Mutant scaling laws reveal that accelerated mutant evolution via gene amplifications requires spatially structured population growth

Natalia L. Komarova[1], Justin Pritchard,[2] and Dominik Wodarz[3]

[1] Department of Mathematics, University of California San Diego, La Jolla, CA 92093
[2] Department of Biomedical Engineering, the Pennsylvania State University, State College, PA 16801
[3] Department of Department of Ecology, Behavior & Evolution, University of California San Diego, La Jolla, CA 92093

Abstract

Evolutionary processes in growing, well-mixed populations have been thoroughly investigated both experimentally and mathematically. More recently, insights have been gained into mutant evolution in expanding, spatially structured populations, using agent-based models and related simulation approaches. These become computationally challenging at high population sizes, leaving gaps in our understanding of evolutionary processes in microbes and cancer. To address this issue, we derive laws of mutant growth in two- and three-dimensional, expanding cell populations, including multi-step mutant accumulation. These laws, which include both neutral and advantageous mutations, can be used to readily predict mutant abundance in large populations under various assumptions. We apply this approach to ask under what conditions intermediate cell types with an elevated mutation rate can substantially contribute to the accumulation of point mutations. In particular, we focus on gene duplications/amplifications that can increase mutation rate due to increased copy number. We find that cells with elevated mutation rates contribute most in a 2D setting, with substantial contributions also occurring during 3D growth. In contrast to that, for well-mixed, exponentially growing populations, intermediate types cannot accelerate mutant production under biologically realistic assumptions. These results can reconcile contradicting experimental evolution studies.

1. Introduction

The accumulation of mutations in expanding cell populations determines the ability of the cells to adapt to changing environments, such as during drug therapy of bacterial infections or tumor cells, or when encountering naturally occurring selection barriers. An extensive theoretical understanding exists about the laws of mutant growth in exponentially expanding populations [1] [2-5] [6-8] , but an equivalent understanding for spatially structured expanding populations remains to be fully worked out. Principles of mutant evolution in spatially structured populations have been elucidated by both computational and experimental approaches [9] [10] [11] [12] [13] [14-24]. Yet, challenges exist on the computational side when large populations are considered, as is necessary in various settings, such as bacterial growth or cancer cell growth. Spatial evolutionary dynamics are typically simulated with stochastic agent-based models, which become slow and unmanageable at larger population sizes, even if more advanced algorithms are used such as the Next Reaction Method [25], Tau-Leaping methods [26] or similar approaches [27-29]. Hence, shortcuts are often made, such as assuming relatively small populations and unrealistically large mutation rates. To obtain a more complete understanding of evolutionary processes in spatially structured populations that grow to large sizes, new approaches are needed. This is not only required for a scientific understanding of evolutionary processes, but also for the purpose of more accurately predicting outcomes that depend on evolutionary processes, such as drug resistance-induced disease relapse. For well mixed systems, hybrid simulation methods[30] have been used, in which only the mutant colonies below a certain size threshold are treated stochastically, while larger sub-populations are simulated deterministically. For spatially explicit simulations, however, the development of such an approach would be highly complex and has not been worked out so far. An alternative way is to develop analytical insights, such as scaling laws, describing how the number of mutants in a growing system scales with the total population based on the system's dimensionality. A scaling law that connects the number of neutral, single-hit mutants with the colony size was derived previously [9], and the theory was also generalized to neutral and advantageous mutants in 2D and 3D systems [19]. Other types of insightful scaling laws that deal with smooth vs rough fronts, allele surfing, and Muller's rachet have been derived in [9, 10, 19, 31].

Our work builds on previous approaches [9] [19] to derive scaling laws for 2- and 3-dimensional population growth for more complex scenarios that involve multi-step evolutionary processes, where intermediate mutants with varying fitness are taken into account. We apply this methodology to predict the number of cells with a target point mutation at population size N,

and in particular investigate to what extent an intermediate cell type with an accelerated mutation rate can contribute to the acquisition of the target mutation. Such an intermediate cell type could be (i) cells with gene duplications or amplifications, where the effective mutation rate is elevated by increasing the number of mutational targets [32], or (ii) a mismatch repair (MMR) deficient cell population [33], which is characterized by a significantly increased mutation rate. The question that is often asked is whether or not these intermediate cell populations can accelerate the accumulation of a target mutation (e.g. a mutation that is adaptive in a stressful environment, a drug resistant mutation, etc). On the one hand, undergoing gene amplifications or acquiring MMR deficiency increases the mutation rate, thus facilitating target mutant production. On the other hand, these transformations comprise an additional rate limiting step in the evolutionary process. Therefore, in the absence of a quantitative theory, one cannot determine whether or not the additional steps that increase the mutation rate could significantly contribute to target mutant production.

Using the growth laws derived here, we find that intermediate cell phenotypes with an accelerated mutation rate can only substantially contribute to the generation of cells with the target point mutation, if population dynamics are characterized by spatial restrictions. This effect is most pronounced for 2D spatially restricted population growth, and is still substantial for 3D spatial population growth. Interestingly, during growth of well-mixed populations (exponential growth), an intermediate cell type with an accelerated mutation rate does not contribute to the emergence of cells with the target point mutation, unless the fitness advantage of the cells with the higher mutation rate is excessively large and/or the population size reaches biologically unrealistic levels. We use these insights to interpret experimental evolution studies that investigated the role of gene amplifications for the accumulation of point mutations in cells. Different and apparently contradicting experimental results have been reported, and our theory provides an explanation to reconcile these observations.

2. Computer simulations

We consider basic stochastic birth-death processes in expanding populations with mutations. The question under investigation is how many cells that have acquired a specific point mutation (which we will refer to as a “target mutation”) are found in an expanding population at a given size N. To address this question, an agent-based model is considered that contains the

following populations, see Fig.1(a,b). (i) Wild-type cells (Z_{wt}), which have not yet acquired any mutation. (ii) Mutant cells (Z_m), which have acquired the target mutation due to a replication error of the wild-type population. (iii) Accelerated cells (Z_a), which have either acquired a gene amplification in the gene under consideration, or a form of genetic instability, thus effectively having a higher rate at which the target mutation can be generated. (iv) Mutated accelerated cells ($Z_{am}+Z_{ma}$), which are cells with the higher mutation rate that have acquired the target mutation. Each type of cell is characterized by a rate of cell division and a rate of cell death, denoted by parameters r and d, respectively, with the appropriate subscript indicating the cell population. Upon division, wild-type cells are assumed to give rise to the target mutation (and hence to cell type Z_m) with a rate μ. They can also acquire the accelerated phenotype with a rate p, giving rise to type Z_a cells. Accelerated cells (Z_a) are assumed to produce the target mutation with an elevated rate $\mu_A=a\mu$, corresponding either to a cell with gene duplication ($a=2$) or with higher copy numbers or genetic instability (larger a), population Z_{am}. In addition, mutated cells (Z_m) can acquire the accelerated phenotype with a rate p, type Z_{ma}. At every time step, each cell has a chance to divide or die, depending on the assigned probabilities. We compare spatially explicit simulations to a well-mixed system. For spatial simulations, if a division event is chosen, a target spot is selected randomly from all the nearest neighbors (e.g. 8 neighbors for a 2D grid); the offspring is placed there if this target spot is empty, and no division occurs otherwise. If a division occurs, the offspring can be altered with probabilities μ, μ_A, or p. Non-spatial simulations are performed with Gillespie simulations of the corresponding processes in the context of exponential growth.

