The ubiquity of clonal interference in cancer and other range expansions

Alexander Stein
Centre for Cancer Genomics and Computational Biology, Barts Cancer Institute, Queen Mary University of London, London, UK and Department of Physics, ETH Zurich, Zürich, Switzerland

Ramanarayanan Kizhuttil
Department of Physics, Indian Institute of Science Education and Research, Kolkata, India

Maciej Bak and Robert Noble*
Department of Mathematics, City, University of London, London, UK
(Dated: March 8, 2024)

Evolution during range expansions shapes the fate of many biological systems including tumours, microbial communities, and invasive species. A fundamental process of interest is the selective sweep, in which an advantageous mutation evades clonal interference and spreads through the population to fixation. However, most theoretical investigations of selective sweeps have assumed constant population size or have ignored spatial structure. Here we use mathematical modelling and analysis to investigate selective sweep probabilities in populations that spread outwards as they evolve. In the case of constant radial expansion speed, we derive probability distributions for the arrival time and location of the first surviving mutant and hence find surprisingly simple approximate and exact expressions for selective sweep probabilities in one, two and three dimensions, which are independent of mutation rate. Namely, the selective sweep probability is approximately $(1 - c_{\mathrm{wt}}/c_{\mathrm{m}})^d$, where c_{wt} and c_{m} are the wildtype and mutant radial expansion speeds, and d the spatial dimension. Using agent-based simulations, we show that our analytical results accurately predict selective sweep frequencies in the two-dimensional spatial Moran process. We further compare our results with those obtained for alternative growth models. Parameterizing our model for human tumours, we find that selective sweeps are predicted to be rare except during very early solid tumour growth, thus providing a general, pan-cancer explanation for findings from recent sequencing studies.

I. INTRODUCTION

Range expansions abound across biological scales. Evolution during cell population range expansions determines the development and spatial heterogeneity of biofilms [1], tumours [2], mosaicism [3], and normal tissue [4]. At the species level, range expansions influenced human evolution [5] and are of growing importance as climate change forces organisms into new habitats [6, 7]. Range expansions alter the course of evolution in distinct, often profound ways that are incompletely understood [8]. Recent theoretical investigations (for example, [9–12]) have typically assumed that proliferation is restricted to the population boundary, precluding full selective sweeps. Alternative modes of dispersal remain largely unexplored.

Somatic evolution provides especially strong motivation for investigating selective sweeps during range expansions. Cancer results from an accumulation of mutations that drive cells to proliferate uncontrollably and to invade surrounding tissue [13]. Some of these are clonal mutations that occur during tumourigenesis or early tumour development [14, 15]. However, recent advances in multi-region and single-cell sequencing have revealed pervasive genetic intratumour heterogeneity [16] and diverse modes of evolution [17, 18]. Although fitter mutant clones arise throughout cancer progression they seldom achieve selective sweeps, except perhaps in very small tumours [15, 17]. With notable exceptions [19–21], mathematical studies of cancer initiation and early evolution have either ignored spatial structure [22–24] or have relied on agent-based models [25–28]. The typical observed pattern of very early fixation events followed by branching or effectively neutral intratumour evolution [17, 18] – which has important clinical implications [29–31] – has eluded a general explanation.

Here we use mathematical analysis to explain why beneficial mutations typically fixate only – if at all – in the very early stages of range expansions. We focus on a simple model with short-range dispersal throughout the population. By solving our model in the canonical case of constant radial expansion speed we derive exact and simple approximate expressions for sweep probabilities in one, two and three dimensions. We confirm the accuracy and robustness of our analytical results using extensive agent-based simulations of a spatial Moran process, and we compare outcomes for alternative growth models. We discuss how these findings shed light on the nature of evolution in range expansions in general, and cancer development in particular.

* robert.noble@city.ac.uk

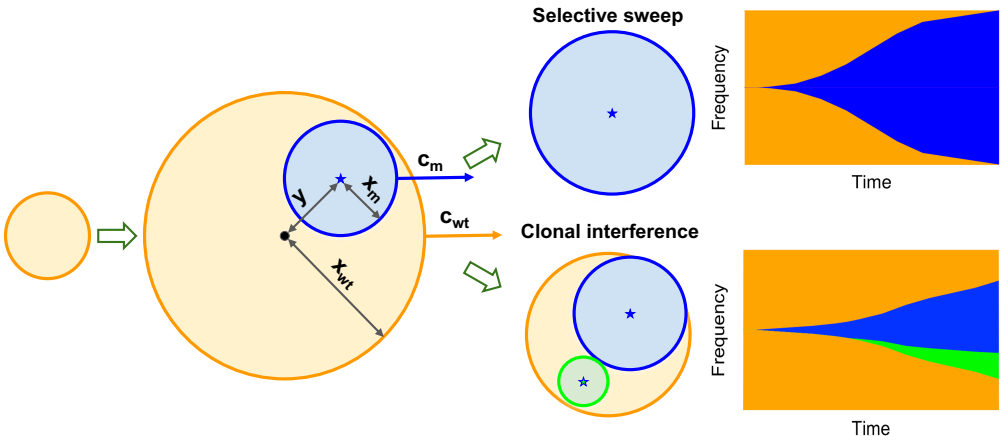

Figure 1. Illustration of the macroscopic model (not to scale) showing the two possible fates of the first surviving mutant (blue) within the wildtype population (orange). In the first case, the mutant reaches every part of the wildtype boundary before further mutants arise in the wildtype, and so achieves a selective sweep. Otherwise, a second (green) mutant arises in the wildtype population and generates clonal interference.

II. MACROSCOPIC MODEL

We consider a wildtype population that starts expanding spherically at time $t = 0$, such that its radius x_{wt} grows at speed c_{wt}. Focusing on selective sweeps, we consider only advantageous mutations, which we assume spread within the wildtype at speed $c_{\mathrm{m}} > c_{\mathrm{wt}}$ (Figure 1). Mutations occur at per-capita rate $\tilde{\mu}$ with each surviving genetic drift with probability ρ. In our analytic model, it is enough to consider the compound parameter $\mu = \rho\tilde{\mu}$, which is the mutation rate conditioned on survival. For brevity, we will refer to μ as the mutation rate unless otherwise mentioned.

For mathematical and biological reasons (see Discussion), we focus on a model with constant radial expansion speeds; later we compare these results with those that pertain for alternative growth models. Various models link propagation speeds to fitness values and migration rates. The most prominent formula is obtained from a reaction-diffusion equation associated with Fisher [32] and Kolmogorov [33]: $c = 2\sqrt{aD}$, where D is the diffusion coefficient carrying information of migration and a is the difference in the local proliferation rate. More recent studies have sought to refine and generalise this result [34–37]. Our main results apply to any model that generates approximately constant speeds.

The first surviving mutation achieves a complete (or "classic") selective sweep only if it reaches every part of the wildtype expansion front before a second mutant of equal or greater fitness arises within the wildtype (Figure 1). Otherwise, the outcome is clonal interference or possibly a soft selective sweep if two mutations confer the same trait [38]. For simplicity, we neglect mutants with fitness values between those of the wildtype and the first mutant, which would slow the expansion of the first mutant and so reduce rather than negate the selective sweep probability. Neither do we investigate the case of a yet-fitter mutant that arises from the first mutant and achieves a selective sweep.

III. SWEEP PROBABILITY

The unconditional sweep probability is derived in four steps. First, we introduce the random variable X, the radius of the wildtype population when the first mutant arises, and we compute its probability density $f_X(x)$. Second, we introduce the random variable Y, the distance between the wildtype and mutant origins, and we calculate its probability density conditioned on X, namely $f_Y(y|X = x)$. Third, we derive an expression for the conditional sweep probability $\Pr(\text{sweep}|X = x, Y = y)$. Finally, we marginalize out X and Y to obtain the unconditional sweep probability $\Pr(\text{sweep})$. Here, we focus on the three-dimensional case; analogous results in one and two dimensions are presented in the SI Text.

A. Probability that no mutations occur

The following result (proved in SI Text) will be useful in various parts of subsequent derivations.

Claim 1. *The probability that k mutations arise and survive during the time interval $[0, t]$ is Poisson distributed,*

$$P_k = \frac{e^{-\lambda}\lambda^k}{k!} \quad \textit{with} \quad \lambda = \mu \int_0^t N(s)\, ds, \tag{1}$$

where $N(t)$ is the population size. In particular, the probability that no successful mutant arises is $P_0 = e^{-\lambda}$.

Although it is commonly assumed that mutations are coupled to divisions, it is straightforward to translate between per-capita and per-division mutation rates (see SI text).

B. Arrival time of the first mutant

In the absence of mutants, the wildtype population in three dimensions grows as $N_{\mathrm{wt}} = \frac{4}{3}\pi x_{\mathrm{wt}}^3 = \frac{4}{3}\pi(c_{\mathrm{wt}}t)^3$. Applying Claim 1, we obtain the probability that no mutations arise in the time interval $[0, t]$,

$$P_0 = e^{-\mu \int_0^t \frac{4}{3}\pi(c_{\mathrm{wt}}s)^3\, ds} = e^{-\left(\frac{t}{\kappa}\right)^4}, \tag{2}$$

in terms of a characteristic duration $\kappa = \sqrt[4]{\frac{3}{\mu\pi c_{\mathrm{wt}}^3}}$. We identify $1 - P_0$ as the cumulative distribution function for the arrival time T of the first surviving mutant. The probability density of T is then

$$f_T(t) = \frac{d(1-P_0)}{dt} = \frac{4t^3}{\kappa^4} e^{-\left(\frac{t}{\kappa}\right)^4}, \tag{3}$$

which is the Weibull distribution with shape parameter 4 and scale parameter κ. Substituting $x = c_{\mathrm{wt}}t$, we find that $f_X(x)$, the probability density of the radius of the wildtype population X at the time the first surviving mutant arises, is the Weibull distribution with shape parameter 4 and scale parameter $\theta = \sqrt[4]{\frac{3c_{\mathrm{wt}}}{\pi\mu}}$. It follows that $\mathbb{E}[X] \approx 0.91\,\theta$ and $\mathrm{Var}[X] \approx 0.065\,\theta^2$. Analogous calculations in one and two dimensions yield similar Weibull distributions (SI text; Figure 2A-C).

C. Location of the first mutant

Next, we compute the probability density for the distance Y between the first surviving mutant and the centre of the wildtype population. Since mutants arise in proportion to the number of wildtype cells, we have $f_Y(y|X = x)\, dy \propto D(y)$, where $D(y)$ is the number of cells at distance y. $D(y)$ corresponds to the infinitesimal shell, $D(y) = 4\pi y^2\, dy\, \mathbf{1}\{y \le x\}$, where the last term is an indicator function that defines the boundary of the wildtype population. After normalisation, we obtain

$$f_Y(y|X = x) = \frac{3y^2}{x^3}\mathbf{1}\{y \le x\}. \tag{4}$$

We calculate the unconditional probability density of Y by marginalizing out X,

$$f_Y(y) = \int_0^\infty f_Y(y|X = x) f_X(x)\, dx = \frac{3y^2}{\theta^3}\Gamma\left(\frac{1}{4}, \frac{y^4}{\theta^4}\right), \tag{5}$$

where $\Gamma(a, z) = \int_z^\infty t^{a-1}e^{-t}\, dt$ is the incomplete gamma function. We then find $\mathbb{E}[Y] \approx 0.68\,\theta$ and $\mathrm{Var}[Y] \approx 0.070\,\theta^2$. Similar results pertain in one and two dimensions (SI text; Figure 2A-C).

D. Conditional sweep probability

To compute the sweep probability, we first need an expression for the remaining wildtype population, N_{wt}. Therefore, we introduce time measure τ, with $\tau = 0$ when the first mutant arises. Recall that $t = 0$ at the origin of the wildtype population, so $t = \tau + x/c_{\mathrm{wt}}$. Once the mutant population starts expanding, we have $N_{\mathrm{wt}}(\tau) = \tilde{N}_{\mathrm{wt}}(\tau) - \Delta(\tau)$, where $\Delta(\tau)$ is the number of wildtype cells that the mutant has replaced, and $\tilde{N}_{\mathrm{wt}}(\tau) = \frac{4}{3}\pi x_{\mathrm{wt}}^3(\tau)$ is the wildtype population size had there been no mutant. While the mutant is surrounded by the wildtype, we have $\Delta(\tau) = \Delta_1(\tau) = \frac{4}{3}\pi x_{\mathrm{m}}^3(\tau)$. After the mutant breaches the wildtype boundary, $\Delta(\tau)$ is given by the intersection formula of two balls, which we denote $\Delta_2(\tau)$ (see SI text for the explicit formula). Together, we have

$$\begin{aligned} N_{\mathrm{wt}}(\tau) =& (\tilde{N}_{\mathrm{wt}}(\tau) - \Delta_1(\tau))\mathbf{1}\{[0, \tau_1]\} \\ &+ (\tilde{N}_{\mathrm{wt}}(\tau) - \Delta_2(\tau))\mathbf{1}\{[\tau_1, \tau_2]\} \end{aligned} \tag{6}$$

where τ_1 is the time at which the mutant reaches the wildtype boundary and τ_2 is the time at which the mutant has entirely replaced the wildtype. Noting that

$$\tau_1 c_{\mathrm{m}} = (x - y) + \tau_1 c_{\mathrm{wt}} \Rightarrow \tau_1 = \frac{x - y}{c_{\mathrm{m}} - c_{\mathrm{wt}}} \tag{7}$$

and

$$\tau_2 c_{\mathrm{m}} = (x + y) + \tau_1 c_{\mathrm{wt}} \Rightarrow \tau_2 = \frac{x + y}{c_{\mathrm{m}} - c_{\mathrm{wt}}}, \tag{8}$$

we express $N_{\mathrm{wt}}(\tau)$ in terms of x and y (see SI text). We then apply Claim 1 to obtain the conditional sweep probability,

$$\Pr(\mathrm{sweep}|X = x, Y = y) = e^{-\mu \int_0^\infty N_{\mathrm{wt}}(\tau)\, d\tau}. \tag{9}$$

Although this integral can be solved analytically (SI text), the resulting expression is complicated and not very enlightening, and we therefore seek simpler approximations. An especially fruitful approach is to assume $y = 0$, so that $f_Y(y|X = x) = \delta(y)$, where δ is the Dirac delta function. This yields an upper bound on the conditional sweep probability because, due to geometrical symmetry, the time required for a mutant to sweep (τ_2 in eqn. 8) is minimal when $y = 0$. With this approximation, eqn. 9 simplifies drastically as $\tau_1 = \tau_2 = \frac{x}{c_{\mathrm{m}} - c_{\mathrm{wt}}}$, and we do not need to integrate over $\Delta_2(\tau)$. The analytic solution is

$$\Pr(\mathrm{sweep}|X = x, Y = 0) = e^{-\left(\frac{x}{\alpha}\right)^4}, \tag{10}$$

where $\alpha = \sqrt[4]{\frac{3(c_{\mathrm{m}} - c_{\mathrm{wt}})^3}{\pi\mu\left(c_{\mathrm{wt}}^2 - 3c_{\mathrm{wt}}c_{\mathrm{m}} + 3c_{\mathrm{m}}^2\right)}}$ is a characteristic length. Despite being based on a seemingly crude approximation, eqn. 10 is remarkably close to the exact conditional sweep probability (Figure 2D). The corresponding approximations in one and two dimensions (SI text) are likewise simple and useful.

