File S1

Supporting Information

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

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

List of equations from the main text

For ease of reference, we list below the equations from the main text used in Supporting Information. Refer to the main text for explanation of symbols.

Equation (1) from the main text:

$$\mathbb{E}^{\Pi,\theta}\left[\xi_i^{(n)}\right] = \frac{\theta}{2}\sum_{k=2}^{n-i+1} p^{(n),\Pi}[k,i]\cdot k\cdot \mathbb{E}^{\Pi}\left[T_k^{(n)}\right], \quad i\in[n-1]. \tag{S1}$$

Equation (3) from the main text:

$$\hat{\theta}^{\Pi} := \frac{2S}{\mathbb{E}^{\Pi}\left[B^{(n)}\right]}, \tag{S2}$$

Equation (5) from the main text:

$$c_N \approx \frac{2\tilde{\mu}}{\theta^{\Pi}}. \tag{S3}$$

Equation (7) from the main text:

$$\varrho_{(\mathtt{E},\mathtt{B};s)}(\underline{\xi}^{(n)}) := \frac{\sup\left\{L(\Pi,\underline{k}^{(n)},s),\ \Pi\in\Theta_s^{\mathtt{E}}\right\}}{\sup\left\{L(\Pi,\underline{k}^{(n)},s),\ \Pi\in\Theta_s^{\mathtt{B}}\right\}}. \tag{S4}$$

Equation (8) from the main text:

$$\sup_{\Pi\in\Theta_s^{\mathtt{E}}} \mathbb{P}^{\Pi,s}\left\{\varrho_{(\mathtt{E},\mathtt{B};s)}(\underline{\xi}^{(n)}) \le \varrho^*_{(\mathtt{E},\mathtt{B};s)}(a)\right\} \le a. \tag{S5}$$

Equation (9) from the main text:

$$\hat{\theta} = \hat{\theta}(\Pi;s) = \frac{2s}{\mathbb{E}^{\Pi}[B^{(n)}]} \tag{S6}$$

Equation (10) from the main text:

$$G_{(\mathtt{E},\mathtt{B};s)}(\Pi) = \mathbb{P}^{\Pi}\{\varrho_{(\mathtt{E},\mathtt{B};s)}(\underline{\xi}^{(n)}) \le \varrho^*_{(\mathtt{E},\mathtt{B};s)}(a,S)\}, \quad \Pi\in\Theta_s^{\mathtt{B}}. \tag{S7}$$

Equation (12) from the main text:

$$\widetilde{L}(\Pi, \underline{\xi}^{(n)}, s) = \prod_{i=1}^{n-1} e^{-\frac{\hat{\theta}(\Pi,s)}{2}\mathbb{E}^{\Pi}[B^{(n)}]\varphi_i^{(n,\Pi)}} \frac{\left(\frac{\hat{\theta}(\Pi,s)}{2}\mathbb{E}^{\Pi}[B^{(n)}]\varphi_i^{(n,\Pi)}\right)^{\xi_i^{(n)}}}{\xi_i^{(n)}!} \tag{S8}$$

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

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

$$\lambda_{b,k} = \int_{[0,1]} x^{k-2}(1-x)^{b-k}\Lambda(dx), \quad k = 2, .., b. \tag{S9}$$

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.*, 2013; 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

$$\nu(s) = \lim_{N \to \infty} X_N(\lfloor Ns \rfloor). \tag{S10}$$

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 (S1), 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 (S1), 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

$$S_j^{(n),\nu} := T_n^{(n),\nu} + T_{n-1}^{(n),\nu} + \cdots + T_j^{(n),\nu}, \quad j = n, \ldots, 2 \tag{S11}$$

