

Extreme value statistics of mutation accumulation in renewing cell populations: Supplemental Material

Philip Greulich^{1,2} and Benjamin D. Simons^{3,4,5}

¹*Mathematical Sciences, University of Southampton, Southampton, UK*

²*Institute for Life Sciences, University of Southampton, Southampton, UK*

³*Cavendish Laboratory, University of Cambridge, Cambridge, UK*

⁴*Gurdon Institute, University of Cambridge, Cambridge, UK*

⁵*Cambridge Stem Cell Institute, University of Cambridge, Cambridge, UK*

(Dated: September 28, 2018)

EXTREME VALUE STATISTICS FOR $\lambda = 0$

In the case $\lambda = 0$, each cell accumulates mutations independently from each other, according to process (1), main text. This process corresponds to a simple Poisson process in each cell, thus the model is equivalent to N independent and identically distributed (i.i.d.) Poisson variables. In this case, the probability that $m^* = \max(m_1, \dots, m_N) < m_c$, the CDF $P_N^{*(\lambda=0)}(m_c, T)$ is the same as the probability that all individual i.i.d. $m_i < m_c$, $P^{\lambda=0}(m_c)$. Thus,

$$P_N^{*(\lambda=0)}(m_c, T) = [P^{\lambda=0}(m_c, T)]^N = \left[\frac{\Gamma(m_c + 1, \mu T)}{\Gamma(m_c + 1)} \right]^N, \quad (1)$$

where the term in brackets on the right-hand side stands for the CDF of the Poisson process, with the gamma function $\Gamma(x)$ and incomplete gamma function $\Gamma(x, y)$.

Notably, it has been shown that the distribution of the maximum of Poisson variables does not converge to any simple scaling form [1]. This contrasts the behaviour of normally distributed variables x , whose extreme value distribution $P_N^{*(n)}(x_c) = \text{Prob}(x^* < x_c)$ for the maximum, $x^* = \max(x_1, \dots, x_N)$, converges towards the Gumbel distribution for $N \rightarrow \infty$ [2]. This distribution can be written as a scaling form [21]

$$P_N^{*(n)}(X) = e^{-e^{-X}} \text{ with } X = \frac{x_c - \bar{x}}{\sigma_N} \quad (2)$$

where \bar{x} and σ_N are scaling parameters (\bar{x} being also the mode of the distribution). For a Normal distribution with variance σ^2 , for $N \gg 1$ these are [2]

$$\bar{x} \simeq \sqrt{2 \ln N} \sigma, \quad \sigma_N \simeq \frac{\sigma}{\sqrt{2 \ln N}} \quad (3)$$

We can also use the relationship between median \tilde{x} and mode \bar{x} of the Gumbel distribution, $\tilde{x} = \bar{x} - \sigma_N \ln \ln 2$ [3] to express the median as

$$\tilde{x} \simeq \bar{x} - (\sigma \ln \ln 2)(2 \ln N)^{-1/2} \simeq \sqrt{2 \ln N} \sigma, \quad (4)$$

where on the right hand side the term vanishing as $O((\ln N)^{-1/2})$ has been omitted in the scaling $N \rightarrow \infty$.

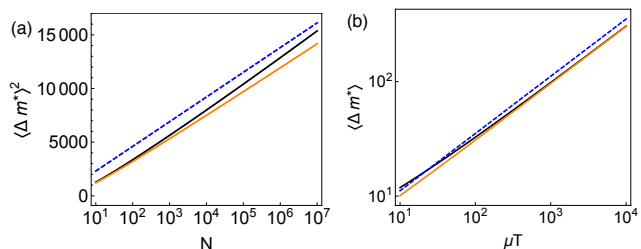


Figure 1. The scaling of the difference of mean maximum mutation number and population mean $\langle \Delta m^* \rangle = \langle m^* \rangle - \langle m_i \rangle$, for Poisson-distributed (black) and normally-distributed random variables (orange), together with the predicted scaling limit of the latter, $\simeq \sqrt{2\mu T \ln N}$ [2] (blue dashed line). The former is according to the mutation accumulation for $\lambda = 0$ and is computed by $\langle \Delta m^* \rangle = \tilde{m} + \sigma_N(\gamma_e + \ln \ln 2)$, with $\gamma_e = 0.5772$ and CDF from Eq. (1). (a) $\langle \Delta m^* \rangle^2$ as a function of N for $\mu T = 500$. (b) $\langle \Delta m^* \rangle$ as a function of μT for $N = 500$.

For $\mu T \gg 1$ we can approximate the statistics of m^* for our model with $\lambda = 0$ – independent Poisson variables m_i , with mean μT – by normally distributed random variables $\Delta m_i = m - \langle m_i \rangle$ with mean zero and variance $\sigma^2 = \mu T$. We now define the scaling variable in terms of the maximum mutation number and its median \tilde{m} and write $\Delta m_c := m_c - \langle m_i \rangle$. Then we have the approximation

$$P_N^*(\Delta m_c) \simeq P_N^{*(n)}(X) = e^{-e^{-X}} \quad (5)$$

$$\text{with } X = \frac{\Delta m_c - \tilde{m}}{\sigma_N} - \ln \ln 2$$

with median and scaling width

$$\tilde{m} \simeq \sqrt{2 \mu T \ln N}, \quad \sigma_N \simeq \sqrt{\frac{\mu T}{2 \ln N}} \quad (6)$$

The same scaling applies to the mean $\langle \Delta m^* \rangle = \tilde{m} + \sigma_N(\gamma_e + \ln \ln 2)$ with the Euler-Mascheroni constant $\gamma_e = 0.5772$, for which the vanishing term can be again neglected for $N \gg 1$.