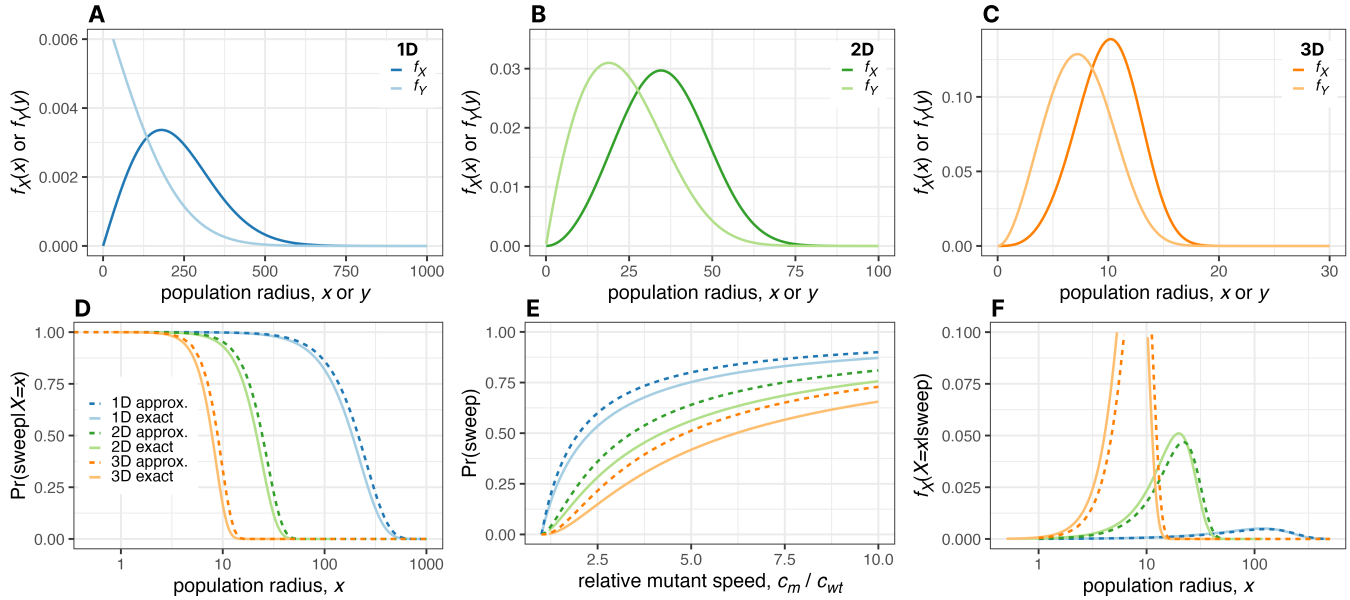

Figure 2. Analytical and numerical solutions of the macroscopic model. **A.-C.** Probability density of the wildtype population radius at the time the first surviving mutant arises (dark curves) and of the distance between the origins of the mutant and wildtype expansions (light curves) in one dimension **(A)**, two dimensions **(B)** and three dimensions **(C)**. **D.** Approximate and exact solutions for the sweep probability conditioned on the wildtype population radius. **E.** Approximate and exact solutions for the unconditional sweep probability. **F.** Approximate and exact solutions for the probability density of the wildtype population radius at the time the first surviving mutant arises, conditioned on this mutant achieving a selective sweep. Except where parameter values are explicitly varied, we set $c_{\mathrm{wt}} = 0.15, \tilde{\mu} = 10^{-5}, \rho = 0.23$ and $c_{\mathrm{m}} = 0.31$. Formulas for all these curves are summarised in Table S2.

E. Unconditional sweep probability

We now have all the necessary ingredients to compute the unconditional sweep probability by marginalizing out X and Y from the conditional sweep probability,

$$\begin{aligned}\Pr(\text{sweep}) &= \\ &\int_0^\infty \int_0^\infty \Pr(\text{sweep}|X=x, Y=y) f_Y(y|X=x) f_X(x)\, dy\, dx \\ &= \int_0^\infty \int_0^x e^{-\mu \int_0^\infty N_{\mathrm{wt}}(\tau)\, d\tau} \frac{3y^2}{x^3} \frac{4x^3 e^{-x^4/\theta^4}}{\theta^4}\, dy\, dx. \qquad (11)\end{aligned}$$

As before, we use the approximation $y = 0$ to obtain the strikingly simple result

$$\begin{aligned}\Pr(\text{sweep}) &\leq \int_0^\infty \Pr(\text{sweep}|X=x, Y=0) f_X(x)\, dx \\ &= \int_0^\infty e^{-\left(\frac{x}{\alpha}\right)^4} \frac{4x^3 e^{-x^4/\theta^4}}{\theta^4}\, dx \\ &= \left(\frac{c_{\mathrm{m}} - c_{\mathrm{wt}}}{c_{\mathrm{m}}}\right)^3. \qquad (12)\end{aligned}$$

This result generalizes to analogous one- and two-dimensional models, with the exponent 3 replaced by the respective spatial dimension (SI text). The approximate expressions are close to the exact solution in one dimension, and to numerical evaluations of the integral in two and three dimensions (Figure 2E). The sweep probability is independent of the mutation rate not only when $y = 0$ but also in the general case of eqn. 11. This follows from a more general result (SI text).

When the expanding wildtype must displace a resident competitor (as in our agent-based simulations, to follow), the sweep probability can be approximated in terms of evolutionary parameters via the FKPP solution as

$$\Pr(\text{sweep}) \approx \left(1 - \sqrt{\frac{r_{\mathrm{m}} - r_{\mathrm{wt}}}{r_{\mathrm{wt}} - r_{\mathrm{re}}}}\right)^d, \qquad (13)$$

where r_{m}, r_{wt} and r_{re} are the proliferation rates of the mutant invader, wildtype invader and resident populations, respectively, and d is the spatial dimension.

F. Conditional arrival time of the first mutant

In biological systems where it is infeasible to track evolutionary dynamics, selective sweeps must be inferred from subsequent genetic data. For example, we might observe a public mutation in a tumour and ask when this mutation occurred. We can use our model to obtain the probability distribution of the radius X at the time the first mutant arose, given that we observe a

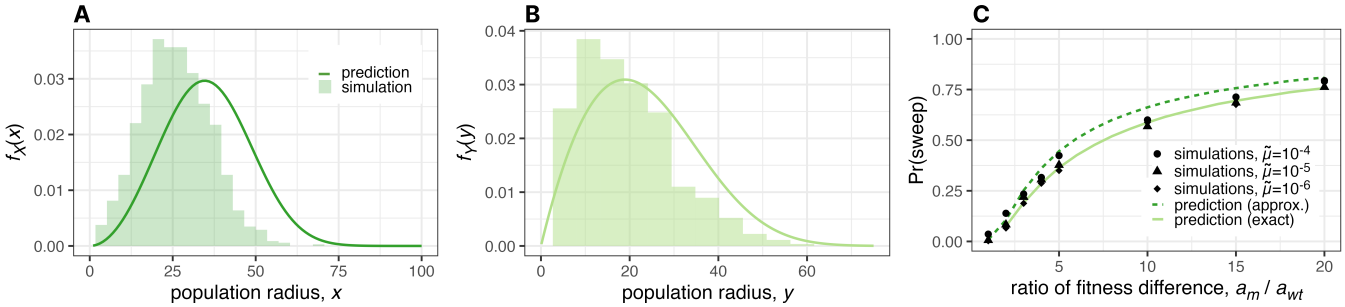

Figure 3. Agent-based simulation results versus macroscopic model predictions. **A.** Probability density of the wildtype population radius at the time the first surviving mutant arises. **B.** Probability density of the distance between the wildtype origin and the location at which the first successful mutant is born. **C.** Unconditional selective sweep probability versus the ratio of the proliferation rate differences $a_{\mathrm{wt}} = r_{\mathrm{n}} - r_{\mathrm{re}}$ and $a_{\mathrm{m}} = r_{\mathrm{m}} - r_{\mathrm{wt}}$. Each histogram in panels **A** and **B** and each data point in **C** is based on 1,000 replicates. Except where parameter values are explicitly varied, we set $m = 0.05, \tilde{\mu} = 10^{-5}, K = 16, r_{\mathrm{re}} = 0.91, r_{\mathrm{wt}} = 1$ and $r_{\mathrm{m}} = 1.3$, leading to speeds $c_{\mathrm{wt}} \approx 0.15$ and $c_{\mathrm{m}} \approx 0.31$ and mutant survival probability $\rho \approx 0.23$. Results for alternative parameter values and less stringent sweep definitions are shown in Figures S1-S3. The conversion from simulation parameters to macroscopic model parameters is explained in Materials and Methods.

selective sweep, simply by applying Bayes' theorem,

$$f_X(x|\text{sweep}) = \frac{\Pr(\text{sweep}|X = x) f_X(x)}{\Pr(\text{sweep})}. \tag{14}$$

Using the approximation $y = 0$, we obtain the Weibull distribution

$$f_X(x|\text{sweep}) \approx \frac{4x^3}{\beta^3\theta^4} e^{-\frac{x^4}{\beta^3\theta^4}} \tag{15}$$

where $\beta = \frac{c_{\mathrm{m}} - c_{\mathrm{wt}}}{c_{\mathrm{m}}}$. This approximation and the corresponding results in one and two dimensions are close to the exact solutions (Figure 2F).

IV. SIMULATIONS

To gauge the robustness of our theoretical predictions against stochastic growth and discretization, we measured the frequency of selective sweeps in a two-dimensional agent-based model.

A. An agent-based realisation of the two-dimensional spatial Moran process

We suppose that the wildtype population invades a habitat initially occupied by a resident competitor, which is a plausible biological scenario for both invasive species [39] and invasive tumours [27, 31]. Our agent-based model thus has three types of individuals: resident, wildtype invader, and mutant invader with local proliferation rates r_{re}, r_{wt} and r_{m}. Localised competition between wildype and resident slows the wildtype expansion and creates potential for selective sweeps. We implemented this model using the *demon* agent-based modelling framework [40] within the *warlock* computational workflow [41], which facilitates running large numbers of simulations on a high-performance computing cluster. We have previously applied the same framework to studying cancer evolution [27, 41]. Further model details are given in Materials and Methods.

The agent-based simulations provide useful insights additional to the macroscopic model because, although the general setup is the same, they differ in several ways. Space in the simulations is divided into discrete patches; the times between birth and dispersal events are exponentially distributed random variables constituting another source of stochasticity; population boundaries are rough, not smooth; the expansion wave front is typically not sharp, and changes shape as the wave progresses; and the mutant will have increased propagation speed when competing with the resident population. Hence we would not expect perfect agreement between the results of the two models.

B. Linking the microscopic and macroscopic models

Our agent-based model approximately resembles a spatial death-birth Moran process (also known as the stepping stone model) [42, 43]. Expansion speeds in the spatial Moran process can in turn be approximated using the Fisher-Kolmogorov-Petrovsky-Piscounov (FKPP) equation [36, 37], which predicts that the mutant will expand within the wildtype with constant radial expansion speed c_{m} dependent on the difference in their local proliferation rates $a_{\mathrm{m}} = r_{\mathrm{m}} - r_{\mathrm{wt}}$ [32, 33, 35]. Analogously, we have a constant expansion speed of the wildtype c_{wt} dependent on $a_{\mathrm{wt}} = r_{\mathrm{wt}} - r_{\mathrm{re}}$. To compare the results of our discrete-space simulations to our continuous-space macroscopic model, we measured the propagation speeds of the wildtype within

the resident, and of the mutant within the wildtype (see Materials and Methods). Further investigations of propagation speeds in this model will be the subject of a further study.

C. Simulation results

Given the considerable differences between the models, the probability density functions resulting from the macroscopic and microscopic models are reassuringly consistent. The radius at the time the first surviving mutant arises is slightly lower in the simulations than in our analytical model (Figure 3A). Similarly, the distribution for the location of the first surviving mutant coincides well except for a small offset of the mean (Figure 3B). Such offsets are expected due to discretization effects and the fact that, in the simulations, the propagation front needs to be established before the expansion can proceed. The sweep probabilities in simulations are nevertheless very close to our analytical predictions and change very little when varying the mutation rate over orders of magnitude, confirming our prediction that the sweep probability is independent of the mutation rate (Figure 3C; Figures S1-S4).

V. ALTERNATIVE GROWTH MODELS

To further demonstrate the robustness of our findings, we compare them to results obtained from alternative models. Instead of permitting mutants to grow within the wildtype population, one might instead assume that dispersal is restricted to the wildtype population boundary. In this case, a complete sweep is impossible and we instead ask whether the first arising mutant envelopes the wildtype. Interestingly, this envelopment probability is independent of the mutation rate and obtains values close to the sweep probabilities of our main model [20] (Figures 4 and S5; SI text). Alternatively, one might consider exponential growth with no competition, and ask whether a single mutant comes to dominate the population. Adapting our methods to this case (see SI text), we find that a single mutant is unlikely to become dominant unless its exponential growth rate is several times larger than that of the wildtype (Figure S6). Although our focus is on expanding populations, we also applied our methods to compute the sweep probability in constant populations (SI text), obtaining a result that depends on both mutation rate and population size (Figure 4), in agreement with prior analyses [38, 44].

VI. APPLICATION TO CANCER

Our findings have especially interesting implications for understanding cancer evolution. Here we consider a wildtype tumour that evolves while invading a resident population of normal cells.