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 $j \in 2, \ldots, n$,

$$F(t) := \int_0^t \nu(u)\, du \quad \text{and} \quad a_j^{(\vartheta)} := \int_0^\infty e^{-\binom{j}{2}F(s)}\, ds, \tag{S12}$$

assuming that the first integral in (S12) 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,

$$\mathbb{E}\big[S_m^{(n),\nu}\big] = \sum_{j=m}^{n} c_m^{(j,n)} a_j^{(\vartheta)}, \tag{S13}$$

where

$$c_m^{(j,n)} := \prod_{\substack{m \leq i \leq n \\ i \neq j}} \frac{\binom{i}{2}}{\binom{i}{2} - \binom{j}{2}} = (-1)^{j-m} \frac{(2j-1)m}{j(j-1)} \frac{\binom{n}{j}\binom{j+m-2}{j}\binom{j}{m}}{\binom{n+j-1}{j}}, \tag{S14}$$

(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,

$$\mathbb{E}^{\Pi}\left[B_i^{(n)}\right] = \sum_{j=2}^{n} W_{i,j}^{(n)} a_j^{(\vartheta)} \tag{S15}$$

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

$$\begin{aligned} W_{i,2}^{(n)} &= \frac{6}{n+1}, \\ W_{i,3}^{(n)} &= \frac{30(n-2i)}{(n+1)(n+2)}, \\ W_{i,j+2}^{(n)} &= \frac{(3+2j)(n-2i)}{j(n+j+1)} W_{i,j+1}^{(n)} - \frac{(1+j)(3+2j)(n-j)}{j(2j-1)(n+j+1)} W_{i,j}^{(n)}. \end{aligned} \tag{S16}$$

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

$$a_j^{(\beta)} = \frac{1}{\beta} \exp\left(\beta^{-1}\tbinom{j}{2}\right) E_1\left(\beta^{-1}\tbinom{j}{2}\right), \tag{S17}$$

where

$$E_1(t) := \int_t^{\infty} \frac{e^{-x}}{x}\, dx = \int_1^{\infty} \frac{e^{-tx}}{x}\, dx \tag{S18}$$

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

$$a_j^{(\gamma)} = \frac{\Gamma\big(1/(\gamma+1)\big)}{(1+\gamma)^{\gamma/(\gamma+1)}} \binom{j}{2}^{-1/(\gamma+1)}. \tag{S19}$$

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,

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

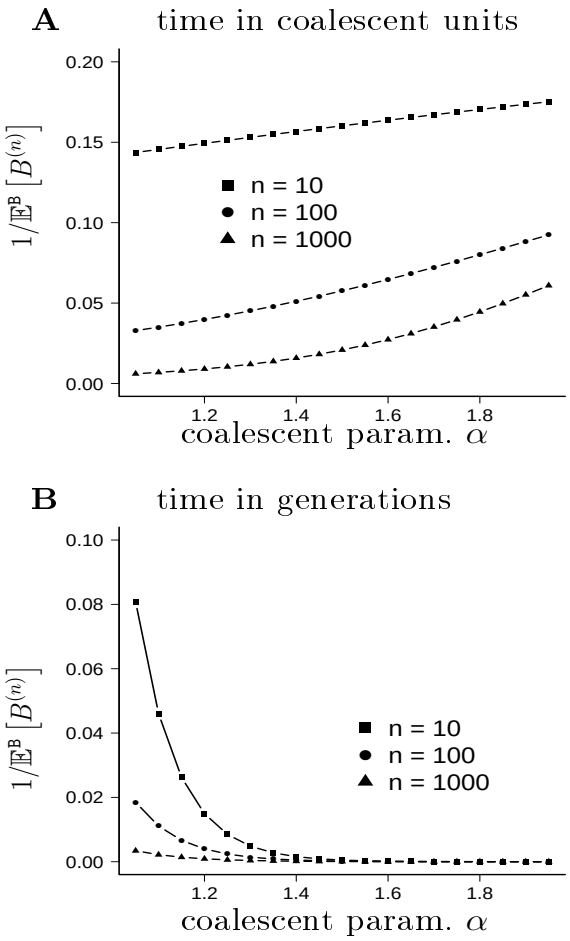
and

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

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 (S13) 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 $1/\mathbb{E}^{\mathtt{B}}\left[B^{(n)}\right]$ as a function of α