To test this approximation, we compute the mean maximum mutation number of the Poisson process from Eq. (1), $\langle m^* \rangle = \sum_{m_c=0}^{\infty} (1 - P_N^*(m_c))$. This is shown as a

function of cell number N and time T in Fig. 1, together with the corresponding value for normally distributed variables with mean and variance μT and the scaling limit of the latter, according to Eq. (6). This illustrates that despite not converging for $N \rightarrow \infty$, the mean maximum of normally distributed variables, with mean and variance μT , serves as a good approximation for the mean maximum of Poisson variables with mean μT , for a wide range of N . While for $N \rightarrow \infty$ (with T fixed), the scaling of Poisson variables is expected to deviate from that of Normal variables [1], for $\mu T \rightarrow \infty$, the mean value $\langle \Delta m^* \rangle$ of Poisson and Normal variables do converge, as can be seen in Fig. 1b, which is due to the convergence of Poisson and Normal distribution for large μT .

MONTE CARLO SIMULATIONS

We use a kinetic Monte Carlo method with random sequential update to simulate the stochastic model described by Eqs. (1) in the main text. The system state is defined by N sites, $i = 1, \dots, N$, each characterized by the mutation number m_i of the cell residing on it. Initially, at time $t = 0$, all m_i are set to $m_i = 0$. In this article, we only consider scenarios with $\mu \leq \lambda$, therefore the following algorithm is described under this assumption (which improves time efficiency). By this algorithm, at each Monte Carlo time step, first the time is updated by $t \rightarrow t + 1/\lambda$ and the following steps are executed N times:

1. Generate two integer pseudo-random numbers $r_1, r_2 \in \{1, 2, \dots, N\}$.
2. Choose sites $i = r_1$ and $j = r_2$.
3. Set $m_j = m_i$.
4. Generate integer pseudo-random number $r_3 \in \{1, 2, \dots, N\}$ and floating point random number $r_4 \in [0, 1]$.
5. Choose site $i = r_3$. If $r_4 < \mu/\lambda$, set $m_i \rightarrow m_i + 1$. Return to step 1.

For $\mu > \lambda$ the algorithm would proceed accordingly.

This algorithm describes the stochastic process asymptotically exactly for large N . It differs, however, from the Gillespie algorithm [4], which is exact also for small N , in some aspects: (i) Random numbers are drawn independently for each process, mutation and loss/replacement. (ii) It does not check whether $j = i$, in which case no division happens, thus the rate λ is effectively reduced by $\lambda \rightarrow \lambda(1 - 1/N)$. (iii) Time is updated by the mean time between processes instead of an exponentially distributed random time step. These simplifications improve the time efficiency, as for $\mu = \lambda$ the time step size is

effectively doubled compared to the Gillespie algorithm. Since we are only interested in a theory for large N , this approximation does not affect our results for the scaling with N and T , yet reaching higher simulation efficiency.

CONSTRUCTION OF THE GENEALOGY

A *mutational path* is defined as the mutational history of a cell on site i , from the cell's birth to its replacement by another cell, by which the information on site i is overwritten. The *genealogy* of the cell population at time $t = T$ is recursively defined as the mutational history of cells at present time T combined with the genealogies of all their mother cells. It can also be seen as the collection of all mutational paths of the progeny of cells at time $t = 0$, removing those mutational paths that end before time $t = T$ without progeny (see Fig. 1a, main text, and Ref. [5]).

The genealogy forms a binary tree, characterized by its branching times t_k , at which a new branch from $k - 1$ independent branches is generated. Thus, during the time period t with $t_k < t < t_{k+1}$, there are k branches. The genealogy can be reconstructed by following the mutational history of cells backwards in time $\hat{t} := T - t$, in form of a *coalescent* process [6–8]. At each time point when a cell division has occurred, the trajectories of two daughter cells are merged (in time direction \hat{t}), forming a new branch of the genealogy, corresponding to their mother cell. This corresponds to a Markov process with coalescence/branching time intervals $\Delta t_k = t_{k+1} - t_k$ distributed exponentially with probability $\text{Prob}(\Delta t_k) = \langle \Delta t_k \rangle^{-1} e^{-\Delta t_k / \langle \Delta t_k \rangle}$ [8].

To determine $\langle \Delta t_k \rangle$ we need to derive the rate at which merging of branches occurs (see also Ref. [8]). The following applies to a Moran process in which a lost cell can be replaced by any other cell. Later we will also consider related lattice models in which only neighboring cells can replace each other. For two branches of the genealogy to merge, a cell division needs to occur *and* both daughter cells need to be part of the genealogy. The total rate of cell divisions in the population is $\lambda_{\text{tot}} = N\lambda$, and the probability that the first selected cell is a branch of the genealogy is k/N while the probability that the second selected cell is also part of the genealogy is $(k - 1)/N$. Hence, the rate at which two branches of the genealogy merge is $\omega = (\lambda N) \times (k - 1)/N \times k/N = \lambda k(k - 1)/N$, and thus

$$\langle \Delta t_k \rangle = \frac{1}{\omega} = \frac{N}{k(k - 1)} \frac{1}{\lambda} . \quad (7)$$

For large enough T , there will be a time point $\hat{T}_{\text{LCA}} := T - t_2$ after which – in backward time direction \hat{t} – only a single branch remains. This single branch is the last common ancestor (LCA). The mean time at which this

occurs is

$$\begin{aligned} \langle \hat{T}_{\text{LCA}} \rangle &:= \langle T - t_2 \rangle = \sum_{k=2}^N \langle \Delta t_k \rangle \\ &= \frac{N}{\lambda} \sum_{k=2}^N \frac{1}{k(k-1)} = \frac{N-1}{\lambda} \simeq \frac{N}{\lambda} \end{aligned} \quad (8)$$

for $N \gg 1$.

