Moran Process Version of the Tug-of-War Model: Complex Behavior Revealed by Mathematical Analysis and Simulation Studies

Adam Bobrowski[1], Marek Kimmel[2*], Monika K. Kurpas[3] and Elżbieta Ratajczyk[1]

[1]Lublin University of Technology, Nadbystrzycka 38A, 20-618 Lublin, Poland

[2]Rice University, Department of Statistics, 6100 Main St., Houston, TX 77005, USA

[3]Silesian University of Technology, Akademicka 16, 44-100 Gliwice, Poland

[*]*Corresponding author:* kimmel@rice.edu

Abstract. In a series of publications McFarland and co-authors introduced the tug-of-war model of evolution of cancer cell populations. The model is explaining the joint effect of rare advantageous and frequent slightly deleterious mutations, which may be identifiable with driver and passenger mutations in cancer. In this paper, we put the Tug-of-War model in the framework of a denumerable-type Moran process and use mathematics and simulations to understand its behavior. The model is associated with a time-continuous Markov Chain (MC), with a generator that can be split into a sum of the drift and selection process part and of the mutation process part. Operator semigroup theory is then employed to prove that the MC does not explode, as well as to characterize a strong-drift limit version of the MC which displays "instant fixation" effect, which was an assumption in the original McFarland's model. Mathematical results are fully confirmed by simulations of the complete and limit versions. They also visualize complex stochastic transients and genealogies of clones arising in the model.

1. Introduction

The Tug-of-War model was developed in a series of papers of McFarland and co-authors [21, 23, 24] to account for existence of mutually counteracting rare advantageous driver mutations and more frequent slightly deleterious passenger mutations in cancer. In its original version it is a state-dependent branching process, analyzed by a range of simulation methods and analytical approximations.

We adopt a different, simpler, approach, in which we reformulate McFarland's original definition to put it into the framework of a Moran model, which we investigate by complementary methods of mathematical analysis and simulation.

In the current study we are not primarily concerned with understanding the genealogies of the individuals such as cancer cells present in the populations. We identify individuals with the same counts of passenger and driver mutations and follow trajectories of the so-defined types. As it will become clear in the sequel, process behavior is quite complicated. Nevertheless, we demonstrate absorption properties of the process with no mutations (Section 4) and use operator semigroup theory to prove two limit cases (Section 7).

2. The model: a population under selection, drift and mutation

We consider a population of a fixed number N of individuals, each of them characterized by a pair of integers (α, β), corresponding to the numbers of drivers and passengers in its genotype, respectively. This pair determines the fitness f of the individual by the formula

$$f = (1+s)^{\alpha}(1-d)^{\beta}, \tag{2.1}$$

where $s > 0$ and $d \in (0,1)$ are certain parameters describing selective advantage of driver mutations over passenger mutations. Thus, the entire population may be identified with the vector

$$\mathfrak{p} = ((\alpha_1, \beta_1), \dots, (\alpha_N, \beta_N))$$

of N pairs of integers, with the accompanying vector

$$\mathsf{f} = (f_1, \dots, f_N)$$

of fitnesses.

The population is under drift and selection pressure: the individual of type (α_i, β_i) lives for an exponential time with parameter

$$\sum (1+s)^{\alpha_j}(1-d)^{\beta_j}$$

where the sum is over all $j = 1, \ldots, N$ such that $(\alpha_j, \beta_j) \neq (\alpha_i, \beta_i)$, and then is replaced by an individual of different type. More specifically, let $n_{\alpha,\beta}$ be the number of individuals of type (α, β) and n be the number of different types of individuals in the population, then the time to the death of each individual of type (α_i, β_i) is the minium of $n-1$ exponential random variables T_{α_j,β_j} where $(\alpha_j, \beta_j) \neq (\alpha_i, \beta_i)$ and T_{α_j,β_j} has parameter $n_{\alpha_j,\beta_j}(1+s)^{\alpha_j}(1-d)^{\beta_j}$. Upon this individual's death, conditional on the minimal time being equal to T_{α_k,α_k}, this individual is replaced by one of the individuals of type (α_k, β_k), each if these individuals being equally likely. This process then continues with $\mathfrak{p}$ modified by replacing its ith coordinate (α_i, β_i) by its kth coordinate (α_k, β_k).

In particular, if all individuals in $\mathfrak{p}$ are pairwise different, the time to the first drift and selection event for the entire population is exponential with parameter $(N-1)\Sigma\mathsf{f}$ where

$$\Sigma\mathsf{f} = \sum_{k=1}^{N} f_k.$$

After this time is over, one individual dies and is replaced by an exact copy of one of the remaining individuals, the probability that the ith individual dies and is replaced by the jth $(j \neq i)$ being $\frac{f_j}{(N-1)\Sigma\mathsf{f}}$. If, on the other hand, all individuals are the same, nothing happens: there are no drift and selection events.

Moreover, each individual may, after an independent exponential time with parameter, say λ, and independently of other individuals, undergo a mutation event, changing its state to either $(\alpha+1, \beta)$ or $(\alpha, \beta+1)$ with (conditional) probabilities $p \in (0,1)$ and q, respectively. In other words, all mutations occur at the epochs of a Poisson process with intensity λ, occurrences of driver mutations on each individual form a colored Poisson process, with probability of coloring equal p, and the occurrences of passenger mutations form a colored Poisson process with probability of coloring equal q. It follows (see the Colouring Theorem on p. 53 in [17]) that driver and passenger mutations form Poisson processes with parameters

$$\nu = \lambda p \qquad \text{and} \qquad \mu = \lambda q,$$

respectively, and these processes are independent of each other, and independent of mutation processes on other individuals. (Although technically we never use this assumption, what we have in mind is the case where p is significantly smaller than $1-p$, so that long strings of passenger mutations are interrupted by rare driver mutations.) In particular, given that initially an individual's fitness is f, after time t its expected fitness is the product of $f, \mathrm{e}^{-\lambda pt}\mathrm{e}^{\lambda p(1+s)t}$ (the contribution of driver mutations) and $\mathrm{e}^{-\lambda qt}\mathrm{e}^{\lambda q(1-d)t}$ (the contribution of passenger mutations), and thus equals

$$f(t) = \mathrm{e}^{\lambda t(sp-dq)} f, \qquad t \geq 0. \tag{2.2}$$

In a number of cells this expected fitness does not grow to infinity or decay to zero; such cells are thus characterized by the following balance condition for the introduced parameters:

$$sp = dq, \quad \text{or} \quad p = \frac{d}{s+d}. \tag{2.3}$$

In other words, the advantage gained by a driver mutation is balanced by the small probability of such event.

In other cells, however, driver mutations, though rare may have a slight edge over the passenger mutations caused by large s. Such cell populations are characterized by

$$sp > dq, \quad \text{or} \quad p > \frac{d}{s+d}. \tag{2.4}$$

In yet different populations, driver mutations will be so rare that the expected total fitness diminishes in time. To characterize such populations, we reverse the inequalities in (2.4).

3. A Markov chain and the related intensity matrix

The population described in Section 2 is modeled by a stochastic process

$$\mathfrak{p}(t), t \geq 0 \tag{3.1}$$

with values in the state-space $\mathfrak{P}$ of N ordered copies of the Cartesian product $\mathbb{N} \times \mathbb{N}$, where $\mathbb{N}$ is the set of natural numbers:

$$\mathfrak{P} \coloneqq (\mathbb{N} \times \mathbb{N})^N.$$

This is just to say that at each time t, the population is an N-tuple of pairs $(\alpha_i(t), \beta_i(t)), i \in \mathcal{N}$ of positive integers, where

$$\mathcal{N} \coloneqq \{1, \ldots, N\}.$$

Since $\mathfrak{P}$ is a countable set, the process $\mathfrak{p}(t), t \geq 0$ may be thought of as a time-continuous Markov chain.