Figure S1: Graphs of $1/\mathbb{E}^{\mathtt{B}}\left[B^{(n)}\right]$, the estimated value of $\theta/2$ per observed mutation when using the Watterson estimator (S2) as a function of α (**A**), compare with (S2); and the estimated value of μ per observed mutation (**B**), using (S3) together with (S2), 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 $\mathbb{E}^{\mathtt{B}}\left[B^{(n)}\right]$ with $N^{\alpha-1}$, when $N = 10^5$.



Approximate robustness 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

$$\begin{aligned}
\mathbb{E}^{\Pi,\theta}\left[\zeta_i^{(n)}\right] &= \mathbb{E}^{\Pi,\theta}\left[\zeta_i^{(n)} 1_{\{|\xi^{(n)}|>0\}}\right] = \mathbb{E}^{\Pi,\theta}\left[\mathbb{E}^{\Pi,\theta}\left[\zeta_i^{(n)} 1_{\{|\xi^{(n)}|>0\}}\,\middle|\,\mathcal{T}\right]\right] \\
&= \mathbb{E}^{\Pi,\theta}\left[\mathbb{P}^{\Pi,\theta}(|\xi^{(n)}|>0\,|\,\mathcal{T})\frac{\mathbb{E}^{\Pi,\theta}\left[\zeta_i^{(n)} 1_{\{|\xi^{(n)}|>0\}}\,\middle|\,\mathcal{T}\right]}{\mathbb{P}^{\Pi,\theta}(|\xi^{(n)}|>0\,|\,\mathcal{T})}\right] \\
&= \mathbb{E}^{\Pi}\left[\left(1-e^{-\frac{\theta}{2}\sum_{i=1}^{n-1} B_i^{(n)}}\right)\frac{\frac{\theta}{2}\cdot B_i^{(n)}}{\frac{\theta}{2}\sum_{i=1}^{n} B_i^{(n)}}\right] \\
&= \mathbb{E}^{\Pi}\left[\frac{B_i^{(n)}}{\sum_{i=1}^{n} B_i^{(n)}}\right] - \mathbb{E}^{\Pi}\left[e^{-\frac{\theta}{2}\sum_{i=1}^{n-1} B_i^{(n)}}\frac{B_i^{(n)}}{\sum_{i=1}^{n} B_i^{(n)}}\right].
\end{aligned} \tag{S20}$$

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 (S20) 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 $k\le m$. We compute

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

□

Robustness 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 with branch lengths $B_i^{(n)}$, 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 (S5). 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 $\hat{\theta}(\Pi, s)$ (S6). 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 (S6). 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 (S6). All estimates from 10^5 iterates.

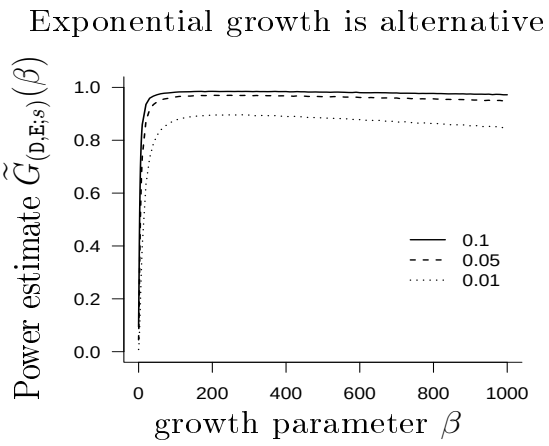
β	$x\%$	θ $(\theta_W(\beta))$	test size
1	10%	7.895425 $\theta_W(1)$	0.10
		5.0	0.10
		7.0	0.10
		9.0	0.10
		11.0	0.10
	5%	7.895425 $\theta_W(1)$	0.05
		5.0	0.05
		7.0	0.05
		9.0	0.05
		11.0	0.05
	1%	7.895425 $\theta_W(1)$	0.01
		5.0	0.01
		7.0	0.01
		9.0	0.01
		11.0	0.01
10	10%	16.33632 $\theta_W(10)$	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 (S6). 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 (S7) based on the approximate likelihood (S8) 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 $\sup\{\tilde{\ell}(\Pi, \underline{\xi}^{(n)}, s), \Pi \in \Theta_s^{\mathtt{B}}\} - \sup\{\tilde{\ell}(\Pi, \underline{\xi}^{(n)}, s), \Pi \in \Theta_s^{\mathtt{E}}\}$ (S4) , with $\tilde{\ell}(\Pi, \underline{\xi}^{(n)}, s)$ the log of the Poisson likelihood function (S8) (no lumping). Values at $\beta = 0$ correspond to the Kingman coalescent. A total of 10^6 replicates for both quantiles and power estimates.

