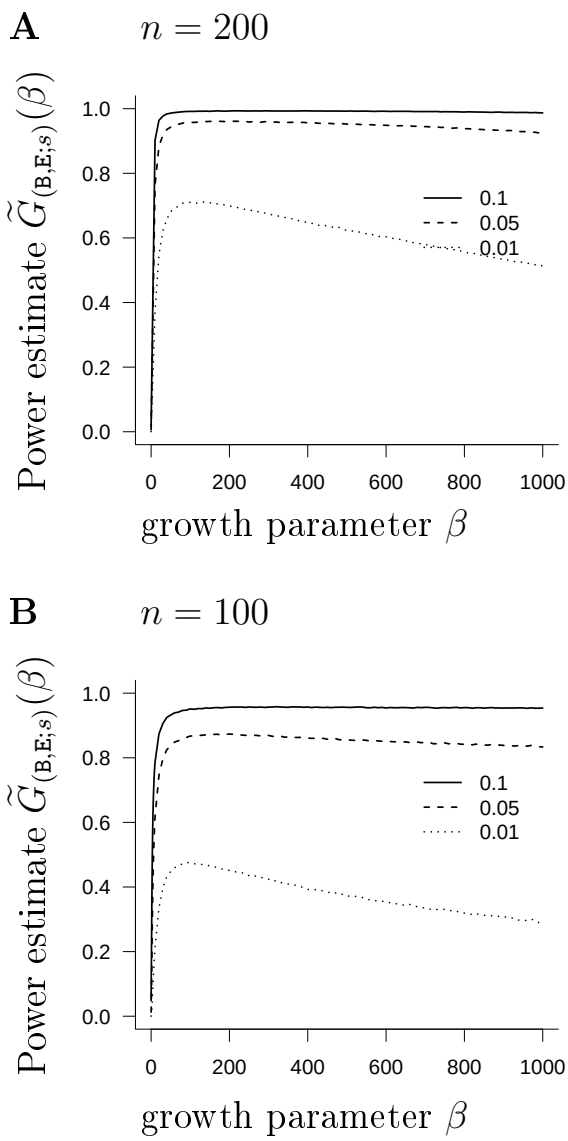


Figure S4: Estimate $\widetilde{G}_{(\mathtt{E},\mathtt{B};s)}(\alpha)$ of power as a function of α for $\alpha \in [1,2]$ when exponential growth ($\mathtt{E}$) is the null hypothesis, Beta$(2-\alpha,\alpha)$-coalescent ($\mathtt{B}$) is the alternative, and the test statistic is $\sup\{\tilde{\ell}(\Pi,\underline{\xi}^{(n)},s),\Pi \in \Theta_s^{\mathtt{E}}\} - \sup\{\tilde{\ell}(\Pi,\underline{\xi}^{(n)},s),\Pi \in \Theta_s^{\mathtt{B}}\}$ (7), with $\tilde{\ell}(\Pi,\underline{\xi}^{(n)},s)$ the log of the Poisson likelihood function (12) (no lumping). Values at $\alpha = 2$ correspond to the Kingman coalescent; number of segregating sites $s = 300$; 10^6 replicates for quantiles and power estimates.