Such Markov chains are conveniently described by means of intensity (Kolmogorov) matrices that gather information on rates (intensities) with which these processes leave a given state and jump to other states (see e.g. [4, 25]; see also our Section 10). We will write the intensity matrix for the process (3.1) as the sum of two intensity matrices representing mutations and drift and selection events, respectively.

To describe the first of these, call it Q_M, ('M' for 'mutations') let D and P ('D' for driver and 'P' for passenger) be the following maps of $\mathbb{N} \times \mathbb{N}$ into itself:

$$D(\alpha, \beta) = (\alpha + 1, \beta) \qquad \text{and} \qquad P(\alpha, \beta) = (\alpha, \beta + 1). \tag{3.2}$$

Moreover, for each $i = 1, \ldots, N$, let $D_i : \mathfrak{P} \to \mathfrak{P}$ be the map in which the ith coordinate (α_i, β_i) of a $\mathfrak{p} \in \mathfrak{P}$ is replaced by $D(\alpha_i, \beta_i)$. Similarly, let $P_i : \mathfrak{P} \to \mathfrak{P}$ be the map in which the ith coordinate (α_i, β_i) of a $\mathfrak{p} \in \mathfrak{P}$ is replaced by $P(\alpha_i, \beta_i)$. In these notations, the intensity $q_{\mathfrak{p},\mathfrak{q}}$ of going from a state $\mathfrak{p} \in \mathfrak{P}$ to a state $\mathfrak{q} \in \mathfrak{P}$ in the mutation process is

$$q_{\mathfrak{p},\mathfrak{q}} = \begin{cases} \lambda p & \text{if } \mathfrak{q} = D_i\mathfrak{p} \text{ for some } i \in \mathcal{N}, \\ \lambda q & \text{if } \mathfrak{q} = P_i\mathfrak{p} \text{ for some } i \in \mathcal{N}, \\ -N\lambda & \text{if } \mathfrak{q} = \mathfrak{p}, \\ 0 & \text{otherwise.} \end{cases} \tag{3.3}$$

Similarly, for $i, j = 1, \ldots, N$ let $R_{ij} : \mathfrak{P} \to \mathfrak{P}$ be the map that replaces the ith coordinate (α_i, β_i) of a $\mathfrak{p} \in \mathfrak{P}$ by its jth coordinate (α_j, β_j), leaving the remaining coordinates intact. For example, if $N = 3$, $R_{1,3}$ maps $((\alpha_1, \beta_1), (\alpha_2, \beta_2), (\alpha_3, \beta_3))$ to $((\alpha_3, \beta_3), (\alpha_2, \beta_2), (\alpha_3, \beta_3))$. The intensity matrix describing drift and selection events, say Q_S, has then the following entries:

$$q_{\mathfrak{p},\mathfrak{q}} = \begin{cases} n_j f_j & \text{if } \mathfrak{q} = R_{i,j}\mathfrak{p} \text{ for some } i, j \in \mathcal{N} \text{ such that } (\alpha_i, \beta_i) \neq (\alpha_j, \beta_j), \\ 0 & \text{otherwise,} \end{cases} \tag{3.4}$$

where n_j is the number of individuals in $\mathfrak{p}$ that are identical to the individual number j and $\mathsf{f} = (f_i)_{i \in \mathcal{N}}$ is the vector of fitnesses of individuals in $\mathfrak{p}$. More specifically,

$$f_i = (1 + s)^{\alpha_i}(1 - d)^{\beta_i}, i \in \mathcal{N}. \tag{3.5}$$

This formula does not cover the case where $\mathfrak{q} = \mathfrak{p}$ because this case requires a bit of preparation. Namely, let $n_{\alpha,\beta}$ denote the number of individuals of type (α,β) so that in particular $\sum_{i,j\in\mathbb{N}} n_{ij} = N$. Then,

$$q_{\mathfrak{p},\mathfrak{p}} = -\sum_{\alpha,\beta\in\mathbb{N}} n_{\alpha,\beta} \sum_{\gamma,\delta\in\mathbb{N};(\gamma,\delta)\neq(\alpha,\beta)} n_{\gamma,\delta}(1+s)^{\gamma}(1-d)^{\delta}. \tag{3.6}$$

We note that in the case where all individuals in a population $\mathfrak{p}$ are different, the formula for $q_{\mathfrak{p},\mathfrak{p}}$ simplifies to:

$$q_{\mathfrak{p},\mathfrak{p}} = -(N-1)\Sigma\mathsf{f}.$$

On the other hand, if all individuals in this population are of the same type, $q_{\mathfrak{p},\mathfrak{p}} = 0$.

Finally, the entries in the intensity matrix Q for the entire chain (3.1) are sums of the entries of matrices Q_M and Q_S:

$$Q = Q_M + Q_S. \tag{3.7}$$

4. Properties of the drift and selection chain

Consider the evolution of a population when mutations are absent, and only drift and selection events, as described above, are possible. This evolution is governed the intensity matrix Q_S with entries given in (3.4) and (3.6).

For definiteness, let $\mathfrak{p} = ((\alpha_i,\beta_i))_{i\in\mathcal{N}}$ be the initial state of this population and assume that all its individuals have different characteristics, i.e. $(\alpha_i,\beta_i) \neq (\alpha_j,\beta_j)$ for $i \neq j$. It is rather easy to see first of all there is only a finite number of states that can be reached from the state $\mathfrak{p}$: more precisely, there are at most N^N such states (including $\mathfrak{p}$ itself). For, since the chain is that of replacing coordinates of $\mathfrak{p}$ by copies of other coordinates, there are only N possibilities for the first coordinate of future states, N possibilities for the second coordinate, and so on.

Second, all these states, except for those with all coordinates equal, i.e. except for

$$\mathfrak{p}_i \coloneqq ((\alpha_i,\beta_i),(\alpha_i,\beta_i),\dots,(\alpha_i,\beta_i)), \qquad i\in\mathcal{N}$$

are transient for this chain. Indeed, for any other state there is a non-zero probability that the number of different individuals in the population will decrease in the next drift and selection event. Since the rules of the chain do not allow jumps from the states with smaller number of different individuals to the states with larger number of different individuals, the process will never come back to the state under consideration. This shows that this state cannot be recurrent, and thus, by the well-known dichotomy (see e.g. [25], Section 3.4) it must be transient. On the other hand, all the states $\mathfrak{p}_i, i \in \mathcal{N}$ are absorbing. Hence, the process starting at a $\mathfrak{p}$ must eventually end up at one of $\mathfrak{p}_i$'s. Certainly, in the case where not all individuals in the original population are different, the fate of the population is similar: it is only the number of possibilities for the end population that is smaller. We summarize our discussion in the following theorem.

Theorem 4.1. *Let $\mathfrak{p} \in \mathfrak{P}$ be a population and let M be the number of different variants in $\mathfrak{p}$. Then, there is a set $\mathcal{M} \subset \mathcal{N}$ such that (a) $M = \#\mathcal{M}$, and (b) $i \neq j, i, j \in \mathcal{M}$ implies $(\alpha_i, \beta_i) \neq (\alpha_j, \beta_j)$. For $i \in \mathcal{M}$ let $\mathfrak{p}_i \in \mathfrak{P}$ be the population in which all individuals are identical to each other and to the ith individual in the original population $\mathfrak{p}$. Then, the drift and selection chain starting at $\mathfrak{p}$ will eventually end up at $\mathfrak{p}_i$ with certain probability $p_i = p_i(\mathfrak{p})$ where $\sum_{i \in \mathcal{M}} p_i = 1$.*

We note in passing that whereas there could be many choices of $\mathcal{M}$, the thesis of our theorem remains the same for all of them.