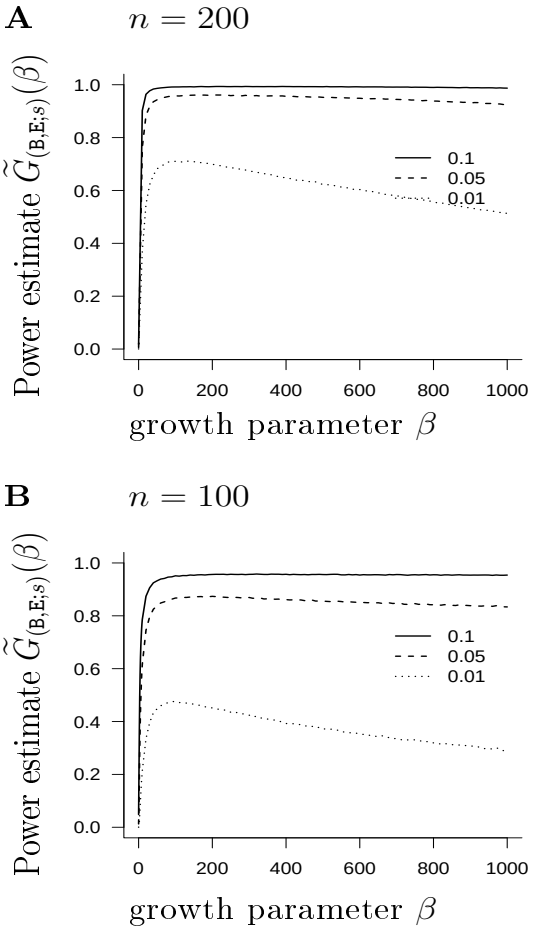
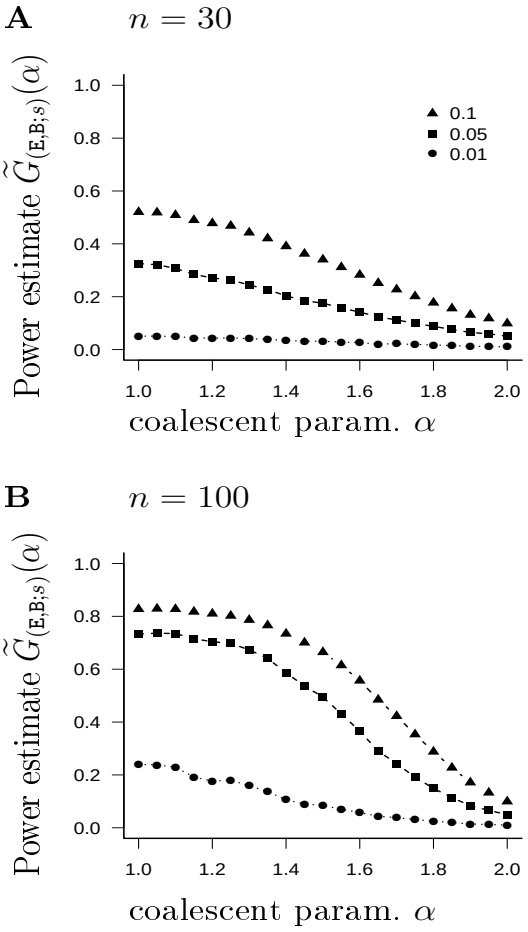
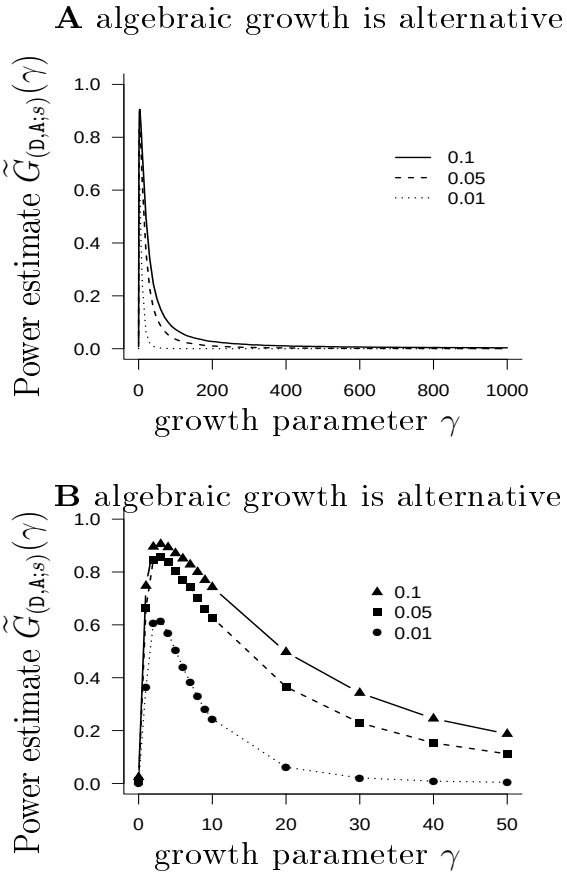


