File S1 The conditioned Markov chain and its properties

This section explains how to obtain the transition probability matrix of the Markov chain conditioned on the event that the population reaches size z before going extinct (reaching size 0), i.e. conditioned on the event $T_z < T_0$. We also explain how to derive further properties of the conditioned Markov chain. We first restrict our Markov chain to the states $0, 1, \ldots, z-1, z$, where 0 and z are absorbing states and $1, \ldots, z-1$ are transient, that is the Markov chain will leave them at some time. We can write the transition probability matrix of the original Markov chain as

$$\mathbf{P} = \begin{pmatrix} \mathbf{Q} & \mathbf{R} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}, \tag{S1}$$

where $\mathbf{Q}$ is a $(z-1) \times (z-1)$ matrix representing the transitions between transient states, $\mathbf{R}$ is a $(z-1) \times 2$ matrix with the transition probabilities from the transient states to the absorbing states z (first column) and 0 (second column), $\mathbf{0}$ is a $2 \times (z-1)$ matrix filled with zeros, and $\mathbf{I}$ is an identity matrix (in this case 2×2).

Following Pinsky and Karlin (2010), we then computed the fundamental matrix $\mathbf{W} = (\mathbf{I} - \mathbf{Q})^{-1}$. W_{ij} gives the expected number of generations a population starting at size i spends at size j before reaching one of the absorbing states. This matrix operation is based on first-step analysis, i.e. on a decomposition of expected quantities according to what happens in the first step (see Pinsky and Karlin 2010, Section 3.4 for details).

The probabilities of absorption in either of the two absorbing states can then be computed as $\mathbf{U} = \mathbf{WR}$. The first column of $\mathbf{U}$ contains the success probabilities $\mathsf{Pr}(T_z < T_0 | N_0 = i)$ shown in Figure 2. Using the success probabilities and Bayes' formula, we then computed the transition probabilities of the Markov chain conditioned on $T_z < T_0$:

$$Q^c_{ij} = \mathsf{Pr}(N_{t+1} = j | N_t = i, T_z < T_0) = \frac{Q_{ij} \cdot \mathsf{Pr}(T_z < T_0 | N_0 = j)}{\mathsf{Pr}(T_z < T_0 | N_0 = i)}. \tag{S2}$$

As z is the only absorbing state of this new Markov chain, the full transition probability matrix is

$$\mathbf{P^c} = \begin{pmatrix} \mathbf{Q^c} & \mathbf{R^c} \\ \mathbf{0} & 1 \end{pmatrix}, \tag{S3}$$

where $\mathbf{R^c}$ contains the transition probabilities from the transient states to z. These probabilities are chosen such that each row sums to 1. In this case, $\mathbf{0}$ stands for a $1 \times (z-1)$ vector filled with zeros. We used this transition probability matrix to simulate the population dynamics conditioned on success.

To further study the conditioned Markov chain, we computed its fundamental matrix $\mathbf{W^c} = (\mathbf{I} - \mathbf{Q^c})^{-1}$. W^c_{ij} gives the number of generations a population starting at size i spends at size j before reaching z, conditioned on reaching

z before going extinct. These are the values shown in Figure 4. Note that in these plots we did not include the first generation, which the population necessarily spends at its founder size.

We also computed the expected number of surviving offspring per individual at population size i under the conditioned population dynamics (Figure S1):

$$\frac{1}{i}\sum_{j=1}^{z} j \cdot P_{ij}^{c}. \tag{S4}$$

This is an approximation because our Markov chain is restricted to population sizes up to z whereas actual populations would be able to grow beyond z. However, in the range of population sizes that is most relevant for our study, i.e. at small and intermediate population sizes, Equation (S4) should give an accurate approximation of the expected number of surviving offspring per individual.