This theorem is a reflection of the fact that drift and selection chain strives to reduce the number of variants in the population by removing randomly selected variants and replacing them by other variants; in the absence of other forces, and mutation in particular, the chain's operation in the long run leads to fixation of one the variants. What this theorem does not express openly is that drift and selection chain favors variants with larger fitness. The latter information, besides being visible in formula (3.4), is hidden in the probabilities p_i featuring in Theorem 4.1; roughly speaking, the larger the fitness of an individual, the larger is the probability of fixation of its variant. Notably, even though selection favors variants with larger fitness, it acts together with genetic drift which may 'blindly', by chance, remove better fit variants from the population. Hence, the fact that a variant with larger fitness is favored by selection results in a higher probability of its fixation, and not in the inevitability of its fixation.

In what follows we will see this principle expressed in explicit formulae for p_i's in the cases $N = 2$ and $N = 3$, considered here for the sake of illustration. In our calculations it will be convenient, for the sake of shortening our equations and making figures readable, to identify a population $\mathfrak{p}$, which is an N-dimensional vector of pairs of integers, with the N-dimensional vector of corresponding fitnesses calculated by formula (3.5), the latter vector being half as long as the former. Although it is possible, by an appropriate choice of parameters, to have two different individuals with the same fitnesses, i.e. to have $(\alpha_i, \beta_i) \neq (\alpha_j, \beta_j)$ and at the same time $(1+s)^{\alpha_i}(1-d)^{\beta_i} = (1+s)^{\alpha_j}(1-d)^{\beta_j}$, such an identification should not lead to misunderstandings.

For $N = 2$ the probabilities p_1 and p_2 are easily calculated explicitly: unless it is already uniform, a population $\mathfrak{p}$ with fitness (f_1, f_2), after an exponential time with parameter $f_1 + f_2$, becomes $\mathfrak{p}_1$ with probability $\frac{f_1}{f_1+f_2}$ or $\mathfrak{p}_2$ with probability $\frac{f_2}{f_1+f_2}$. This shows that $p_1 = \frac{f_1}{f_1+f_2}$ and $p_2 = 1 - p_1 = \frac{f_2}{f_1+f_2}$.

Somewhat surprisingly, already for $N = 3$ the formulae for p_i's are more complicated, and do not follow the perhaps expected pattern $p_i = \frac{f_i}{\Sigma \mathsf{f}}$. Before we see that, however, we note the following important property of the chain under consideration: Let us call two states $\mathsf{f} = (f_1, \dots, f_N)$ and $\mathsf{f}' = (f_1', \dots, f_N')$ *associated* if there is a permutation Π of the set $\mathcal{N}$ such that $f_i' = f_{\Pi(i)}, i \in \mathcal{N}$. The property we want to note is as follows.

Theorem 4.2. *The drift and selection processes starting at two associated states are analogous.*

What we mean by that proposition is that (a) the times to the first drift and selection events for either of two associated states f and f' have the same distribution, (b) the probability that in such an event the ith coordinate of f is replaced by the jth, is the same as the probability that the $\Pi(i)$th coordinate of f' is replaced by the $\Pi(j)$th, and (c) if in these drift and selection events the ith coordinate of f is replaced by the jth, and the $\Pi(i)$th coordinate of f' is replaced by the $\Pi(j)$th then the states after these events are again associated. These statements are clear from the description of the drift and selection chain, and combined together prove Theorem 4.2.

We are now ready to find p_is for $N = 3$. We think of a process that starts at an $\mathsf{f} = (f_1, f_2, f_3)$. Figure 1 illustrates the fact that in order to reach the state $\mathsf{f}_1 = (f_1, f_1, f_1)$ this process must go through (f_1, f_1, f_3), (f_1, f_2, f_1) or one of their associates. The first of these states is reached directly with probability $\frac{f_1}{2\Sigma \mathsf{f}}$. This state or one of its associates may also be reached indirectly, via (f_1, f_3, f_3), with probability $\frac{f_3}{2\Sigma \mathsf{f}} \frac{f_1}{f_1+f_3}$. Thus, the probability of reaching (f_1, f_1, f_3) or one of its associates is

$$\frac{f_1}{2\Sigma \mathsf{f}} \frac{f_1 + 2f_3}{f_1 + f_3}.$$

Then, before reaching f_1 from one of these associated states, the process may visit an associate of (f_1, f_3, f_3), and this may happen $k \geq 0$ times. Since the properties of the processes starting from associated states are analogous, the probability of reaching f_1 from one of associates of

(f_1, f_1, f_3) is

$$\frac{f_1}{f_1+f_3}\sum_{k=0}^{\infty}\left(\frac{f_3}{f_1+f_3}\right)^k\left(\frac{f_1}{f_1+f_3}\right)^k = \frac{f_1}{f_1+f_3}\frac{1}{1-\frac{f_1f_3}{(f_1+f_3)^2}}.$$

Therefore, the probability that f_1 will be reached through (f_1, f_1, f_3) or its associate is $\frac{f_1^2}{2\Sigma\mathsf{f}}\frac{f_1+2f_3}{f_1^2+f_3^2+f_1f_3}$ and so

$$p_1 = \frac{f_1^2}{2\Sigma\mathsf{f}}\left[\frac{f_1+2f_2}{f_1^2+f_2^2+f_1f_2}+\frac{f_1+2f_3}{f_1^2+f_3^2+f_1f_3}\right], \tag{4.1}$$

because the case where the process goes through associates of (f_1, f_2, f_1) is symmetrical. Using symmetry again, we obtain

$$p_2 = \frac{f_2^2}{2\Sigma\mathsf{f}}\left[\frac{f_2+2f_1}{f_1^2+f_2^2+f_1f_2}+\frac{f_2+2f_3}{f_2^2+f_3^2+f_2f_3}\right]$$

and

$$p_3 = \frac{f_3^2}{2\Sigma\mathsf{f}}\left[\frac{f_3+2f_1}{f_1^2+f_3^2+f_1f_3}+\frac{f_3+2f_2}{f_2^2+f_3^2+f_2f_3}\right].$$

As remarked above, these formulae illustrate the fact that the drift and selection process, besides striving to minimize the number of variants, tries also to maximize the total fitness of the population by selecting against the variants with small fitness.

Analogous formulae for the case $N = 4$ were also obtained, using Maple, but even after simplification, they were too long to be informative; each of them occupied half a page. Hence, in the absence of explicit formulae, we content ourselves with the following theorem which shows that drift and selection events 'on average' increase the total fitness of population.

Theorem 4.3. *Let f' be the state of the process right after drift and selection event of a population f. Then*

$$E\,\Sigma\mathsf{f}' \geq \Sigma\mathsf{f},$$

where E denotes expected value.

Proof. Each event of replacing the ith coordinate of f by its jth coordinate is paired by the event in which the jth coordinate is replaced by the ith coordinate. The first of these events takes place with probability $f_j/|q_{\mathfrak{p},\mathfrak{p}}|$, where $q_{\mathfrak{p},\mathfrak{p}}$ is the diagonal element of the generator matrix in Equ. (3.6). Accordingly, $\Sigma\mathsf{f}' - \Sigma\mathsf{f} = f_j - f_i$, and the second event's characteristics are symmetrical. Therefore, $E\,\Sigma\mathsf{f}' - \Sigma\mathsf{f}$ equals

$$\sum_{i<j}\left[\frac{(f_j-f_i)f_j}{|q_{\mathfrak{p},\mathfrak{p}}|}+\frac{(f_i-f_j)f_i}{|q_{\mathfrak{p},\mathfrak{p}}|}\right] = \frac{1}{|q_{\mathfrak{p},\mathfrak{p}}|}\sum_{i<j}(f_i-f_j)^2 \geq 0,$$

completing the proof. □

Next, we turn our attention to the situation where in a homogeneous population a new, possibly better fitted, variant shows up. In other words, the vector of fitnesses is of the form $(f_1, f_2, f_2, \dots, f_2)$ where $f_1 \gg f_2$. We are interested in the probability that variant with fitness f_1 will take over the entire population, i.e. in the probability that the drift and selection process will be absorbed in the state $(f_1, f_1, \dots, f_1)$.