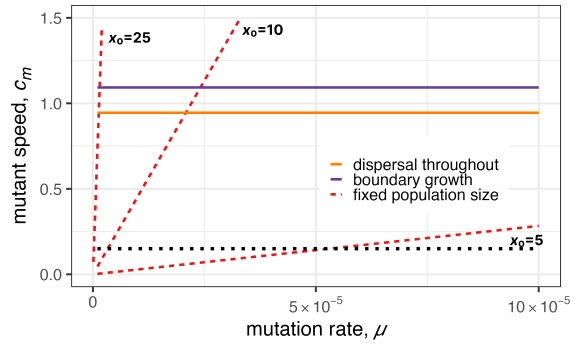

Figure 4. Comparison of sweep probabilities between different growth models in 3D. We show the tuples (μ, c_{m}) that yield $\Pr(\text{sweep}) = 0.5$ for models with proliferation restricted to the boundary (solid purple line), population size fixed at radius x_0 (dashed red lines), and our main model in which mutants can expand within the expanding wildtype population (solid orange line). In the two range expansion models, the wildtype speed is set to $c_{\mathrm{wt}} = 0.15$ (dotted black line).

A. Arrival of the first driver mutation during tumour growth

Our model has only three parameters: the mutation rate conditional on survival, μ, and the wildtype and mutant propagation speeds, c_{wt} and c_{m}. To estimate c_{wt} for human tumours, we follow a similar procedure to prior studies [20, 45]. Consider a tumor that grows to volume V between 1 and 10 cm^3 in time T between 5 and 20 years. The propagation speed can then be estimated as

$$\tilde{c} = \frac{r}{T} = \frac{\sqrt[3]{\frac{3}{4\pi}V}}{T},$$

which equates to between 1 and 40 μm per day. Given that the diameter of a typical cancer cell is $l \approx 20\,\mu$m [46] and the generation time (cell cycle time) is $\tau_G \approx 4$ days [47], we can switch units to obtain $c = \tilde{c} \times \tau_G / l$, which is between 0.15 and 7 cell diameters per generation.

The rate of acquiring advantageous (driver) mutations is usually estimated at around $\tilde{\mu} = 10^{-5}$ per cell per generation [45, 48]. For the survival probability ρ, we assume values between 0.09 and 0.5, in agreement with inferred values in colorectal tumours [49].

Together, these parameter values lead to a typical length scale

$$\theta = \sqrt[4]{\frac{3c_{\mathrm{wt}}}{\pi\mu}} \approx 10 \text{ to } 50 \text{ cells}.$$

The expected values of X and Y are then $\mathbb{E}[X] = 9$ to 45 cell diameters and $\mathbb{E}[Y] = 7$ to 35 cell diameters. It follows that a tumour, having acquired sufficient mutations to grow, will likely gain further driver mutations already during early development.

B. Ubiquity of clonal interference in cancer

During tumour growth, we predict very low sweep probabilities unless the mutant grows much faster within the tumour than the tumour grows into surrounding tissue (Figure 2E, 3C). When sweeps do occur, they will begin and end in the very early stages of tumour growth. Assuming c_{m} is at most ten times c_{wt}, we find that the expected tumour radius when a selective sweep occurred, given that a sweep did occur, is no more than 50 cells, corresponding to a population size N of no more than 400,000 cells. The time for the sweep to be completed is then $\tau_2 \approx 40$ generations, at a population size of approximately $800,000$ cells. These numerical estimates are highly insensitive to varying parameter values because $\mathbb{E}[X]$, $\mathbb{E}[Y]$ and the expected value of $f_X(x|\text{sweep})$ are proportional to the characteristic length θ, which varies with the fourth root of μ and c_{wt}. Hence, in three dimensions, changing μ or c_{wt} by a factor of 100 changes our estimates by only a factor of $\sqrt[4]{100} \approx 3$.

VII. DISCUSSION

Here we have used mathematical modelling and analysis to determine the expected frequency of "classic" selective sweeps versus incomplete or soft sweeps during range expansions. We find that this frequency is generally expected to be low, even for mutations with very high selection coefficients. Moreover, when the wildtype and mutant radial expansion speeds are constant, the sweep probability can be expressed solely in terms of those speeds, which can in turn be related, through the FKPP equation or other standard models, to the selection coefficient, dispersal rates, and other basic parameters. Why is the sweep probability independent of the mutation rate? An intuitive explanation is that if the mutation rate is higher then, on the one hand, the first advantageous mutation is likely to arise in a smaller population – meaning that it has less distance to travel to achieve a sweep – but, on the other hand, competing mutations will also tend to arise sooner than in the case of a lower mutation rate. These two effects exactly cancel out under the assumption of constant radial growth speeds.

We make three arguments to justify our focus on this particular model of range expansions. First, our model corresponds to the continuum approximation of standard mathematical models of range expansions and spatial population genetics: the spatial Moran process (or stepping stone model [43]) and the biassed voter model (which is equivalent to an Eden growth model extended to allow local dispersal and competition throughout the population) [35]. These models are well understood, intuitive, and easy to parameterise. Second, because our model is relatively permissive to selective sweeps, it provides useful upper bounds for selective sweep probabilities in more complex scenarios, as further explained below. Third, in much the same way as the Moran and Wright-Fisher processes are the most useful, tractable models of evolution in constant-sized, non-spatial populations, so the constant-radial-speed model yields the clearest rule-of-thumb results for range expansions. For instance, if the radial expansion speed at which a mutant spreads within the wildtype population is twice the speed at which the wildtype expands then the probability of this mutant achieving a selective sweep can be approximated simply as $(1 - 1/2)^2 = 1/4$ in two dimensions and $1/8$ in three dimensions.

Some alternative models have been considered previously. Antal and colleagues [20] used a macroscopic model similar to ours to investigate the case in which mutations arise only at the boundary of a range expansion. Given that selective sweeps are then impossible, the interesting outcome is when the mutant envelops the wildtype. Ralph & Coop [38] and Martens and colleagues [19, 44] instead considered constant-sized populations, and found that selective sweeps are likely only if the population width is much smaller than a characteristic length scale, which depends on the mutation rate, the dispersal rate, the effective local population density, and the strength of selection. In SI Text we compare our results to those of these prior studies and consider other alternative growth laws. The general conclusions are that selective sweeps are predicted to be rare except in small populations and that our model provides a useful upper bound on the sweep probability in range expansions. Selective sweeps will be even less frequent when the wildtype radial expansion speed increases over time (as is typical in the middle stage of tumour progression) or when mutant expansion is restricted to the tumour boundary (which may be the case in some tumours [27, 28, 50]).

A selective sweep can occur only if the rate of spread of an advantageous mutation exceeds the expansion speed of the wildtype population. This scenario is plausible, for example, in the case of an invasive species that is still adapting to its new conditions and whose range expansion is slowed by the need to modify its environment (niche construction) or by interspecific interactions [39]. If the invader must outcompete a resident then the sweep probability can be approximated via the FKPP solution in terms of proliferation rates as in eqn. 13. The dimensional exponent in this result implies that selective sweeps are most likely to occur in species invading essentially linear habitats, such as

coastlines.

Our three-dimensional results are particularly relevant to understanding the nature of solid tumour evolution. There are two ways in which tumours might acquire the multiple clonal drivers (advantageous mutations) observed in sequencing studies. The first route is via sequential population bottlenecks. Suppose that early tumour development comprises not one but several successive range expansions. Growth repeatedly stalls due to constraints such as hypoxia, immune control, and physical barriers. Driver mutations enable subclones to escape these constraints and invade new territory, each time purging genetic diversity so that the final, prolonged expansion originates from a single highly transformed cell. This episodic model is conventional and has been particularly well characterised in recent studies of colorectal cancer [51, 52] and breast cancer [53]. The second way to acquire clonal drivers is via selective sweeps during the final range expansion. Our results suggest that the first process is the more important. Selective sweeps of even extremely strong drivers are highly unlikely to occur in tumours containing more than a few hundred thousand cells, equivalent to less than a cubic millimeter in volume. Pervasive genetic heterogeneity and parallel evolution [54] are therefore revealed to be straightforward consequences of spatial structure, expected to be ubiquitous across solid tumours, irrespective of mutation rate and degree of genomic instability.

VIII. MATERIALS AND METHODS

A. Numerical Integration

We performed numerical integration using the MATLAB function 'trapz'. For values of x close to 0 and beyond 2θ, $f_X(x)$ is relatively very small and its contribution to the integrals is negligible. Hence, to avoid numerical errors without compromising precision, we set the lower and upper bounds of integration to 0.001θ and 3θ when integrating over $f_X(x)$. Similarly, we set 0.001θ as the lower bound of the integration over y. We set interval widths to 0.001θ for $x \leq 0.5$, and 0.01θ for $x > 0.5$.

B. Agent-based simulations

Individuals in our agent-based model are subdivided into well-mixed demes on a regular two-dimensional grid. The demes have identical carrying capacities, K, and are initially filled with residents, except that a single wildtype invader is introduced at the centre of the grid. At each time step, an individual is chosen at random, with probability weighted by fitness, to be replaced by two offspring. Each offspring then either migrates, with probability m, to a neighbouring deme in a randomly chosen direction, or remains in its parent deme. Local density dependence is implemented by imposing a very high death rate whenever a deme is above carrying capacity. Mutation is coupled to wildtype reproduction. To improve computational efficiency, resident individuals are not permitted to disperse; we verified with additional simulations (to be published in a later study) that this asymmetry has only a minor effect on the wildtype expansion speed. Further model details have been published previously [27, 41]. All simulations were performed on City, University of London's Hyperion cluster.

For all simulations, we set the deme size to $K = 16$ and the migration probability $m = 0.05$. For such small m, the probability to survive drift can be estimated as the probability to become locally fixed in one deme. Since the within-deme dynamics approximate a Moran process, this probability is $\rho = \frac{1-r_{\mathrm{wt}}/r_{\mathrm{m}}}{1-(r_{\mathrm{wt}}/r_{\mathrm{m}})^K}$, where r_{wt} and r_{m} are the proliferation rates [55]. We measured propagation speeds by simulating the expansion of a wildtype into a large resident population, or of a mutant into a large wildtype population, in the absence of mutation and then applying linear regression to the growth curve of the effective radius, defined as the square root of the population size divided by π. For each parameter set, we took the mean speed from ten replicates, for which the standard deviation was consistently below 2% of the mean. We set the proliferation rates to $r_{\mathrm{re}} = 0.91$ and $r_{\mathrm{wt}} = 1.0$ for the resident and wildtype, respectvively, leading to $a_{\mathrm{wt}} = r_{\mathrm{wt}} - r_{\mathrm{re}} = 0.09$. The measured wildtype speed was then $c_{\mathrm{wt}} \approx 0.15$. We varied the mutant proliferation rate r_{m} from 1.1 to 3.0, so that $a_{\mathrm{m}} = r_{\mathrm{m}} - r_{\mathrm{wt}}$ ranged from 0.1 to 2.0. The measured mutant speeds are presented in Table I.

a_{m}	0.1	0.2	0.3	0.4	0.5	0.6	0.7
c_{m}	0.14	0.23	0.31	0.38	0.46	0.53	0.60
a_{m}	0.8	0.9	1.0	1.1	1.2	1.3	1.4
c_{m}	0.68	0.75	0.82	0.89	0.96	1.03	1.11
a_{m}	1.5	1.6	1.7	1.8	1.9	2.0	
c_{m}	1.17	1.24	1.31	1.39	1.46	1.52	

Table I. Correspondence between difference in fitness a_{m} and mutant propagation speed c_{m}, measured in simulations.

C. Code availability

The code and data needed to reproduce our results can be found at https://zenodo.org/records/10775693.

[1] Oskar Hallatschek, Pascal Hersen, Sharad Ramanathan, and David R Nelson. Genetic drift at expanding frontiers promotes gene segregation. *Proceedings of the National Academy of Sciences*, 104(50):19926–19930, 2007.

[2] Zaira Seferbekova, Artem Lomakin, Lucy R Yates, and Moritz Gerstung. Spatial biology of cancer evolution. *Nature Reviews Genetics*, 24(5):295–313, 2023.

[3] Leslie G Biesecker and Nancy B Spinner. A genomic view of mosaicism and human disease. *Nature Reviews Genetics*, 14(5):307–320, 2013.

[4] Iñigo Martincorena, Amit Roshan, Moritz Gerstung, Peter Ellis, Peter Van Loo, Stuart McLaren, David C Wedge, Anthony Fullam, Ludmil B Alexandrov, Jose M Tubio, et al. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science*, 348(6237):880–886, 2015.

[5] Luigi L Cavalli-Sforza, Paolo Menozzi, and Alberto Piazza. Demic expansions and human evolution. *Science*, 259(5095):639–646, 1993.

[6] Gretta T Pecl, Miguel B Araújo, Johann D Bell, Julia Blanchard, Timothy C Bonebrake, I-Ching Chen, Timothy D Clark, Robert K Colwell, Finn Danielsen, Birgitta Evengård, et al. Biodiversity redistribution under climate change: Impacts on ecosystems and human well-being. *Science*, 355(6332):eaai9214, 2017.

[7] Emily V Moran and Jake M Alexander. Evolutionary responses to global change: lessons from invasive species. *Ecology Letters*, 17(5):637–649, 2014.

[8] Laurent Excoffier, Matthieu Foll, and Rémy J Petit. Genetic consequences of range expansions. *Annual Review of Ecology, Evolution, and Systematics*, 40:481–501, 2009.

[9] Oskar Hallatschek and David R Nelson. Life at the front of an expanding population. *Evolution*, 64(1):193–206, 2010.

[10] Kirill S Korolev, Melanie JI Müller, Nilay Karahan, Andrew W Murray, Oskar Hallatschek, and David R Nelson. Selective sweeps in growing microbial colonies. *Physical biology*, 9(2):026008, 2012.

[11] Diana Fusco, Matti Gralka, Jona Kayser, Alex Anderson, and Oskar Hallatschek. Excess of mutational jackpot events in expanding populations revealed by spatial luria–delbrück experiments. *Nature communications*, 7(1):12760, 2016.

[12] Serhii Aif, Nico Appold, Lucas Kampman, Oskar Hallatschek, and Jona Kayser. Evolutionary rescue of resistant mutants is governed by a balance between radial expansion and selection in compact populations. *Nature Communications*, 13(1):7916, 2022.

[13] Douglas Hanahan and Robert A Weinberg. The hallmarks of cancer. *cell*, 100(1):57–70, 2000.

[14] Eric R Fearon and Bert Vogelstein. A genetic model for colorectal tumorigenesis. *cell*, 61(5):759–767, 1990.

[15] Moritz Gerstung, Clemency Jolly, Ignaty Leshchiner, Stefan C Dentro, Santiago Gonzalez, Daniel Rosebrock, Thomas J Mitchell, Yulia Rubanova, Pavana Anur, Kaixian Yu, et al. The evolutionary history of 2,658 cancers. *Nature*, 578(7793):122–128, 2020.