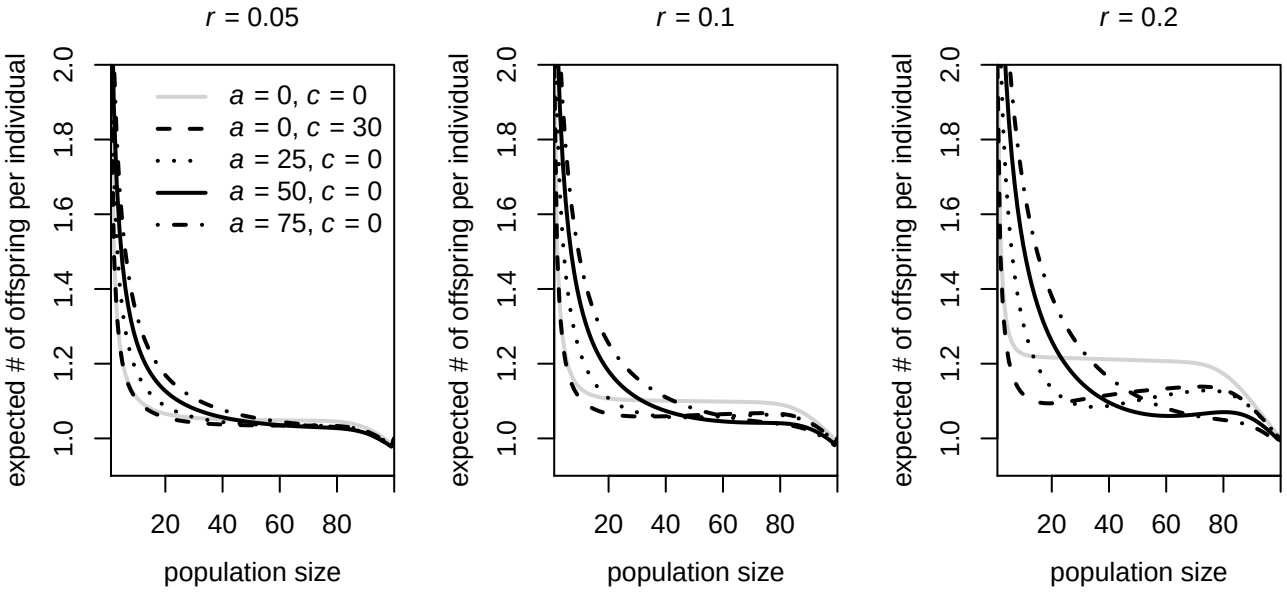

Figure S1 The expected number of surviving offspring per individual in successful populations (see Equation (S4)) as a function of the current population size without an Allee effect (gray line), with a weak Allee effect (dashed black line), or with a strong Allee effect (other black lines) and different critical sizes. The subplots differ in the growth parameter r. $k_1 = 1000$.

File S2 Details on the simulation of genealogies

The genealogies are constructed by tracing the ancestry of the sampled genetic material backwards in time. Apart from multiple and simultaneous mergers, another special feature is that the genealogical process takes into account explicitly that individuals are diploid and bi-parental and thus avoids logical inconsistencies that may occur when independently simulating the genealogies at different loci (Wakeley *et al.* 2012). However, this realism comes at a computational cost and in cases where we are only interested in average levels of genetic diversity, i.e. for the analyses underlying Figures 5, 7, and S3 we resorted to independently simulating the genealogies at the different loci.

The current state of the ancestry is defined by a set of lineage packages for each population (source population and newly founded population). Such a lineage package contains all the genetic material that is traveling within the same individual at that time point. It has two sets of slots, one set for each genome copy. Each set has a slot for each locus. If the genetic material at a certain locus and genome copy is ancestral to the sample, the slot is occupied by a node, otherwise it is empty.

The ancestral history starts with $2 \cdot n_s$ lineage packages in the newly founded population. Initially all slots in the lineage packages are occupied by nodes. From there, the ancestry is modeled backwards in time until at each locus there is just one node left. Given the state of the ancestry in generation t, the state in generation $t-1$ is generated as follows: Backward in time, each generation starts with a migration phase (Figure S2). All lineage packages that are currently in the newly founded population choose uniformly without replacement one of the Y_t migrants from the source population, or one of the $N_t - Y_t$ residents. Note that in our simulations with a single founding event, $Y_0 = N_0$ and $Y_t = 0$ for all $t > 0$, whereas in the case of multiple introductions, Y_t can be positive also at $t > 0$. According to their choice in this step, lineage packages either remain in the newly founded population or are transferred to the source population.

Then each lineage package splits into two because the two genome copies (sets of slots) each derive from a possibly different parent (see Figure S2). Lineage packages that do not contain ancestral material are discarded immediately. For each of the remaining lineage packages, recombination is implemented by independently constructing a stochastic map $R : \{1, \ldots, n_l\} \to \{0, 1\}$ such that

$$R(1) = \begin{cases} 0 & \text{with probability } \frac{1}{2} \\ 1 & \text{with probability } \frac{1}{2} \end{cases} \tag{S5}$$

and then

$$R(n+1) = \begin{cases} R(n) & \text{with probability } 1 - \rho_n \\ 1 - R(n) & \text{with probability } \rho_n \end{cases} \tag{S6}$$

is drawn recursively for $n \in \{1, \ldots, n_l - 1\}$. The recombination probability ρ_n between loci n and $n+1$ was 0.5 for all analyses in this study. A node at locus l in the new lineage package is placed into the first genome copy if $R(l) = 0$ and

into the second genome copy if $R(l) = 1$.

After each lineage package underwent splitting and recombination, all resulting lineage packages uniformly pick one of the N_{t-1} or k_0 individuals as ancestor, depending on whether they are in the newly founded or in the source population, this time with replacement (see Figure S2). Lineage packages that chose the same ancestor are merged. If there is more than one node at the same genome copy and slot, a coalescent event takes place.

