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 $\mathbf{U} = \mathbf{W}\mathbf{R}$. The first column of $\mathbf{U}$ contains the success probabilities $\Pr(T_z < T_0 | N_0 = i)$ shown in Figure 2. Using the success probabilities and Bayes' formula, we then computed the transition probabilities of the Markov chain conditioned on $T_z < T_0$:

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

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.

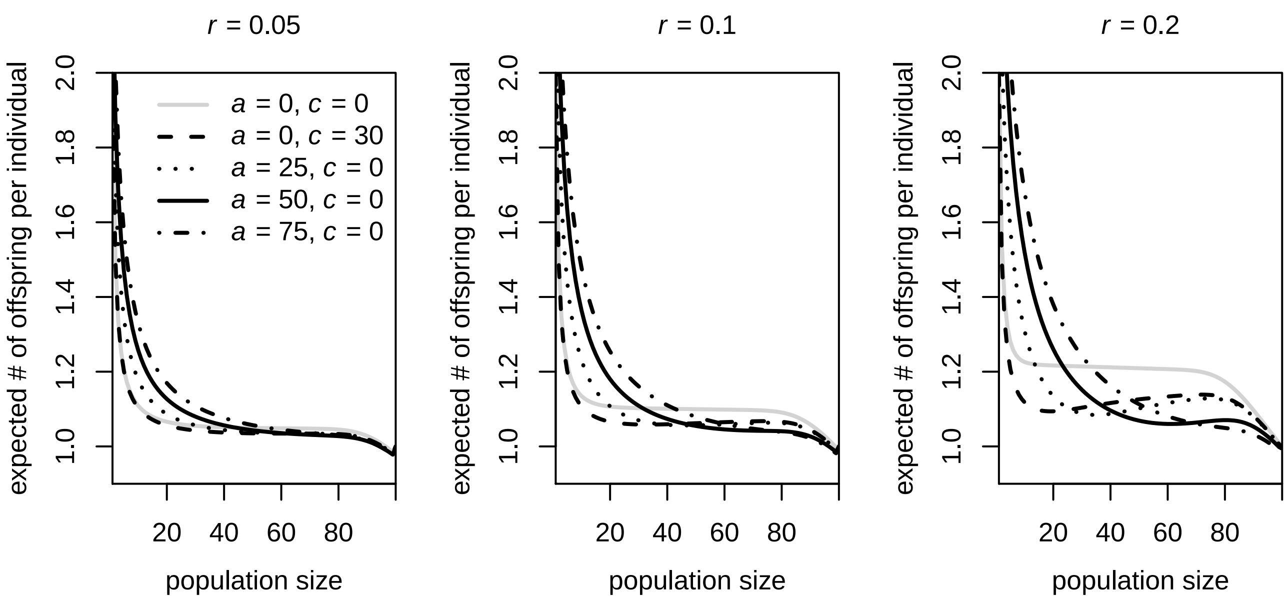

Figure S1 The expected number of surviving offspring per individual in successful populations (see Equation (S4)) as a function of the current population size without an Allee effect (gray line), with a weak Allee effect (dashed black line), or with a strong Allee effect (other black lines) and different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