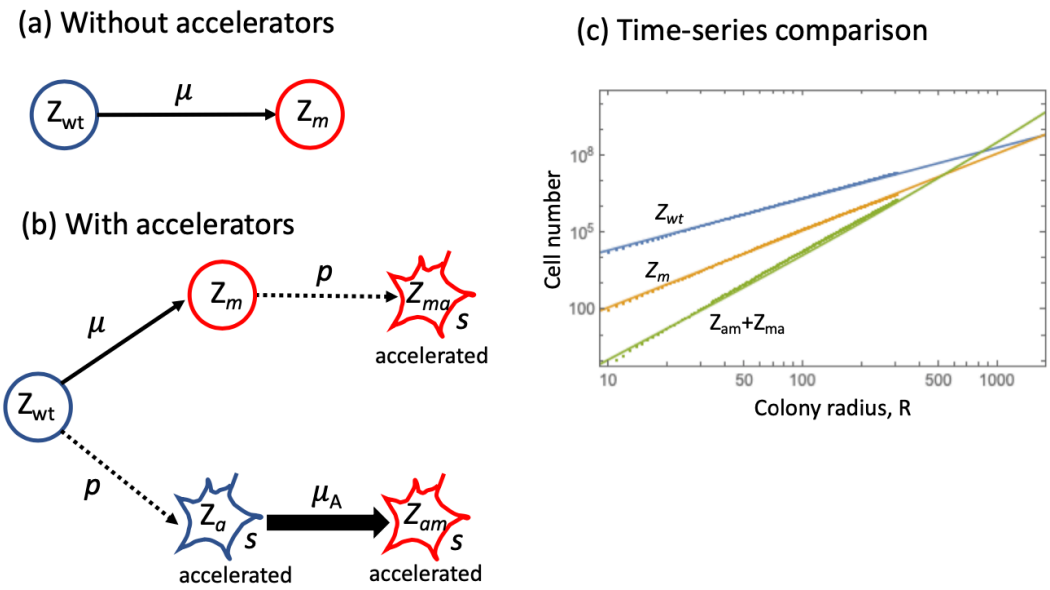

Fig.1. Comparing two pathways to a mutation. (a-b) A schematic showing the pathways leading to the generation of a mutation: (a) in the absence and (b) in the presence of the accelerated phenotype. The accelerated phenotype is denoted by a star shape, and the mutant of interest is colored red. The accelerated mutation rate is denoted by μ_A. (c) The number of mutants generated directly (Z_m, shown in orange) is plotted together with the number of double-hit mutants generated in the presence of an accelerated type (Z_{am}+Z_{ma}, green) and the number of wild type cells (Z_{wt}, blue) in an expanding 2D colony. Simulations (average of 3,612 independent simulations): dots, theory: lines. Parameters are $s=0.01$, $\mu=p=10^{-4}$, $\mu_A=2\mu$.

The simulation is started from a single wild-type cell (placed in the center of the grid in the spatial setting). As the simulation proceeds, the population expands according to surface growth (see Fig 2 for the example of a 2D surface growth). During cell divisions, mutated as well as amplified cells are generated with their corresponding rates, and these undergo range expansions that either continue to grow or become enclosed by competing cells over time. In the example shown in Figure 2, both target-mutant (Z_m) and amplified cells (Z_a) are produced by wild-type cells and undergo range expansions (teal and beige coloring, respectively). Either population can then generate mutated accelerated cells (Z_{am}+Z_{ma}, yellow), which also undergo range expansion dynamics.

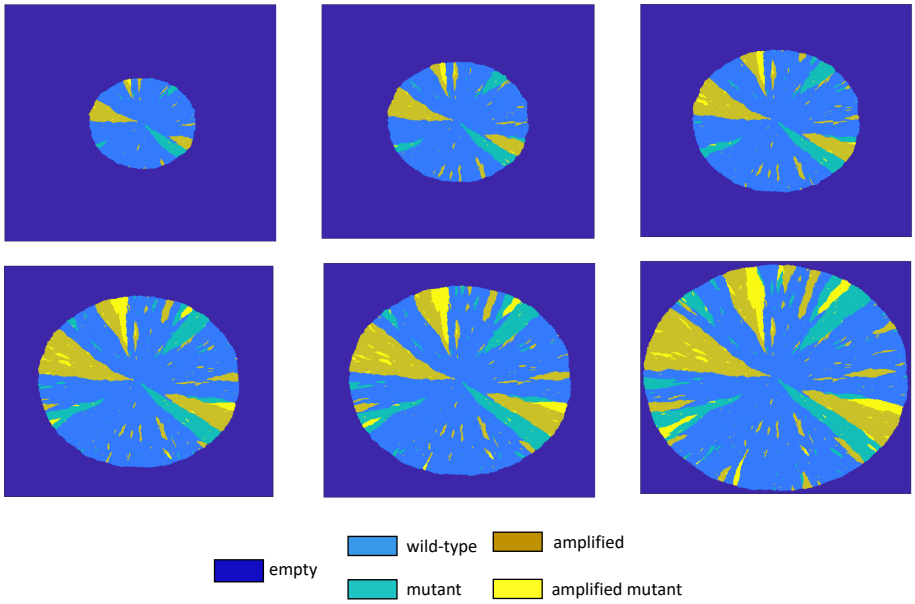

Fig.2. Spatial configuration of a 2D expanding and evolving colony of cells, a typical simulation. The individual graphs represent different snap shots in time, going from left to right top row, followed by left to right bottom row. Color codes are given below the pictures. Parameters are as follows: $r=0.985$; $d=0.0005$; $s=1$; $p=1.5x10^{-4}$; $\mu=1x10^{-4}$; $a=2$.

To obtain statistics about the number of mutants at a given population size, the simulation is run repeatedly up to a threshold size N, and the average number of target mutants at population size N is determined (i.e. the sum of populations $Z_m+Z_{am}+Z_{ma}$). Two situations are compared (see the schematic in Fig.1): simulations in which gene duplications can occur ($p>0$, panel (b)), and simulations where this process is disabled ($p=0$, panel (a)), such that the target mutation can only be generated from the wild-type cell population. For tumor or bacterial cell populations, the number of cells in growing colonies often exceeds what is computationally feasible to simulate in stochastic agent-based models. For illustration purposes, we ran computer simulations for relatively small population sizes and larger mutation rates. An example is seen in Fig 1(c), which shows the population of wild-type cells (blue) as well as cells that acquire the target mutations directly (Z_m, yellow), and the amplified cells that also harbor the target mutation (green). Stochastic simulations results (averaged over a number of independent runs), are shown by dots and show populations up to ~10^7 cells. We observe that the single-hit mutants (Z_m) seem to increase faster than the wild-type cells, and the double-hit mutants ($Z_{am}+Z_{ma}$) appear to have an even higher slope in the log-log plot. We next derive mathematical laws of mutant population growth and test them against such simulations. These derived laws

are then used to investigate the role of accelerated cells in mutant evolution under biologically realistic population sizes and mutation rates.