Because genetic drift is strong in small populations, many pairs of lineages will already encounter their common ancestor within the newly founded population. The lineages that did not coalesce until time 0 must all be in the source population which is assumed to be of constant size k_0 at all times. To simulate the remainder of the ancestry, we could proceed by going backwards generation-by-generation. However, as the source population is large it would take a long time until all lineages find their most recent common ancestor (MRCA) and in most generations nothing would happen. Furthermore, nodes within the same lineage package typically become separated by recombination relatively fast. For the sake of computational efficiency, we therefore use an approximative algorithm to simulate the remainder of the ancestral history. This efficient simulation mode excludes multiple and simultaneous mergers, events that should be very rare for a reasonably large source population size k_0. Independently for each locus, we determine the remaining number of nodes n_{total} and draw the number of generations T until the next coalescent event from a geometric distribution with success probability

$$p_{merge} = \frac{\binom{n_{total}}{2}}{2k_0}. \tag{S7}$$

We merge two randomly chosen nodes, reduce n_{total} by 1, and update the current time to $t-T$. We repeat this procedure until eventually there is only one node left, the MRCA of all sampled genetic material at the respective locus. Throughout the simulation, we store all information needed to provide the topology and branch lengths (in number of generations) for the genealogies at each locus.

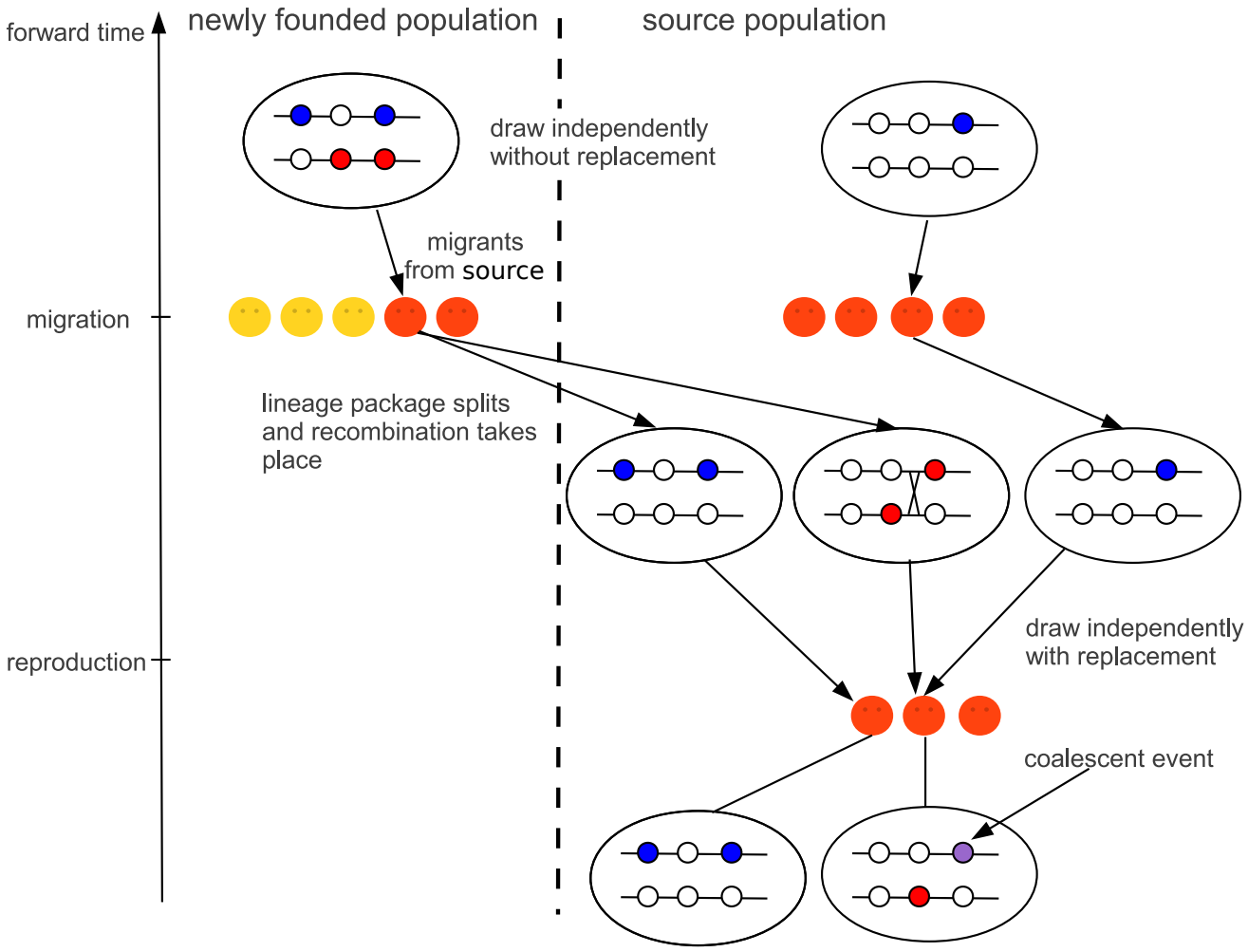

Figure S2 Illustration of the backward-in-time simulation of genealogies.