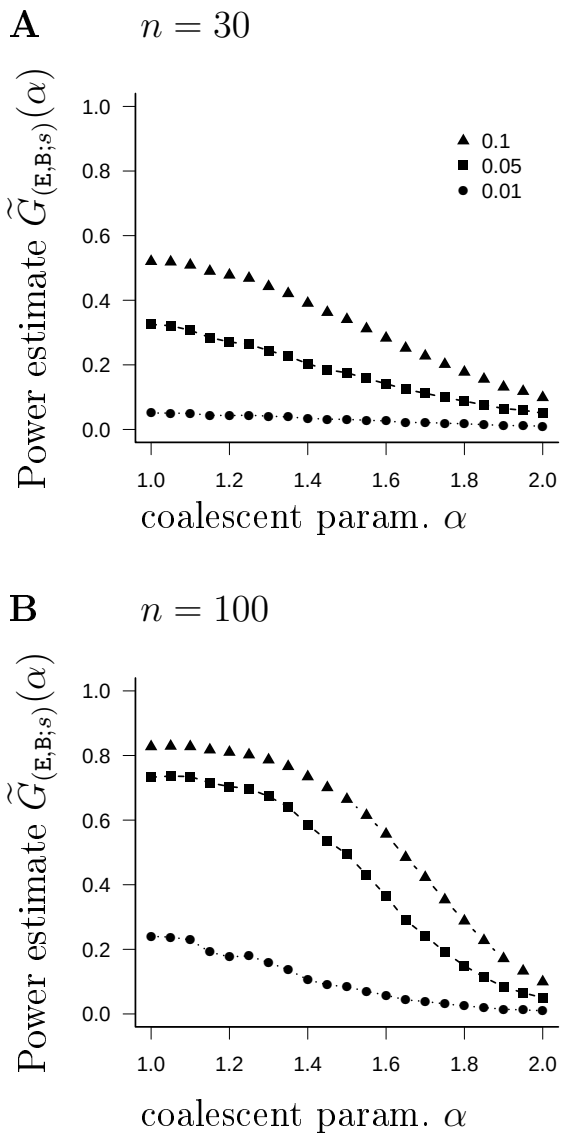


Figure S5: Estimate $\widetilde{G}_{(\mathtt{D},\mathtt{A};s)}(\gamma)$ of power (10) between algebraic growth and the Dirac Lambda-coalescent when the test statistic is $\sup\{\tilde{\ell}(\Pi, \underline{\xi}^{(n)}, s), \vartheta \in \Theta_s^{\mathtt{D}}\} - \sup\{\tilde{\ell}(\Pi, \underline{\xi}^{(n)}, s), \Pi \in \Theta_s^{\mathtt{A}}\}$ (7), with $\tilde{\ell}(\Pi, \underline{\xi}^{(n)}, s)$ the log of the Poisson likelihood function (12) (no lumping); with $n = 100$ and number of segregating sites $s = 50$. The test sizes are as shown in the legend. The interval hypotheses are $\Theta_s^{\mathtt{A}} \equiv \{\gamma : \gamma \in \{0, 1, 2, \ldots, 10, 20, 30, \ldots, 1000\}\}$ and $\Theta_s^{\mathtt{D}} \equiv \{\psi : \psi \in \{0.01, 0.02, \ldots, 0.1, 0.15, 0.2, \ldots, 0.95\}\}$. Values at $\gamma = 0$ correspond to the Kingman coalescent. Expected values were computed exactly, and quantiles and power estimated from 10^5 replicates. Reverting the hypotheses shows a very similar pattern (results not shown). In **B**, we 'zoom in' on the range $0 \leq \gamma \leq 50$.