3. Growth laws of accelerated cells in the spatial model

Computer simulations present evidence that the existence of cells with accelerated mutation rates may contribute significantly to the evolutionary process in spatially distributed systems, to a larger degree than they do in models with exponential growth. They also clearly demonstrate the need for a more general approach, which would allow for a systematic exploration of the evolutionary dynamics under a wide range of parameters. Therefore, we turned to developing scaling laws that govern mutant growth in expanding populations.

Scaling laws for single mutations. Consider the simplest situation described by the selection-mutation diagram: $Y_0 \xrightarrow{u_1} Y_1$, where Y_0 denotes the wild-type, Y_1 denotes the population of single mutants (with relative fitness $(1+s_1)$), and the mutation rate is given by u_1. We will compare two types of spatial growth: a 2D range expansion, where a colony grows in a circular shape, and a 3D range expansion (a spherical colony). We will use variable R to characterize the colony size. Denote by R the radius of the colony, where distance is measured in terms of cell diameters. If the total number of cells is denoted by Y_0, we have $Y_0 \propto \pi R^2$ in 2D and $Y_0 \propto \frac{4}{3}\pi R^3$ in 3D; in particular, in the absence of cell death, the density of cells is one and the proportionality sign becomes "=". Assuming so-called surface growth [34], where only cell divisions on the surface contribute to expansion, the mean temporal growth of the population can be described by the following ODEs:

$$(2D) \quad \dot{Y}_0 = b_2 Y_0^{1/2}, \ (3D) \ \dot{Y}_0 = b_3 Y_0^{1/3} ,$$

where b_i with $2 \leq i \leq 3$ are constants that include information about the growth rate of cells and also depend on the dimensionality of the system. Consequently, the total number of cells grows according to the following power laws:

$$(2D) \quad Y_0 \propto t^2, \ (3D) \ Y_0 \propto t^3 ,$$

and the linear colony size always grows linearly with time, $R \propto t$.

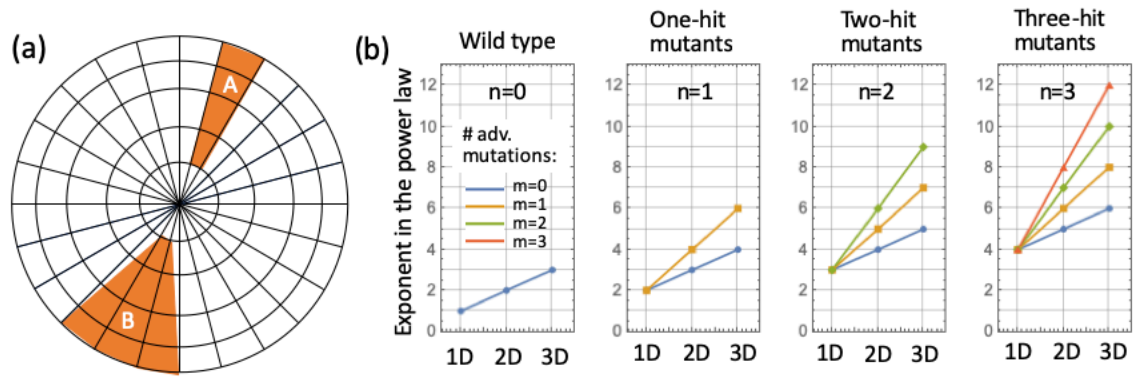

Fig.3. Scaling laws for mutant growth. (a) A schematic showing the spread of a mutant in a 2D range expansion geometry. Mutant A is neutral and its expected domain follows the radial lines. Mutant B is advantageous and its domain boundaries cross radial lines, increasing the arc angle of the mutant region. (b) The scaling laws of equation (1). Each panel shows the exponent in the power law of growth of the population of *n*-hit mutants in 1D, 2D, and 3D geometries. A cell is considered to have an advantage if it has an increase of fitness with respect to the type that generates it (not just the wild type). For the one-, two-, and three-hit mutant scenarios, all the mutants could be neutral (*m*=0, blue line), or one of them could be advantageous (*m*=1, yellow lines). If two (*m*=2) or three (*m*=3) mutations are advantageous, green and red colors are used, respectively. According to the theory, the order in which mutations occur does not matter for the scaling laws.

To include mutant generation, we will assume that mutations are relatively infrequent, such that a single mutation typically gives rise to a clone that does not merge with other clones, and the total mutant number is significantly smaller than the colony size. As a result, it is possible to study mutant clones in isolation. The schematic in figure 3(a) shows a neutral (A) and an advantageous (B) clone growing in a colony undergoing a 2D range expansion. Denote by $n^s(r_1,r_2)$ the expected size of a mutant clone with relative fitness advantage *s*, that was created at colony size r_1, and measured at size r_2. Further, denote by M_0(r) the size of the advancing front, through which the growth occurs: $M_0(r) \propto 2\pi r$ in 2D, and $M_0(r) \propto 4\pi r^2$ in 3D. Then the expected number of mutants with fitness advantage s_1, in a colony of size *R* can be approximated by: $Y_1^{s_1}(R) = \int_0^R u_1 M_0(r) n^{s_1}(r,R)dr$, where u_1 is the rate at which mutations are generated.

The expression for the clone size depends on the system dimensionality as well as the mutant fitness relative to the wild type, which is given by $(1+s_1)$ with $s_1 \geq 0$. It is derived in the supplement for dimensionalities *D=1,2,3*. As a result, we obtain a scaling law for the number of mutants in a colony of linear size *R,* with $s_1 \geq 0$:

$$Y_1^{s_1}(R) = 2^{D-1}\pi u_1 (D-1)! \sum_{k=0}^{D-1} \frac{R^{2D-k} s_1^{D-1-k}}{k!\,(D-k)!\,(2D-k)}, \qquad (1)$$

where D=2 or D=3 is the system's dimensionality. In particular, for neutral mutants (s_1=0) the only nonzero term on the right corresponds to $k=D-1$, and we have

$$(2D) \quad Y_1^0(R) = \frac{2}{3}\pi u_1 R^3, \qquad (3D) \quad Y_1^0(R) = \pi u_1 R^4$$

Table 1 shows the power of the leading term for different species in different dimensionalities. For completeness, this table also includes the case D=1, as well as disadvantageous mutants. A more detailed discussion for these laws can be found in [19]; here it suffices to say that each disadvantageous mutant is assumed to give rise to a clone of a constant size, $n^s(r_1,r_2)$=const, which leads to the results presented in Table 1.

Single-mutant scaling laws	1D	2D	3D
Wild type or disadvantageous mutant ($s_1<0$)	t^1	t^2	t^3
Neutral mutant ($s_1=0$)	t^2	t^3	t^4
Advantageous mutant ($s_1>0$)	t^2	t^4	t^6

Table 1. The leading power of time in the growth laws of different populations in expanding colonies with single mutations, in 1D, 2D, and 3D. See also Fig.3(b), $n=0$.

<u>Scaling laws for double mutations and beyond.</u> Our goal is to investigate the effect of cells with an increased mutation rate on the evolutionary dynamics. Before such cells can contribute, they have to be created from the wild-type cells, which implies the existence of an additional mutation event. Therefore, we need to develop the growth laws of double-hit mutants. Assume the following selection-mutation diagram: $Y_0 \xrightarrow{u_1} Y_1 \xrightarrow{u_2} Y_2$, where, as before, Y_0 denotes the wild-type, Y_1 denotes single mutants (with relative fitness $(1+s_1)$), and Y_2 denotes double mutants (with relative fitness $(1+s_2)$). For now we fix the order of events, and assume that type Y_2 is generated by a mutation in type Y_1. Note that here and below, for each consecutive mutation we measure the resulting fitness with respect to the type that hosts the mutation, and not with respect to the wild type. The rates of mutation are given by u_1 and u_2 for the first and the second hit, respectively.