Branching times in low dimensions

The genealogy of the Moran process corresponds to a coalescing random walk backwards in time \hat{t} with rate λ in a network for which all sites are connected with each other. A realistic scenario would be that in a tissue only nearby cells can replace each other, mediated via cell-cell signalling. In this case it is sensible to consider a small range of loss/replacement, by embedding the dynamics on a finite regular d -dimensional square lattice on which loss/replacement only occurs between neighbors (cf. Refs. [9, 10]). For such a model, the genealogy in backward time-direction \hat{t} is a coalescent random walk on a d -dimensional lattice, for which analytical results have been obtained by Bramson and Griffeath [11] for asymptotically large times \hat{t} in any dimension d . They showed that the number of random walkers (branches) at time t , $k(t)$, diminish as

$$\frac{k(\hat{t})}{N} \simeq \begin{cases} \frac{1}{(\pi\lambda\hat{t})^{1/2}} & \text{for } d = 1 \\ \frac{\ln(\lambda\hat{t})}{\pi\lambda\hat{t}} & \text{for } d = 2 \\ \frac{1}{\gamma_d\lambda\hat{t}} & \text{for } d > 2 \end{cases} \quad (9)$$

asymptotically, for large time $\lambda\hat{t} \gg 1$ and $N \gg 1$, where γ_d depends on the dimension and corresponds to the probability that a random walker on a d -dimensional lattice never returns to its starting point [11]. For example, $\gamma_{d=3} \approx 0.66$ (see [12] as follows from [13–16]) and $\gamma_{d=\infty} = 1$. The merging rate is then $-\partial_{\hat{t}}k(\hat{t})$ and the mean merging time, which corresponds to the mean branching time of the genealogy is $\langle \Delta t_{k(\hat{t})} \rangle = -(\partial_{\hat{t}}k(\hat{t}))^{-1}$. With this we get

$$\langle \Delta t_k \rangle \simeq \begin{cases} \frac{2(\pi\lambda\hat{t})^{3/2}}{N\pi\lambda} = \frac{2N^2}{k^3\pi\lambda} & \text{for } d = 1 \\ \frac{(\pi\lambda\hat{t})^2}{(\ln(\lambda\hat{t})-1)N\pi\lambda} \approx \frac{N \ln(\lambda\hat{t})}{k^2\pi\lambda} & \text{for } d = 2 \\ \frac{(\gamma_d\lambda\hat{t})^2}{N\gamma_d\lambda} = \frac{N}{k^2\gamma_d\lambda} & \text{for } d > 2 \end{cases} \quad (10)$$

for large $N \gg 1$ and $\lambda\hat{t} \gg 1$ (we approximated $\ln(\lambda\hat{t}) - 1 \approx \ln(\lambda\hat{t})$). We see that the result for dimensions $d = \infty$ (with $\gamma_{d=\infty} = 1$) is consistent with branching times of the Moran process, since $\langle \Delta t_k \rangle = N/\lambda k(k-1) \simeq N/\lambda k^2$ for $k \gg 1$. This results holds, as the Moran process corresponds to the dynamics on an infinite-dimensional lattice.

The same scaling applies for all $d > 2$, though with different pre-factors γ_d . For a two-dimensional system the result is the same in leading order, but with a logarithmic correction in time. Nonetheless, for large $\lambda\hat{t}, \lambda T \gg 1$, we can approximate $\ln(\lambda\hat{t}) = \ln(\lambda T) + \ln(1 + (\hat{t} - T)/T) = \ln(\lambda T) + O((\lambda T)^{-1})$ which asymptotically depends only on λT . Thus, $\langle \Delta t_k \rangle \sim \lambda T N / (k^2 \lambda)$, which scales with N and k as the in the case $d > 2$. For $d = 1$, however, the branching times scale significantly differently. This affects the time to the LCA, \hat{T}_{LCA} , as

$$\langle \hat{T}_{\text{LCA}} \rangle = \sum_{k'=2}^{\infty} \langle \Delta t_{k'} \rangle \simeq \begin{cases} \frac{2(\zeta(3)-1)}{\pi\lambda} N^2 & \text{for } d = 1 \\ \sim N/\lambda & \text{for } d = 2 \\ \frac{1}{\gamma_d\lambda} N & \text{for } d > 2 \end{cases} \quad (11)$$

where $\zeta(x)$ is the Riemann zeta function. For $d = 2$ no pre-factor is analytically available, since due to the term $\ln(\lambda\hat{t})$ no closed form of the series can be found. It is notable that for $d = 1$ the LCA is at a much further time point in the past than for $d > 1$, for the same cell number N . This means that when scaling the median, \tilde{m} , and mean, $\langle \Delta m^* \rangle$, with time T , the saturating plateau is reached far later ($\sim N^2/\lambda$) than for $d > 2$ ($\sim N/\lambda$, compare Fig. 1b, main text). Furthermore, we expect that the scaling of \tilde{m} and $\langle \Delta m^* \rangle$ with N will be different for $d = 1$, which will be studied in the following sections.