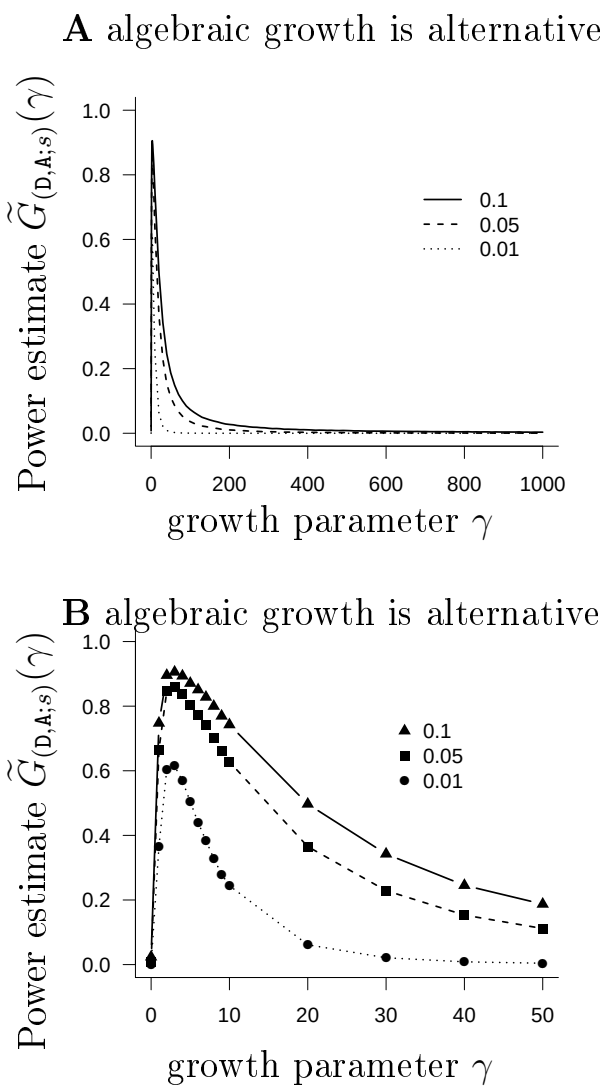
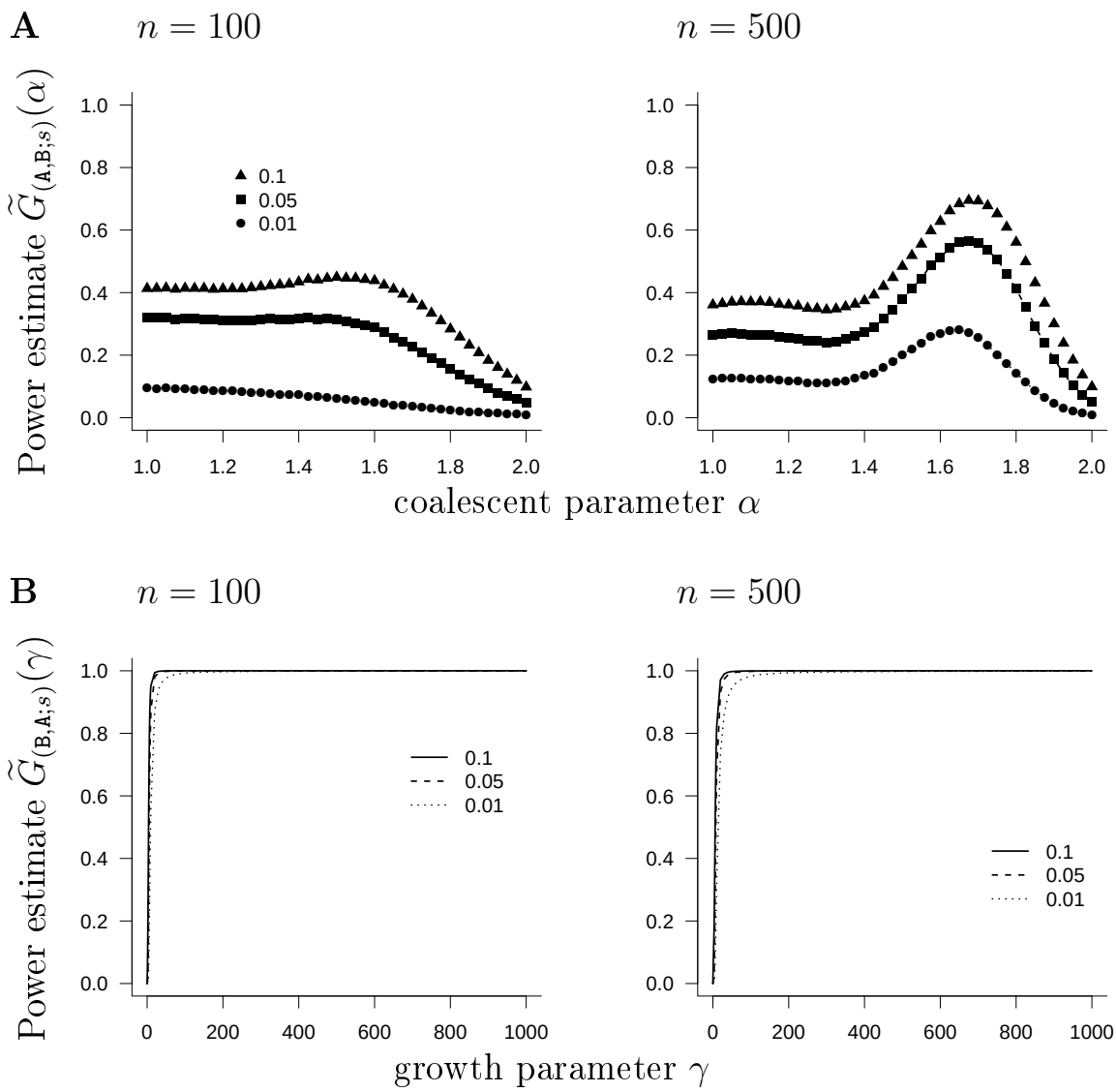


Figure S6: Estimate $\widetilde{G}_{(\mathtt{A},\mathtt{B};s)}(\alpha)$ (**A**) and $\widetilde{G}_{(\mathtt{B},\mathtt{A};s)}(\gamma)$ (**B**) of power (10) between algebraic growth and the Beta$(2-\alpha,\alpha)$-coalescent when the test statistic is $\sup\{\tilde{\ell}(\Pi,\underline{\xi}^{(n)},s),\Pi\in\Theta_s^{\Pi_0}\}-\sup\{\tilde{\ell}(\Pi,\underline{\xi}^{(n)},s),\Pi\in\Theta_s^{\Pi_1}\}$ (7), with $\tilde{\ell}(\Pi,\underline{\xi}^{(n)},s)$ the log of the Poisson likelihood function (12) (no lumping); with number of leaves n as shown and number of segregating sites $s=50$. The test sizes are as shown in the legend. The interval hypotheses are $\Theta_s^{\mathtt{A}}\equiv\{\gamma:\gamma\in\{0,1,2,\ldots,10,20,30,\ldots,1000\}\}$ and $\Theta_s^{\mathtt{B}}\equiv\{\alpha:\alpha\in\{1,1.025,\ldots,2\}\}$. Values at $\gamma=0$ and $\alpha=2$ correspond to the Kingman coalescent. Expected values were computed exactly, and quantiles and power estimated from 10^5 replicates. In **A**, the Beta$(2-\alpha,\alpha)$-coalescent is the alternative hypothesis; in **B**, algebraic growth is the alternative.



