File S2. Extended Moran model simulations

In this Supporting Information we compare a set of tree statistics obtained from the ancestral process $\left(\mathcal{A}^{\psi,\rho}_{\lfloor t/c_N(0)\rfloor,k}\right)_{t\in\mathbb{R}_{\geq 0}}$ from the underlying extended Moran model to those obtained from the coalescent process $\left(\Pi^{\psi,\rho}_{t,k}\right)_{t\geq 0}$ for different values of $\rho\in\mathbb{R}_{\geq 0}$, $\psi\in(0,1)$ with $\gamma=1.5$ (i.e., in the regime where sweepstake reproductive events dominate and the corresponding ancestral process is a time-inhomogeneous psi-coalescent). Coalescent simulations follow the algorithm outlined in the main text.

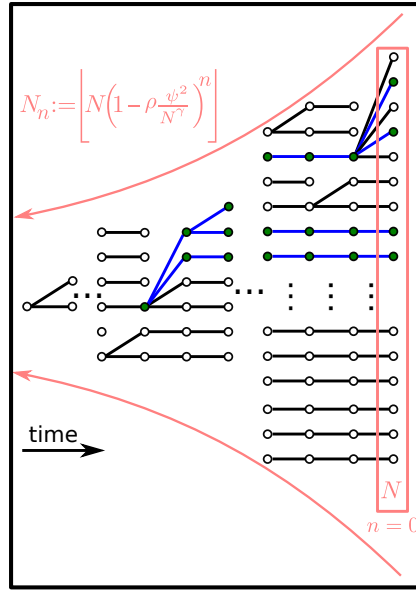The rational behind the extended Moran model simulations is outlined in Figure A.

Figure A – Illustration of genealogical relation in the extended Moran model with exponential growth. Circles denote individuals; time increases from left to right with the rightmost individuals constituting the present population. Green circles denote sampled individuals. Connecting lines between circles show the genealogical relationship between individuals; the genealogy of the sampled individuals is highlighted with blue. Pink lines illustrate that population size increases (forward in time) exponentially with rate ρ.

Starting from the present (i.e., $n=0$) and choosing k random samples, these were followed successively backwards in time thus creating a coalescent tree from the leaves to the root. At each time step the population size was adjusted (shrinking backward in time; eq. 6) and the type of the reproductive event (eq. 5) along with the corresponding number of offspring $U_N(n)$ (eq. 4) were (randomly) determined . Finally, to determine whether a merger event has occurred – i.e., if one of the k active lineages had found its parent in the previous time step – we drew a Hypergeometric random number Y with parameters k, N_n-k and $U_N(n)$. When $y>1$, i.e., when at least a 2-merger event has occurred, the number of active lineages k in the next time step becomes $k+1-y$, and remains constant otherwise. This process is repeated until $k=1$, that is until all samples have found their most recent common ancestor.

For both processes the total tree length T_{tot}, the time to the most recent common ancestor $TMRCA$, and the ratio of the sum of the length of all branches with i descendants over the total tree length T_i/T_{tot} were recorded. While the number of samples was limited to $k=4$ for computational reasons, the match between coalescent and Moran model simulations was almost perfect across the entire range of ψ (Fig. B-D). Only for very small ψ we observed some slight deviations observed.
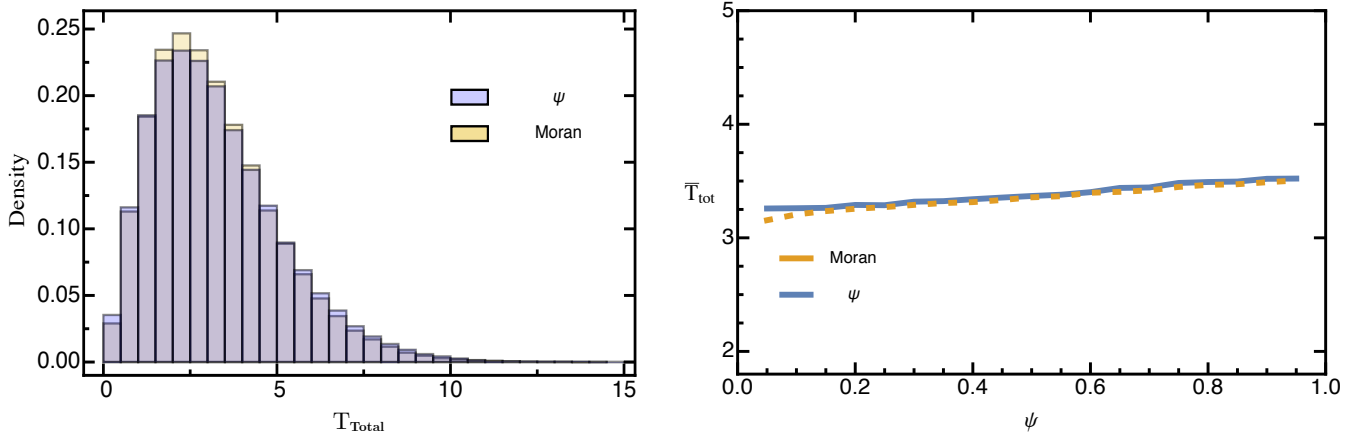