Theorem 4.4. *Let $N \geq 2$. The probability that the drift and selection process starting at $(f_1, f_2, \dots, f_2) \in \mathfrak{P}$ is eventually absorbed at $(f_1, f_1, \dots, f_1)$ equals*

$$h_1 = \frac{1}{1 + \left(\frac{f_2}{f_1}\right) + \cdots + \left(\frac{f_2}{f_1}\right)^{N-1}}. \tag{4.2}$$

Proof. Let $X(t)$ be the number of individuals of fitness f_1 in the population $\mathfrak{p}(t)$. Because of Theorem 4.2, $X(t), t \geq 0$ is a time-continuous Markov chain with values in the set $\mathcal{N}_0 \coloneqq \{0\} \cup \mathcal{N}$, and the elements of its intensity matrix are as follows:

$$q_{i,j} = \begin{cases} 0 & \text{if } i \in \{0, N\}, \\ i(N-i)f_2 & \text{if } j = i-1, i \notin \{0, N\}, \\ -(f_1+f_2)i(N-i) & \text{if } j = i, i \notin \{0, N\}, \\ i(N-i)f_1 & \text{if } j = i+1, i \notin \{0, N\}, \\ 0 & \text{otherwise.} \end{cases} \tag{4.3}$$

We are interested in the probability that $X(t), t \geq 0$ will be absorbed at the state N, given that $X(0) = 1$. This probability is the second coordinate in the vector $(h_0, \dots, h_N)$ of so-called hitting probabilities for the absorbing state $\{N\}$ which, by Theorem 3.3.1 in [25] satisfy the following system of equations:

$$\sum_{j=0}^{N} q_{i,j} h_j = 0, \qquad i \notin \{0, N\}$$

with 'boundary conditions' $h_0 = 0, h_N = 1$. In other words

$$h_i = c h_{i-1} + d h_{i+1}, \tag{4.4}$$

where $c = \frac{f_2}{f_1+f_2}$ and $d = \frac{f_1}{f_1+f_2}$ for $i = 1, \dots, N-1$.

To solve this system, as in p. 16 of [25] or [11] p. 192, we introduce $u_i = h_{i-1} - h_i$ for $i = 1, \dots, N$. Then the recurrence relation (4.4) becomes $u_{i+1} = \frac{c}{d} u_i = \frac{f_2}{f_1} u_i$, $i = 1, \dots, N-1$.

Therefore, by induction,

$$u_i = \left(\frac{f_2}{f_1}\right)^{i-1} u_1, \qquad i = 1, \dots, N. \tag{4.5}$$

It follows that $\sum_{i=1}^N u_i = \left(1 + \left(\frac{f_2}{f_1}\right) + \cdots + \left(\frac{f_2}{f_1}\right)^{N-1}\right) u_1$. On the other hand, by the definition of u_is and the boundary conditions for h_is, $\sum_{i=1}^N u_i = h_0 - h_N = -1$. Hence we obtain that $h_1 = -u_1 = \frac{1}{1+\left(\frac{f_2}{f_1}\right)+\cdots+\left(\frac{f_2}{f_1}\right)^{N-1}}$, as desired. □

We complement this theorem with three remarks. First, we note that (4.5) hides an explicit formula for h_i for any $i \in \mathcal{N}$, i.e. for the probability that a subpopulation of i individuals of fitness f_1 will take over the entire population. For, since $\sum_{j=1}^i u_j = h_0 - h_i = -h_i$, this formula, when combined with (4.2), renders

$$h_i = \frac{1 + \left(\frac{f_2}{f_1}\right) + \cdots + \left(\frac{f_2}{f_1}\right)^{i-1}}{1 + \left(\frac{f_2}{f_1}\right) + \cdots + \left(\frac{f_2}{f_1}\right)^{N-1}}, \qquad i \in \mathcal{N}.$$

Second, observe that formula (4.2) in the case $N = 2$ agrees with the formula for p_1, and in the case $N = 3$ can be obtained from (4.1) by replacing f_3 by f_2.

Third, in our proof of (4.2) we never used the assumption that f_1 is larger than f_2. However, since N, being the size of the considered population, is typically rather large, for $f_1 \le f_2$, the probability of (4.2) is small. This means that new variants without sufficient selective advantage are simply washed away from the population.

On the other hand, given $r \in (0, 1)$ (to play the role of a probability), think of f_1 as of chosen so large as compared to f_2 that

$$\frac{f_2}{f_1} < 1 - r. \tag{4.6}$$

Then $1 + \left(\frac{f_2}{f_1}\right) + \cdots + \left(\frac{f_2}{f_1}\right)^{N-1} < \frac{1}{1-\frac{f_2}{f_1}} < \frac{1}{r}$. Therefore, for such f_1 (and f_2), $h_1 > r$. In other words, by enlarging f_1 sufficiently, we may make the probability of mutant's fixation as large as we wish.

We complete this section with information on the expected time to allele's fixation.

Theorem 4.5. *Let $N \geq 2$. The expected time for the drift and selection process starting at $(f_1, f_2, \dots, f_2) \in \mathfrak{P}$ to be eventually absorbed at $(f_1, f_1, \dots, f_1)$ or $(f_2, f_2, \dots, f_2)$ equals*

$$k_1 = \frac{h_1}{f_1} \sum_{i=1}^{N-1} \left(\sum_{j=1}^{i} \frac{1}{j(N-j)} \left(\frac{f_2}{f_1}\right)^{i-j} \right), \tag{4.7}$$

where h_1 is defined by formula (4.2).

Proof. The time of interest is the second coordinate in the vector $(k_0, \dots, k_N)$ of hitting times for the absorbing set $\{0, N\}$ which, by Theorem 3.3.3 in [25] satisfy the following system of equations:

$$-\sum_{j=0}^{N} q_{i,j} k_j = 1, \qquad i \notin \{0, N\}$$

with 'boundary conditions' $k_0 = 0, k_N = 0$, where $q_{i,j}$'s are defined in (4.3). In other words

$$k_i = c k_{i-1} + d k_{i+1} + e_i, \tag{4.8}$$

where $c = \frac{f_2}{f_1+f_2}, d = \frac{f_1}{f_1+f_2}$ and $e_i = \frac{1}{i(N-i)(f_1+f_2)}$ for $i = 1, \dots, N-1$.

To solve this system, as in the previous theorem, we introduce $v_i = k_{i-1} - k_i$ for $i = 1, \dots, N$. Then the recurrence relation (4.8) becomes $v_{i+1} = \frac{c}{d} v_i + \frac{e_i}{d} = \frac{f_2}{f_1} v_i + \frac{1}{i(N-i)f_1}$, $i = 1, \dots, N-1$. It follows that

$$v_i = \left(\prod_{j=1}^{i-1} \frac{f_2}{f_1} \right) \left(v_1 + \frac{1}{f_1} \sum_{j=1}^{i-1} \frac{1}{j(N-j) \prod_{k=1}^{j} \frac{f_2}{f_1}} \right)$$

for $i = 2, \dots, N$. We see that

$$\sum_{i=1}^{N} v_i = \frac{v_1}{h_1} + \sum_{i=2}^{N} \left(\left(\frac{f_2}{f_1}\right)^{i-1} \frac{1}{f_1} \sum_{j=1}^{i-1} \frac{1}{j(N-j) \left(\frac{f_2}{f_1}\right)^{j}} \right)$$
$$= \frac{v_1}{h_1} + \frac{1}{f_1} \sum_{i=1}^{N-1} \left(\sum_{j=1}^{i} \frac{1}{jf(N-j)} \left(\frac{f_2}{f_1}\right)^{i-j} \right).$$

On the other hand, by the definition of v_i's and the boundary conditions for k_i's, $\sum_{i=1}^{N} v_i = k_0 - k_N = 0$. Hence

$$k_1 = -v_1 = \frac{h_1}{f_1} \sum_{i=1}^{N-1} \left(\sum_{j=1}^{i} \frac{1}{j(N-j)} \left(\frac{f_2}{f_1}\right)^{i-j} \right),$$

completing the proof. ∎

5. Asymptotic behavior of $\{P_S(t), t \geq 0\}$

In the Supplement 10, we demonstrated the existence of a Markov chain related to the intensity matrix $Q = Q_M + Q_S$ of (3.7), i.e. the chain encompassing the mutation and drift "components" of the model we consider.