Mean misclassification probabilities and posterior probabilities for ABC approach - alternative parameter choices

Table S4: Approximations of mean posterior probabilities and misclassification probabilities for the ABC model comparison between E and B for different growth parameter ranges or tolerance rates. The nSFS is used as summary statistics. $\beta_{\max}$ denotes the maximal growth rate used in the growth model, n_{cv} denotes the number of cross-validations; 'lump' indicates which mutation classes are lumped into one class. An expected number $s = 75$ of mutations are assumed.

$\beta_{\max}$	lump	n_{cv}	tolerance	$\mathbb{E}^{\mathtt{B}}\left[\pi(\mathtt{E}\|\underline{\zeta})\right]$	$\mathbb{E}^{\mathtt{E}}\left[\pi(\mathtt{B}\|\underline{\zeta})\right]$	$\mathbb{E}^{\mathtt{B}}\left[\pi(\varrho^{\mathcal{B}}_{(\mathtt{E},\mathtt{B})} \geq 1\|\underline{\zeta})\right]$	$\mathbb{E}^{\mathtt{E}}\left[\pi(\varrho^{\mathcal{B}}_{(\mathtt{E},\mathtt{B})} \leq 1\|\underline{\zeta})\right]$
10^3	10+	24000	0.01	0.24	0.11	0.18	0.04
"	"	"	"	0.24	0.11	0.18	0.04
10^3	50+	12000	0.01	0.22	0.09	0.18	0.03
"	"	"	"	0.23	0.09	0.19	0.03
10^3	100+	1200	0.01	0.22	0.09	0.19	0.03
"	"	12000	"	0.22	0.08	0.20	0.02
10^3	no	12000	0.01	0.30	0.14	0.23	0.04
"	"	"	"	0.30	0.14	0.23	0.04
500	10+	24000	0.01	0.26	0.13	0.20	0.05
500	50+	12000	0.01	0.24	0.10	0.20	0.04
500	100+	1200	0.01	0.26	0.09	0.22	0.03
100	10+	24000	0.01	0.31	0.21	0.23	0.12
100	50+	12000	0.01	0.27	0.18	0.20	0.10
10^3	10+	24000	0.0025	0.20	0.11	0.15	0.05
10^3	50+	12000	0.0025	0.19	0.08	0.15	0.03
10^3	100+	1200	0.0025	0.18	0.08	0.16	0.03
10^3	no	1200	0.0025	0.25	0.13	0.20	0.05

Table S5: Approximations of mean posterior probabilities and misclassification probabilities for the ABC model comparison between E and B for tolerance $x = 0.0025$ and sample size $n = 200$ and assumed expected number $s = 15$ of mutations. The nSFS is used as summary statistics. n_{cv} denotes the number of cross-validations 'lumped' indicates which mutation classes are lumped into one class.

lump	n_{cv}	$\mathbb{E}^{\mathtt{B}}\left[\pi(\mathtt{E}\|\underline{\zeta})\right]$	$\mathbb{E}^{\mathtt{E}}\left[\pi(\mathtt{B}\|\underline{\zeta})\right]$	$\mathbb{E}^{\mathtt{B}}\left[\pi(\varrho^{\mathcal{B}}_{(\mathtt{E},\mathtt{B})} \geq 1\|\underline{\zeta})\right]$	$\mathbb{E}^{\mathtt{E}}\left[\pi(\varrho^{\mathcal{B}}_{(\mathtt{E},\mathtt{B})} \leq 1\|\underline{\zeta})\right]$
10	24000	0.28	0.24	0.23	0.14
50	12000	0.31	0.26	0.25	0.14
100	12000	0.33	0.27	0.28	0.15
no	12000	0.34	0.26	0.29	0.15