File S3 Results based on total length of the genealogy

In the main text, we use average pairwise coalescence times to assess genetic diversity. Here we show the corresponding results for the average total length of the genealogy G_{total}, a measure related to the number of segregating sites or the number of alleles in a sample. To measure the proportion of variation maintained, we divided G_{total} by $4k_0 \cdot \sum_{i=1}^{2n_s-1} \frac{1}{i}$, the expected total length of the sample genealogy if all lineages would have been sampled in the source population (Wakeley 2009, p. 76). The results (Figure S3 and S4) were qualitatively similar to the results based on average pairwise coalescence times (see Figures 5 and 6), except that the proportion of variation maintained more slowly approached one with increasing founder population size.

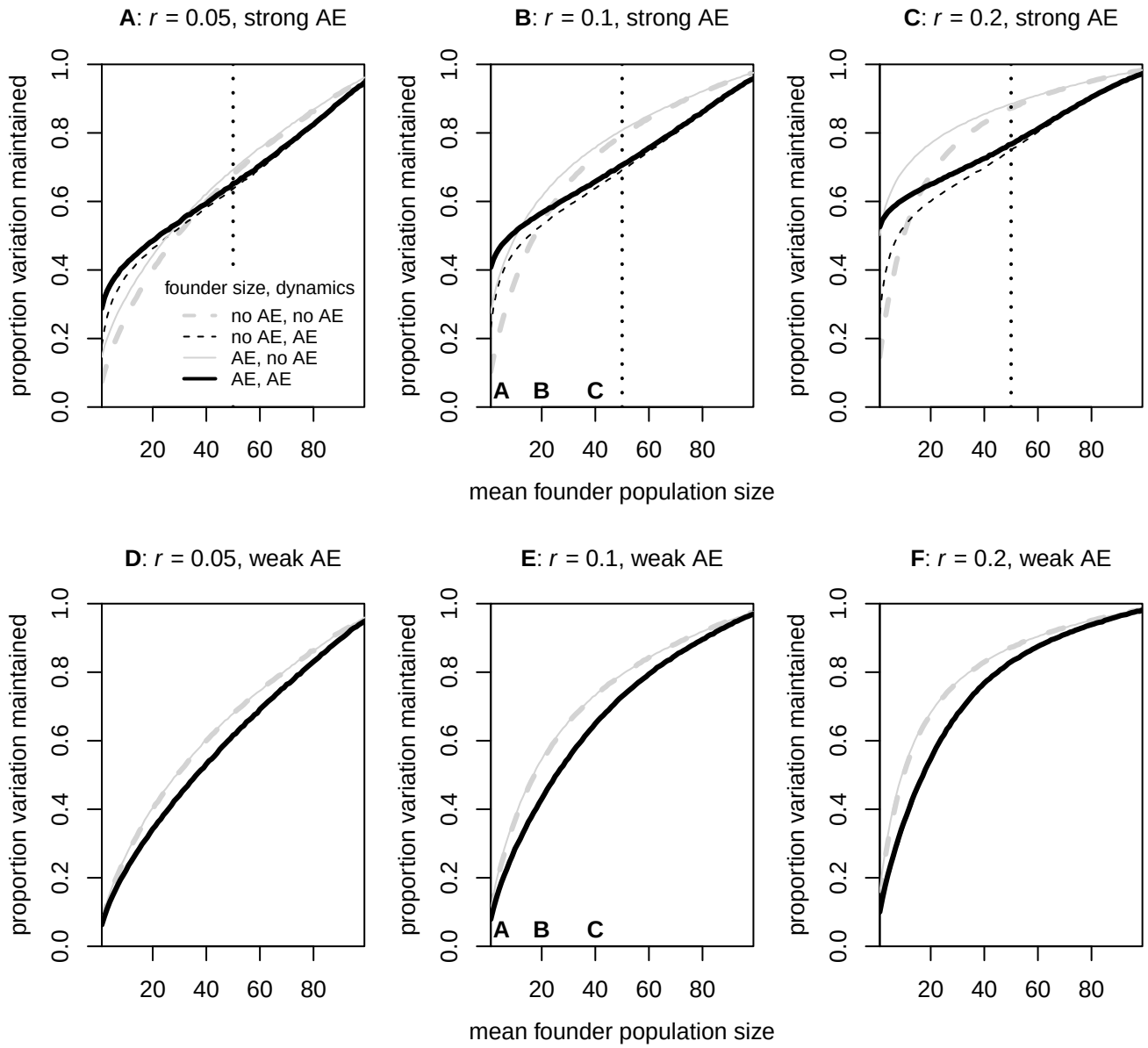

Figure S3 Average proportion of genetic variation from the source population (based on the average total length of sample genealogies) that is maintained by an introduced population upon reaching size z. The subplots differ in the value of the growth rate parameter r and in the type of Allee effect. In the upper row, Allee-effect populations have a strong demographic Allee effect with $a = 50$ (indicated by dotted vertical line) and $c = 0$. In the lower row, Allee-effect population have a weak Allee effect with $a = 0$ and $c = 30$. The values on the x-axes correspond to the mean of the original founder-size distribution. The four sets of populations in each subplot serve to disentangle the genetic effects resulting from the shift in founder population sizes and those from the altered post-introduction population dynamics. Dashed lines: founder population size drawn from the success-conditioned distribution without an Allee effect. Solid lines: founder population sizes drawn from the success-conditioned distribution with an Allee effect (strong in the upper row, and weak in the lower row). Black lines: success-conditioned population dynamics with an Allee effect. Gray lines: success-conditioned population dynamics without an Allee effect. The letters A, B, and C in subplots (B) and (E) refer to the subplots in Figures 3 and 4, where we examined for $r = 0.1$ and the respective (mean) founder population sizes how the Allee effect influences the conditioned distribution of founder population sizes and the conditioned population dynamics. Across all points, standard errors were between 0.0004 and 0.0015, and the corresponding standard deviations between 0.689 and 0.203.

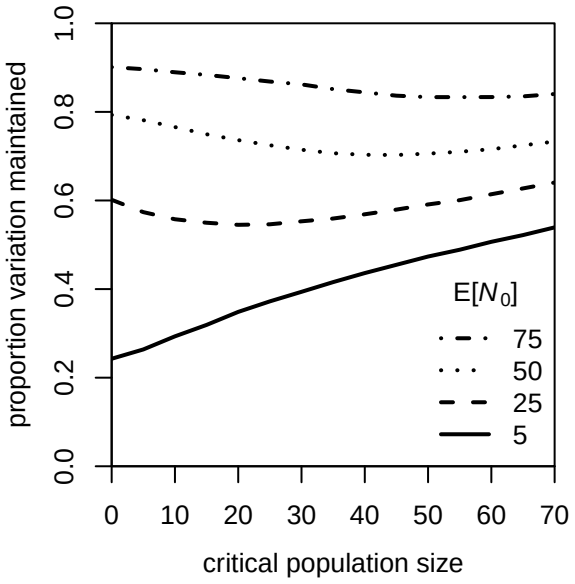