The above theory can be extended to include double-hit mutants (see Supplement for details). We assume that double-hit mutants are generated by the single-hit mutants at the edge of

expansion. Their expected number is given by: $Y_2^{s_1,s_2}(R) = \int_0^R u_2 M_1^{s_1}(r) n^{s_2}(r,R) dr$, where $M_1^{s_1}(r)$ is the number of single mutants of fitness advantage s_1 at the edge of expansion for a colony of linear size r, and $n^{s_2}(r,R)$ is the size of double-mutant colony of fitness advantage s_2 that is generated at size r and measured at size R. We obtain the following expression for the expected number of double-mutants:

$$Y_2^{s_1,s_2}(R) = u_1\, u_2 \pi 2^{D-1}\big((D-1)!\big)^2 \sum_{k=0}^{D-1}\sum_{m=0}^{D-1} \frac{R^{3D-k-m} s_1^{D-1-k} s_2^{D-1-m}}{k!\,m!\,(2D-k-m)!\,(3D-k-m)}. \qquad (2)$$

For example, when both mutations are neutral (s_1=s_2=0), the only nonzero contribution comes from the term with k=m=D-1, and we obtain

$$(2D)\quad Y_2^{0,0}(R) = \frac{1}{4}\pi u_1 u_2 R^4, \qquad (3D)\quad Y_2^{0,0}(R) = \frac{2}{5}\pi u_1 u_2 R^5$$

In the Supplement, we provide an explicit formula for any number of consecutive mutations, some or all of which can be advantageous. There are several consequences of the scaling laws that are derived, which we list here.

1. A population of n-hit neutral mutants under D dimensional range expansion grows as R^{D+n}, where R is the whole colony radius; the power scaling with time is the same, t^{D+n}.
2. Suppose that among n sequential hits, m hits confer a nonzero advantage with respect to the previous cell type, and the rest of the mutations do not alter the fitness properties of cells that harbor those mutations. Surprisingly, the order in which these advantageous mutations are acquired does not influence the scaling law for the number of n-hit mutants, which grow as

   $$M \sim R^{D+n+m(D-1)}; \qquad (1)$$

   the power scaling with time is the same, $t^{D+n+m(D-1)}$. The exponents of the scaling law for different values of n and m are presented in figure 3(b).
3. Suppose that individuals in a population of n-hit mutants contain m mutations that confer selective advantage to cells (relative to the cell type that acquired the mutation), and denote the corresponding selection coefficient values by $s_1,\ldots,s_m$ (where they are listed in a non-decreasing order). If they all have roughly the same order of magnitude (s), then the n-hit mutant population behaves as if neutral, until the colony size reaches R~1/s, at which time it switches to the growth specified in (2). In the case where selective advantage coefficients are different, the growth law increases the power by one at sizes $1/s_m$, $1/s_{m-1}$,…, $1/s_1$.

Using this methodology, one can investigate the relative contributions of different pathways to the evolutionary processes in growing populations.

4. Using derived growth laws to obtain insights into evolutionary processes in spatially growing populations

Here we show how the analytical expressions for mutant numbers derived above can be used to gain insights into complex evolutionary processes in spatially growing populations. In particular, we ask under what conditions an intermediate population with an increased rate of mutation accumulation can enhance the presence of cells carrying the target point mutation. We consider both gene duplications / amplifications and mismatch repair-deficient cells as the intermediate cell populations.

Spatially structured growth is required for <u>gene duplications to significantly accelerate mutant production</u>. To quantify the role of cells with gene duplications for target mutant emergence, we compare the following populations (see Fig. 1): on the one hand, population Z_m in panel (a) is the total number of mutants generated in the absence of mutation acceleration; on the other hand, the sum of the populations $Z_{am}+Z_{ma}$ represents the number of mutants that are associated with the possibility of accelerating the mutational process. Here, cells Z_{am} first undergo a mutation that allows for an accelerated mutation rate, and then acquire the target mutation at a higher rate, μ_A. Cells Z_{ma} are created by first acquiring the target mutation and then developing the characteristics of accelerated cells by a second hit. Figure 1(c) compares numerical simulations with the theoretical results for these populations, where for Z_m, we used $Y_1^{s_1}(R)$ in formula (1) with $u_1=\mu$ and $s_1=s$; for Z_{ma} we used $Y_2^{s_1,s_2}(R)$ in formula (2) with $u_1=\mu$, $u_2=p$, $s_1=0$, $s_2=s$; and for Z_{am} we used $Y_2^{s_1,s_2}(R)$ with $u_1=p$, $u_2=\mu_A$, $s_1= s$, $s_2=0$. We can see that the theoretical approximation works well (see further information on validation and applicability in the Supplement).

Using the derived growth laws to predict mutant dynamics for population sizes that are too large to track by computer simulation, we can predict at what colony sizes the cells with duplicated genes make a substantial contribution to target mutant emergence, compared to the number of mutations generated directly by a single point mutation hit from the wild type. Let us define the quantity $Q=(Z_{am}+Z_{ma})/Z_m$, as the relative contribution of duplicated cells to the total number of

mutants created. This quantity has the following straightforward interpretation. Suppose that in the absence of duplicated cells ($p=0$) the number of cells that contains the target mutation in a colony of a given size is Z_m. Then in the presence of duplicated cells ($p>0$), the colony of the same size will contain $(1+Q)Z_m$ cells with the target mutation. An alternative measure of the contribution of duplicated cells is $\tilde{Q} = Z_{am}/Z_m$, which simply compares the number of cells that acquired the mutation after duplication, with the number of cells that acquired the mutation directly from the wild type. This measure is very similar, and in fact we have $\frac{1}{2}Q < \tilde{Q} < Q$, which means that it is somewhere between 50% and 100% of Q. If the contribution of duplications in a system is predicted to be significant by measure Q, it is also significant by measure $\tilde{Q}$. It if is negligible by measure Q, it is also negligible by measure $\tilde{Q}$.