Before embarking on the study of limit versions of the semigroup $\{P(t), t \geq 0\}$ related to this chain, let us rephrase the results of Section 4 to find the limit

$$\lim_{t\to\infty} P_S(t).$$

To this end, given $\mathfrak{p} \in \mathfrak{P}$, consider

$$e_{\mathfrak{p}} \in \ell^1 \tag{5.1}$$

defined by $e_{\mathfrak{p}}(\mathfrak{q}) = 0, \mathfrak{q} \neq \mathfrak{p}, e_{\mathfrak{p}}(\mathfrak{p}) = 1$. Then $P_S(t)e_{\mathfrak{p}}$ is the $\mathfrak{p}$th row of transition probability matrix $P_S(t)$, composed of probabilities $p_{\mathfrak{p},\mathfrak{q}}(t)$ that drift and selection chain starting at $\mathfrak{p}$ will be at $\mathfrak{q}$ at time $t \geq 0$. As explained in Section 4 at most N^N probabilities in this row are non-zero, and as $t \to \infty$ even all of these at most N^N probabilities tend to zero, save for $M \leq N$ of them corresponding to populations where all individuals are identical to each other and to one of the members of the original population, where $M = M(\mathfrak{p})$ is the number of variants in the population $\mathfrak{p}$. Each of the latter M probabilities, on the other hand, converges to one of the probabilities $p_i(\mathfrak{p})$ described in Theorem 4.1. Since $\sum_{i\in\mathcal{M}} p_i = 1$, by Scheffés' Theorem (see e.g. [4]), $P_S(t)e_{\mathfrak{p}}$ converges to the vector with probabilities $p_i(\mathfrak{p})$, in the norm of ℓ^1. Here is a consequence of this remark.

Theorem 5.1. *Let the matrix* $\Pi = (\pi_{\mathfrak{p},\mathfrak{q}})_{\mathfrak{p},\mathfrak{q}\in\mathfrak{P}}$ *be defined as follows. For each* $\mathfrak{p}$ *we choose a subset* $\mathcal{M} \subset \mathcal{N}$ *as described in Theorem 4.1, and let*

$$\pi_{\mathfrak{p},\mathfrak{q}} = \begin{cases} p_i(\mathfrak{p}) & \text{if } \mathfrak{q} = \mathfrak{p}_i \text{ for an } i \in \mathcal{M}, \\ 0 & \text{otherwise} \end{cases}$$

(where $p_i(\mathfrak{p})$*s are defined in Theorem 4.1). Then, for any* $x \in \ell^1$,

$$\lim_{t\to\infty} P_S(t)x = \Pi x \; (:= x \cdot \Pi). \tag{5.2}$$

Proof. By the reasoning presented above, (5.2) holds for $x = e_{\mathfrak{p}}, \mathfrak{p} \in \mathfrak{P}$. By linearity, this formula extends to all combinations of $e_{\mathfrak{p}}$'s. Since such combinations are dense in ℓ^1, a three epsilon argument based on $\|P_S(t)\| = 1, t \geq 0$ completes the proof. □

6. Asymptotic behavior of $\{P(t), t \geq 0\}$: intuitions

The semigroup $\{P(t), t \geq 0\}$ describes the chain in which both mutations and drift and selection events take place. As such, it describes not only a tug-of-war between driver and passenger mutations but also a competition between selections and mutations, these population genetic forces counteracting each other. But, it is one of the main characteristics of drift and selection chain (see (3.4) and (3.6)) that the rate at which new drift and selection events come about grows with the total fitness of the population. It follows that in some regions of $\mathfrak{P}$ the rate of mutation is larger than the rate of drift and selection events and in other regions the former is smaller than the latter. Hence, in some regions selection will be more expressed, and in other regions effects of mutations will be more apparent.

There are no clear boundaries between these regions, no man's lands lie between them, and random forces may lead via these no man's lands from one region to another. Nevertheless, the three main regions, denoted R_0, R_l and R_u, may be characterized as follows.

6.1. R_0 **region.** The central region R_0 contains populations in which drift and selection events occur at a rate that is of the same order as the rate of mutations. By suitable scaling of parameters, this is the region where

$$(N-1)\Sigma\mathsf{f} \approx N\lambda.$$

The expression above is inaccurate, since on the left-hand side, instead of the exact expression for the intensity of time to the drift and selection event, we placed a simplified one, true only when all individuals are different. However, this is sufficient for the present purposes which is to define a region in which mutations have force comparable to drift and selection. We will carry out a more accurate analysis using the limit process (see Lemma 9.1 and the text preceding and following it).

An individual member of a population in this region collects new driver and passenger mutations over time: being characterized initially, at time $t = 0$, by (α, β), by the time $t > 0$ it becomes of the type $(\alpha + m, \beta + n)$ (provided it is still alive), but if assumption (2.3) is satisfied, the quotient $\frac{\text{new fitness}}{\text{old fitness}} = (1+s)^m(1-d)^n$ is roughly 1. In other words, between drift and selection events, all individuals travel the path where $n \approx -\frac{\log(1+s)}{\log(1-d)}m$; individual fitness does not change much in time. Travels along such paths are of course interrupted by deaths of individuals which are replaced by copies of other individuals. Genetic drift may thus sweep away rare variants, but some kind of statistical equilibrium is obtained between mutations introducing new variants and drift that continually reduces variability (see e.g. [5, 6, 7, 8]).

However, random fluctuations may force a population out of this region of balanced genetic forces to one of the following two regions, where one force prevails against the other. A population may, even the more, be forced out of this region by a temporary or permanent change of parameters a, d and p, so that e.g. condition (2.4) rather than (2.3) is satisfied.

6.2. R_l **region.** In the lower region R_l we have

$$(N-1)\Sigma \mathsf{f} \ll N\lambda.$$

Because of that assumption, drift and selection events are very seldom as compared to mutation events. This means that individuals live for relatively long times, and over periods of their lives accumulate mutations that distinguish them more and more from other individuals. In other words, the members of the population are rather loosely linked, and evolve quite independently of each other.

6.3. R_u **region.** The upper region R_u is characterized by

$$(N-1)\Sigma \mathsf{f} \gg N\lambda.$$

Here the situation is quite different: these are the mutations that are relatively rare as compared to the drift and selection events. Each individual lives for a short time and before it is able to collect a significant number of mutations distinguishing it from other individuals, dies and is replaced by another individual. As a result, very quickly the population becomes uniform: there is practically only one variant in it (i.e., one variant is fixed) as in Theorems 4.1 and 5.1, members of this population could be descendants of a rare but strong variant or of a week but frequent one in the initial population (see Theorem 4.4). In the next two sections we will be able to say more on how mutation process looks like in such populations.

7. Asymptotic behavior of $\{P(t), t \geq 0\}$ in the upper and lower regions

In this section, we provide a more rigorous mathematical argument, based on the theory of convergence of semigroups, for the intuitions of Sections 6.2 and 6.3. However, this argument still needs to be preceded by the following heuristic reasoning.