THE EXTREME VALUE CDF OF A BRANCHING RANDOM WALK FOR TIME-VARYING DIFFUSION CONSTANT

In the main text we show that the statistics of m^* for $\lambda > 0$ can be approximated by a branching random walk (BRW) with time-varying diffusion constant. According to Fang and Zeitouni [17] (see Eqs. 2-4 therein), if the diffusion constant is decreasing with time τ while the branching rate is constant, the maximum Δm^* of a continuous-time BRW in the variable Δm with mean zero (in Ref. [17] called ‘‘branching brownian motion’’) follows the CDF

$$P^*(\Delta m_c, \tau) = f(\Delta m_c - \tilde{m}(\tau)) \quad (12)$$

with

$$\tilde{m}(\tau) = \sqrt{2} \left[\int_0^1 \sigma_{\tau}(s) ds \right] \tau \times \left(1 - O(\tau^{-2/3}) \right), \quad (13)$$

where $s = \tau'/\tau$, Δm_c is the threshold variable, and $\sigma_{\tau}^2(s) = \partial_{\tau'}(\sigma^2(s))$ is the variance increment rate of the random walk of an individual branch. Since for a diffusive random walk $\sigma^2(\tau') = 2D\tau'$, σ_{τ}^2 corresponds to twice the diffusion constant, $\sigma_{\tau}^2 = 2D(\tau)$. In the context of our model, the random variable is $\Delta m_b := m_b - \mu T$, where m_b is the mutational history along each branch b

of the genealogy, and the time scale is defined through

$$\tau(t) := \nu_k (t - t_k) + \tau_k \quad (14)$$

with

$$\tau_k := \sum_{k'=2}^{k-1} \nu_{k'} \Delta t_{k'}, \quad \nu_k := (\langle \Delta t_k \rangle k)^{-1} \quad (15)$$

for the largest $t_k < t$ with $k > 2$, while $\tau_{k \leq 2} := 0$. Here, ν_k is the branching rate of individual branches, and thus τ measures time in units of branching times per branch. Therefore, the probability of branching at time t_{k+1} is $\nu_k e^{-\Delta t_k \nu_k} dt = e^{-(\tau_{k+1} - \tau_k)} d\tau$, corresponding to a unit branching rate in the time scale of τ . It further follows that $\tau_N = \sum_{k'=2}^{N-1} k^{-1} = H_{N-1} - 1 \simeq \ln N$ where H_n are the harmonic numbers.

In this time scale, the variance increment of an individual random walk in the variable Δm_b is $\sigma_\tau^2(k) = \mu/\nu_k$, and thus the diffusion constant $D_k := \sigma_\tau^2(k)/2 = \mu/2\nu_k$ depends on the random variable k , but not explicitly on τ . In order to use formula (13), however, we need to express the dependence on k by a dependence on τ . To this end, we note that the stochastic dynamics of Δm_b along a single branch of the BRW, $P_b(\Delta m_b, \tau)$, is described by the diffusion equation (heat equation)

$$\partial_\tau P_b(\Delta m_b, \tau) = D_k \partial_{\Delta m_b}^2 P_b(\Delta m_b, \tau) . \quad (16)$$

Taking the ensemble average over branch numbers k at time τ , we have

$$\begin{aligned} \partial_\tau P_b(\Delta m_b, \tau) &= \langle D_k \partial_{\Delta m_b}^2 P_b(\Delta m_b, \tau) \rangle_k \\ &\approx \langle D_k \rangle_k \partial_{\Delta m_b}^2 P_b(\Delta m_b, \tau) \end{aligned} \quad (17)$$

where we have used that $P_b(\Delta m_b, \tau)$ does not explicitly depend on k at time τ . Thus, the underlying diffusion process with k -dependent diffusion constant D_k can be approximated by a time-dependent diffusion constant $D(\tau) := \langle D_k \rangle_k |_\tau$.

To find an explicit expression for $D(\tau)$, the probability distribution of k , $p(k, \tau)$ is required. This corresponds to the probability distribution of a simple binary-splitting, continuous-time branching process starting with two branches. To this end, we note that the future branching events of one branch are independent of the other branch, therefore we can consider this as two independent branching processes $b = 1, 2$ starting with a single branch each, also known as *Yule process*. Each branch b has branch number k_b and the total branch number is $k := k_1 + k_2$. For a Yule process it has been shown that the probability distribution of branches converges for large times τ to an exponential probability distribution with mean $\bar{k}_b = e^\tau$, $p(k_b, \tau) = e^{-k_b/\bar{k}_b}/\bar{k}_b$ [18]. To separate the parameter dependence from numerical constants, we will assume in the following that it is possible to factorize $\nu_k =: \alpha_{N,\lambda} \tilde{\nu}_k$ so that only $\alpha_{N,\lambda}$ depends on

the model parameters N and λ , while $\tilde{\nu}_k$ depends only on $k = k_1 + k_2$ and comprises constant factors. This can always be done for our model, as is shown below. Thus, with $D_k = \mu/2\nu_k$ we have, for large τ ,

$$\begin{aligned} D(\tau) &= \left\langle \left\langle \frac{\mu}{2\alpha_{N,\lambda} \tilde{\nu}_{k_1+k_2}} \right\rangle_{k_1} \right\rangle_{k_2} \Big|_\tau \\ &\approx \frac{\mu}{2\alpha_{N,\lambda}} \sum_{k_2=1}^{\infty} \left[\sum_{k_1=1}^{\infty} \frac{p(k_1, \tau)}{\tilde{\nu}_{k_1+k_2}} \right] p(k_2, \tau) \\ &\approx \frac{\mu}{2\alpha_{N,\lambda}} \int_1^{\infty} \left[\int_1^{\infty} \frac{p(k_1, \tau)}{\tilde{\nu}_{k_1+k_2}} dk_1 \right] p(k_2, \tau) dk_2 \\ &= \frac{\mu}{2\alpha_{N,\lambda}} \int_2^{\infty} \frac{1}{\tilde{\nu}_k} \int_1^{k-1} e^{-\tau} e^{-(k_1+(k-k_1))} e^{-2\tau} dk_1 , \end{aligned} \quad (18)$$