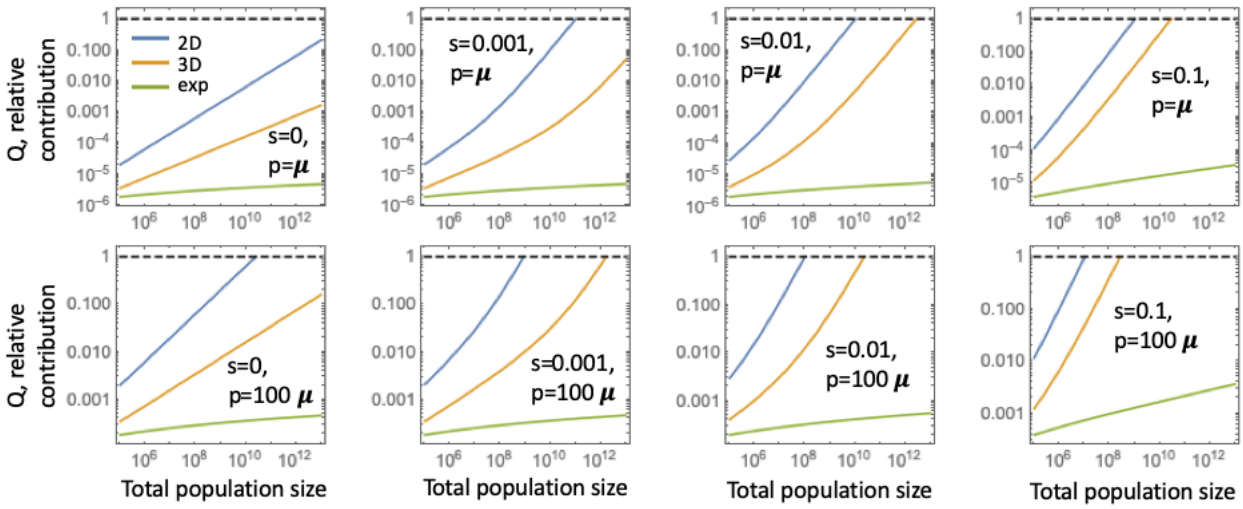

Fig.4. Gene duplication: the relative contribution of the accelerated type to the total number of mutants, $Q=(Z_{am}+Z_{ma})/Z_m$ (as predicted by the model), is plotted as a function of the total population size. The three lines in each panel correspond to the 2D (blue), 3D (orange), and mass-action (green) systems. We assume that the accelerated mutation rate is $\mu_A=2\mu$ (duplications). In the top row, the duplications are generated at rate $p=\mu$, and in the bottom row $p=100\mu$. The fitness parameter, s, increases from left to right as $s=0$, $s=0.001$, $s=0.01$, and $s=0.1$. The dashed horizontal line represents $Q=1$. $\mu_A=10^{-7}$.

Assuming that cells with the gene duplication have twice the mutation rate compared to cells without ($\mu_A=2\mu$), Figure 4 compares the behavior of spatial mutant colonies in the cases of 2D and 3D range expansion with a mass-action (exponential growth) system; the methodology for the latter case is standard and is provided in the Supplement. Plotted is the relative contribution of the duplicated cells to the total number of mutants created, $Q=(Z_{am}+Z_{ma})/Z_m$, for different colony sizes (the horizontal axes). We can see that the relative contribution increases with the colony size. This happens because the net growth rate of the double-mutants is faster; in

particular, in the spatial systems it corresponds to an increased exponent in the power law. We also observe that the relative contribution of the duplicated cells is always the largest in 2D (blue), followed by 3D (orange), and both of these are substantially larger than that for the exponentially growing colony (green).

The striking result is the vast difference in the contribution of the duplicated cells, between the spatial and non-spatial (exponentially growing) systems. In the case of the parameters used in Fig.4, the increase in the number of target mutants due to gene duplication in the non-spatial model never reaches 1%, even for extremely large populations and assuming that gene duplication happens at a rate that exceeds the base mutation rate a hundredfold and confers a 10% advantage to the cell (see the bottom right panel). In well-mixed populations, gene duplications can only contribute to the accumulation of target point mutations if the fitness of duplicated cells is very large, upwards of 40% (Figure S2). On the other hand, in spatial systems, the increase in the number of target mutants due to duplications can easily reach a 100% for much smaller population sizes and much smaller selective advantages of duplicated cells. If gene duplications occur much more often (100-fold in Fig 4, bottom) than point mutations, duplicated cells make substantial contribution to target mutant evolution even if the duplicated cells are neutral with respect to the wild-type cells in 2D growth. In the Supplement we present additional graphs where we vary the base mutation rate over a wide range, revealing the same patterns.

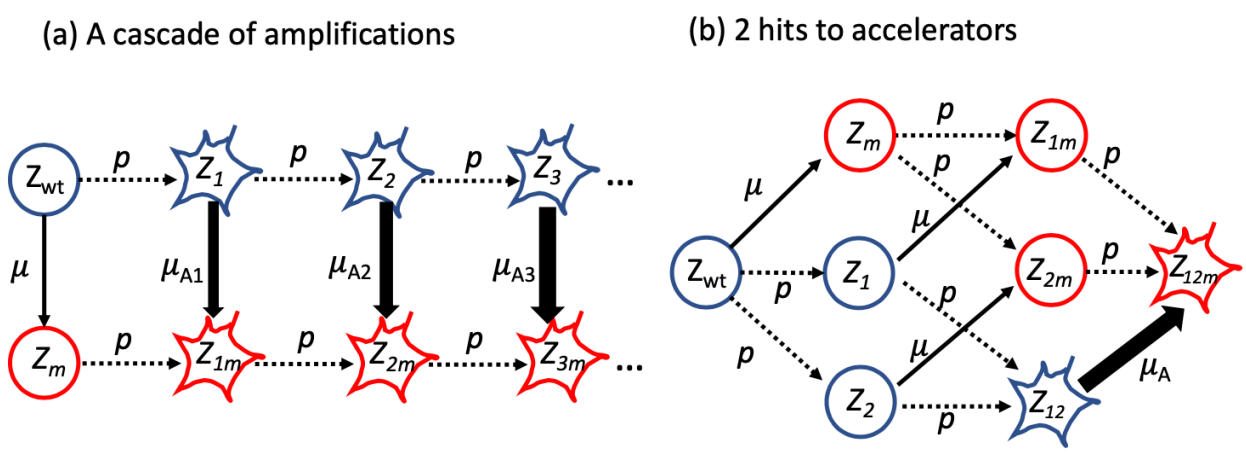

Fig.5. Other selection-mutation networks with mutation accelerations. The accelerated phenotype is denoted by a star shape, and the mutant of interest is colored red. (a) A cascade if amplifications; type Z_k with $k=1,2,\ldots$ has experienced k gene duplication events in the target gene; type Z_{km} contains cells with a target mutation that have undergone k gene duplication events in the target gene. The increasingly accelerated mutation rates are denoted by μ_{A1} , μ_{A2} , … (b) Two hits to acceleration. Types Z_1 and Z_2 have the first and the second copy of the MMR gene inactivated, respectively, while the accelerated type Z_{12} has both copies inactivated. Types Types Z_{1m} , Z_{2m} , and Z_{12m} are the same as types Z_1 and Z_2, and Z_{12} except they also contain the target mutation.

Further amplifications can significantly contribute to mutant production in spatial populations. Beyond duplications, cells can further amplify gene copy number. The schematic of figure 5(a) shows a cascade of amplifications (gene duplications), where each duplication event leads to an increase in the accelerated mutation rate. Here, type Z_1 corresponds to type Z_a of figure 1, while types Z_2, Z_3, etc represent cells that have experienced 2,3, etc consecutive duplication events. Type Z_{1m} contains cells that have undergone a single duplication event and also acquired a target mutation (corresponding to the sum $Z_{am}+Z_{ma}$ in Figure 1(b)), while types Z_{2m} , Z_{3m} , etc are types with multiple duplication events that also harbor the target mutation.

It is interesting to determine whether or not further duplication events can significantly increase the production of target mutants, beyond the type Z_{1m} . Applying the scaling laws derived here (see Supplement for details) we show that in 2D and 3D systems, further duplication events could make a difference for biologically realistic parameter ranges, if each subsequent event induces additional fitness advantage, see figure S7 of the Supplement.