[16] Marco Gerlinger, Andrew J Rowan, Stuart Horswell, James Larkin, David Endesfelder, Eva Gronroos, Pierre Martinez, Nicholas Matthews, Aengus Stewart, Patrick Tarpey, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *New England journal of medicine*, 366(10):883–892, 2012.

[17] Alexander Davis, Ruli Gao, and Nicholas Navin. Tumor evolution: Linear, branching, neutral or punctuated? *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer*, 1867(2):151–161, 2017.

[18] Samra Turajlic, Andrea Sottoriva, Trevor Graham, and Charles Swanton. Resolving genetic heterogeneity in cancer. *Nature Reviews Genetics*, 20(7):404–416, 2019.

[19] Erik A Martens, Rumen Kostadinov, Carlo C Maley, and Oskar Hallatschek. Spatial structure increases the waiting time for cancer. *New journal of physics*, 13(11):115014, 2011.

[20] Tibor Antal, PL Krapivsky, and MA Nowak. Spatial evolution of tumors with successive driver mutations. *Physical Review E*, 92(2):022705, 2015.

[21] Chay Paterson, Martin A Nowak, and Bartlomiej Waclaw. An exactly solvable, spatial model of mutation accumulation in cancer. *Scientific reports*, 6(1):39511, 2016.

[22] Peter Armitage and Richard Doll. The age distribution of cancer and a multi-stage theory of carcinogenesis. *British journal of cancer*, 91(12):1983–1989, 2004.

[23] Chay Paterson, Hans Clevers, and Ivana Bozic. Mathematical model of colorectal cancer initiation. *Proceedings of the National Academy of Sciences*, 117(34):20681–20688, 2020.

[24] Michael D Nicholson, David Cheek, and Tibor Antal. Sequential mutations in exponentially growing populations. *PLOS Computational Biology*, 19(7):e1011289, 2023.

[25] Ruping Sun, Zheng Hu, Andrea Sottoriva, Trevor A Graham, Arbel Harpak, Zhicheng Ma, Jared M Fischer, Darryl Shibata, and Christina Curtis. Between-region genetic divergence reflects the mode and tempo of tumor evolution. *Nature genetics*, 49(7):1015–1024, 2017.

[26] Jeffrey West, Ryan O Schenck, Chandler Gatenbee, Mark Robertson-Tessi, and Alexander RA Anderson. Normal tissue architecture determines the evolutionary course of cancer. *Nature communications*, 12(1):2060, 2021.

[27] Robert Noble, Dominik Burri, Cécile Le Sueur, Jeanne Lemant, Yannick Viossat, Jakob Nikolas Kather, and Niko Beerenwinkel. Spatial structure governs the mode of tumour evolution. *Nature ecology & evolution*, 6(2):207–217, 2022.

[28] Xiao Fu, Yue Zhao, Jose I Lopez, Andrew Rowan, Lewis Au, Annika Fendler, Steve Hazell, Hang Xu, Stuart Horswell, Scott TC Shepherd, et al. Spatial patterns of tumour growth impact clonal diversification in a computational model and the tracerx renal study. *Nature ecology & evolution*, 6(1):88–102, 2022.

[29] Noemi Andor, Trevor A Graham, Marnix Jansen, Li C Xia, C Athena Aktipis, Claudia Petritsch, Hanlee P Ji, and Carlo C Maley. Pan-cancer analysis of the extent and consequences of intratumor heterogeneity. *Nature medicine*, 22(1):105–113, 2016.

[30] Nicholas McGranahan and Charles Swanton. Clonal heterogeneity and tumor evolution: past, present, and the future. *Cell*, 168(4):613–628, 2017.

[31] Robert Noble, John T Burley, Cécile Le Sueur, and Michael E Hochberg. When, why and how tumour clonal diversity predicts survival. *Evolutionary applications*, 13(7):1558–1568, 2020.

[32] Ronald Aylmer Fisher. The wave of advance of advantageous genes. *Annals of eugenics*, 7(4):355–369, 1937.

[33] Andrey Nikolaevich Kolmogorov, Ivan Georgievich Petrovsky, and Nikolay Semenovich Piskunov. A study of the diffusion equation with increase in the amount of substance, and its application to a biological problem. *Bullentin of Moscow State University*, 1937.

[34] Eric Brunet and Bernard Derrida. Shift in the velocity of a front due to a cutoff. *Physical Review E*, 56(3):2597, 1997.

[35] James Dickson Murray. *Mathematical Biology: I: An introduction*, volume 17. Springer, 2002.

[36] Oskar Hallatschek and Kirill S Korolev. Fisher waves in the strong noise limit. *Physical review letters*, 103(10):108103, 2009.

[37] Bahram Houchmandzadeh and Marcel Vallade. Fisher waves: An individual-based stochastic model. *Physical Review E*, 96(1):012414, 2017.

[38] Peter Ralph and Graham Coop. Parallel adaptation: one or many waves of advance of an advantageous allele? *Genetics*, 186(2):647–668, 2010.

[39] Jens-Christian Svenning, Dominique Gravel, Robert D Holt, Frank M Schurr, Wilfried Thuiller, Tamara Münkemüller, Katja H Schiffers, Stefan Dullinger, Thomas C Edwards Jr, Thomas Hickler, et al. The influence of interspecific interactions on species range expansion rates. *Ecography*, 37(12):1198–1209, 2014.

[40] Robert Noble. demon: deme-based oncology model. https://github.com/robjohnnoble/demonmodel, 2023.

[41] Maciej Bak, Blair Colyer, Veselin Manojlović, and Robert Noble. Warlock: an automated computational workflow for simulating spatially structured tumour evolution. *arXiv preprint arXiv:2301.07808*, 2023.

[42] Richard Durrett, Jasmine Foo, and Kevin Leder. Spatial moran models, ii: cancer initiation in spatially structured tissue. *Journal of mathematical biology*, 72:1369–1400, 2016.

[43] Motoo Kimura and George H Weiss. The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics*, 49(4):561, 1964.

[44] Erik A Martens and Oskar Hallatschek. Interfering waves of adaptation promote spatial mixing. *Genetics*, 189(3):1045–1060, 2011.

[45] Ivana Bozic, Tibor Antal, Hisashi Ohtsuki, Hannah Carter, Dewey Kim, Sining Chen, Rachel Karchin, Kenneth W Kinzler, Bert Vogelstein, and Martin A Nowak. Accumulation of driver and passenger mutations during tumor progression. *Proceedings of the National Academy of Sciences*, 107(43):18545–18550, 2010.

[46] Babita Shashni, Shinya Ariyasu, Reisa Takeda, Toshihiro Suzuki, Shota Shiina, Kazunori Akimoto, Takuto Maeda, Naoyuki Aikawa, Ryo Abe, Tomohiro Osaki, et al. Size-based differentiation of cancer and normal cells by a particle size analyzer assisted by a cell-recognition pc software. *Biological and Pharmaceutical Bulletin*, 41(4):487–503, 2018.

[47] Siân Jones, Wei-dong Chen, Giovanni Parmigiani, Frank Diehl, Niko Beerenwinkel, Tibor Antal, Arne Traulsen, Martin A Nowak, Christopher Siegel, Victor E Velculescu, et al. Comparative lesion sequencing provides insights into tumor evolution. *Proceedings of the National Academy of Sciences*, 105(11):4283–4288, 2008.

[48] Marc J Williams, Benjamin Werner, Timon Heide, Christina Curtis, Chris P Barnes, Andrea Sottoriva, and Trevor A Graham. Quantification of subclonal selection in cancer from bulk sequencing data. *Nature genetics*, 50(6):895–903, 2018.

[49] Benjamin Werner, Jack Case, Marc J Williams, Ketevan Chkhaidze, Daniel Temko, Javier Fernández-Mateos, George D Cresswell, Daniel Nichol, William Cross, Inmaculada Spiteri, et al. Measuring single cell divisions in human tissues from multi-region sequencing data. *Nature communications*, 11(1):1035, 2020.

[50] Maya A Lewinsohn, Trevor Bedford, Nicola F Müller, and Alison F Feder. State-dependent evolutionary models reveal modes of solid tumour growth. *Nature Ecology & Evolution*, 7(4):581–596, 2023.

[51] Marc D Ryser, Diego Mallo, Allison Hall, Timothy Hardman, Lorraine M King, Sergei Tatishchev, Inmaculada C Sorribes, Carlo C Maley, Jeffrey R Marks, E Shelley Hwang, et al. Minimal barriers to invasion during human colorectal tumor growth. *Nature communications*, 11(1):1280, 2020.

[52] William Cross, Michal Kovac, Ville Mustonen, Daniel Temko, Hayley Davis, Ann-Marie Baker, Sujata Biswas, Roland Arnold, Laura Chegwidden, Chandler Gatenbee, et al. The evolutionary landscape of colorectal tumorigenesis. *Nature ecology & evolution*, 2(10):1661–1672, 2018.

[53] Artem Lomakin, Jessica Svedlund, Carina Strell, Milana Gataric, Artem Shmatko, Gleb Rukhovich, Jun Sung Park, Young Seok Ju, Stefan Dentro, Vitalii Kleshchevnikov, et al. Spatial genomics maps the structure, nature and evolution of cancer clones. *Nature*, 611(7936):594–602, 2022.

[54] Alexander M Frankell, Michelle Dietzen, Maise Al Bakir, Emilia L Lim, Takahiro Karasaki, Sophia Ward, Selvaraju Veeriah, Emma Colliver, Ariana Huebner, Abigail Bunkum, et al. The evolution of lung cancer and impact of subclonal selection in tracerx. *Nature*, pages 1–9, 2023.

[55] Martin A Nowak. *Evolutionary dynamics: exploring the equations of life*. Harvard university press, 2006.

Supporting Information Text

1. Proof of Claim 1

Proof of claim 1. By definition of the per capita mutation rate, we have $\mu N(t_1)dt$ successful mutations in the infinitesimal time interval $[t_1, t_1 + dt]$. We sum over the time interval $[0, t]$ and obtain $\lambda = \int_0^t \mu N(t)\, dt$. We assume that mutations occur independently of each other (infinite alleles) such that the stochastic process is a non-homogeneous Poisson process with mean λ. □

2. Connection between per-capita and per-division mutation rate

In our macroscopic model, we assume a constant per-capita mutation rate $\tilde{\mu}$. Often, it is assumed that mutations are coupled to division. Here, we show how a per-capita mutation rate connects to a per-division mutation rate. Suppose that after each division one of the two daughter cells obtains a driver mutation with probability p_μ. (The probability that both daughter cells obtain a driver mutation scales with p_μ^2, which can be neglected for sufficiently small p_μ.) Suppose cells proliferate at rate b and die at rate d. Those rates may depend on the time or location. If division rates are constant over time and space, then the per-division mutation probability p_μ yields a constant per-capita mutation rate $\tilde{\mu} = p_\mu b$. The death rate may still differ in time and space.

3. Exact solution for the conditional sweep probability

Here, we provide the full expression and exact solution for the conditional sweep probability $\Pr(\text{sweep}|X = x, Y = y)$ for the three-dimensional case (eqn. 9 in the main text). From the main text, we have

$$\Pr \text{sweep}|X = x, Y = y = e^{-\mu \int_0^t N_{\text{wt}}(\tau)\, d\tau}. \tag{1}$$

were N_{wt} is the size of the remaining wildtype population and following eqn. 6 from the main text the integral can be divided into

$$\int_0^\infty N_{\text{wt}}(\tau)\, d\tau = \int_0^{\tau_1} \tilde{N}_{\text{wt}}(\tau) - \Delta_1(\tau)\, d\tau + \int_{\tau_1}^{\tau_2} \tilde{N}_{\text{wt}}(\tau) - \Delta_2(\tau)\, d\tau. \tag{2}$$

Here $\tilde{N}_{\text{wt}}$ and $\Delta_1(m)$ are the population sizes of the wildtype and mutant in case of undisturbed radial growth such that

$$\begin{aligned} \tilde{N}_{\text{wt}} &= \frac{4}{3}\pi x_{\text{wt}}^3(\tau) \\ \Delta_1(\tau) &= \frac{4}{3}\pi x_{\text{m}}^3(\tau). \end{aligned} \tag{3}$$

Furthermore, $\Delta_2(\tau)$ is given by the intersecting volume of two balls with radii x_{wt} and x_{m} at distance y. Following ref. (1), we have

$$\Delta_2(\tau) = \frac{\pi}{12y}(x_{\text{wt}} + x_{\text{m}} + y)^2(y^2 + 2yx_{\text{m}} - 3x_{\text{m}}^2 + 2yx_{\text{wt}} + 6x_{\text{wt}}x_{\text{m}} - 3x_{\text{wt}}^2). \tag{4}$$

We can express x_{wt} and x_{m} in terms of the wildtype radius at the time the first mutant occurred x and the time τ from the emergence of the first mutant using that

$$\begin{aligned} x_{\text{wt}}(\tau) &= c_{\text{wt}} t = c_{\text{wt}}(\tau + x/c_{\text{wt}}) \\ x_{\text{m}}(\tau) &= c_{\text{m}}\tau. \end{aligned} \tag{5}$$

Now, we are ready to integrate over $N_{\text{wt}}(\tau)$ and obtain

$$\int_0^\infty N_{\text{wt}}(\tau)\, d\tau = \frac{\pi\left(15x^4\left(3c_{\text{m}}^2 - 3c_{\text{m}}c_{\text{wt}} + c_{\text{wt}}^2\right) - y^4\left(23c_{\text{m}}^2 + 31c_{\text{m}}c_{\text{wt}} + 23c_{\text{wt}}^2\right) - 210c_{\text{m}}^2x^2y^2\right)}{45(c_{\text{m}} - c_{\text{wt}})^3}. \tag{6}$$

Inserting this result into eqn. 1 gives us the conditional sweep probability. The exact solution reduces to the solution in the main text with $y = 0$, for which

$$\int_0^\infty N_{\text{wt}}(\tau)\, d\tau = \frac{1}{\mu}\left(\frac{x}{\alpha}\right)^4 \quad \text{with} \quad \alpha = \sqrt[4]{\frac{3(c_{\text{m}} - c_{\text{wt}})^3}{\pi\mu(c_{\text{wt}}^2 - 3c_{\text{wt}}c_{\text{m}} + 3c_{\text{m}}^2)}}. \tag{7}$$

4. The sweep probability is independent of the mutation rate

We show that the unconditional selective sweep probability $\Pr(\text{sweep})$ (eqn. 11 in the main text) is independent of the mutation rate. We start by defining homogeneous functions and stating a lemma for homogeneous functions. We then show how this lemma proves independence of the unconditional sweep probability and eventually prove the lemma. Here, we focus on the three-dimensional case whereas the independence for the two- and one-dimensional case is shown later.