Figure S4: Estimate $\widetilde{G}_{(\mathtt{E},\mathtt{B};s)}(\alpha)$ of power as a function of α for $\alpha \in [1,2]$ when exponential growth ($\mathtt{E}$) is the null hypothesis, Beta$(2-\alpha,\alpha)$-coalescent ($\mathtt{B}$) is the alternative, and the test statistic is $\sup\{\tilde{\ell}(\Pi, \underline{\xi}^{(n)}, s), \Pi \in \Theta_s^{\mathtt{E}}\} - \sup\{\tilde{\ell}(\Pi, \underline{\xi}^{(n)}, s), \Pi \in \Theta_s^{\mathtt{B}}\}$ (S4), with $\tilde{\ell}(\Pi, \underline{\xi}^{(n)}, s)$ the log of the Poisson likelihood function (S8) (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.



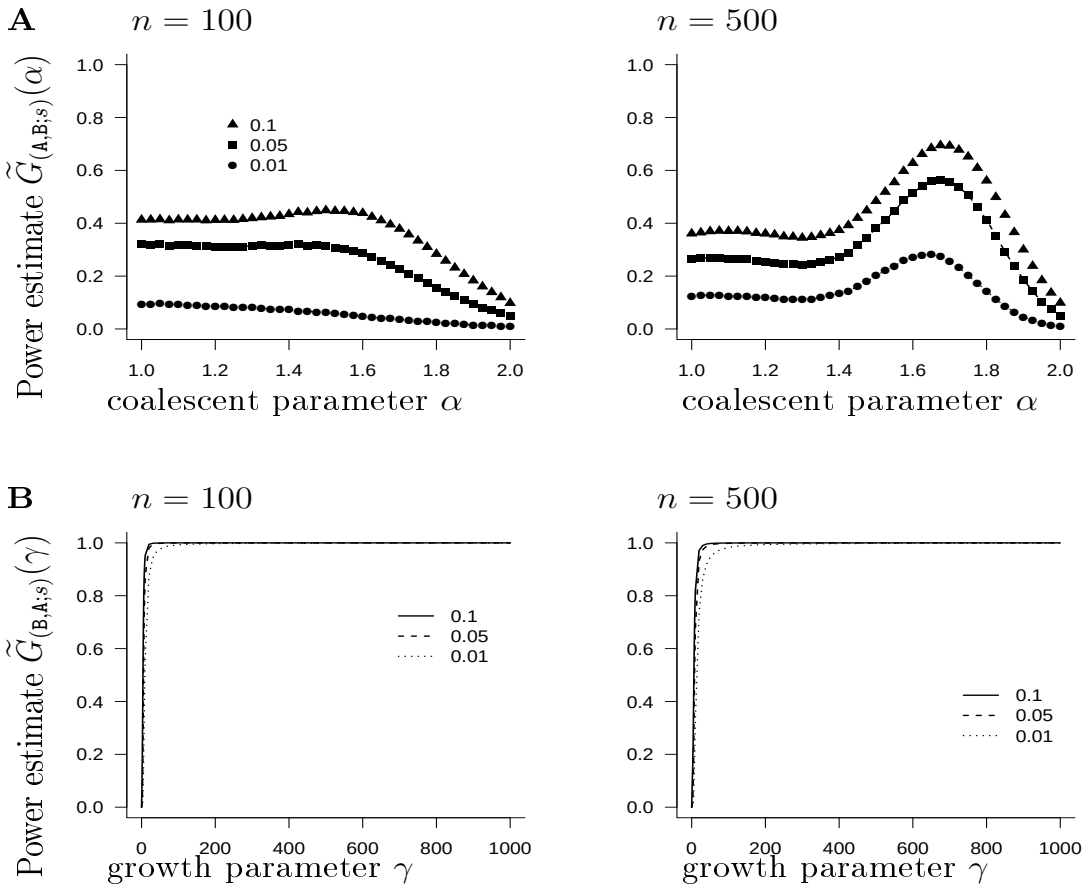
Estimate of power comparing $\Theta_s^{\mathtt{A}}$ and $\Theta_s^{\mathtt{D}}$

Figure S5: Estimate $\widetilde{G}_{(\mathtt{D},\mathtt{A};s)}(\gamma)$ of power (S7) between algebraic growth and the Dirac Lambda-coalescent when the test statistic is $\sup\{\tilde{\ell}(\Pi, \underline{\xi}^{(n)}, s), \vartheta \in \Theta_s^{\mathtt{D}}\} - \sup\{\tilde{\ell}(\Pi, \underline{\xi}^{(n)}, s), \Pi \in \Theta_s^{\mathtt{A}}\}$ (S4), with $\tilde{\ell}(\Pi, \underline{\xi}^{(n)}, s)$ the log of the Poisson likelihood function (S8) (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 $0 \leq \gamma \leq 50$.



Estimate of power comparing $\Theta_s^{\mathtt{A}}$ and $\Theta_s^{\mathtt{B}}$

Figure S6: Estimate $\widetilde{G}_{(\mathtt{A},\mathtt{B};s)}(\alpha)$ (**A**) and $\widetilde{G}_{(\mathtt{B},\mathtt{A};s)}(\gamma)$ (**B**) of power (S7) between algebraic growth and the Beta$(2-\alpha,\alpha)$-coalescent when the test statistic is $\sup\{\tilde{\ell}(\Pi,\underline{\xi}^{(n)},s),\Pi\in\Theta_s^{\Pi_0}\}-\sup\{\tilde{\ell}(\Pi,\underline{\xi}^{(n)},s),\Pi\in\Theta_s^{\Pi_1}\}$ (S4), with $\tilde{\ell}(\Pi,\underline{\xi}^{(n)},s)$ the log of the Poisson likelihood function (S8) (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 $n_{\mathrm{r}} = 200,000$ 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 ($\varrho^{\mathcal{B}}_{(\mathtt{E},\mathtt{B})} = 0.69$ 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 $\varrho^{\mathcal{B}}_{(\mathtt{E},\mathtt{B})}$ given the Atlantic cod mtDNA data