Spatial population structure is required for the inactivation of an MMR gene (a two-hit event) can contribute significantly to mutant production. MMR deficiency (leading the small-scale genetic instability) is a biologically distinct intermediate cell population that can

accelerate the rate of target mutant production (Fig 5b). In contrast to gene duplications, the generation of the MMR-deficient phenotype requires two mutational events in most cases, because the mismatch repair mechanism needs to be inactivated. In this setting, type Z_{12} is a cell with both copies of an MMR gene inactivated. It can be produced from the wild type by means of two consecutive inactivation events (through type Z_1 or Z_2), and is characterized by a significantly (orders of magnitude) increased rate, at which the target mutation can be generated. Hence, although it takes more steps to generate the MMR-deficient phenotype compared to a gene duplication, the gain in the mutation rate is a lot higher. The question is whether for realistic population sizes, this mechanism can contribute to mutant production in expanding populations.

Using our scaling laws (see Supplement for details) we can show that, both in 2D and 3D systems, the contribution of MMR-deficient cells can be similar to or larger than direct target mutations. This is in contrast to the non-spatial, exponentially growing systems, where the effect of these longer pathways is negligible, see Fig. S5 of the Supplement.

5. Discussion

Stochastic computer simulations that involve evolutionary processes are computationally very ``expensive," and are difficult to perform under realistic population sizes. In the case of spatially expanding populations that produce neutral or advantageous mutants, we developed an approach that is based on newly derived scaling laws for these populations. Using these laws, we can study evolutionary processes that involve neutral and/or advantageous mutations, and quantify the role played by various processes in the evolutionary transformation experienced by populations. In particular, this methodology allows reasoning about the relative likelihood of different pathways and suggest which events may be causal, given the patterns that are observed. This can be done under realistic assumptions about the mutation rates and population sizes, which is difficult to do using traditional methods of stochastic agent-based modeling.

The scaling laws derived here predict the expected size of a mutant population created by an expanding colony of cells/organisms, both in 2D and 3D range expansion. Results include one-hit mutants, as well as mutants that are created by a longer chain of mutation events, which

may include both neutral and advantageous hits. These theoretical results were applied to several biological questions that range from gene duplication/amplification to systems that involve MMR deficiency. An emerging theme was that spatial population structure might be crucial for intermediate cell genotypes with elevated mutation rates to contribute substantially to target mutant generation. The effect was strongest for 2D spatial population structures, still substantial for 3D spatial population structures, but largely negligible for well-mixed populations.

Our work adds to an extensive and expanding literature on evolutionary dynamics in spatially structured populations. In recent years, spatial evolutionary dynamics have garnered significant mathematical interest, both computationally [35] and theoretically. Agent-based models have been used to show that the number of mutants at a given threshold size in spatially growing populations are larger than at the same size in a well-mixed system, due to the emergence of jackpot mutations [9]. Further work has shown that this pattern can be more complex for disadvantageous mutants and depend on the parameter in which the disadvantage is expressed [19]. These studies also defined mutant scaling laws for single-hit mutants, which formed the basis for our work., see also [10] [11] [12]. Other theoretical work[36] examined a spatially distributed population where mutations arose and spread to neighboring locations, calculating the time distribution required for an individual to acquire a certain number of mutations. Deterministic equations for stochastic spatial games were developed under the assumption of long-range interactions [37]. Another study compared a death-birth process on a lattice with the dynamics of the replicator equation, revealing that spatial factors significantly altered the evolutionary outcomes [38]. These studies build on a rich tradition of mathematical research into spatial evolutionary stochastic processes, utilizing both analytical approaches and partial differential equation (PDE) modeling to study trait evolution in spatial contexts [39-43]. Spatial evolutionary processes have also been investigated specifically in the context of cancer, with implications for tumor progression and expansion [22], for spatial heterogeneity and tumor sampling [44], and for the evolution and management of drug resistance [15, 45].

Modeling results help interpret experimental evolution data on the role of copy number variation. Gene amplifications are known to be important drivers of evolutionary processes [46, 47]. While amplifications by themselves can be costly to cells, they can help to increase fitness under stressful conditions. An extensively studied and debated system are Escherichia coli cells with a mutation in the lac operon (*lac*$^-$) [48-51], which renders them almost unable to grow on

lactose as the carbon and energy source (only about 2% enzyme activity compared to wild-type). When a large bacterial population is plated on lactose, populations of revertant cells (*lac*$^+$) emerge. While there has been debate about the mechanisms underlying this evolution, a picture has emerged, in which gene amplifications play a major role. It is thought that the originally plated cells contain small subpopulations with a lac operon duplication. This allows the cells to grow slowly on lactose, and subsequent further gene amplifications allow for more efficient growth. At the same time, the increased gene copy number results in a higher chance to generate a revertant mutant that allows the bacteria to grow on lactose even without the gene amplification. Since gene amplifications by themselves are costly to cells, the increased copy numbers are subsequently lost. This represents the type of evolutionary process that we studied mathematically here (except that we did not explicitly take into account loss of amplifications, and the target mutation could be any mutation in the amplified sequence, not just a mutation that functionally replaces an amplification). The same kind of patterns have also been documented for Salmonella enterica [32]. In all these experiments, the bacteria grew as lawns, which represent 2D spatial growth dynamics. According to our model, this is the population structure in which it is most likely to detect cells with point mutations that were generated in cells with gene amplifications. Similar evolutionary processes are thought to occur under other conditions of stress, including antibiotic resistance evolution, heat stress, or salt stress [52, 53].

The role of gene duplications and amplifications for the rate of mutant evolution in *E. coli* has also been investigated in experiments in which bacteria grew in liquid medium[54]. This represents a more mixed and less spatial system. A synthetic genetic reporter system in *E. coli* was used to distinguish between copy number and point mutations, and focused on the early phase of the evolutionary dynamics. The system involved an intact endogenous *galK* gene with a random promoter system. When *E. coli* is grown in the presence of galactose, there is selection pressure for increased *galK* expression, which can occur either through gene amplification or through point mutation. Among the experimental conditions investigated in this study, the low galactose environment [54] is the most applicable to our model, because the point mutations tracked in amplified cells did not increase the bacterial fitness beyond the advantage already attained through amplification. In this system it was found that bacteria carried either one or the other type of alteration, but not both together. The explanation given for this observation was based on clonal interference between adaptive amplifications and point mutations[54]. It is interesting to point out, however, that our model also predicts that lack of co-occurrence of gene amplifications and point mutations in this system because the bacteria were

grown in liquid medium. In a mixed system, our model suggests that to observe the co-occurrence of point mutations and gene amplifications, the population size would have to grow to unrealistically large sizes, or the fitness advantage of the amplified cells would have to be extremely large. Hence, it is possible that the lack of point mutations arising in amplified cells is the result of the basic growth laws of amplified cells in a mixed system (as opposed to a spatial system), without the need to invoke further complexities. It would be interesting to repeat this study in a setting where bacterial growth is more spatially restricted.