A. Homogeneous functions.

Definition 1. *A function $f : \mathbb{R}^n \to \mathbb{R}$ is called homogeneous of degree k if*

$$f(sx_1, sx_2, ..., sx_n) = s^k f(x_1, x_2, ..., x_n) \qquad [8]$$

for every $x_1, x_2, ..., x_n$ and $s \neq 0$.

Noteworthy, polynomials in which all terms have the same polynomial degree k are homogeneous functions with degree k. For example, $5x^2y^2 + \pi x^4$ is homogeneous in x and y with degree 4.

Lemma 1. *Let $\nu > 0$. Let $H(x, y)$, $P(x, y)$ and $Q(x, y)$ be homogeneous functions in x and y with degrees $h, p, q > 0$ respectively and $Q(x, y) \neq 0$ for all $x > 0$ and $y > 0$. Then, the integral*

$$I = \int_0^\infty \int_0^x e^{-\nu H(x,y)} \frac{P(x, y)}{Q(x, y)} \, dy \, dx \qquad [9]$$

is proportional to $\nu^{\frac{q-p-2}{h}}$.

B. The sweep probability is independent of the mutation rate.

To apply the lemma on the unconditional sweep probability, we set $\nu = \mu$ and

$$\begin{aligned} H(x, y) &= \int_0^\infty N_{\mathrm{wt}}(\tau) \, d\tau + \frac{3c_{\mathrm{wt}}}{\pi} x^4 \\ P(x, y) &= 3y^2 \times 4x^3 \\ Q(x, y) &= \frac{3c_{\mathrm{wt}}}{\pi} x^3. \end{aligned} \qquad [10]$$

It is straightforward that $P(x, y)$ and $Q(x, y)$ are homogeneous in x and y with degrees $p = 5$ and $q = 3$. Looking at eqn. 6, it is also apparent that $H(x, y)$ is homogeneous with degree 4. Using those expressions and applying lemma 1, we have

$$\Pr(\text{sweep}) = \mu \times I \propto \mu \times \mu^{\frac{3-5-2}{4}} = \mu^0, \qquad [11]$$

which proves the independence of the mutation rate.

We found that the approximate solution using $f_Y(y|X = x) = \delta(y)$ is also independent of the mutation rate indicating a more general result. Looking at the derivation above, this becomes clear. Consider fixed equations for $\Pr(sweep|X = x, Y = y)$ and $f_X(x)$. For any density function $f_Y(y|X = x)$ that is homogeneous in x and y with degree -1, we can define functions $P(x, y)$ and $Q(x, y)$ such that their difference in degrees is $p - q = -2$. After application of the lemma, it follows independence.

C. Proof of Lemma 1.

Proof of lemma 1. We show the relation by making the substitutions

$$\hat{x} = \nu^{1/h} x \quad \text{and} \quad \hat{y} = \nu^{1/h} y$$

such that

$$\begin{aligned} H(\hat{x}, \hat{y}) &= \nu H(x, y) \\ P(\hat{x}, \hat{y}) &= \nu^{p/h} P(x, y) \\ Q(\hat{x}, \hat{y}) &= \nu^{q/h} Q(x, y) \end{aligned}$$

and

$$dx = \nu^{-1/h} d\hat{x} \quad \text{and} \quad dy = \nu^{-1/h} d\hat{y}.$$

Inserting the substitutions in eqn. 9, we obtain

$$\begin{aligned} I &= \int_0^\infty \int_0^x e^{-\nu H(x,y)} \frac{P(x, y)}{Q(x, y)} \, dy \, dx \\ &= \int_0^\infty \int_0^{\hat{x}} e^{-H(\hat{x},\hat{y})} \frac{\nu^{-p/h} P(\hat{x}, \hat{y})}{\nu^{-q/h} Q(\hat{x}, \hat{y})} \nu^{-1/h} \nu^{-1/h} \, d\hat{y} \, d\hat{x} \\ &= \nu^{-2/h} \nu^{(q-p)/h} \int_0^\infty \int_0^{\hat{x}} e^{-H(\hat{x},\hat{y})} \frac{P(\hat{x}, \hat{y})}{Q(\hat{x}, \hat{y})} \, d\hat{y} \, d\hat{x} \end{aligned}$$

The remaining integral is no longer dependent on parameter ν and shows the desired scaling with $\nu^{\frac{q-p-2}{h}}$. □

5. Sweep probability in one dimension

We consider a population that is expanding in one dimension in two directions with constant speed. If x_{wt} describes the length from the origin of the population to the leading edge, then the population size is $N_{\mathrm{wt}} = 2x_{\mathrm{wt}}(t) = 2c_{\mathrm{wt}}t$. We follow the derivations described in the main text.

A. Arrival time and location of the first mutant. We start by computing the probability density for the time at which the first mutant arises. Using claim 1, we compute the probability that no mutant occurs until time t to be

$$P_0 = e^{-\mu \int_0^t 2c_{\mathrm{wt}} t' dt'} = e^{-\mu c_{\mathrm{wt}} t^2} = e^{-\left(\frac{t}{\kappa_{1\mathrm{D}}}\right)^2} \quad \text{with} \quad \kappa_{1\mathrm{D}} = \sqrt{\frac{1}{\mu c_{\mathrm{wt}}}}. \tag{12}$$

The probability density function for the timing of the first surviving mutant is then

$$f_T(t) = \frac{d(1-P_0)}{dt} = \frac{2t}{\kappa_{1\mathrm{D}}^2} e^{-\left(\frac{t}{\kappa_{1\mathrm{D}}}\right)^2} \tag{13}$$

With a change of variables $t = \frac{x}{c_{\mathrm{wt}}}$, we obtain the probability density for the population radius X at the time the first surviving mutant arises, which is

$$f_X(x) = \frac{2x}{\theta_{1\mathrm{D}}^2} e^{-\left(\frac{x}{\theta_{1\mathrm{D}}}\right)^2} \quad \text{with} \quad \theta_{1\mathrm{D}} = \sqrt{\frac{c_{\mathrm{wt}}}{\mu}}. \tag{14}$$

Since all distances y from the wildtype origin have the same amount of proliferating cells, the probability density for the location of the first surviving mutant Y conditional on the wildtype radius $X = x$ is a homogeneous. After consideration of boundary conditions and normalization, we have

$$f_Y(y|X = x) = \frac{1}{x} 1\{y \leq x\}, \tag{15}$$

The unconditional probability density of y is computed by

$$f_Y(y) = \int_0^\infty f_X(x) f_Y(y|X = x)\, dx = \frac{1}{\theta_{1\mathrm{D}}} \Gamma\left(\frac{1}{2}, \frac{y^2}{\theta_{1\mathrm{D}}^2}\right). \tag{16}$$

B. Exact sweep probability. In one dimension, we obtain a closed-form expression for the exact solution of the sweep probability. Suppose a mutant is born at distance y once the wildtype radius reached length x. Again, we set $\tau = 0$ at the time the mutant is born. Following the main text, the remaining wildtype population is given by

$$N_{\mathrm{wt}} = (2x_{\mathrm{wt}} - 2x_{\mathrm{m}})\, \mathbf{1}[0, \tau_1] + (x_{\mathrm{wt}} - x_{\mathrm{m}} + y)\, \mathbf{1}[\tau_1, \tau_2] \tag{17}$$

where $\tau_1 = \frac{x-y}{c_{\mathrm{m}} - c_{\mathrm{wt}}}$ is the time at which the mutant population reaches the first leading edge of the wildtype and $\tau_2 = \frac{x+y}{c_{\mathrm{m}} - c_{\mathrm{wt}}}$ is the time at which the mutant reaches the second leading edge of the wildtype such that the sweep is completed. Next, we substitute $x_{\mathrm{wt}} = x + c_{\mathrm{wt}}\tau$ and $x_{\mathrm{m}} = c_{\mathrm{m}}\tau$ and then apply claim 1 to compute the probability that no further mutant arises before the sweep is completed, which leads to

$$\Pr(\text{sweep}|X = x, Y = y) = e^{-\mu \int_0^\infty N_{\mathrm{wt}}(\tau)\, d\tau} = e^{-\frac{x^2+y^2}{\alpha_{1\mathrm{D}}^2}} \quad \text{with} \quad \alpha_{1\mathrm{D}} = \sqrt{\frac{c_{\mathrm{m}} - c_{\mathrm{wt}}}{\mu}}. \tag{18}$$

Then, we marginalize out Y, which gives us

$$\Pr(\text{sweep}|X = x) = \int_0^\infty \Pr(\text{sweep}|X = x, Y = y) f_Y(y|X = x)\, dy = \frac{\sqrt{\pi}\alpha_{1\mathrm{D}}}{2x} e^{-\left(\frac{x}{\alpha_{1\mathrm{D}}}\right)^2} \operatorname{erf}\left(\frac{x}{\alpha_{1\mathrm{D}}}\right), \tag{19}$$

where $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\, dt$ is the Gaussian error function. The exact formula for the unconditional sweep probability is obtained by marginalizing out X. The result is

$$\Pr(\text{sweep}) = \int_0^\infty \Pr(\text{sweep}|X = x) f_X(x)\, dx = \frac{\beta' \cot^{-1}\left(\sqrt{1+\beta'}\right)}{\sqrt{1+\beta'}}, \tag{20}$$

where $\beta' = \frac{c_{\mathrm{m}} - c_{\mathrm{wt}}}{c_{\mathrm{wt}}}$ is the speed difference relative to the wildtype speed. We notice that the exact solution for the sweep probability is independent of the mutation rate.

C. Approximate sweep probability. We use the approximation explained in the main text and set $y = 0$, which translates into a probability density $f_Y(y|X = x) = \delta(y)$. Under this assumption, the conditional sweep probability is given by

$$\Pr(\text{sweep}|X = x) = \Pr(\text{sweep}|X = x, Y = 0) = e^{-\left(\frac{x}{\alpha_{1\text{D}}}\right)^2}. \quad [21]$$

and the unconditional sweep probability becomes

$$\Pr(\text{sweep}) = \int_0^\infty \Pr(\text{sweep}|X = x) f_X(x)\, dx = \frac{c_\text{m} - c_\text{wt}}{c_\text{m}}. \quad [22]$$

Using Bayes theorem, we compute the probability distribution of X given a sweep has occurred,

$$f_X(X = x|\text{sweep}) = \frac{2x}{\theta_{1\text{D}}^2 \beta} e^{-\frac{x^2}{\theta_{1\text{D}}^2 \beta}}, \quad [23]$$

where $\beta = \frac{c_\text{m} - c_\text{wt}}{c_\text{m}}$. The distribution of X is still described by a Weibull distribution with shape parameter 2. The difference is only a scaling factor of the characteristic length $\theta_{1\text{D}} \to \theta_{1\text{D}}\sqrt{\beta}$.

6. Sweep probability in two dimensions

We consider a population that is expanding radially in two dimensions at constant speed. If x_wt is the radius of the wildtype population, then the population size is $N = \pi x_\text{wt}^2 = \pi(c_\text{wt} t)^2$. We follow the derivations in the main text.

A. Arrival time and location of the first mutant. We start and compute the probability density for the time at which the first mutant arises. Using claim 1, we compute the probability that no mutant occurs until time t to be

$$P_0 = e^{-\mu \int_0^t \pi c_\text{wt}^2 t'^2\, dt'} = e^{-\mu \frac{\pi}{3} c_\text{wt}^2 t^3} = e^{-\left(\frac{t}{\kappa_{2\text{D}}}\right)^3} \quad \text{with} \quad \kappa_{2\text{D}} = \sqrt[3]{\frac{3}{\mu \pi c_\text{wt}^2}}. \quad [24]$$

Taking the derivative gives us the probability density for the arrival time of the first surviving mutant,

$$f_T(t) = \frac{d(1 - P_0)}{dt} = \frac{3t^2}{\kappa_{2\text{D}}^3} e^{-\left(\frac{t}{\kappa_{2\text{D}}}\right)^3}. \quad [25]$$

We perform a change of variables $t = \frac{x}{c_\text{wt}}$ to obtain the probability density for the radius of the wildtype population X at the time the first mutant arises that is

$$f_X(x) = \frac{3x^2}{\theta_{2\text{D}}^3} e^{-\left(\frac{x}{\theta_{2\text{D}}}\right)^3}, \quad \text{with} \quad \theta_{2\text{D}} = \sqrt[3]{\frac{3c_\text{wt}}{\pi \mu}}. \quad [26]$$

To compute the probability density for the distance Y between the first mutant and the centre of the wildtype population, we notice that there at distance y, there are $2\pi y\, dy$ dividing cells at distance y. Given that the first mutant occurred when the wildtype population had radius $X = x$, we have $f_Y(y|X = x) \propto y^2$. After taking care of the boundary condition and normalization, we get

$$f_Y(y|X = x) = \frac{2y}{x^2} 1\{y \le x\}, \quad [27]$$

The unconditional probability density of y is obtained by

$$f_Y(y) = \int_0^\infty f_Y(y|X = x) f_X(x)\, dx = \frac{2y}{\theta_{2D}^2} \Gamma\left(\frac{1}{3}, \frac{y^3}{\theta_{2D}^3}\right). \quad [28]$$

B. Exact sweep probability. We start by finding the expression for the remaining wildtype population N_wt given that $X = x$ and $Y = y$. As in the three- and one-dimensional cases, we have the remaining wildtype population given by two formulas such that

$$N_\text{wt}(\tau) = (\tilde{N}_\text{wt}(\tau) - \Delta_1(\tau))\mathbf{1}\{[0, \tau_1]\} + (\tilde{N}_\text{wt}(\tau) - \Delta_2(\tau))\mathbf{1}\{[\tau_1, \tau_2]\} \quad [29]$$

with

$$\begin{aligned} \tilde{N}_\text{wt}(\tau) &= \pi x_\text{wt}^2 \\ \Delta_1(\tau) &= \pi x_\text{m}^2 \end{aligned} \quad [30]$$

given by the area of discs with radii x_wt and x_m and

$$\begin{aligned} \Delta_2(\tau) =& x_\text{wt}^2 \cos^{-1}\left(\frac{y^2 + x_\text{wt}^2 - x_\text{m}^2}{2y x_\text{wt}}\right) + x_\text{m}^2 \cos^{-1}\left(\frac{y^2 + x_\text{m}^2 - x_\text{wt}^2}{2y x_\text{m}}\right) \\ &- \frac{1}{2}\big((-y + x_\text{wt} + x_\text{m})(y + x_\text{wt} - x_\text{m})(y - x_\text{wt} + x_\text{m})(y + x_\text{wt} + x_\text{m})\big)^{1/2}. \end{aligned} \quad [31]$$

is the intersecting area of two circles with radii $x_{\rm wt}$, $x_{\rm m}$ at distance y that we took from ref. (2). The formulas for the times τ_1 and τ_2 remain the same to the one- and two-dimensional case that are $\tau_1 = \frac{x-y}{c_{\rm m}-c_{\rm wt}}$ and $\tau_2 = \frac{x+y}{c_{\rm m}-c_{\rm wt}}$. Again, we replace $x_{\rm wt} = x + c_{\rm wt}\tau$ and $x_{\rm m} = c_{\rm m}\tau$.