Table S6: Approximations of mean posterior probabilities and misclassification probabilities for the ABC model comparison between E and B for tolerance $x = 0.001$ and sample size $n = 200$, assumed expected number $s = 15$ of mutations and alternative prior ranges and distributions. The nSFS is used as summary statistics. n_{cv} denotes the number of cross-validations 'lumped' indicates which mutation classes are lumped into one class. For growth rate β, the prior is uniformly distributed on $\{\beta_{\min}, \beta_{\min} + 10, \ldots, \beta_{\max}\}$. For coalescent parameter α, the prior is uniformly distributed on $[\alpha_{\min}, \alpha_{\max}]$

lump	n_{cv}	$\beta_{\min}, \beta_{\max}$	$\alpha_{\min}, \alpha_{\max}$	$\mathbb{E}^{\mathtt{B}}\left[\pi(\mathtt{E}\|\underline{\zeta})\right]$	$\mathbb{E}^{\mathtt{E}}\left[\pi(\mathtt{B}\|\underline{\zeta})\right]$	$\mathbb{E}^{\mathtt{B}}\left[\pi(\varrho^{\mathcal{B}}_{(\mathtt{E},\mathtt{B})} \geq 1\|\underline{\zeta})\right]$	$\mathbb{E}^{\mathtt{E}}\left[\pi(\varrho^{\mathcal{B}}_{(\mathtt{E},\mathtt{B})} \leq 1\|\underline{\zeta})\right]$
10	24000	0,100	1.5,2	0.39	0.34	0.30	0.23
50	12000	0,100	1.5,2	0.38	0.31	0.31	0.18
10	24000	100,1000	1,1.5	0.33	0.28	0.29	0.14
50	12000	100,1000	1,1.5	0.36	0.32	0.31	0.18

Table S7: Approximations of the misclassification probabilities for the ABC model comparison between models E, B, D for tolerance $x = 0.005$, sample size $n = 200$ and $s \in \{15, 75\}$. The folded nSFS was used as summary statistics. We use the abbreviation $mc(\Pi_1|\Pi_2) := \mathbb{E}^{\Pi_2}\left[\pi(\min_{\Pi \neq \Pi_1} \varrho^{\mathcal{B}}_{(\Pi_1,\Pi)} \geq 1|\underline{\zeta}^{(n)})\right]$, $\Pi_1, \Pi_2 \in \{\mathtt{E}, \mathtt{B}, \mathtt{D}\}$.

s	lump	n_{cv}	$mc(\mathtt{E}\|\mathtt{B})$	$mc(\mathtt{D}\|\mathtt{B})$	$mc(\mathtt{B}\|\mathtt{E})$	$mc(\mathtt{D}\|\mathtt{E})$	$mc(\mathtt{B}\|\mathtt{D})$	$mc(\mathtt{E}\|\mathtt{D})$
15	10+	24000	0.27	0.07	0.12	0.01	0.62	0.01
15	50+	12000	0.39	0.06	0.08	0.01	0.60	0.03
15	no	12000	0.42	0.07	0.08	0.01	0.64	0.04
75	10+	24000	0.19	0.04	0.05	0.00	0.09	0.00
75	50+	12000	0.24	0.04	0.04	0.00	0.09	0.00

Table S8: Approximations of the misclassification probabilities for the ABC model comparison between models A, B, D for tolerance $x = 0.005$, sample size $n = 200$ and $s \in \{15, 75\}$. The folded nSFS was used as summary statistics. We use the abbreviation $mc(\Pi_1|\Pi_2) := \mathbb{E}^{\Pi_2}\left[\pi(\min_{\Pi \neq \Pi_1} \varrho^{\mathcal{B}}_{(\Pi_1,\Pi)} \geq 1|\underline{\zeta}^{(n)})\right]$, $\Pi_1, \Pi_2 \in \{\mathtt{A}, \mathtt{B}, \mathtt{D}\}$.

s	lump	n_{cv}	$mc(\mathtt{A}\|\mathtt{B})$	$mc(\mathtt{D}\|\mathtt{B})$	$mc(\mathtt{B}\|\mathtt{A})$	$mc(\mathtt{D}\|\mathtt{A})$	$mc(\mathtt{B}\|\mathtt{D})$	$mc(\mathtt{A}\|\mathtt{D})$
15	10+	24000	0.01	0.06	0.01	0.04	0.15	0.53
15	50+	12000	0.01	0.06	0.01	0.04	0.18	0.52
75	10+	24000	0.00	0.03	0.01	0.06	0.09	0.25
75	50+	12000	0.00	0.03	0.01	0.05	0.14	0.27

ABC analysis of the cytochrome *b* mtDNA data of Árnason (2004)