The result that spatial population structure greatly promotes the ability of gene amplifications to contribute to the generation of point mutations is important for a detailed interpretation and comparison of experimental results across different systems. It also has important implications for cancer. Since the growth of tumors (especially solid tumors) is characterized by strong spatial restrictions, gene amplifications likely accelerate the accumulation of point mutations in the amplified genes, which can in turn drive the adaptation of tumor cells to overcome natural selective barriers or to become resistant to therapies. While gene amplifications readily occur in cancer genomes [55], this might be especially important for extrachromosomal DNA [56, 57] (ecDNA). Oncogene amplifications occurs on ecDNA, and this can promote drug resistance and disease progression through further evolutionary processes, including the acquisition of point mutations in amplified genes.

Apart from gene amplifications, we also considered MMR-deficient cells as an intermediate state with an elevated mutation rate. While MMR-deficiency occurs in bacteria, a more relevant system in the context of our model could be MMR-deficiency in colorectal cancer. This has been shown to accelerate colorectal cancer evolution through microsatellite instability. The MMR-deficient phenotype in this system typically requires the inactivation of the MMR gene, which involves two separate mutational events. While this slows down the evolutionary pathway via MMR-deficiency, our model predicts that in the context of spatially structured cell growth, the MMR-deficient cells that are generated in a growing population can still substantially contribute to the generation of target mutations, but only if the MMR-deficient cells have a selective advantage. This is likely to be true for colorectal cancer. Interestingly, the gene inactivation leading to MMR-deficiency also leads to a reduction in the rate at which cells undergo apoptotic cell death, which provides these cells with an immediate selective advantage that can outweigh fitness reduction through the accelerated generation of deleterious mutants

[33]. According to our model, this intrinsic fitness advantage of MMR-deficient cells, coupled with the spatial growth of colorectal tumors, allows MMR-deficient cells to drive tumor evolution (and this would not occur if these conditions were violated).

6. Conclusions

Through the derivation of mutant scaling laws in spatially structured and expanding populations, we have provided a tool to study various evolutionary processes during spatial population growth without the restrictions posed by computationally intensive agent-based models. Hence, we can answer evolutionary questions in the context of large populations and low mutation rates, which is relevant for bacterial and cancer cell populations. We have applied this methodology to study the role of intermediate cell types with accelerated mutation rates for the acquisition of target point mutations, particularly the role of gene duplications / amplifications and MMR-deficient cells. This methodology, however, is more broadly applicable, and can be used to answer questions concerned with the evolution of complex traits in growing populations. This has not only implications for evolutionary biology, but also for biomedical questions such as the evolution of therapy resistance, immune escape, and aspects of disease development.

References

[1] Luria, S.E. & Delbruck, M. 1943 Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* **28**, 491-511.
[2] Zheng, Q. 1999 Progress of a half century in the study of the Luria-Delbruck distribution. *Mathematical biosciences* **162**, 1-32.
[3] Dewanji, A., Luebeck, E.G. & Moolgavkar, S.H. 2005 A generalized Luria-Delbruck model. *Mathematical biosciences* **197**, 140-152.
[4] Komarova, N.L., Wu, L. & Baldi, P. 2007 The fixed-size Luria–Delbruck model with a nonzero death rate. *Mathematical biosciences* **210**, 253-290.
[5] Kepler, T.B. & Oprea, M. 2001 Improved inference of mutation rates: I. An integral representation for the Luria–Delbrück distribution. *Theoretical population biology* **59**, 41-48.
[6] Goldie, J.H. & Coldman, A.J. 1983 A model for resistance of tumor cells to cancer chemotherapeutic agents. *Mathematical biosciences* **65**, 291-307.
[7] Goldie, J.H. & Coldman, A.J. 1998 *Drug resistance in cancer: mechanisms and models*, Cambridge University Press.
[8] Komarova, N.L. & Wodarz, D. 2005 Drug resistance in cancer: principles of emergence and prevention. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 9714-9719. (doi:0501870102 [pii] 10.1073/pnas.0501870102).
[9] Fusco, D., Gralka, M., Kayser, J., Anderson, A. & Hallatschek, O. 2016 Excess of mutational jackpot events in expanding populations revealed by spatial Luria-Delbruck experiments. *Nature communications* **7**, 12760. (doi:10.1038/ncomms12760).
[10] Gralka, M. & Hallatschek, O. 2019 Environmental heterogeneity can tip the population genetics of range expansions. *Elife* **8**. (doi:10.7554/eLife.44359).
[11] Gralka, M., Stiewe, F., Farrell, F., Mobius, W., Waclaw, B. & Hallatschek, O. 2016 Allele surfing promotes microbial adaptation from standing variation. *Ecology letters* **19**, 889-898. (doi:10.1111/ele.12625).
[12] Otwinowski, J. & Krug, J. 2014 Clonal interference and Muller's ratchet in spatial habitats. *Physical biology* **11**, 056003. (doi:10.1088/1478-3975/11/5/056003).
[13] Lavrentovich, M.O., Wahl, M.E., Nelson, D.R. & Murray, A.W. 2016 Spatially Constrained Growth Enhances Conversional Meltdown. *Biophysical journal* **110**, 2800-2808. (doi:10.1016/j.bpj.2016.05.024).
[14] Waclaw, B., Bozic, I., Pittman, M.E., Hruban, R.H., Vogelstein, B. & Nowak, M.A. 2015 A spatial model predicts that dispersal and cell turnover limit intratumour heterogeneity. *Nature* **525**, 261-264. (doi:10.1038/nature14971).
[15] Aif, S., Appold, N., Kampman, L., Hallatschek, O. & Kayser, J. 2022 Evolutionary rescue of resistant mutants is governed by a balance between radial expansion and selection in compact populations. *Nature communications* **13**, 7916.
[16] Yu, Q., Gralka, M., Duvernoy, M.-C., Sousa, M., Harpak, A. & Hallatschek, O. 2021 Mutability of demographic noise in microbial range expansions. *The ISME journal* **15**, 2643-2654.