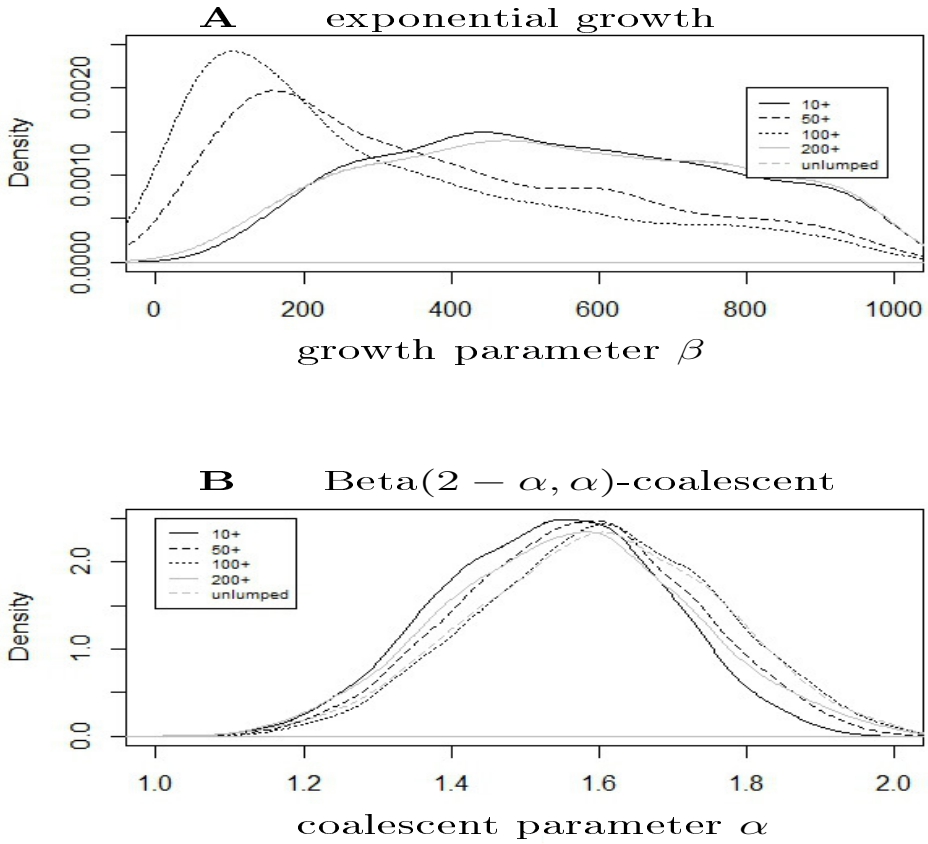
lumping number	10+	50+	100+	200+	no
tolerance 0.005	7.79	2.23	2.09	2.97	2.98
tolerance 0.00125	10.35	2.97	2.23	6.87	7.13

JEFFREYS (1961) suggested interpreting Bayes factors according to the $\log_{10}$ scale. Lumping at 10 (Table **S9**) then gives at least 'substantial' ($1/2 < \log_{10}(\varrho^{\mathcal{B}}_{(\mathtt{E},\mathtt{B})}) \leq 1$) 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' $(2 < 2\log_e(\varrho^{\mathcal{B}}_{(\mathrm{E},\mathrm{B})}) < 6)$ 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