where we substituted $k = k_1 + k_2$, so that $k_2 = k - k_1$ was eliminated. We also made a continuous approximation for k_b , as $p(k_b, \tau)$ is only accurate for large k_b . Then

$$\begin{aligned} D(\tau) &\approx \frac{\mu}{2\alpha_{N,\lambda}} \int_2^{\infty} \frac{k-2}{\tilde{\nu}_k} e^{-e^{-\tau}k} e^{-2\tau} dk \\ &=: A^d(N, \mu, \lambda) C^d(\tau) \end{aligned} \quad (19)$$

with

$$\begin{aligned} A^d(N, \mu, \lambda) &:= \frac{\mu}{\alpha_{N,\lambda}} \\ C^d(\tau) &:= \int_2^{\infty} \frac{k-2}{2\tilde{\nu}_k} e^{-e^{-\tau}k} e^{-2\tau} dk . \end{aligned} \quad (20)$$

A^d comprises the dependence on the parameters N, μ, λ , while C^d depends only on rescaled time τ . Both may, however, depend on the dimension d of the underlying lattice model ($d = \infty$ for the Moran process), through $\alpha_{N,\lambda}$ and $\tilde{\nu}_k$.

The median \tilde{m} of the CDF

Using the above factorization of $D(\tau)$, an expression for \tilde{m} , the median of the CDF $P_N^*(\Delta m_c)$, can be obtained from Eq. (13),

$$\begin{aligned} \tilde{m} &= \left[\sqrt{2} \int_0^\tau \sigma_\tau(\tau') d\tau' \right] \left(1 + O(\tau^{-2/3}) \right) \\ &\approx 2 \int_0^\infty \sqrt{D(\tau')} d\tau' = C_{\tilde{m}}^d \sqrt{A^d(N, \mu, \lambda)} \\ &\text{with } C_{\tilde{m}}^d = \int_0^\infty 2\sqrt{C^d(\tau')} d\tau' , \end{aligned} \quad (21)$$

where we let $\tau \rightarrow \infty$, valid for large $N \approx e^\tau$, neglecting terms of $O(\tau^{-2/3})$. With $\alpha_{N,\lambda} \tilde{\nu}_k = \nu_k = 1/k \langle \Delta t_k \rangle$, we have the factorization $\langle \Delta t_k \rangle = (k \alpha_{N,\lambda} \tilde{\nu}_k)^{-1}$ for all d , which, according to the asymptotic results, Eq. (11), is

fulfilled for

$$\alpha_{N,\lambda} \simeq \begin{cases} \frac{\lambda}{N^2} & \text{for } d = 1 \\ \frac{\lambda}{N} & \text{for } d = 2 \\ \frac{\lambda}{N} & \text{for } d > 2 \end{cases} \quad \tilde{\nu}_k \simeq \begin{cases} \frac{\pi}{2} k^2 & \text{for } d = 1 \\ \frac{\pi k}{\ln(\lambda \hat{t}(\tau))} & \text{for } d = 2 \\ \gamma_d k & \text{for } d > 2 \end{cases} \quad (23)$$

with $\hat{t}(\tau) = T - t(\tau)$ and $t(\tau)$ being the inverse of the monotonic relationship $\tau(t)$ defined in Eq. (14). Hence, we get the scaling of \tilde{m} with model parameters, using Eq. (21) and the definition for $A(N, \mu, \lambda)$, Eq. (20),

$$\tilde{m} \simeq \begin{cases} \mathcal{C}_{\tilde{m}}^1 N \left(\frac{\mu}{\lambda}\right)^{1/2} & \text{for } d = 1 \\ \mathcal{C}_{\tilde{m}}^2 \left(\frac{\mu N}{\lambda}\right)^{1/2} & \text{for } d = 2 \\ \mathcal{C}_{\tilde{m}}^{d>2} \left(\frac{\mu N}{\lambda}\right)^{1/2} & \text{for } d > 2 \end{cases} \quad (24)$$

where for $d=2$ we used that for large λT , $\ln(\lambda \hat{t}) \approx \ln(\lambda T)$, is independent of λ, μ, N . For the model studied in the main text, the formula for $d > 2$ applies, since the Moran process corresponds to dynamics on an infinite-dimensional lattice. This results in the formula $\tilde{m} \approx \mathcal{C}_{\tilde{m}} (\mu N / \lambda)^{1/2}$, given in Eq. (7), main text, for which $\mathcal{C}_{\tilde{m}} := \mathcal{C}_{\tilde{m}}^{d=\infty}$. For dimensions $d = 1, 2, 3$ the estimates from Eq. (24) are shown in Fig. 2, together with numerical results from Monte Carlo simulations. The constants $\mathcal{C}_{\tilde{m}}^d$ were fitted to the data (linear regression). We note that the theoretical estimates are in excellent agreement with simulations for $d = 1, 3$. For $d = 2$ the agreement is good for large N ; for low N , nonetheless, the neglected correction factor $\ln(\lambda \hat{t})$ cannot be considered constant. It needs to be noted that values for $d = 1$ are larger also because the time to the LCA, T_{LCA} scales as N^2 which leads to much larger values of $\langle \Delta m^* \rangle$ for the same N .