Let us consider a subregion S of R_u where total fitness of populations, in addition to being 'much larger' than $N\lambda$, is approximately constant: there is a $\kappa > 0$ such that

$$\frac{\Sigma \mathsf{f}(\mathfrak{p})}{N\lambda} \approx \kappa \gg 1, \qquad \mathfrak{p} \in S, \tag{7.1}$$

where $\mathsf{f} = \mathsf{f}(\mathfrak{p})$ is the fitness vector for $\mathfrak{p}$. Assume also that for each $\mathfrak{p}$ in S with the fitness vector $\mathsf{f} = (f_1, \dots, f_N)$ one may find a $\mathfrak{p}' \in \mathfrak{P}$ with fitness vector $\mathsf{f}' = (f_1', \dots, f_N')$ where $f_i' \approx \frac{f_i}{\Sigma \mathsf{f}} \approx \frac{f_i}{\kappa}$. Then $\mathfrak{p}'$ belongs to R_0 and the intensities $q_{\mathfrak{p},\mathfrak{q}}$ of the drift and selection chain (see (3.4) and (3.6)) in the region S are related to the intensities $q_{\mathfrak{p}',\mathfrak{q}'}$ of the corresponding points $\mathfrak{p}', \mathfrak{q}' \in R_0$ as follows:

$$q_{\mathfrak{p},\mathfrak{q}} \approx \kappa q_{\mathfrak{p}',\mathfrak{q}'}, \qquad \mathfrak{p}, \mathfrak{q} \in S.$$

At the same time, intensities of the mutation chain (see (3.3)) do not change in the transfer from $\mathfrak{p}$ to $\mathfrak{p}'$.

It follows that instead of thinking of the chain in S governed by $\mathfrak{A} = \mathfrak{S} + \mathfrak{M}$ we may think of the chain in R_0 governed by

$$\mathfrak{A}_\kappa := \kappa \mathfrak{S} + \mathfrak{M}. \tag{7.2}$$

Arguing as at the end of Section 10, we check that for each $\kappa > 0$, $\mathfrak{A}_\kappa$ is the generator of a semigroup, say $\{P_\kappa(t), t \geq 0\}$, of Markov operators. Thus, our task is that of characterizing the limit

$$\lim_{\kappa \to \infty} P_\kappa(t) x$$

where $x \in \ell^1$ is a distribution concentrated in R_0.

This can be done effectively via Kurtz's singular perturbation theorem [12, 19, 20] or Chapter 42 in [3], and the analysis does not require assuming that x is concentrated in R_0. In a simple case needed in our situation Kurtz's theorem says that the limit above exists for all $x \in \ell^1$ and $t > 0$ provided the following two conditions are met:

(i) $\lim_{t \to \infty} P_S(t) x =: \Pi x$ exists for all $x \in \ell^1$.

(ii) $\Pi \mathfrak{M}$ with domain equal to ℓ_0^1 is a generator in ℓ_0^1, where $\ell_0^1 := \mathit{Range}\, \Pi$.

To deduce this statement from Theorem 42.2 in [3] note that for $x \in \mathit{Ker}\, \mathfrak{S} = \mathit{Range}\, \Pi$, $\mathfrak{A}_\kappa x = \mathfrak{M} x$ and that for $y \in \mathcal{D}(\mathfrak{S}), \kappa^{-1} \mathfrak{S}_\kappa \to \mathfrak{S} y$. As proved in Theorem 5.1, the first of these two conditions is satisfied, and the second is clear since $\Pi \mathfrak{M}$ is bounded; this establishes the desired convergence. Moreover, Kurtz's theorem states that in such a case,

$$\lim_{\kappa\to\infty} P_\kappa(t)x = \mathrm{e}^{t\Pi\mathfrak{M}}\Pi x, \qquad t > 0. \tag{7.3}$$

An analogous reasoning shows that instead of thinking about the chain generated by $\mathfrak{S}+\mathfrak{M}$ in the lower region R_l, one may think of the chain in R_0 governed by (compare (7.2))

$$\mathfrak{A}_\varepsilon = \kappa\mathfrak{S} + \mathfrak{M},$$

where now $\kappa \ll 1$.

The limit

$$\lim_{\kappa\to 0} P_\kappa(t)x$$

where $x \in \ell^1$ may be found with the help of the Sova-Kurtz version of the Trotter–Kato convergence theorem for semigroups (see [3, 12, 18]): since for each $x \in \mathcal{D}(\mathfrak{S})$ $\lim_{\kappa\to 0}\mathfrak{A}_\kappa x = \mathfrak{M}x$, and the set $\mathcal{D}(\mathfrak{S})$ is dense in l^1, we have

$$\lim_{\kappa\to 0} P_\kappa(t)x = P_M(t)x. \tag{7.4}$$

Proofs of the two results are deferred to the supplement. The first one seems to be less intuitive, but we may provide an elementary derivation for the finite dimensional case in which it is sufficient to use Laplace transform and matrix calculus.

Conjecture: Given matrix exponent

$$\Phi(t) = \mathrm{e}^{(M+\kappa D)t}$$

such that

$$\mathrm{e}^{Dt} \to \Pi, \ \ t \to \infty$$

we have

$$\mathrm{e}^{(M+\kappa D)t} \to \mathrm{e}^{\Pi M t}\,\Pi, \ \ \kappa \to \infty$$

Proof: Consider the Laplace transform of $\Phi(t)$

$$\hat{\Phi}(s) = (sI - (M+\kappa D))^{-1}$$

We find that

$$\Pi = \lim_{t\to\infty} \mathrm{e}^{Dt} = \lim_{s\downarrow 0} s(sI - D)^{-1} = \lim_{\kappa\to\infty}(I - \kappa D)^{-1}$$

and then

$$\hat{\Phi}(s) = (sI - (M+\kappa D))^{-1} = [sI + (s-1)(I-\kappa D)^{-1}\kappa D - (I-\kappa D)^{-1}M]^{-1}(I-\kappa D)^{-1}$$

and since

$$(I-\kappa D)^{-1}\kappa D = (I-\kappa D)^{-1} - I$$

this converges as $\kappa \to \infty$ to

$$(s\Pi - \Pi + I - \Pi M)^{-1}\Pi.$$

Since $\Pi^2 = \Pi$ the above is equal to $(sI - \Pi M)^{-1}\Pi$, which is the Laplace transform of $\mathrm{e}^{\Pi M t}\,\Pi$, as desired.

8. Interpretation of (7.3) and (7.4)

This section is devoted to interpreting the limit theorems just obtained.

8.1. **Interpretation of** (7.3)**.** Let us start by characterizing the space ℓ^1_0 of point (ii) of the previous section. To this end, for $\alpha, \beta \in \mathbb{N}$, let

$$\mathfrak{p}_{\alpha,\beta} \in \mathfrak{P}$$

be the population of N identical individuals, each of type (α,β), and let $\mathbb{Y} \subset \ell^1$ be the subspace spanned by $e_{\mathfrak{p}_{\alpha,\beta}}, \alpha, \beta \in \mathbb{N}$ (recall (5.1)). In other words, $\mathbb{Y}$ is composed of vectors of the form

$$x = \sum_{\alpha,\beta\in\mathbb{N}} \xi_{\alpha,\beta} e_{\mathfrak{p}_{\alpha,\beta}}$$

where $\sum_{\alpha,\beta} |\xi_{\alpha,\beta}| < \infty$.