To investigate which model class (exponential growth E, algebraic growth A, Beta$(2-\alpha, \alpha)$-coalescents B, Dirac coalescents D) fits better to the data, we use the ABC model comparison approach given the (lumped) nfSFS of the observed mitochondrial locus. The exponential growth model class is specified by an uniform prior for growth parameter β on $\{0, 1, 2, \ldots, 1000\}$, the algebraic growth class by an uniform prior for growth parameter γ on $\{0, 1, 2, \ldots, 1000\}$. The class of Beta n-coalescents is specified by an uniform prior on $\{1, 1.01, \ldots, 2\}$ for the coalescent parameter α, the class of Dirac coalescents by an uniform prior on $\{0.01, 0.02, \ldots 0.99\}$ for the coalescent parameter ψ (we omit the star-shaped coalescent $\psi = 1$ because the observed SFS has not only singleton mutations, thus directly violating this model). We used two tolerance levels of 0.005 and 0.00125 and perform $n_{\text{r}} = 200,000$ simulations for each model class. See Table S9 for the approximated Bayes factors $\varrho^{\mathcal{B}}_{\mathtt{E},\mathtt{B}}$ for the model comparison of the growth model and the Beta n-coalescent model using different lumps of the nfSFS as summary statistics. The Bayes factors $\varrho^{\mathcal{B}}_{\mathtt{A},\Pi}, \varrho^{\mathcal{B}}_{\mathtt{D},\Pi}$ for $\Pi \in \{\mathtt{E}, \mathtt{B}\}$ have maximal values of $\approx 0.01, 0.001$ under all lumpings and both tolerances. The observed data fits slightly better to the growth model than to the Beta coalescent class, but not so much better that we could discard the Beta n-coalescents as possible genealogy models for this locus. The latter point is also highlighted by results for an ABC model comparison between only model classes E and B where all lumpings but 100+ again (slightly) favour the growth

model, but for 100+ lumping this is reversed ($\varrho^{\mathcal{B}}_{\mathtt{E,B}} = 0.69$ for tolerance 0.005). The Dirac coalescents and the algebraic growth model show neglectible support for all lumpings and thus we discard them as potential models.

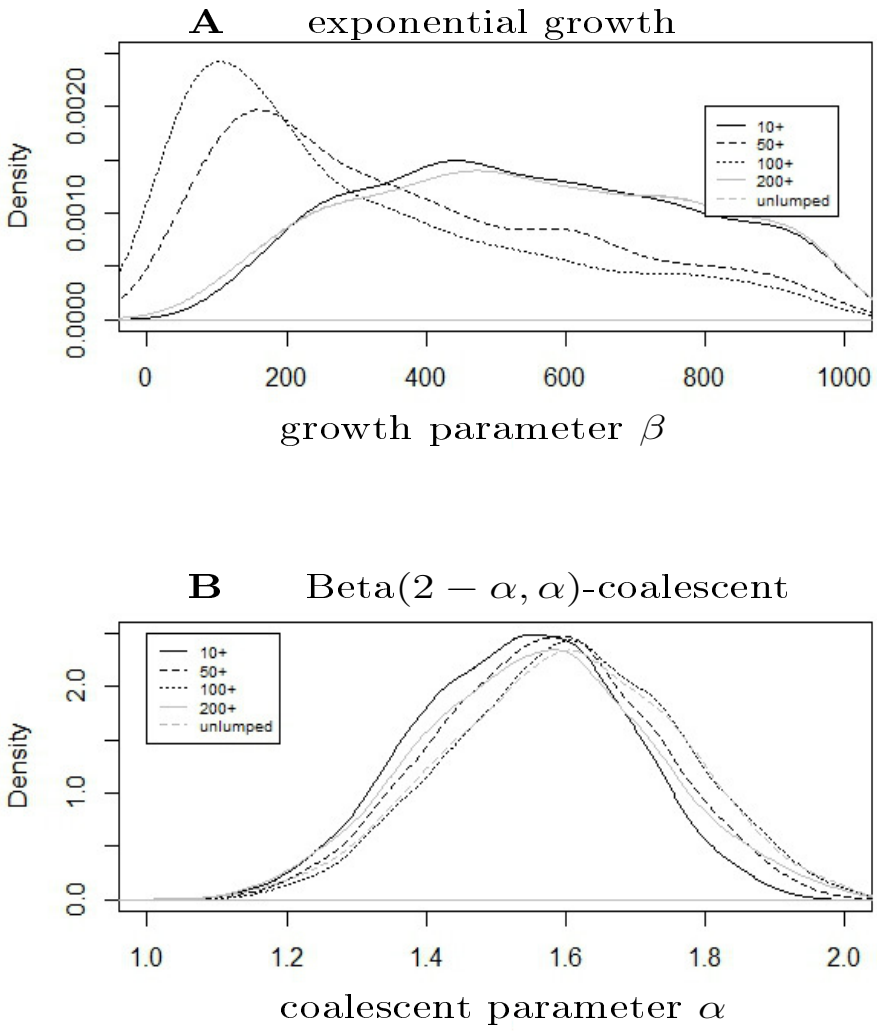
Table S9: Approximated Bayes factor $\varrho^{\mathcal{B}}_{\mathtt{E,B}}$ given the Atlantic cod mtDNA data

lumping number	10+	50+	100+	200+	no
tolerance 0.005	7.79	2.23	2.09	2.97	2.98
tolerance 0.00125	10.35	2.97	2.23	6.87	7.13