In general, the integrals $\mathcal{C}_{\tilde{m}}^d$ cannot be computed with the scaling information for low dimensions, Eq. (23), alone, since the integration requires information for small \hat{t} which is not available from the asymptotic results of Eq. (11). However, we know that the factors $\mathcal{C}_{\tilde{m}}^d$ are just numerical constants and do not depend on the parameters, therefore, Eq. (24) contains the full information about the asymptotic parameter dependence of \tilde{m} for any dimension d .

Nonetheless, $\mathcal{C}_{\tilde{m}} = \mathcal{C}_{d=\infty}^{\tilde{m}}$ can be computed for the Moran process, $d = \infty$, for which $\nu_k = (\langle \Delta t_k \rangle k)^{-1} = (k-1)\lambda/N$ is known (cf. Eq. (7)). First, we calculate $C^\infty(\tau)$ from Eq. (20),

$$\begin{aligned} C^\infty(\tau) &= \frac{1}{2} e^{-2\tau} \int_2^\infty \frac{e^{-e^{-\tau} k} (k-2)}{k-1} dk \\ &= \frac{1}{2} e^{-2\tau} \left[\int_2^\infty e^{-e^{-\tau} k} dk - \int_1^\infty \frac{e^{-e^{-\tau} (k+1)}}{k} dk \right] \\ &= \frac{1}{2} (e^{-2e^{-\tau} - \tau} - e^{-e^{-\tau} - 2\tau} \Gamma(0, e^{-\tau})) \quad (25) \end{aligned}$$

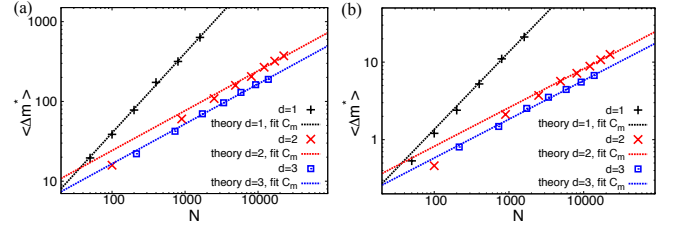


Figure 2. Mean maximum mutation number ahead of the mean, $\langle \Delta m^* \rangle$ as a function of N for $T > \hat{T}_{LCA}$ (achieved through scaling $T = 10N/\lambda$, see also Fig. 2, main text) for cells embedded on a d -dimensional lattice. The time T is rescaled for each N according to $T = 10N/\lambda$ for $d = 2, 3$ and $T = 10(N/\lambda)^2$ for $d = 1$ (note that therefore values are much larger for $d = 1$). Shown are the results of Monte Carlo simulations (points), and theoretical predictions from the BRW approximation, Eq. (24) (dashed lines) with fitted numerical constant $\mathcal{C}_{\tilde{m}}^{\text{fit}}$ for (a) $\mu = \lambda$ and (b) $\mu = 0.001\lambda$. Color coding as in figure key.

where in the right integral we performed the shift $k \rightarrow k+1$, and $\Gamma(x, y)$ is the incomplete gamma function. Now we can compute $\mathcal{C}_{\tilde{m}}$, according to Eq. (22), numerically (by Mathematica, with global adaptive integration strategy) to get for $\tau \rightarrow \infty$

$$\mathcal{C}_{\tilde{m}} \approx 1.79 \quad (26)$$

The use of the asymptotic distribution $p(k, \tau)$ instead of the exact one, however, leads to deviations which do not disappear for large τ , since there are contributions for small k and τ' to the integral, Eq. (22).

The tail of the CDF

To find an asymptotic expression for the tail of the CDF, we use that for $\bar{P}_N(\Delta m_c) = 1 - P_N^*(\Delta m_c) \ll 1$ the Fisher-KPP equation, Eq. (5) in the main text, can be linearized, so that terms of order $\bar{P}_N^2(\Delta m_c)$ can be neglected. This condition is fulfilled in the tail of the distribution, for $\Delta m_c \gg \tilde{m}(N)$. Writing for convenience $x := \Delta m_c$, the linearized Fisher-KPP equation (Eq. (5), main text) for the complementary CDF, \bar{P}_N , with time-dependent diffusion constant $D(\tau)$, becomes

$$\partial_\tau \bar{P}_N(x, \tau) \approx D(\tau) \partial_x^2 \bar{P}_N(x, \tau) + \bar{P}_N(x, \tau) \quad (27)$$

Following Ref. [19], we define the function $\Phi(x, \tau) = \bar{P}_N(x, \tau) e^{-\tau}$. Substituting this, we get the equation

$$e^\tau [\partial_\tau \Phi(x, \tau)] \approx e^\tau [D(\tau) \partial_x^2 \Phi(x, \tau)] \quad (28)$$