Figure B – The left figure shows a comparison of $\overline{T}_{tot}$ based on simulated trajectories of $\left(\mathcal{A}^{\psi,\rho}_{\lfloor t/c_N(0)\rfloor,4}\right)_{t\geq 0}$ and $\left(\Pi^{\psi,\rho}_{t,4}\right)_{t\geq 0}$ for $\rho=0.1$, $\gamma=1.5$, and different values of ψ. The right figure depicts the corresponding empirical distributions of $T^{(i)}_{tot}$ for $\psi=0.1$. Averages were taken over 100,000 replicates. Other parameters: $N=100,000$.
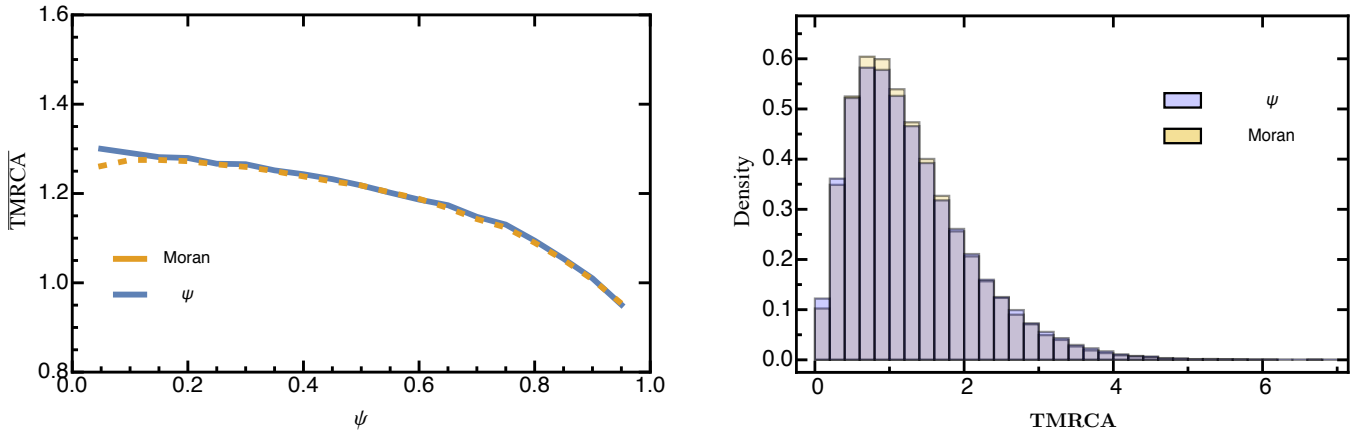
Figure C – The left figure shows a comparison of $\overline{T}_{MRCA}$ based on simulated trajectories of $\left(\mathcal{A}^{\psi,\rho}_{\lfloor t/c_N(0) \rfloor, A} \right)_{t \geq 0}$ and $\left(\Pi^{\psi,\rho}_{t,A} \right)_{t \geq 0}$ for $\rho = 0.1$, $\gamma = 1.5$, and different values of ψ. The right figure depicts the corresponding empirical distributions of $T^{(i)}_{MRCA}$ for $\psi = 0.1$. Averages were taken over 100,000 replicates. Other parameters: $N = 100,000$.
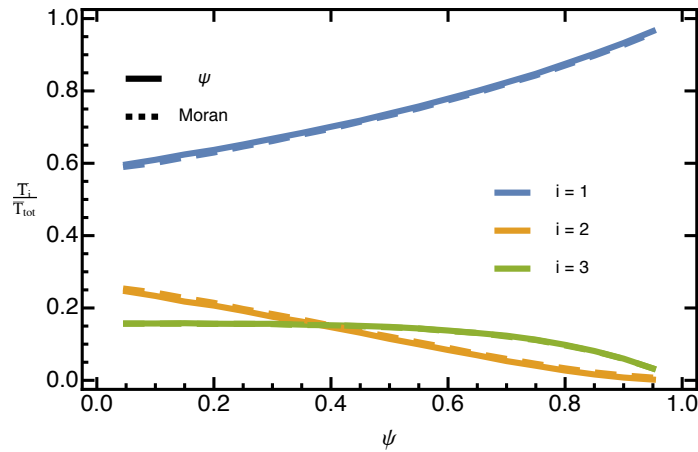

Figure D – A comparison of $(T_i/\overline{T}_{tot})_{i=1,2,3}$ based on simulated trajectories of $\left(\mathcal{A}^{\psi,\rho}_{\lfloor t/c_N(0) \rfloor, A} \right)_{t \geq 0}$ and $\left(\Pi^{\psi,\rho}_{t,A} \right)_{t \geq 0}$ for $\rho = 0.1$, $\gamma = 1.5$, and different values of ψ. Averages were taken over 100,000 replicates. Other parameters: $N = 100,000$.

These results highlight two important points. First and reassuringly these results indicate that for large initial population sizes N the extended Moran model can be described by its limiting ancestral process (i.e., the psi-coalescent). Second, simulating the population dynamics forward in time is computationally intensive and prohibitively slow. However, since the ancestral process accurately captures the forward-in-time population dynamics, the coalescent process can be used to simulate the process quickly and efficiently over a large parameter space and for large sample sizes.