Lemma 8.1. *We have*

$$\ell^1_0 = \mathbb{Y}.$$

Proof. Each $e_{\mathfrak{p}_{\alpha,\beta}}$ obeys $P_S(t)e_{\mathfrak{p}_{\alpha,\beta}} = e_{\mathfrak{p}_{\alpha,\beta}}$, because $\mathfrak{p}_{\alpha,\beta}$ is an absorbing state for the selection/mutation chain. By the definition of Π it follows that $\Pi e_{\mathfrak{p}_{\alpha,\beta}} = e_{\mathfrak{p}_{\alpha,\beta}}$, i.e. that $e_{\mathfrak{p}_{\alpha,\beta}} \in \ell^1_0, \alpha, \beta \in \mathbb{N}$. Conversely, in the argument preceding Theorem (5.1) we have shown that for any $\mathfrak{p} \in \mathfrak{P}$, $\Pi e_{\mathfrak{p}}$ is a convex combination of (a finite number of) vectors $\mathfrak{p}_{\alpha,\beta}$, hence is a member of $\mathbb{Y}$. Since $e_{\mathfrak{p}}$'s span the entire ℓ^1, this completes the proof. $\square$

This lemma shows that the state-space of the limit process (which mimics the selection/mutation/drift process in the regions of high total fitness) is composed of populations in which all individuals are identical: this state-space is

$$\mathfrak{U}\mathfrak{P} := \{\mathfrak{p} \in \mathfrak{P}; \mathfrak{p} = \mathfrak{p}_{\alpha,\beta} \text{ for some } \alpha, \beta \in N\}.$$

The generator, $\Pi\mathfrak{M}$, of this process is of interesting form. The value of $\mathfrak{M}$ on $x \in \ell_0^1$ usually does not belong to ℓ_0^1. This is because $\mathfrak{M}$ describes mutations: since each and every member of a uniform population may increase the number of driver and/or passenger mutations, after some time the population may contain different variants. However, the generator of the process under consideration is a composition of $\mathfrak{M}$ and Π, and the latter operator maps $\mathfrak{M}x$ back to ℓ_0^1. This corresponds to immediate intervening of drift and selection force, which makes the population uniform again, although possibly not quite the same as previously.

Referring back to (3.3) we see $\Pi\mathfrak{M}$ is the generator of a Markov chain in $\mathfrak{U}\mathfrak{P}$ which may be described as follows. Staring at a $\mathfrak{p}_{\alpha,\beta}$ the process stays there for an exponential time with parameter $N\lambda$. After this time is over, a randomly selected individual (each individual being chosen equally likely) changes its type to $(\alpha + 1, \beta)$ with probability p, or to $(\alpha, \beta + 1)$ with probability q. Then drift and selection either eliminates the new variant or allows it to take over the entire population. According to Theorem 4.4, the variant with new driver mutation is fixed with probability

$$p_{\text{driv}} = \frac{1}{1 + \frac{1}{1+s} + \cdots + \frac{1}{(1+s)^{N-1}}} = \frac{s}{1+s}\left(1 - \left(\frac{1}{1+s}\right)^N\right)^{-1},$$

Similarly, the variant with new passenger mutation is fixed with probability

$$p_{\text{pass}} = \frac{d}{1-d}\left(\left(\frac{1}{1-d}\right)^N - 1\right)^{-1}.$$

If the new variant is eliminated from the population, everything goes back to the state from before mutation. Thus, the time to effective change is exponential with parameter

$$N\lambda(pp_{\text{driv}} + qp_{\text{pass}}),$$

and after this time $\mathfrak{p}_{\alpha,\beta}$ becomes $\mathfrak{p}_{\alpha+1,\beta}$ with conditional probability $\frac{pp_{\text{driv}}}{pp_{\text{driv}}+qp_{\text{pass}}}$ or $\mathfrak{p}_{\alpha,\beta+1}$ with conditional probability $\frac{qp_{\text{pass}}}{pp_{\text{driv}}+qp_{\text{pass}}}$.

If $\mathfrak{U}\mathfrak{P}$ is identified with $\mathbb{N} \times \mathbb{N}$, i.e. if each population $\mathfrak{p}_{\alpha,\beta} \in \mathfrak{U}\mathfrak{P}$ is identified with its type (α, β), the process described above is seen to be the pair of two independent Poisson processes on $\mathbb{N} \times \mathbb{N}$: the driver Poisson process with intensity $N\lambda p p_{\rm driv}$ (increasing the α-coordinate) and the passenger Poisson process with intensity $N\lambda q p_{\rm pass}$ (increasing the β-coordinate).

Notably, in contrast to the processes of mutations in single individuals, where intensity of driver mutations is much smaller than that of passenger mutations, here the situation is quite the opposite: these are the driver mutations that are typically more frequent than the passenger mutations. For, we have $p_{\rm pass} \le N^{-1}$; on the other hand, arguing as in the vicinity of (4.6), we see that it suffices to take s so large that

$$\frac{1}{1+s} < 1 - r$$

to have

$$p_{\rm driv} > r$$

where $r \in (0, 1)$ is given in advance.

The latter phenomenon has its source in the intervening selection process, described above, which eliminates the vast majority of passenger mutations from the population.

Before completing this section, we take a last look at (7.3) and note that this formula informs us also that even though the 'true' initial distribution is a member of ℓ^1 and needs not belong to ℓ^1_0, the drift and selection process intervenes so rapidly that before the process of mutations starts the population becomes uniform (this is described by the vector Πx). Again, if, for example, in the initial population there is one dominant variant and a single new variant with larger fitness then the latter variant may be fixed with probability given in Theorem 4.4.

8.2. **Intepretation of (7.4).** Interpretation of (7.4) is much simpler. This formula simply says that in the lower region drift and selection events are so rare that in fact may be disregarded: the chain behaves nearly as if there were no selection or drift. As a result each individual evolves independently of the others, in agreement with intuitions set forth in Section 6.2.

9. Simulations

As stated in the Introduction, we consider a population of a fixed number N of individuals, each of them characterized by a pair of integers (α, β), corresponding to the numbers of drivers and passengers in its genotype, respectively. This pair determines the fitness f of the individual by the formula

$$f = (1 + s)^\alpha (1 - d)^\beta,$$

where $s > 0$ and $d \in (0,1)$ are parameters describing selective advantage of driver mutations over passenger mutations. Thus, the entire population may be identified with the vector

$$\mathfrak{p} = ((\alpha_1, \beta_1), \ldots, (\alpha_N, \beta_N))$$

of N pairs of integers, with the accompanying vector

$$\mathsf{f} = (f_1, \ldots, f_N)$$

of fitnesses.

In each step of simulation, the decision is made whether the next event is the death and replacement or mutation event. Let us denote by T_m and T_s the exponentially distributed and independent random times to the mutation event and to a death/replacement event, respectively. We simulate both times and the next event occurs at time $t + \min(T_m, T_s)$, where t is the current time. According to the rules of our process, with mutation rate per cell λ

$$T_m \sim \exp(N\lambda)$$

while, following Equ. (3.6)

$$T_s \sim \exp\left(\sum_{\alpha,\beta \in \mathbb{N}} n_{\alpha,\beta} \sum_{\gamma,\delta\in\mathbb{N}:\ (\gamma,\delta)\neq(\alpha,\beta)} n_{\gamma,\delta}(1+s)^{\gamma}(1-d)^{\delta} \right)$$

where $n_{\alpha,\beta}$ denote the number of individuals of type (α,β), so that $\sum_{i,j\in\mathbb{N}} n_{ij} = N$.

Mutation If $T_m < T_s$, the next event is mutation. Given this, the index of individual undergoing mutation is drawn from discrete uniform distribution on $\{1, \ldots, N\}$ and the event changes the state of the individual to either $(\alpha+1, \beta)$ or $(\alpha, \beta+1)$ with probabilities $p \in (0,1)$ and $q = 1 - p$ respectively. Fitness of the mutated individual is recalculated accordingly.

Death and replacement If $T_m \geq T_s$, the next event is death and replacement. Suppose that K types of individuals are present, with respective counts $n(k)$, $k = 1, \ldots, K$, summing up to N. Following Equ. (3.4), individual i with state (α_i, β_i) is replaced by individual j with state (α_j, β_j), such that $(\alpha_j, \beta_j) \neq (\alpha_i, \beta_i)$ with probability proportional to fitness of j, $f_i = (1+s)^{\alpha_j}(1-d)^{\beta_j}$. The replaced individual inherits the state and fitness from the replacement.

Trends in fitness We present a set of stochastic simulations illustrating the richness of possible behaviors of the Tug-of-War process, in its complete and limit versions. One of the issues that we attend to is *criticality*, understood here as the trend of the fitness trajectories, upward, downward or neutral. There are two mathematical facts that provide guidance:

- Mutation vs. selection coefficient balance Equ. (2.4), which indicates the trend in fitness absent drift is determined by the sign of $ps - qd$.
- Theorem 4.3, which states that the death and replacement events absent mutation lead to a positive trend in fitness.