Dividing by e^τ yields the normal time-dependent diffusion equation, which has the general solution [20]

$$\Phi(x, \tau) = \int_{-\infty}^\infty \Phi(x', 0) e^{-\frac{(x-x')^2}{2\sigma_{\text{eff}}^2}} (2\pi\sigma_{\text{eff}}^2)^{-1/2} dx' \quad (29)$$

where $\sigma_{\text{eff}}^2 = \sigma_{\text{eff},1}^2 + \sigma_{\text{eff},2}^2$ is the sum of the variances of the two individual sub-genealogies $b = 1, 2$ originating from the two initial branches of the genealogy. Thus,

$$\sigma_{\text{eff}} = \sqrt{4 \int_0^\tau D(\tau') d\tau'} = \sqrt{A^d(N, \mu, \lambda)} C_\sigma^d \quad (30)$$

with $C_\sigma^d := 2\sqrt{\int_0^\infty C^d(\tau') d\tau'}$ (taking again the limit $\tau \rightarrow \infty$), depending on dimension d . With Eq. (23) we then get

$$\sigma_{\text{eff}} = \begin{cases} C_\sigma^1 N \left(\frac{\mu}{\lambda}\right)^{1/2} & \text{for } d = 1 \\ C_\sigma^2 \left(\frac{\mu N}{\lambda}\right)^{1/2} & \text{for } d = 2 \\ C_\sigma^{d>2} \left(\frac{\mu N}{\lambda}\right)^{1/2} & \text{for } d > 2 \end{cases}, \quad (31)$$

which has the same parameter dependence as \tilde{m} , yet with other pre-factors. For the Moran process, the formula for $d = \infty$ applies, and we can again integrate Eq. (25) numerically (according to the definition of C_σ^d above), giving

$$C_\sigma := C_\sigma^{d=\infty} \approx 0.57. \quad (32)$$

Finally, re-substituting Δm_c for x , with the initial condition $\phi(\Delta m_c, 0) = \bar{P}_N(\Delta m_c, \tau = 0) = 1 - \theta(\Delta m_c)$, and using that $e^\tau = N$, we get

$$\begin{aligned} \bar{P}_N(\Delta m_c, \tau) &\approx \frac{N}{2} \left(1 - \text{erf}\left(\frac{\Delta m_c}{\sigma_{\text{eff}}}\right)\right) \\ &\approx \frac{N \sigma_{\text{eff}} e^{-\frac{\Delta m_c^2}{2\sigma_{\text{eff}}^2}}}{\sqrt{2\pi} \Delta m_c} \end{aligned} \quad (33)$$

and thus, for $\Delta m_c \gg \tilde{m}$

$$P_N^*(\Delta m_c, \tau) \approx 1 - \frac{N \sigma_{\text{eff}} e^{-\frac{\Delta m_c^2}{2\sigma_{\text{eff}}^2}}}{\sqrt{2\pi} \Delta m_c}, \quad (34)$$

with σ_{eff} according to Eq. (30). Here, we used the approximation $N \approx e^\tau$. The form of Eq. (34) corresponds to the tail of a non-normalized Normal distribution with mean zero and variance σ_{eff}^2 and is Eq. (8) in the main text.

THE SCALING OF Δm^* FOR $N \rightarrow \infty$

If N is large and T is fixed, so that $\hat{T}_{\text{LCA}} > T$, the cell population as a whole does not possess an LCA, but the genealogy fragments into k sub-trees, corresponding to sub-populations $l = 1, \dots, k$ which accumulate mutations independently from each other. k is the number of branches of the genealogy in reverse time $\hat{t} = T$ which is estimated by Eq. (9). Each sub-population has n_l individuals, with $\sum_{l=1}^k n_l = N$. Since the sub-populations

are not related to each other, each their maximum mutation numbers m_l^* are i.i.d. random numbers. Thus, the probability that $m^* < m_c$, $P_N^*(m_c)$ ($m^* = \max(1, \dots, N)$ of the whole population), is equal to the probability that each independent $m_l^* < m_c$, $P_{n_l}^*(m_c)$ [22]. Therefore,

$$P_N^*(m_c) = [P_{n_l}^*(m_c)]^k. \quad (35)$$

Each sub-population has got a maximum m_l^* for which we can apply the BRW approximation. Hence, m_l^* is distributed according to the Fisher-KPP solution, with the tail from Eq. (34) for large τ . For the moment, we assume the limit $\tau \gg 1$ and discuss below when this is justified. Under this assumption the tail of $P_{n_l}^*(m_c)$ is Gaussian with variance σ_{eff}^2 from Eq. (30), and thus the distribution of the extreme values, Eq. (35) approaches a Gumbel distribution for large N and μT , Eq. (5), with an effective number of random variables, k . Hence, we have [2]

$$\tilde{m} \simeq \sqrt{2 \ln k} \sigma_{\text{eff}} \quad (36)$$

$$\sigma_N \simeq \sigma_{\text{eff}} / \sqrt{2 \ln k}.$$

With k according to Eq. (9) and $\sigma_{\text{eff}} = C_\sigma^d \sqrt{\mu \langle n_l \rangle / \lambda} \simeq C_\sigma^d \sqrt{\mu T}$ from Eq. (31), this gives

$$\begin{aligned} \tilde{m} &\simeq \begin{cases} C_\sigma^d \sqrt{2 \mu T \ln \frac{N}{\sqrt{\pi \lambda T}}} & \text{for } d = 1 \\ C_\sigma^d \sqrt{2 \mu T \ln \frac{N \ln(\lambda T)}{\pi \lambda T}} & \text{for } d = 2 \\ C_\sigma^d \sqrt{2 \mu T \ln \frac{N}{\gamma_d \lambda T}} & \text{for } d = 3 \end{cases}, \quad (37) \\ \sigma_N &\simeq \begin{cases} C_\sigma^d \sqrt{\frac{\mu T}{2 \ln \frac{N}{\sqrt{\pi \lambda T}}}} & \text{for } d = 1 \\ C_\sigma^d \sqrt{\frac{\mu T}{2 \ln \frac{N \ln(\lambda T)}{\pi \lambda T}}} & \text{for } d = 2 \\ C_\sigma^d \sqrt{\frac{\mu T}{2 \ln \frac{N}{\gamma_d \lambda T}}} & \text{for } d = 3 \end{cases}. \end{aligned}$$

For the simple (infinite-dimensional) Moran process we have $\gamma_{d=\infty} = 1$ and the resulting term corresponds to Eq. (9) in the main text. For $d = 3$, $\gamma_{d=3} \approx 0.66$ (see [12], as follows from [13–16]).