Applying claim 1 on the remaining wildtype population, we obtain the conditional sweep probability

$$\Pr(\text{sweep}|X=x, Y=y) = \int_0^\infty \Pr(\text{sweep}|X=x, Y=y), \tag{32}$$

which can be solved numerically. The unconditional sweep probability is then given by the integral

$$\begin{aligned}\Pr(\text{sweep}) &= \int_0^\infty \int_0^\infty \Pr(\text{sweep}|X=x, Y=y) f_Y(y|X=x) f_X(x)\, dy\, dx,\\ &= \int_0^\infty \int_0^x e^{-\mu \int_0^\infty N_{\rm wt}(\tau) d\tau} \frac{2y}{x^2} \frac{3x^2 e^{-x^2/\theta_{\rm 2D}^3}}{\theta_{\rm 2D}^3}\, dy\, dx,\end{aligned} \tag{33}$$

which we solved numerically.

C. Independence of the mutation rate. We apply lemma 1 on the integral form of Pr(sweep) by defining $\nu = \mu$ and

$$\begin{aligned}H(x,y) &= \int_0^\infty N_{\rm wt}(\tau) d\tau + x^3 \frac{3c_{\rm wt}}{\pi},\\ P(x,y) &= 2y \times 3x^2,\\ Q(x,y) &= x^2 \frac{\pi}{3\text{cwt}},\end{aligned} \tag{34}$$

which are homogeneous and have degrees $h = 3$, $p = 3$ and $q = 2$. It follows that

$$\begin{aligned}\Pr(\text{sweep}) &= \mu \times \int_0^\infty \int_0^x e^{-\mu H(x,y)} \frac{P(x,y)}{Q(x,y)}\, dy\, dx\\ &\propto \mu \times \mu^{\frac{2-3-2}{3}} = \mu^0,\end{aligned} \tag{35}$$

which proves independence of he mutation rate.

D. Approximate sweep probability. We use the approximation explained in the main text and set $y = 0$, which translates into a probability density $f_Y(y|X=x) = \delta(y)$. Under this assumption, the conditional sweep probability is given by

$$\Pr(\text{sweep}|X=x) = \Pr(\text{sweep}|X=x, Y=0) = e^{-\left(\frac{x}{\alpha_{\rm 2D}}\right)^3}, \quad \text{with} \quad \alpha_{\rm 2D} = \sqrt[3]{\frac{3(c_{\rm m}-c_{\rm wt})^2}{\pi\mu(2c_{\rm m}-c_{\rm wt})}} \tag{36}$$

and the unconditional sweep probability becomes

$$\Pr(\text{sweep}) = \left(\frac{c_{\rm m}-c_{\rm wt}}{c_{\rm m}}\right)^2 \tag{37}$$

Using Bayes theorem, we compute the probability distribution of X given a sweep has occurred,

$$f_X(X=x|\text{sweep}) = \frac{3x^2}{\theta_{\rm 2D}^3 \beta^2} e^{-\frac{x^3}{\theta_{\rm 2D}^3 \beta^2}} \tag{38}$$

where $\beta = \frac{c_{\rm m}-c_{\rm wt}}{c_{\rm m}}$. The distribution of X is still described by a Weibull distribution with shape parameter 3. The difference is only a scaling factor of the characteristic length $\theta_{\rm 2D} \to \theta_{\rm 2D}\beta^{\frac{2}{3}}$.

7. Sweep probabilities in other growth models

We focused on constant radial expansion speeds for multiple reasons explained in the main text. However, to obtain more generality we also computed sweep probabilities in other growth models. Here, we consider (i) constant populations in which the mutant propagates with constant speed (ii) radial growth of the wildtype and the mutant where proliferation is restricted the population boundary and (iii) exponential growth. We briefly discuss (iv) sigmoidal growth leading to Lotka-Volterra type of dynamics.

A. Constant populations. We consider a constant population of spherical form with radius x_0 leading to total population size $N_0 = \frac{4}{3}\pi x_0^3$. The probability distribution of the wildtype radius X at the time the fist mutant arises can be formally written as the delta function,

$$f_X(x) = \delta(x - x_0). \tag{39}$$

The probability density for the distance Y between the first mutant and the centre of the wildtype population conditioned on $X = x$ is the same as for expanding populations. However, we can replace the random variable $X = x$ by the fixed radius x_0 such that

$$f(y) = f(y|X = x_0) = \frac{3y^2}{x_0^3}\mathbf{1}\{y < x_0\}. \tag{40}$$

Next, we calculate the sweep probability conditioned on X and Y. We denote the remaining wildtype population once a mutant arose by $N_{\mathrm{wt}}(\tau)$ and analogously to the previous calculation introduce the time measure τ that starts with the emergence of the first mutant. We have

$$N_{\mathrm{wt}} = (\tilde{N}_{\mathrm{wt}}(\tau) - \Delta_1(\tau))\mathbf{1}\{[0, \tau_1]\} + (\tilde{N}_{\mathrm{wt}}(\tau) - \Delta_2(\tau))\mathbf{1}\{[\tau_1, \tau_2]\} \tag{41}$$

where the terms are identical to those described in section 3 except for constant wildtype radius x_0 such that

$$\begin{aligned} \tilde{N}_{\mathrm{wt}} &= \frac{4}{3}\pi x_0^3 \\ \Delta_1(\tau) &= \frac{4}{3}\pi x_{\mathrm{m}}^3 \\ \Delta_2(\tau) &= \frac{\pi}{12y}(x_0 + x_{\mathrm{m}} + y)^2(y^2 + 2yx_{\mathrm{m}} - 3x_{\mathrm{m}}^2 + 2yx_0 + 6x_0x_{\mathrm{m}} - 3x_0^2) \end{aligned} \tag{42}$$

where $x_{\mathrm{m}} = x_{\mathrm{m}}(\tau) = c_{\mathrm{m}}\tau$. Using claim 1, the conditional sweep probability is computed by

$$\Pr(\text{sweep}|Y = y) = e^{-\mu\int_0^\infty N_{\mathrm{wt}}(\tau)\,d\tau} \quad \text{with} \quad \int_0^\infty N_{\mathrm{wt}}(\tau)\,d\tau = \frac{\pi\left(45x_0^4 - 210x_0^2y^2 - 23y^4\right)}{45c_{\mathrm{m}}}. \tag{43}$$

Integrating the conditional sweep probability together with $f_Y(y)$ does not yield a closed-form expression.

For analytical insight, let us assume again that the mutant originates in the centre of the wildtype population, $f_Y(y) = \delta(y)$. We then obtain

$$\Pr(\text{sweep}) = \Pr(\text{sweep}|Y = 0) = e^{-\frac{\mu\pi x_0^4}{c_m}}. \tag{44}$$

Because the timing of the first mutation plays no role in constant populations, our reasoning for the independence of the mutation rate no longer holds, and we obtain a sweep probability dependent on the mutation rate. We compared this sweep probability for different x_0 to sweep probabilities in expanding populations (Figure 4 and S5).

Performing the same analysis in one and two dimensions, we obtain

$$\begin{aligned} \Pr(\text{sweep}) &= e^{-\frac{\mu x_0^2}{2c_{\mathrm{m}}}} \quad \text{in 1D} \\ \Pr(\text{sweep}) &= e^{-\frac{2\mu\pi x_0^3}{3c_{\mathrm{m}}}} \quad \text{in 2D} \end{aligned} \tag{45}$$

Comparison with Ralph & Coop: Motivated by parallel adaptation on the species level, Ralph & Coop investigated selective sweeps in constant populations using a similar approach (3). However, whereas we assume spherical wildtype populations, Ralph & Coop consider a general shapes and use scaling arguments. To derive a concrete expression, they ignore boundary effects and obtain "the expected number of other mutations to arise in an area of diameter a in the time it takes the wave to cover that area". Simplifying eqn. (4) in ref. (3) and keeping their notation, we have

$$\frac{2a^3\lambda}{v} \quad \text{in 1D} \quad \text{and} \quad \frac{\pi a^3\lambda}{v} \quad \text{in 2D}. \tag{46}$$

The number of mutations arising in an area over time is a Poisson process (claim 1). Thus, we can interpret this number as mean of a Poisson distribution. A selective sweep has occurred if no other mutation arose and thus the sweep probability reads

$$\begin{aligned} \Pr(\text{sweep}) &= P_0 = e^{-\frac{2a^2\lambda}{v}} \quad \text{in 1D}, \\ \Pr(\text{sweep}) &= P_0 = e^{-\frac{\pi a^3\lambda}{v}} \quad \text{in 2D}. \end{aligned} \tag{47}$$

We identify v as the speed of the mutant c_{m}, λ as the local mutation rate conditioned on survival $\mu = \tilde{\mu}\rho$ and a as the maximal travelled distance by the first mutant, which is x_0 in the case we set $y = 0$. Finally, we find the solution of Ralph & Coop (eqn. 47) to be identical to our solution (eqn. 45) up to multiplication by a constant that can be explained by differing boundary conditions. The conclusion in either case is that full selective sweeps are highly unlikely for sufficiently large population radius x_0.

Comparison with Martens et al.: Martens and colleagues investigated the likelihood of clonal interference and its impact on the speed of evolution in spatially structured constant-size populations (4), and then studied the implications for understanding cancer initiation (5). Two modes of evolution are considered: (i) acquisition via subsequent selective sweeps and (ii) acquisition with parallel arising mutations which interfere with each other. To distinguish between these two modes, the authors compare the timescale for a surviving mutation to occur, t_{mut}, and the timescale for a mutation to sweep through the entire constant wildtype population, t_{fix}. Equating these two timescales leads to a critical length of the wildtype population:

$$\begin{aligned} L_c &= \left(\frac{c_0}{2s_0\mu}\right)^{\frac{1}{2}} \quad \text{in 1D} \\ L_c &= \left(\frac{c_0}{2s_0\mu}\right)^{\frac{1}{3}} \quad \text{in 2D.} \end{aligned} \qquad [48]$$

The authors argue that clonal interference is very likely in populations of size $L \gg L_c$, whereas we should expect selective sweeps when $L \ll L_c$. To compare this result with our model, we put L_c into eqn. 45. Therefore, we match parameters by $c_0 \leftrightarrow c_{\mathrm{m}}$, $2s\mu \leftrightarrow \rho\tilde{\mu} = \mu$ and $x_0 \leftrightarrow L$ leading to

$$\begin{aligned} \Pr(\text{sweep}) &= e^{-\frac{x_0^2}{2L_c^2}} \quad \text{in 1D} \\ \Pr(\text{sweep}) &= e^{-\frac{\pi x_0^3}{3L_c^3}} \quad \text{in 2D.} \end{aligned} \qquad [49]$$

Indeed, we have $\Pr(\text{sweep}) \to 0$ for $x_0 \gg L_c$ and $\Pr(\text{sweep}) \to 1$ for $x_0 \ll L_c$. We conclude that the result of Martens et al. are in agreement with our result for constant population sizes.

B. Exponential growth. We consider an exponentially growing population, $N(t) = e^{rt}$. Applying claim 1, the probability that no mutants occur until time t is given by

$$P_0 = e^{-\mu \int_0^t e^{rt'}\, dt'} = e^{-\frac{\mu}{r}(e^{rt}-1)}. \qquad [50]$$

Dropping the -1, this equation coincides with the solution for the stochastic birth-death process obtained by Durrett (eqn. [25] in Ref. (6)) up to the constant V_0 that is caused by fluctuations in early growth history and is neglected within our deterministic growth assumption. We proceed to calculate the probability density of T by

$$f(t) = \frac{d(1-P_0}{dt} = \mu e^{rt} e^{-\frac{\mu}{r}e^{rt}}, \qquad [51]$$

Instead of asking for the radius, it is more sensible to ask for the population size of the wildtype at the arrival of the first mutant. We write $N_x = e^{rt}$, such that $t = \frac{1}{r}\ln(N_x)$. After substitution, we have

$$f_{N_x}(N_x) = \frac{\mu}{r} e^{-\frac{\mu}{r}N_x}, \qquad [52]$$

which is an exponential distribution. The probability for the radius X can then be calculated by assuming spherical growth starting from one cell, $N_x = \frac{4}{3}\pi x + 1$. For mathematical convenience, we neglect the $+1$ term. The probability distribution of X reads

$$f_X(x) = \frac{3x^2}{\theta_{\mathrm{exp}}^3} e^{-\frac{x^3}{\theta_{\mathrm{exp}}^3}} \quad \text{with} \quad \theta_{\mathrm{exp}} = \sqrt[3]{\frac{3r}{4\pi\mu}}. \qquad [53]$$

Noteworthy, exponential growth does not come with an underlying spatial model and migration of the mutant is unclear. Nevertheless, no such assumptions are required to compute the location at which the first mutant arises. The conditional probability distribution of Y remains the same to our main model that is

$$f_Y(y|X=x) = \frac{3y^2}{x^3} 1\{y \le x\}. \qquad [54]$$

To obtain the unconditional probability for y, we can marginalize out X giving us

$$\begin{aligned} f_Y(y) &= \int_0^\infty f_Y(y|X=x) f_X(x)\, dx \\ &= \frac{9y^2}{\theta_{\mathrm{exp}}^3} \int_y^\infty \frac{e^{-\frac{x^3}{\theta_{\mathrm{exp}}^3}}}{x}\, dx \\ &= \frac{3y^2}{\theta_{\mathrm{exp}}^3} \mathrm{Ei}\left(\frac{x^3}{\theta_{\mathrm{exp}}^3}\right), \end{aligned} \qquad [55]$$

where $\mathrm{Ei}(z) = \int_0^\infty \frac{1}{t} e^{-t}\, dt$ is te exponential integral function.