Jeffreys (1961) suggested interpreting Bayes factors according to the $\log_{10}$ scale. Lumping at 10 (Table S9) then gives at least 'substantial' ($1/2 < \log_{10}(\varrho^{\mathcal{B}}_{\mathtt{E,B}}) \leq 1$) evidence against the Beta$(2-\alpha, \alpha)$-coalescent in favor of exponential growth. Using Kass and Raftery (1995) suggestion of considering Bayes factors on $2\log_e$ scale gives 'positive' ($2 < 2\log_e(\varrho^{\mathcal{B}}_{\mathtt{E,B}}) < 6$) evidence in favor of exponential growth, based on lumping at 10.

Additionally to the ABC model comparison, we also evaluate which parameters fit best to the observed nfSFS at the mitochondrial locus. We omit the class of Dirac coalescents and algebraic growth models from further analysis since the observed frequency spectrum clearly does not fit to this model class. For each other model class used, we record the prior parameters from the 0.5% of the $n_{\mathrm{r}} = 200,000$ simulations that have the smallest ℓ^2 distance to the observed nfSFS (summary statistics). This gives an approximate sample of the posterior distribution of $\pi(\alpha|$ observed $\underline{\zeta}^{(n)})$ resp. $\pi(\beta|$ observed $\underline{\zeta}^{(n)})$. Again, we used the lumped nSFS as summary statistics. Figure S7 shows the posterior distributions for different lumping numbers.

Figure S7: Approximate posterior density of the coalescent parameter from ABC fitting of the (**A**) growth, and (**B**) Beta n-coalescent model classes to the observed nfSFS in the Atlantic cod data. Denote by α the Beta n-coalescent parameter, β the growth rate. Priors were uniform on both sets.



ABC quality control for the Árnason (2004) data

We follow the recommendation from the `R` package abc (Csilléry *et al.*, 2012) and perform three checks of quality for the presented ABC approach. We focus on the lumping which gives the clearest distinction, namely the lumping of all classes with mutation counts 10 or higher (class 10+). All checks are performed using the `R` package abc

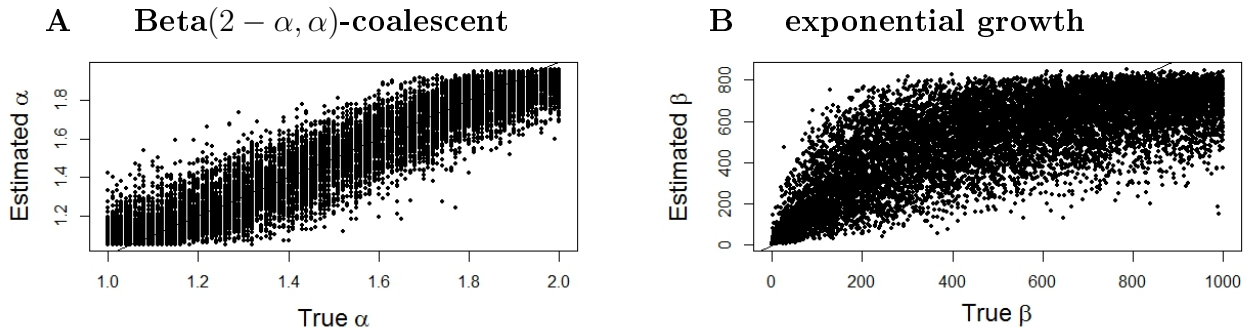
To assess the general ability to distinguish between the two model classes in the setting (i.e., number of observed mutations and sample size) given by the Atlantic cod mtDNA data from Árnason (2004), we again employ a leave-one-out cross-validation as described in Methods. See Tables S10 for the results.