Figure S4 The role of the critical population size a for the average proportion of genetic variation from the source population (based on the average total length of sample genealogies) that is maintained by an introduced population upon reaching size $z = 100$. The average founder population size $\mathbf{E}[N_0]$ is held fixed at a different value for each of the four curves. Each point represents the average over 20,000 successful populations. Standard deviations were between 0.085 and 0.158 and standard errors between 0.0006 and 0.0012. $c = 0, r = 0.1$.

File S4 Methodology for estimating the critical population size

We generated 1000 pseudo-observed data sets and 100,000 simulated data sets, each with independent introductions to 200 locations. The critical population sizes were drawn from a uniform distribution on [0,100]. We fixed the other parameters of the population dynamics ($k_0 = 10,000, k_1 = 1000, r = 0.1, c = 0$) and assumed them to be known with certainty. We further assumed that the original distribution of founder population sizes was Poisson with mean 20, and sampled the founder population sizes independently for each location from the conditioned distribution of founder population sizes for the respective critical population size. Given the selected founder population size, we simulated the population dynamics at each location from the conditioned Markov chain until the population reached size 200, i.e. twice the largest possible critical population size.

At this point, we sampled $n_l = 20$ individuals at both genome copies, resulting in 40 copies of each locus from a given location. We generated genealogies for 20 freely recombining loci. To obtain a more differentiated picture of patterns of genetic variation and capture as much information as possible, we did not use the average pairwise coalescence times or total lengths of the genealogy as before. Instead, we took the means and variances across loci of the entries of the site-frequency spectrum (SFS) ξ_i, i.e. the number of mutations that appear in i chromosomes in the sample for $i \in 1, 2, \ldots, 39$. To compute these summary statistics, we first took the combined length of all branches B_i that have i descendants in the sample and assumed the infinite-sites model such that the number of mutations on a branch of length b is Poisson-distributed with parameter $\mu \cdot b$. Given sufficiently many loci, which we assume we have, we do not actually need to simulate mutations along the branches. Instead, we can use the branch lengths to directly estimate the means and variances across loci of the ξ_i as

$$\hat{\mathbf{E}}[\xi_i] = \mu \cdot \bar{B}_i \tag{S8}$$

and, using the law of total variance,

$$\hat{\mathbf{Var}}[\xi_i] = \hat{\mathbf{E}}\left[\mathbf{Var}[\xi_i|B_i]\right] + \hat{\mathbf{Var}}\left[\mathbf{E}[\xi_i|B_i]\right] = \mu \cdot \bar{B}_i + \mu^2 \cdot s^2(B_i), \tag{S9}$$

where the $\bar{B}_i$ are the average branch lengths across the n_l loci and the $s^2(B_i)$ are the corresponding empirical variances. We assumed throughout that $\mu = 0.001$.