In the exponential growth model, there is no competition such that the wildtype population will never go extinct and the mutant will never become fixed. We can however compute whether the mutant reaches a certain frequency χ before another mutant arises. We assume deterministic growth of the wildtype and the mutant and random time for the emergence of the mutant. We have

$$\chi = \frac{N_{\mathrm{m}}}{N_{\mathrm{m}} + N_{\mathrm{wt}}}. \tag{56}$$

Now, define t_1 to be the time at which the first mutant occurs and let t_2 be the time at which the mutant has grown to frequency χ. Then

$$\chi = \frac{e^{r_{\mathrm{m}} t_2}}{e^{r_{\mathrm{m}} t_2} + e^{r_{\mathrm{wt}}(t_1+t_2)}}. \tag{57}$$

Here, t_1 is a random variable that follows the probability density described by eqn. 51, and t_2 can be computed by solving the equation for χ giving us

$$t_2 = \frac{r_{\mathrm{wt}} t_1 - \ln\left(\frac{1-\chi}{\chi}\right)}{r_{\mathrm{m}} - r_{\mathrm{wt}}}. \tag{58}$$

The conditional sweep probability (or more precisely the probability for the first mutant to reach frequency χ without interference of another mutant) is given by

$$\Pr(\text{sweep}|T = t_1) = e^{-\lambda} \quad \text{with} \quad \lambda = \mu \int_0^{t_2} e^{r_{\mathrm{wt}}(t_1+\tau)}\, d\tau = \frac{\mu}{r_{\mathrm{wt}}} e^{r t_1} \left(e^{r_{\mathrm{wt}} t_2} - 1\right), \tag{59}$$

where $t_2 = t_2(t_1)$ is a function of t_1 described above. Eventually, the unconditional sweep probability is computed by

$$\begin{aligned} \Pr(\text{sweep}) &= \int_0^{\infty} \Pr(\text{sweep}|T = t_1) \times f_T(t_1)\, dt_1 \\ &= \int_0^{\infty} e^{-\frac{\mu}{r_{\mathrm{wt}}} e^{r t_1}\left(e^{r_{\mathrm{wt}} t_2(t_1)} - 1\right)} \times \mu e^{r t_1} e^{-\frac{\mu}{r} e^{r t_1}}\, dt_1 \end{aligned} \tag{60}$$

which we solved numerically and present in Figure S6. Notably, the probability is dependent on the mutation rate.

C. Boundary growth without proliferation in the interior. Antal and colleagues studied aspects of the evolutionary dynamics in boundary-driven growth where cell proliferation is restricted to the boundary only (7). More properties of this system have further been studied extensively (8). Whereas we assume turnover of the entire population, Antal et al. consider turnover only at the boundary and focus on the three-dimensional case. A full selective sweep is impossible in this model since individuals located away from the boundary neither proliferate nor die. Instead, we can consider the probability that a mutant will envelop the wildtype and thus becomes the only proliferating population. Using this interpretation, eqn. (17) in ref. (7) provides the unconditional sweep probability that is

$$\Pr \text{sweep} = \frac{9 + \beta^2}{\beta^2} \frac{2}{1 + e^{3\pi/\beta}}, \tag{61}$$

with $\beta = \sqrt{v^2 - 1}$ and we can identify $v = \frac{c_{\mathrm{m}}}{c_{\mathrm{wt}}}$. The unconditional sweep probability is independent of the mutation rate just like in our model. Furthermore, Antal et al. find similar expressions for the arrival time of the first mutant $f_T(t)$ (eqn. 15 in (7)) and the conditional sweep probability $\Pr(\text{sweep}|X = x)$ (eqn. 16 in ref. (7)) as well as the size of the wildtype population when the first mutant occurs (eqn. 25 in ref. (7) provides the cumulative density function). The unconditional sweep probability in the case of boundary restricted proliferation is lower compared to the sweep probability allowing proliferation throughout the tumor including the interior (Figure 4 and S5).

D. Sigmoidal growth. Many biological systems do not grow infinitely but reach a carrying capacity. There are a number of common models described by differential equations that start with initially experiential growth that eventually converges to a carrying capacity (e.g. logistic, Gompertzian (9)). The most common sigmoidal growth model is the logistic growth. If there are at least two types of populations with different growth parameters, the logistic growth model is naturally extended to a competitive Lotka-Volterra system. Those models have long been investigated and closed-form expressions are usually not known. Furthermore, the implementation of competition can take multiple forms. A fitness advantage in Lotka-Volterra systems can be obtained by an increased growth rate, an increased carrying capacity or frequency dependent competition factors that are subject of studies in evolutionary game theory (e.g. (10)). A study of selective sweeps in sigmoidal growth models thus comes with further complications that cannot be solved with methods presented in this paper alone.

Nevertheless, until the first mutant arises, there is only one population type (and thus no complicated competition between different types). Thus, we can compute the time and location of the first surviving mutant. We consider a population that grows logistically according to $N(t) = \frac{K}{1+Ke^{-rt}}$. We apply claim 1, and obtain the probability no mutation occurs until time t,

$$P_0 = e^{-\frac{\mu K}{r} \ln\left(\frac{e^{rt}+K}{K+1}\right)}. \tag{62}$$

The probability density is then obtained taking the derivative,

$$f_T(t) = \frac{d(1-P_0)}{dt} = \frac{\mu K (K+1)^{\frac{\mu K}{r}} e^{rt}}{(e^{rt}+K)^{\frac{\mu K}{r}+1}}. \quad [63]$$

The probability density for the population size N_x at the time the first mutant arises can be obtained by substituting $N_x = \frac{K}{1+Ke^{-rt}}$ yielding

$$f_{N_x}(N_x) = \frac{\mu\,(K+1)^{\frac{\mu K}{r}}\,(K-N_x)^{\frac{\mu K}{r}-1}}{rK^{\frac{2\mu K}{r}-1}}. \quad [64]$$

Assuming spherical growth, $N_x = \frac{4}{3}\pi x^3$, the probability density for the radius X can be obtained by substituting $N_x = \frac{4}{3}\pi x^3$. We obtain

$$f_X(x) = \frac{\mu\,(K+1)^{\frac{\mu K}{r}}\left(K - \frac{4\pi x^3}{3}\right)^{\frac{\mu K}{r}-1} 4\pi x^2}{rK^{\frac{2\mu K}{r}-1}}. \quad [65]$$

The distance Y of the first mutant from the wildtype's origin conditioned on the radius $X = x$ remains the same as before,

$$f_Y(y|X=x) = \frac{3y^2}{x^3} 1\{y \le x\}. \quad [66]$$

To obtain the unconditional density for Y, we can marginalize out X,

$$f_Y(y) = \int_0^\infty f_Y(y|X=x) f_X(x)\, dx. \quad [67]$$

which gives us

$$\frac{12\pi y^2 \mu (K+1)^{\frac{\mu K}{r}}}{rK^{\frac{2\mu K}{r}-1}} \int_y^\infty \frac{\left(K - \frac{4\pi x^3}{3}\right)^{\frac{\mu K}{r}-1}}{x}\, dx. \quad [68]$$

The integral can be evaluated numerically.

E. Sweep probability at 50% . To compare the models in the space of parameters (μ, c_{m}), we can compute the tuples (μ, c_{m}) at which the sweep probability is 50%. The tuples can be expressed as curve $c_{\mathrm{m}}(\mu)$ for which $\Pr(\text{sweep}) = 0.5$ leading to the following expressions for the three-dimensional expansions:

- Growth throughout (eqn.[14] in the main text): $c_{\mathrm{m}}(\mu) \approx 6.3\, c_{\mathrm{wt}}$
- Boundary proliferation (eqn. 61): $c_{\mathrm{m}}(\mu) \approx 7.2\, c_{\mathrm{wt}}$
- constant population (eqn. 44): $c_{\mathrm{m}}(\mu) = \frac{\mu \pi x_0^4}{\ln(2)}$.

We solved the first two expressions numerically. The plots are shown in the main text (Figure 4).

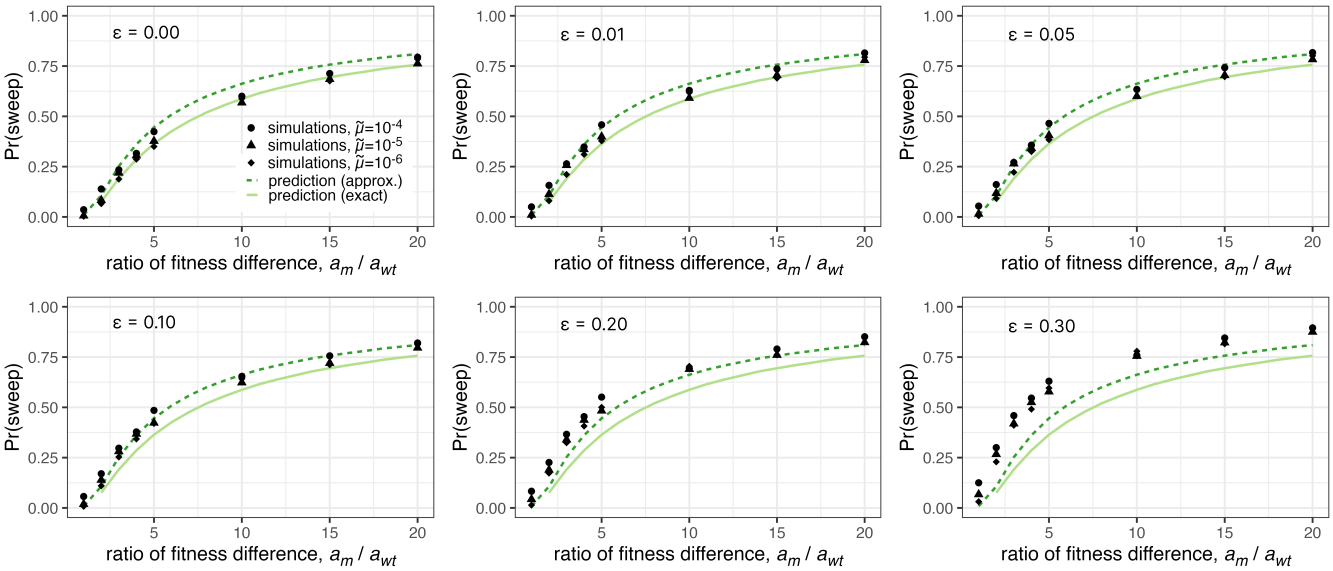

Fig. S1. Simulation results of selective sweep frequencies with different cutoffs in 2D. A selective sweep might be incomplete because of other minority populations. Here, we show simulation results using a weakened definition of a selective sweep. A selective sweep is said to be successful if the mutant's frequency reaches $1-\epsilon$. The case $\epsilon = 0.00$ corresponds to a complete selective sweep as shown and discussed in the main text. Parameters were set to $m = 0.05$, $K = 16$, $\tilde{\mu} = 10^{-4}, 10^{-5}, 10^{-6}$, $r_{\text{re}} = 0.91$, $r_{\text{wt}} = 1.0$. We vary over r_{m} between 1.1 and 3.0 corresponding to different a_{m}. Conversion between simulation and macroscopic model parameters is described in the Materials and Methods.

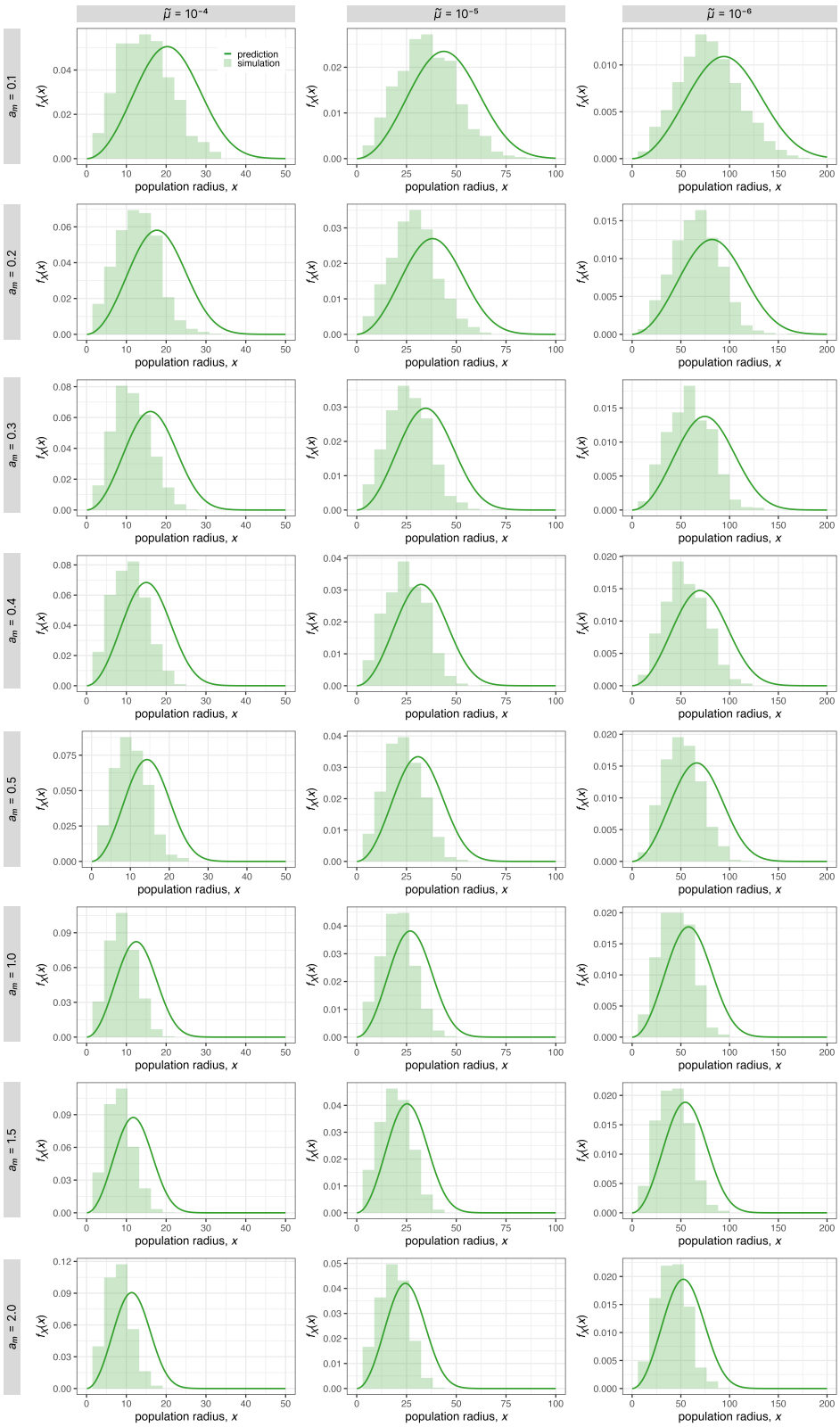

Fig. S2. Comparison of simulation results with our analytic predictions of $f_X(x)$ in 2D for different parameter combinations. We use $m = 0.05$, $K = 16$, $r_{\text{re}} = 0.91$, $r_{\text{wt}} = 1.00$. We vary over the mutation rate μ and r_{m} corresponding to different a_m. Conversion between simulation and macroscopic model parameters is described in the Materials and Methods.