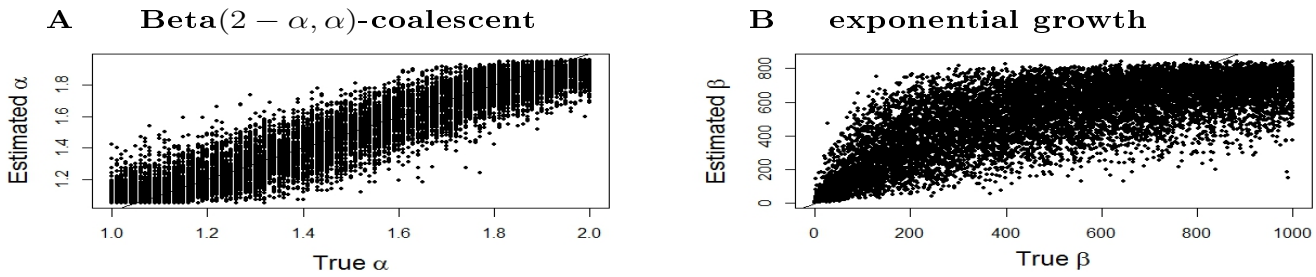
To assess the general ability to distinguish between the two model classes in the setting (i.e., number of observed mutations and sample size) given by the Atlantic cod mtDNA data from Árnason (2004), we again employ a leave-one-out cross-validation as described in Methods. See Table **S10** for the 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 $\mathbb{E}^{\Pi_{\text{row}}}\left[\pi(\Pi_{\text{col}}|\underline{\zeta})\right]/\mathbb{E}^{\Pi_{\text{row}}}\left[\pi(\min_{\Pi\neq\Pi_{\text{col}}}\varrho^{\mathcal{B}}_{\Pi_{\text{col}},\Pi}\geq 1|\underline{\zeta}^{(n)})\right]$.

	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

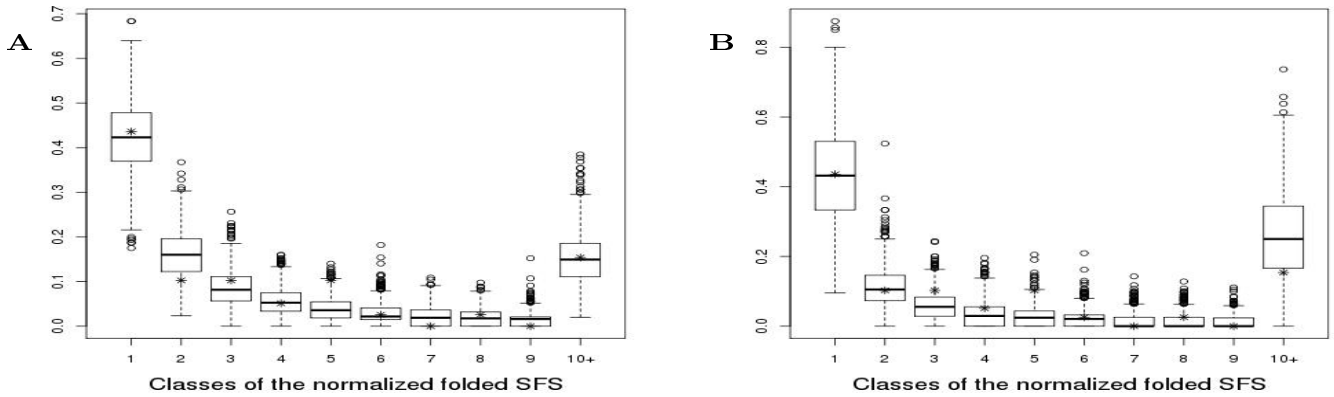
To assess the quality to distinguish the parameters within one model class, we again use leave-one-out cross-validations ($n_{\text{cv}} = 12,000$). The parameter of each simulation chosen for cross-validation is estimated as the median of the 0.5% of simulations with the smallest ℓ^2 distance to the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

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.

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

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

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, 2013 Statistical properties of the site-frequency spectrum associated with lambda-coalescents. Genetics **195**: 1037–1053.

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.

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

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.

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

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

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.

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

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.

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

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.

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.

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

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