We further summarized the data for each SFS entry $i \in 1, 2, \ldots, 39$ by computing the averages and empirical standard deviations of the quantities in (S8) and (S9) across locations. To investigate how the quality of the estimation depends on the number of independent locations available, we took into account either only 10, 25, 50, 100, or all 200 of them to compute these statistics. Using the pls script from abctoolbox (Wegmann *et al.* 2010) and the pls package in R (Mevik and Wehrens 2007), we then conducted partial least squares regression on the first 10,000 simulated data sets to condense

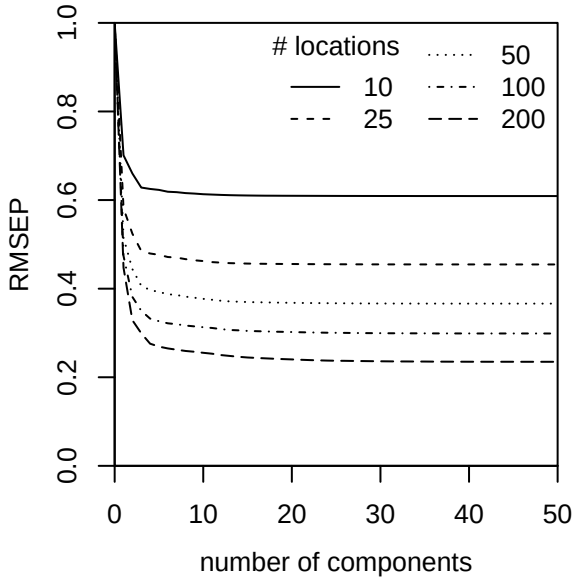

Figure S5 Root mean squared error of prediction (RMSEP) as a function of the number of PLS components for various numbers of locations.

the information contained in the 156 summary statistics to a smaller number of components. To decide on the number of components, we examined plots of the root mean squared error of prediction (RMSEP) as a function of the number of components (Figure S5). For none of the different numbers of locations did the RMSEP change substantially beyond 20 components. Thus, we decided to include 20 components as summary statistics for ABC.

We used these 20 PLS components as summary statistics for parameter estimation with the R package abc (Csilléry *et al.* 2012). We chose a tolerance of 1 % and used the option "loclinear" implementing the local linear regression method (Beaumont *et al.* 2002). To avoid estimated parameter values that fall outside the prior, we estimated $\ln(a/(100-a))$ and then back-transformed the estimated values. For each pseudo-observed data set, we thus used the 100,000 simulated data sets to approximate the posterior distribution of the critical population size given a uniform prior on [0,100]. For each data set, we stored the mean of the posterior, which we take as our point estimator, and the 50 % and 95 % credibility intervals. We observed that the quality of parameter inference improved with an increasing number of locations (Figures 8 and S6). An examination of the percentage of pseudo-observed data sets for which the true parameter value falls into the respective 50 % or 95 % credibility interval suggests that ABC approximates Bayesian inference reasonably well in this case (Figure S7).

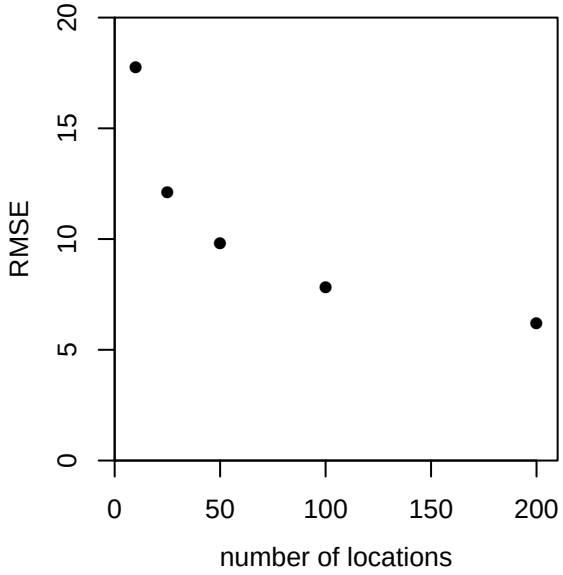

Figure S6 The root mean squared error (RMSE) of the estimated critical population size as a function of the number of independent locations used for the estimation.

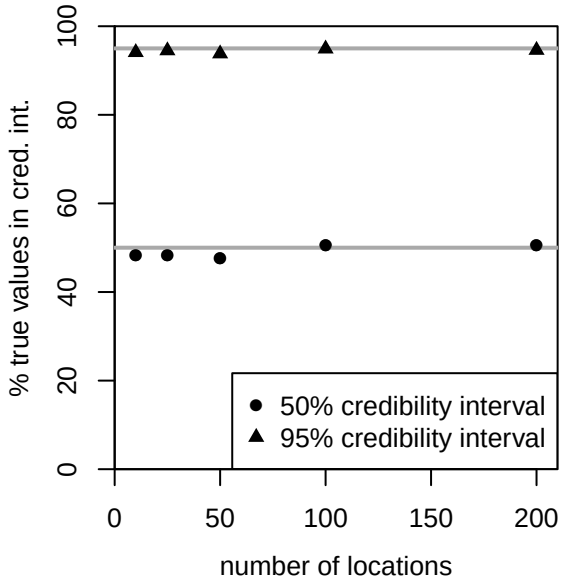

Figure S7 Percentage of true parameter values that fall within the 50% and 95% credibility interval, an indicator for how well Approximate Bayesian Computation approximates Bayesian inference. The gray lines are at 50% and 95%.