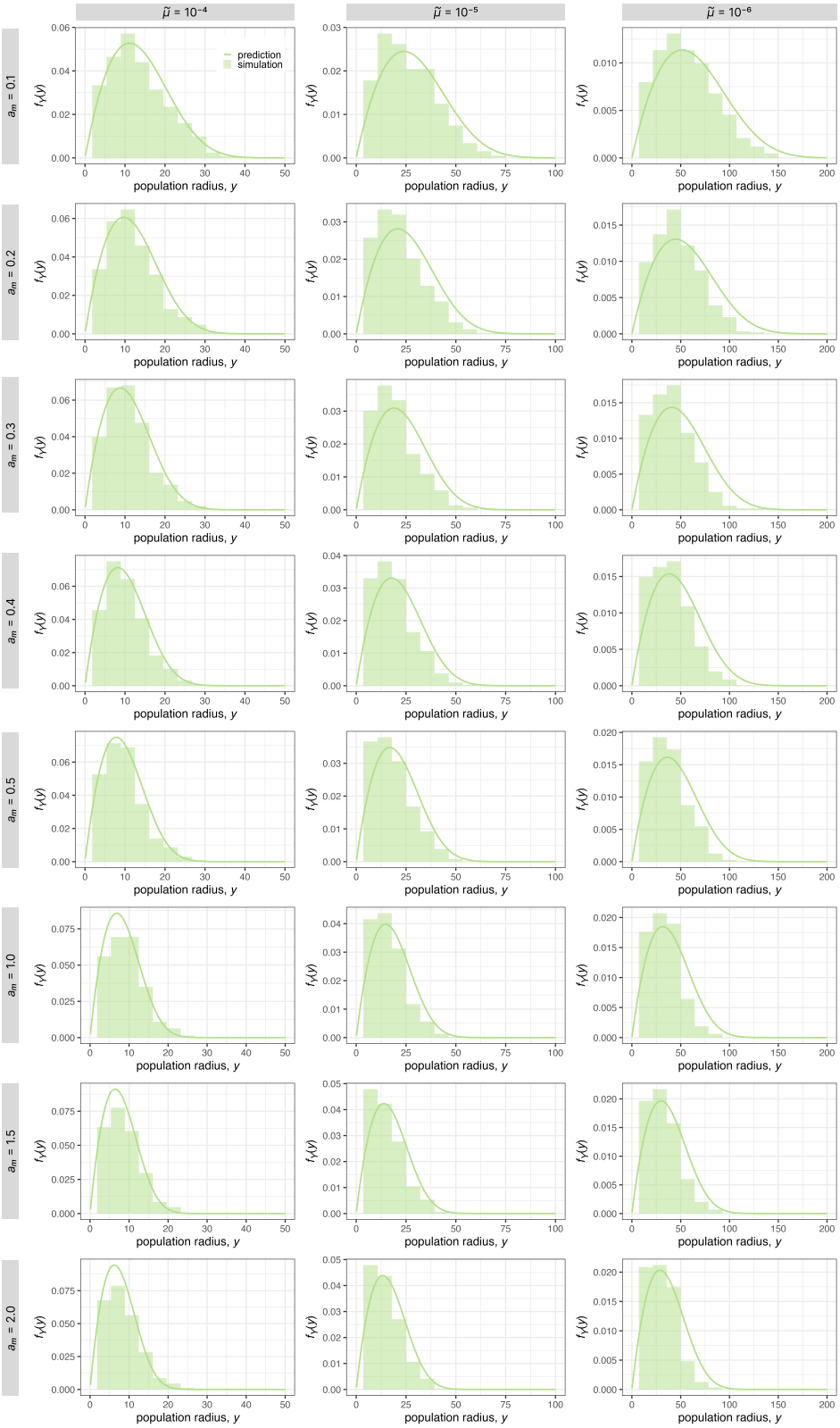

Fig. S3. Comparison of simulation results with our analytic predictions of f_Y in 2D for different parameter combinations. We use $m = 0.05$, $K = 16$, $r_{\text{re}} = 0.91$, $r_{\text{wt}} = 1.00$. We vary over μ and r_{m} corresponding to different a_{m}. Conversion between simulation and macroscopic model parameters is described in the Materials and Methods.

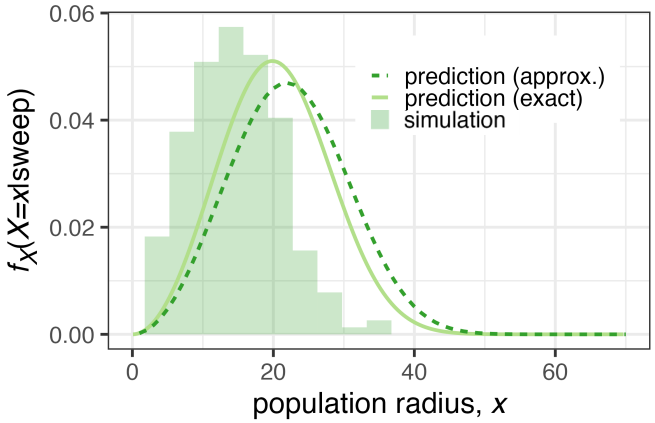

Fig. S4. Comparison of simulation results with our analytic predictions of $f_X(x|\text{sweep})$ in 2D. We use $m = 0.05$, $K = 16$, $r_{\text{re}} = 0.91$, $r_{\text{wt}} = 1.00$, $r_{\text{m}} = 1.3$ and $\tilde{\mu} = 10^{-5}$ leading to speeds $c_{\text{wt}} = 0.15$ and $c_{\text{m}} = 0.31$ and survival probability $\rho = 0.23$ (see Materials and Methods).

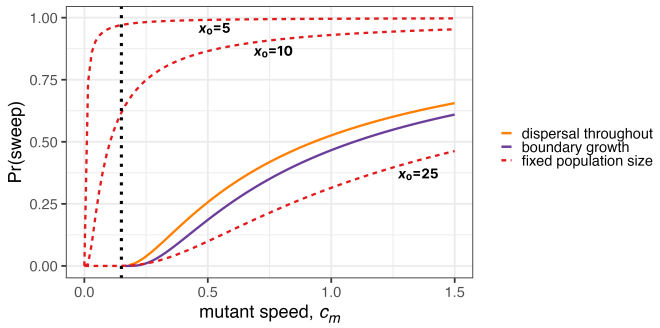

Fig. S5. Comparison of sweep probabilities between different growth models in 3D. We show the sweep probability over the mutant speed for models with proliferation restricted to the boundary (solid purple line), population size fixed at radius x_0 (dashed red lines), and our main model in which mutants can expand within the expanding wildtype population (solid orange line). In the two range expansion models, the wildtype speed is set to $c_{\text{wt}} = 0.15$ (dotted black line).

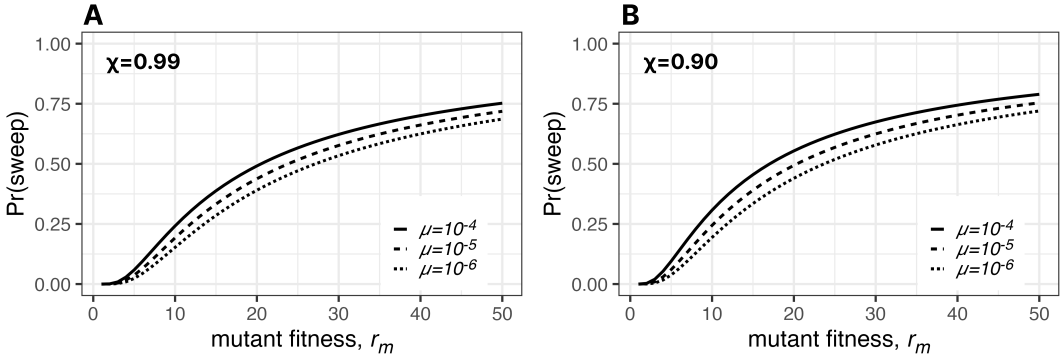

Fig. S6. Sweep probability in exponentially growing populations. The wildtype growth rate is set to $r_{\text{wt}} = 1.0$ and we vary over the mutation rate μ that is conditioned on survival of drift. We say a sweep has occurred once the mutant has reached frequency χ, which we set to $\chi = 0.99$ in **A** and to $\chi = 0.90$ in **B**.

	1D	2D	3D
$f_X(x)$	$\frac{2x}{\theta_{1D}^2} e^{-\frac{x^2}{\theta_{1D}^2}}$	$\frac{3x^2}{\theta_{2D}^3} e^{-\frac{x^3}{\theta_{2D}^3}}$	$\frac{4x^3}{\theta^4} e^{-\frac{x^4}{\theta^4}}$
$f_Y(y\|X=x)$	$\frac{1}{x} 1\{y \leq x\}$	$\frac{2y}{x^2} 1\{y \leq x\}$	$\frac{3y^2}{x^3} 1\{y \leq x\}$
$f_Y(y)$	$\frac{1}{\theta_{1D}} \Gamma\left(\frac{1}{2}, \frac{y^2}{\theta_{1D}^2}\right)$	$\frac{2y}{\theta_{2D}^2} \Gamma\left(\frac{1}{3}, \frac{y^3}{\theta_{2D}^3}\right)$	$\frac{3y^2}{\theta^3} \Gamma\left(\frac{1}{4}, \frac{y^4}{\theta^4}\right)$
Pr(sweep$\|X=x, Y=y$)	$e^{-\frac{x^2+y^2}{\alpha_{1D}^2}}$	numerical evaluation	$e^{-\mu\left(Ax^4+By^4+Cx^2y^2\right)}$
$\Pr$(sweep$\|X=x$)	$\frac{\sqrt{\pi}\alpha_{1D}}{2x} e^{-\left(\frac{x}{\alpha_{1D}}\right)^2}$ erf $\left(\frac{x}{\alpha_{1D}}\right)$	numerical evaluation	numerical evaluation
Pr(sweep$\|X=x, Y=0$)	$e^{-\left(\frac{x}{\alpha_{1D}}\right)^2}$	$e^{-\left(\frac{x}{\alpha_{2D}}\right)^3}$	$e^{-\left(\frac{x}{\alpha}\right)^4}$
Pr(sweep) [approx.]	$\frac{c_{\mathsf{m}}-c_{\mathsf{wt}}}{c_{\mathsf{m}}}$	$\left(\frac{c_{\mathsf{m}}-c_{\mathsf{wt}}}{c_{\mathsf{m}}}\right)^2$	$\left(\frac{c_{\mathsf{m}}-c_{\mathsf{wt}}}{c_{\mathsf{m}}}\right)^3$
Pr(sweep) [exact]	$\frac{\beta'}{\sqrt{1+\beta'}} \cot^{-1}\left(\sqrt{1+\beta'}\right)$	numerical evaluation	numerical evaluation
$f_X(X=x\|$sweep) [approx.]	$\frac{2x}{\theta_{1D}^2\beta} e^{-\frac{x^2}{\theta_{1D}^2\beta}}$	$\frac{3x^2}{\theta_{2D}^3\beta^2} e^{-\frac{x^3}{\theta_{2D}^3\beta^2}}$	$\frac{4x^3}{\theta^4\beta^3} e^{-\frac{x^4}{\theta^4\beta^3}}$
$f_X(X=x\|$sweep) [exact]	$\frac{\Pr(\text{sweep}\|X=x) f_X(x)}{\Pr(\text{sweep})}$	numerical evaluation	numerical evaluation

Table S1. Summary of analytical results of our main model for 1D, 2D and 3D.

Model	Sweep probability	Reference
boundary growth and cell proliferation throughout the population	$\int_0^\infty \int_0^x e^{-\mu \int_0^\infty N_{\mathsf{wt}}(\tau)\, d\tau} \frac{3y^2}{x^3} \frac{4x^3 e^{-x^4/\theta^4}}{\theta^4}\, dy\, dx$ $< \left(\frac{c_{\mathsf{m}} - c_{\mathsf{wt}}}{c_{\mathsf{m}}}\right)^3$	here
boundary growth and cell proliferation restricted to the boundary	$\frac{8+\frac{c_{\mathsf{m}}}{c_{\mathsf{wt}}}}{\frac{c_{\mathsf{m}}}{c_{\mathsf{wt}}}-1} \frac{2}{1+\exp\left(3\pi/(\frac{c_{\mathsf{m}}}{c_{\mathsf{wt}}}-1)\right)}$	Ref. (7)
constant population	$e^{-\frac{\mu \pi x_0^4}{c_{\mathsf{m}}}}$	Ref. (3, 4) and here
exponential growth	$\int_0^\infty e^{-\frac{\mu}{r_{\mathsf{wt}}} e^{rt_1}\left(e^{r_{\mathsf{wt}} t_2(t_1)}-1\right)}$ $\times \mu e^{rt_1} e^{-\frac{\mu}{r} e^{rt_1}}\, dt_1$	here

Table S2. Summary of sweep probabilities for different growth models in 3D.

References

1. EW Weisstein, Sphere-sphere intersection. from mathworld–a wolfram web resource. (https://mathworld.wolfram.com/Sphere-SphereIntersection.html) (year?) Accessed: 22/11/2023.
2. EW Weisstein, Circle-circle intersection. from mathworld–a wolfram web resource. (https://mathworld.wolfram.com/Circle-CircleIntersection.html) (year?) Accessed: 22/11/2023.
3. P Ralph, G Coop, Parallel adaptation: one or many waves of advance of an advantageous allele? *Genetics* **186**, 647–668 (2010).
4. EA Martens, O Hallatschek, Interfering waves of adaptation promote spatial mixing. *Genetics* **189**, 1045–1060 (2011).
5. EA Martens, R Kostadinov, CC Maley, O Hallatschek, Spatial structure increases the waiting time for cancer. *New journal physics* **13**, 115014 (2011).
6. R Durrett, *Branching process models of cancer*, Mathematical Biosciences Institute Lecture Series. (Springer), (2015).
7. T Antal, P Krapivsky, M Nowak, Spatial evolution of tumors with successive driver mutations. *Phys. Rev. E* **92**, 022705 (2015).
8. KS Korolev, et al., Selective sweeps in growing microbial colonies. *Phys. biology* **9**, 026008 (2012).
9. S Benzekry, et al., Classical mathematical models for description and prediction of experimental tumor growth. *PLoS computational biology* **10**, e1003800 (2014).
10. W Huang, C Hauert, A Traulsen, Stochastic game dynamics under demographic fluctuations. *Proc. Natl. Acad. Sci.* **112**, 9064–9069 (2015).