Table S10: Approximations of the mean posterior probabilities and misclassification probabilities (based on $n_{\text{cv}} = 12,000$ cross-validations) for the ABC model comparison between models `E`, `A`, `B`, `D` for tolerance $x = 0.005$, sample size $n = 1278$ and mutation rate estimated via Watterson's estimator from $s = 39$ observed mutations. The lumped nfSFS (10+) was used as summary statistics. The entries are listed as $\mathbb{E}^{\Pi_{row}}\left[\pi(\Pi_{col}|\underline{\zeta})\right]/\mathbb{E}^{\Pi_{row}}\left[\pi(\min_{\Pi\neq\Pi_{col}} \varrho^{\mathcal{B}}_{\Pi_{col},\Pi} \geq 1|\underline{\zeta}^{(n)})\right]$

	E	A	B	D
E	0.79/0.88	0.00/0.00	0.21/0.12	0.00/0.00
A	0.00/0.00	0.24/0.25	0.06/0.01	0.70/0.74
B	0.24/0.18	0.01/0.00	0.71/0.79	0.03/0.02
D	0.00/0.00	0.03/0.03	0.08/0.03	0.90/0.94

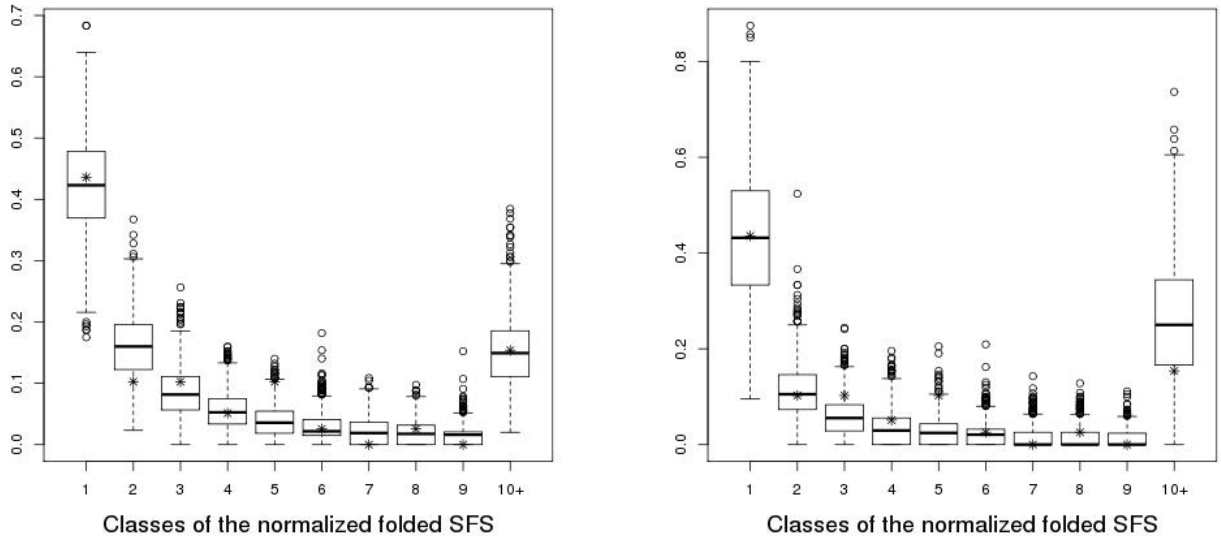
To assess the quality to distinguish the parameters within one model class, we again use leave-one-out cross-validations ($n_{\text{cv}} = 12,000$). The parameter of each simulation chosen for cross-validation is estimated as the median of the 0.5% of simulations with the smallest ℓ^2 distance to the chosen simulation. Figure S8 shows the resulting scatter plots of the parameters of the chosen simulations and the corresponding estimations.

Figure S8: Scatter plots of estimated vs. true parameters of $n_{\text{cv}} = 12000$ cross-validated simulations in the (**A**) Beta coalescent; (**B**) exponential growth.



To see whether the posterior distributions given the cod mtDNA data from Árnason (2004) define models under which the observed data is reproducible, we perfomed posterior predictive checks by simulating the 10+ lumped nfSFS under the posterior distribution (i.e., simulating once from each parameter set of each of the 1,000 accepted simulations) for each model class and compare these with the nfSFS observed. See Figure S9 for the results within each nfSFS class. To assess the minimal l^2 distance of the simulations using the posterior parameter distributions from the observed nfSFS, we simulated 5 replications under the posteriors. The minimal l^2 distance was 0.04 under the posterior growth model and 0.06 under the posterior Beta coalescent model.

Figure S9: Posterior predictive checks with 1,000 simulations of the nfSFS under the approximate posterior distributions given the cod data from Árnason (2004) for the (**A**) Beta coalescent model class; (**B**) growth model class. Asterisks denote the observed values in the data.



The quality checks reveal that we can not distinguish well within the model classes of exponential growth and of Beta coalescents, but moderately between them. Additionally, the ABC approach distinguishes well between these two classes on one hand and the (non-fitting) other two classes `A`, `D`. The posterior predictive checks reveal that both model classes can produce the observed values in each class of the nfSFS, but do not match well in l^2 to the actual observed nfSFS. Neither model class thus captures the observed nfSFS well.