Now we note that τ is the effective, rescaled time of a BRW leading to a subpopulation l with n_l cells. Thus, $\tau \approx \ln(n_l) \approx \ln(N/k)$, according to Eq. (14) and (15) ff. For fixed T it is therefore not assured that our approximation, Eq. (37) holds. To check this, the estimates from (37) are shown in Fig. 3, together with numerical results from Monte Carlo simulations, for $\lambda T = 1000$, $\mu = \lambda$ and $\mu = 0.001\lambda$, whereby C_σ^d has been fitted. With $\tau \approx \ln(N/k)$ we have $\tau \approx 4$ for $d = 1$, $\tau \approx 6$ for $d = 2$, and $\tau \approx 7$ for $d = 3$, so that deviations from the limit $\tau \gg 1$ can be expected. Nonetheless, for $\mu = \lambda$, we see an excellent agreement of theory and simulation data for $d = 1, 2, 3$ and for the Moran process (main text, Fig. 3a). This can be attributed to the fact that for large μT the individual m_i are indeed approximately distributed

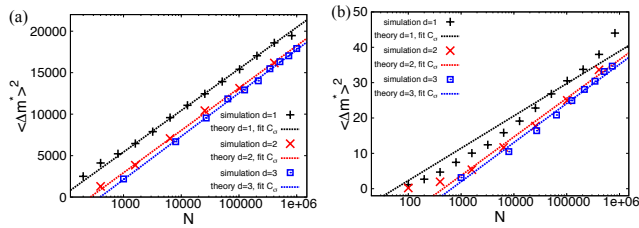


Figure 3. Squared mean maximum mutation number ahead of the population mean, $\langle \Delta m^* \rangle^2$ as a function of N , for fixed $T = 1000/\lambda$, with cells embedded on a d -dimensional lattice. Shown are the results of Monte Carlo simulations (points) and the theory (dashed lines), with fit parameter C_σ , for $d = 1$ (black), $d = 2$ (red), and $d = 3$ (blue), according to Eqs. (37). (a) $\mu = \lambda$, (b) $\mu = 0.001\lambda$.

as a normal distribution, so that the Gumbel distribution (Eq. (2) ff.) is a good approximation for the distribution of m^* . For $\mu = 0.001\lambda$, the agreement is still reasonable for $d = 2$ and $d = 3$ ($\tau \approx 6, 7$), but not for $d = 1$ which has lower $\tau \approx 4$, closer to one, while our approximation is valid only for $\tau \gg 1$.

-
- [1] M. R. Leadbetter, G. Lindgren, and H. Rootzén, *Extremes and related properties of random sequences and processes* (Springer, New York, 1983).
 - [2] J.-P. Bouchaud and M. Marc, *J. Phys. A: Math. Gen.* **30**, 7997 (1997).
 - [3] J. Galambos, *The Asymptotic Theory of Extreme Order Statistics* (Wiley, 1978).
 - [4] D. T. Gillespie, *J. Phys. Chem.* **81**, 2340 (1977).

- [5] R. R. Hudson, in *Oxford Surveys in Evolutionary Biology*, edited by D. Futuyama and J. Antonovics (1991), p. 1, 7th ed.
- [6] J. Kingman, *Stochastic Processes and their Applications* **13**, 235 (1982).
- [7] J. Kingman, *Journal of Applied Probability* **19**, 27 (1982).
- [8] J. Wakely, in *Coalescent Theory: An Introduction* (W. H. Freeman, 2008).
- [9] A. M. Klein, D. P. Doupe, P. H. Jones, and B. D. Simons, *Phys. Rev. E* **77**, 031907 (2008).
- [10] A. M. Klein and B. D. Simons, *Development* **138**, 3103 (2011).
- [11] M. Bramson and D. Griffeath, *Zeitschrift fuer Wahrscheinlichkeitstheorie und Verwandte Gebiete* **53**, 183 (1980).
- [12] E. W. Weisstein, *Polya's Random Walk Constants* (2018).
- [13] G. N. Watson, *Quart. J. Math., Oxford Ser. 2* **10**, 266 (1939).
- [14] W. H. McCrea and F. J. W. Whipple, *Proceedings of the Royal Society of Edinburgh* **60**, 281 (1940).
- [15] C. Domb, *Mathematical Proceedings of the Cambridge Philosophical Society* **50**, 586 (1954).
- [16] M. L. Glasser and I. J. Zucker, *Proc. Natl. Acad. Sci.* **74**, 1800 (1977).
- [17] M. Fang and O. Zeitouni, *J. Stat. Phys.* **149**, 1 (2012).
- [18] D. G. Kendall, *J. London. Math. Society* **41**, 385 (1966).
- [19] M. Cencini, C. Lopez, and D. Vergni, *Lect. Notes Phys.* **636**, 187 (2003).
- [20] J. Crank, *The Mathematics of Diffusion* (Clarendon Press, Oxford, 1975).
- [21] Note1, for convenience, we do not use different names for the CDFs of variables which just differ by a shift or scaling, such as m_c , Δm_c and X .
- [22] Note2, note that in the main text this probability is for convenience named P_l^* .