File S5 Consequences of a neglected Allee effect

Using the ABC framework again, we explored the consequences of neglecting the Allee effect when estimating other demographic parameters: the founder population size N_0, the growth parameter r, and the number of generations since the founding event. We generated 2000 pseudo-observed data sets from our stochastic model, 1000 without an Allee effect and 1000 with an Allee effect and a critical population size of 50. As the basis for estimation in ABC, we used 100,000 data sets that were simulated from a model without an Allee effect. To also explore the consequences of neglecting stochasticity, we considered two versions of the model without an Allee effect: our stochastic model with $a = 0$ and a modified version where we removed as much stochasticity as possible. That is, the population size in the next generation was not drawn from a Poisson distribution, but was set to $\mathbf{E}[N_{t+1}]$ if this value was an integer. Otherwise, we randomly set N_{t+1} to the next smallest or next largest integer with the respective probabilities chosen such that Equation (1) was fulfilled.

The priors for the demographic parameters of interest were as follows: $r \sim \text{unif}([0.05, 0.1])$ and $n_g \sim \text{unif}(\{30, \ldots, 500\})$, where unif stands for the uniform distribution. To generate values for N_0, we first drew $Y \sim \text{unif}([\ln(5), \ln(80)])$, and then set N_0 to e^Y, rounded to the next integer. The other parameters were fixed: $k_0 = 10,000, k_1 = 1000, \mu = 0.001, n_s = 10$. For each data set, we retried simulating with the same parameter combination until we obtained a successful population with $N_{n_g} \geq n_s$. We generated 100 independent genealogies for samples of size n_s taken at time n_g and computed means and variances of the entries of the site-frequency spectrum as described in File S4. Using partial least squares regression on the first 10,000 simulated data sets, we reduced this information to 20 components that served as summary statistics for ABC. As above, we used the R package abc (Csilléry *et al.* 2012) with a tolerance of 1 % and the option "loclinear'. In Figure S8, we compare the quality of parameter estimation across the four possible combinations of whether or not the true model includes an Allee effect and whether the model used for estimation was stochastic or deterministic.

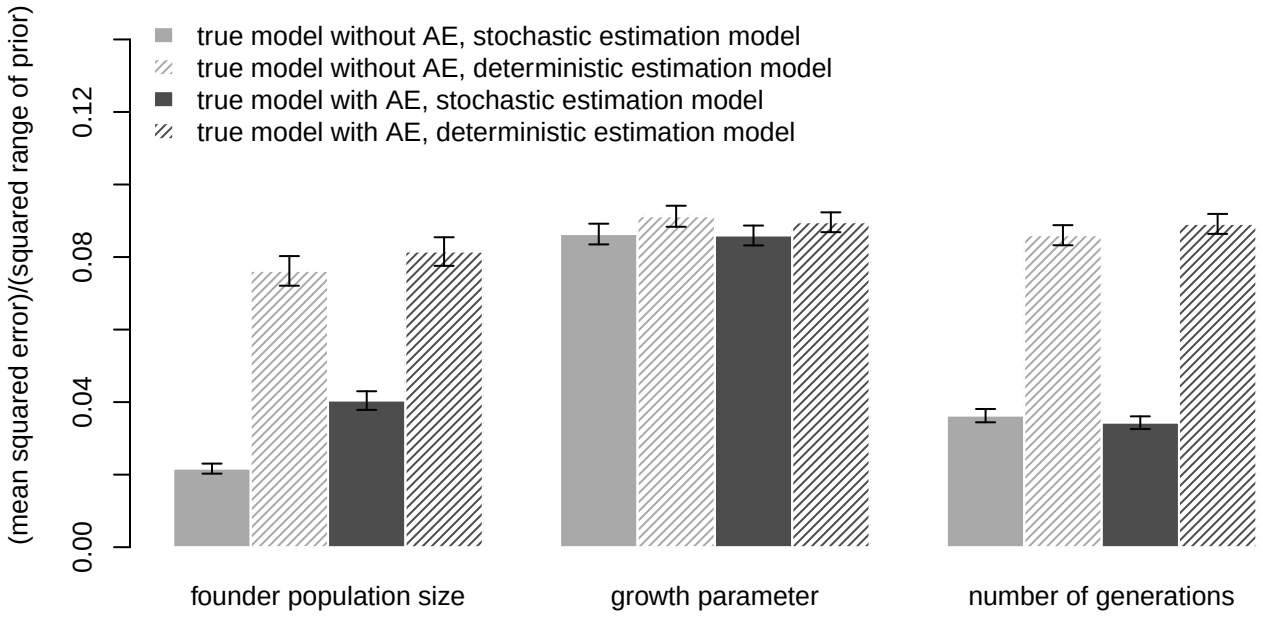

Figure S8 Mean squared error (MSE) $\pm$ its standard error, relative to the squared range of the prior for different demographic parameters in an Approximate Bayesian Computation analysis that neglects the Allee effect.