[17] Paulose, J. & Hallatschek, O. 2020 The impact of long-range dispersal on gene surfing. *Proceedings of the National Academy of Sciences* **117**, 7584-7593.
[18] Birzu, G., Matin, S., Hallatschek, O. & Korolev, K.S. 2019 Genetic drift in range expansions is very sensitive to density dependence in dispersal and growth. *Ecology letters* **22**, 1817-1827.
[19] Wodarz, D. & Komarova, N.L. 2020 Mutant evolution in spatially structured and fragmented expanding populations. *Genetics* **216**, 191-203.
[20] Schreck, C.F., Fusco, D., Karita, Y., Martis, S., Kayser, J., Duvernoy, M.-C. & Hallatschek, O. 2023 Impact of crowding on the diversity of expanding populations. *Proceedings of the National Academy of Sciences* **120**, e2208361120.
[21] Kayser, J., Schreck, C.F., Gralka, M., Fusco, D. & Hallatschek, O. 2019 Collective motion conceals fitness differences in crowded cellular populations. *Nature ecology & evolution* **3**, 125-134.
[22] Noble, R., Burri, D., Le Sueur, C., Lemant, J., Viossat, Y., Kather, J.N. & Beerenwinkel, N. 2022 Spatial structure governs the mode of tumour evolution. *Nature ecology & evolution* **6**, 207-217.
[23] Colyer, B., Bak, M., Basanta, D. & Noble, R. 2024 A seven-step guide to spatial, agent-based modelling of tumour evolution. *Evolutionary applications* **17**, e13687.
[24] West, J., Schenck, R.O., Gatenbee, C., Robertson-Tessi, M. & Anderson, A.R. 2021 Normal tissue architecture determines the evolutionary course of cancer. *Nature communications* **12**, 2060.
[25] Gibson, M.A. & Bruck, J. 2000 Efficient exact stochastic simulation of chemical systems with many species and many channels. *The journal of physical chemistry A* **104**, 1876-1889.
[26] Cao, Y., Gillespie, D.T. & Petzold, L.R. 2006 Efficient step size selection for the tau-leaping simulation method. *Journal of Chemical Physics* **124**. (doi:Artn 044109 10.1063/1.2159468).
[27] Gillespie, D.T. 2007 Stochastic simulation of chemical kinetics. *Annual Review of Physical Chemistry* **58**, 35-55. (doi:10.1146/annurev.physchem.58.032806.104637).
[28] Salis, H. & Kaznessis, Y. 2005 Accurate hybrid stochastic simulation of a system of coupled chemical or biochemical reactions. *Journal of Chemical Physics* **122**. (doi:Artn 054103 10.1069/1.1835951).
[29] Wilkinson, D.J. 2011 *Stochastic modelling for systems biology*, CRC press.
[30] Rodriguez-Brenes, I.A., Komarova, N.L. & Wodarz, D. 2019 The role of telomere shortening in carcinogenesis: A hybrid stochastic-deterministic approach. *Journal of theoretical biology* **460**, 144-152. (doi:10.1016/j.jtbi.2018.09.003).
[31] Granelli-Piperno, A., Shimeliovich, I., Pack, M., Trumpfheller, C. & Steinman, R.M. 2006 HIV-1 selectively infects a subset of nonmaturing BDCA1-positive dendritic cells in human blood. *Journal of immunology* **176**, 991-998.
[32] Slechta, E.S., Bunny, K.L., Kugelberg, E., Kofoid, E., Andersson, D.I. & Roth, J.R. 2003 Adaptive mutation: general mutagenesis is not a programmed response to stress but results from rare coamplification of dinB with lac. *Proceedings of the National Academy of Sciences* **100**, 12847-12852.
[33] Madden-Hennessey, K., Gupta, D., Radecki, A.A., Guild, C., Rath, A. & Heinen, C.D. 2022 Loss of mismatch repair promotes a direct selective advantage in human stem cells. *Stem cell reports* **17**, 2661-2673.

[34] Rodriguez-Brenes, I.A., Komarova, N.L. & Wodarz, D. 2013 Tumor growth dynamics: insights into evolutionary processes. *Trends in ecology & evolution* **28**, 597-604. (doi:10.1016/j.tree.2013.05.020).
[35] Tomassini, M. 2005 *Spatially structured evolutionary algorithms: artificial evolution in space and time*, Springer.
[36] Foo, J., Leder, K. & Schweinsberg, J. 2020 Mutation timing in a spatial model of evolution. *Stochastic Processes and their Applications* **130**, 6388-6413.
[37] Hwang, S.H., Katsoulakis, M. & Rey-Bellet, L. 2013 Deterministic equations for stochastic spatial evolutionary games. *Theoretical Economics* **8**, 829-874.
[38] Evilsizor, S. & Lanchier, N. 2016 Evolutionary games on the lattice: death-birth updating process.
[39] Champagnat, N. & Méléard, S. 2007 Invasion and adaptive evolution for individual-based spatially structured populations. *J Math Biol* **55**, 147-188.
[40] Roca, C.P., Cuesta, J.A. & Sánchez, A. 2009 Effect of spatial structure on the evolution of cooperation. *Physical Review E* **80**, 046106.
[41] Ovaskainen, O., Finkelshtein, D., Kutoviy, O., Cornell, S., Bolker, B. & Kondratiev, Y. 2014 A general mathematical framework for the analysis of spatiotemporal point processes. *Theoretical ecology* **7**, 101-113.
[42] Ibsen-Jensen, R., Chatterjee, K. & Nowak, M.A. 2015 Computational complexity of ecological and evolutionary spatial dynamics. *Proceedings of the National Academy of Sciences* **112**, 15636-15641.
[43] Chen, Y.-T. 2018 Wright–Fisher diffusions in stochastic spatial evolutionary games with death–birth updating.
[44] Opasic, L., Zhou, D., Werner, B., Dingli, D. & Traulsen, A. 2019 How many samples are needed to infer truly clonal mutations from heterogenous tumours? *BMC cancer* **19**, 1-11.
[45] Strobl, M.A., Gallaher, J., West, J., Robertson-Tessi, M., Maini, P.K. & Anderson, A.R. 2022 Spatial structure impacts adaptive therapy by shaping intra-tumoral competition. *Communications medicine* **2**, 46.
[46] Andersson, D.I. & Hughes, D. 2009 Gene amplification and adaptive evolution in bacteria. *Annual review of genetics* **43**, 167-195.
[47] Elliott, K.T., Cuff, L.E. & Neidle, E.L. 2013 Copy number change: evolving views on gene amplification. *Future microbiology* **8**, 887-899.
[48] Hastings, P., Bull, H.J., Klump, J.R. & Rosenberg, S.M. 2000 Adaptive amplification: an inducible chromosomal instability mechanism. *Cell* **103**, 723-731.
[49] Hersh, M.N., Ponder, R.G., Hastings, P. & Rosenberg, S.M. 2004 Adaptive mutation and amplification in Escherichia coli: two pathways of genome adaptation under stress. *Research in Microbiology* **155**, 352-359.
[50] Hendrickson, H., Slechta, E.S., Bergthorsson, U., Andersson, D.I. & Roth, J.R. 2002 Amplification–mutagenesis: Evidence that “directed” adaptive mutation and general hypermutability result from growth with a selected gene amplification. *Proceedings of the National Academy of Sciences* **99**, 2164-2169.
[51] Kugelberg, E., Kofoid, E., Reams, A.B., Andersson, D.I. & Roth, J.R. 2006 Multiple pathways of selected gene amplification during adaptive mutation. *Proceedings of the National Academy of Sciences* **103**, 17319-17324.

[52] Yona, A.H., Manor, Y.S., Herbst, R.H., Romano, G.H., Mitchell, A., Kupiec, M., Pilpel, Y. & Dahan, O. 2012 Chromosomal duplication is a transient evolutionary solution to stress. *Proceedings of the National Academy of Sciences* **109**, 21010-21015.
[53] Kondrashov, F.A. 2012 Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proceedings of the Royal Society B: Biological Sciences* **279**, 5048-5057.
[54] Tomanek, I. & Guet, C.C. 2022 Adaptation dynamics between copy-number and point mutations. *Elife* **11**, e82240.
[55] Albertson, D.G. 2006 Gene amplification in cancer. *TRENDS in Genetics* **22**, 447-455.
[56] Lange, J.T., Rose, J.C., Chen, C.Y., Pichugin, Y., Xie, L., Tang, J., Hung, K.L., Yost, K.E., Shi, Q. & Erb, M.L. 2022 The evolutionary dynamics of extrachromosomal DNA in human cancers. *Nature genetics* **54**, 1527-1533.
[57] Wu, S., Bafna, V., Chang, H.Y. & Mischel, P.S. 2022 Extrachromosomal DNA: an emerging hallmark in human cancer. *Annual Review of Pathology: Mechanisms of Disease* **17**, 367-386.