File S6 A conditioned diffusion process

Our results in the main text indicate that the growth rate under the conditioned population dynamics depends mostly on the absolute value and not so much on the sign of the growth rate under the unconditioned population dynamics. For mathematically interested readers, we now explore a simple model where this fact can be proven easily. We consider a diffusion process on the interval $[0, 1]$ with constant infinitesimal mean $\mu(x) = \mu$ and constant infinitesimal variance $\sigma^2(x) = \sigma^2$. We will show that the associated diffusion process conditioned on hitting 1 before 0 is independent of the sign of μ and that its infinitesimal mean increases with $|\mu|$.

Our task is to compute the infinitesimal mean and variance of the conditioned diffusion process. Following the formulas given by Karlin and Taylor (1981, p. 263), the infinitesimal mean of the conditioned diffusion process is

$$\mu^*(x) = \mu(x) + \frac{s(x)}{S(x)} \cdot \sigma^2(x), \tag{S10}$$

where $S(x)$ is the scale function and $s(x)$ is its derivative. Using the definitions of these functions (e.g. Karlin and Taylor 1981, p. 262) and plugging in the parameters of our diffusion, we obtain

$$s(x) = \exp\left(-\int_0^x \frac{2\mu(\eta)}{\sigma^2(\eta)} d\eta\right) = \exp\left(-\frac{2\mu x}{\sigma^2}\right) \tag{S11}$$

and

$$S(x) = \int_0^x s(\eta) d\eta = \frac{\sigma^2}{2\mu} \cdot \left[1 - \exp\left(-\frac{2\mu x}{\sigma^2}\right)\right]. \tag{S12}$$

Substituting Equation (S11) and (S12) into Equation (S10), we obtain

$$\mu^*(x) = \mu \cdot \frac{\exp(2\mu x/\sigma^2) + 1}{\exp(2\mu x/\sigma^2) - 1} =: f(\mu), \tag{S13}$$

a function that is symmetric about 0, i.e. $f(-\mu) = f(\mu)$, and increases with the absolute value of μ (Figure S9). The variance $\sigma^{2*}(x)$ equals the original variance $\sigma^2(x)$.

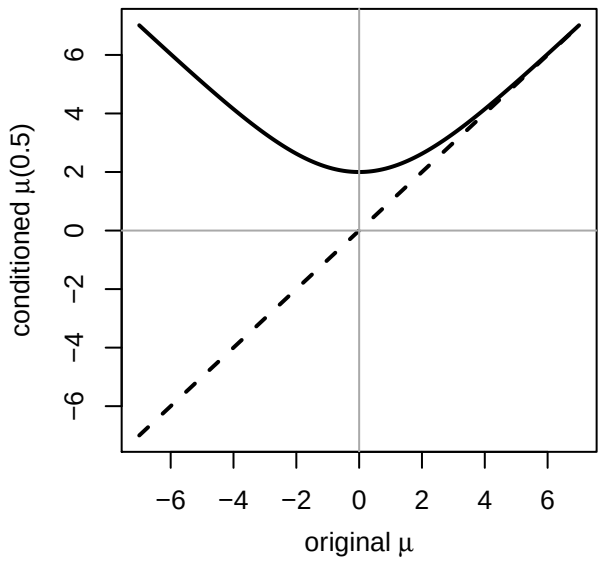

Figure S9 The infinitesimal mean of the conditioned diffusion process at 0.5 (solid line), i.e. in the middle of the interval, as a function of the infinitesimal mean of the original process. On the dashed line, the infinitesimal means of original and conditioned process would be equal.

File S7 Source code and analysis scripts

164 File S7 is available for download as a zip-folder at

`http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.114.167551/-/DC1`.

References

Beaumont, M. A., W. Y. Zhang, and D. J. Balding, 2002 Approximate Bayesian computation in population genetics. Genetics 162: 2025–2035.

Csilléry, K., O. Francois, and M. G. B. Blum, 2012 abc: an R package for Approximate Bayesian Computation (ABC). Methods Ecol. Evol. 3: 475–479.

Karlin, S., and H. M. Taylor, 1981 *A Second Course in Stochastic Processes*. Academic Press, New York.

Mevik, B.-H., and R. Wehrens, 2007 The pls package: Principal component and partial least squares regression in R. J. Stat. Softw. 18.

Pinsky, M., and S. Karlin, 2010 *An introduction to stochastic modeling*. Academic Press.

Wakeley, J., 2009 *Coalescent Theory - An introduction*. Roberts & Company, Greenwood Village, Colorado.

Wakeley, J., L. King, B. S. Low, and S. Ramachandran, 2012 Gene genealogies within a fixed pedigree, and the robustness of Kingman's coalescent. Genetics 190: 1433–1445.

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