Fitness trajectory in the complete process is the result of the interaction of the two trends. If $ps > qd$, then the influx of advantageous mutations prevails and in addition, the drift works towards their fixation. As a result the fitness increases rather fast. The effect is subtler when $ps \leq qd$. If the influx of disadvantagous mutations prevails but is not too strong, drift affords to purge the deleterious mutants before they may be fixed. A strong influx is needed to flip this trend.

Complete process We follow the interplay between the fluxes ps of advantageous and qd of deleterious mutations, but also between mutation and drift. The latter can be varied for example by adjusting the parameter $L = N\lambda$. Figure 2 depicts distribution of individual fitness averaged over N individuals, in 30 independent runs of the model with a range of parameters. Panel A depicts the case with a $ps \gg qd$ (for exact parameter values, see the Figure legend), resulting exponential-like growth of fitness. Panel B shows the case $ps = qd$, with the effect being a slow increase of fitness for most runs and a very slow decrease for some. Panel C shows the case of slightly negative trend in mutations $ps < qd$. Panel D demonstrates that if $ps = 0$, then drift may efficiently keep purging recurrent deleterious mutants. Panels E and F are showing that in case of increased flux of mutants (large L) small changes in the value of s parameter, from s=0.01 in Fig. 2E to s=0.05 in Fig. 2F, may cause a fraction of average fitness trajectories display an upward trend despite large amount of highly deleterious mutations.

Subsequent figures depict runs of a single trajectory of fitness in the process with a range of parameters. Figure 3 depicts one of the average trajectories of Fig. 2A. Panel A depicts the average fitness of population. Mutation events are marked with red (driver) and blue (passenger) asterisks. Let us notice that major jumps in population fitness arise as a result of death-replacement events, more so than of the mutational events, since new arising mutants are frequently purged by death-replacement. Panel B depicts time succession patterns of clones started by driver mutations colored according to fitness of given clone. Panel C depicts genealogies of the clones initiated by drivers and passengers. Lines between nodes represent driver (red) and passenger (blue) mutations. Green circles mark clones alive at the end of simulation ($t = 20$). The vertical axis in panel **C** does not coincide with the time or even with the strict order of clone appearance. However, it is consistent with the ancestor-descendant relationship.

Figures 4, 5, 6, 7, 8 and 9 depict single-trajectory plots corresponding all other cases in Fig. 3. Figures corresponding to cases with high mutation rates lack the third panel, since the genealogies of clones become to dense to follow with an increased mutation rate.

Limiting process

Suppose that all individuals have the same fitness $(1+s)^\alpha(1-d)^\beta$. The difference between expected fitness right after mutation/drift event and the fitness before this event equals

$$(1+s)^\alpha(1-d)^\beta(psp_{\mathrm{driv}} - qdp_{\mathrm{pass}}). \tag{9.1}$$

In interpreting this relation we encounter an apparent paradox: in a certain range of parameters, an increase of d, that is, a decrease of fitness of passenger mutants, leads to a decrease of the studied difference. In order to explain this paradox we need to consider the function

$$e(d) = dp_{\mathrm{pass}}. \tag{9.2}$$

Lemma 9.1. *The function e initially increases and then decreases with d.*

Proof. As d increases from 0 to 1, $x := \frac{1}{1-d}$ increases from 1 to ∞. Hence, it suffices to check monotonicity of

$$g(x) := \frac{(x-1)^2}{x^{N+1}-x}, \qquad x \in [0,\infty).$$

Since

$$g'(x) = \frac{(x-1)((1-N)x^{N+1} + (N+1)x^N - x - 1)}{(x^{N+1}-x)^2} \tag{9.3}$$

monotonicity of g is determined by the sign of

$$h(x) := (1-N)x^{N+1} + (N+1)x^N - x - 1, \qquad x \geq 1. \tag{9.4}$$

Here $h(1) = 0$ and

$$h'(x) = (1-N)(N+1)x^N + N(N+1)x^{N-1} - 1, \tag{9.5}$$

so that in particular $h'(1) = N > 0$. Moreover,

$$h''(x) = (N-1)N(N+1)x^{N-2}(1-x) < 0, \qquad x > 0, \tag{9.6}$$

proving that $h'(x)$ strictly decreases from N to $-\infty$ in the interval $[1, \infty)$. Hence, h increases from $h(1) = 0$ to a maximum point, and then starts to decrease. Since $\lim_{x\to\infty} h(x) = -\infty$, we conclude that h is initially positive, and then, beyond a certain points, say x_0 becomes negative and stays negative for all $x > x_0$. It follows that g increases up to x_0 and then starts to decrease, as claimed. □

Consider first the scenario in which e decreases with the increase of d, and thus the difference (9.1) increases. Here, everything seems to agree with our intuition: A decrease in fitness of passenger mutants causes the probability of their fixation to drop and thus if the fitness of driver mutations is the same, the overall population fitness grows faster, because the influence of passenger mutations is smaller.

However, in a certain range of d, an increase of d (a decrease of fitness of passenger mutations) causes an increase of e, and thus a decrease of (9.1). This is because an increase in d causes a decrease of p_{pass} but this is accompanied by an increase of the first factor in (9.2). It is possible that a change in d causes a much smaller change in p_{pass} than in d itself, and thus may result in the overall growth of e. In other words, even though the probability of fixation of a passenger mutation is lower, if such a variant is fixed fitness will drop radically. From this point of view, the observation that a decrease in the fitness of passenger mutants may lead to a decrease of (9.1) is not surprising.

To summarize, analysis of e, the expected drop of the population fitness given that a passenger mutant was fixed, is the key to understanding of the apparent paradox we encountered. It is more informative than the probability p_{pass} alone.

The influence of the function

$$s \mapsto s p_{\text{driv}}$$

is monotonous; the larger is the fitness of driver mutants, the faster is the growth of the fitness of the entire population.

10. Discussion

Following the earlier work of McFarland et al. [21, 22, 23, 24] we build a model of early cancer development which accounts for the influence of two types of mutations: rare drivers with higher fitness, and more frequent passengers with smaller fitness. Mathematically, the model is a continuous-time Markov chain with state-space composed of N-tuples of pairs of non-negative integers. Here, N is the number of individuals (cells) in the population under study, and is assumed constant; the first coordinate of each pair (individual/cell) is the number of accumulated driver mutations, whereas the second is the number of mutations of passenger type; the resulting individual's fitness is given by (2.1).

The model may be seen as describing competition of two population genetic forces: selection combined with drift, on one side, and mutations, on the other. Interestingly, the mathematical theory of semigroups of operators, our main tool, allows analysing consequences of these two forces separately, and to infer properties of the full model from the properties of its two components.

The main effect of the first of these components, related to selection combined with drift, is that a population that may initially be heterogeneous, becomes increasingly homogeneous with time. For the associated Markov chain this means that after a random time the process reaches an absorbing state in which all individuals have the same counts of passenger and driver mutations. The corresponding probabilities of fixation of a mutant and the expected times to fixation are calculated in Section 4.

Mutations, on the other hand, introduce new variants to the data at the epochs of a Poisson process; either selectively advantageous drivers, or disadvantageous passengers.

Mathematical analysis of analytical and stochastic properties of the processes related to the two main factors described above allows concluding that they may be combined, and that the new Markov chain that encompasses mutation and selection, on one had, as well as mutations, on the other, is non-explosive. In other words, the underlying stochastic process is a well-defined honest Markov chain.

The resulting process is difficult to analyze. Insights can be obtained using a simpler limit model, presented in Sections 7 and 8, and simulations, see Section 9.

The limit theorem is obtained using the theory of convergence of semigroups of operators [3], and corresponds to the scenario in which the total fitness of the population exceeds certain threshold. The model then predicts that drift and selection events are much more frequent than mutation events. Under such scenario, when a new mutant arises, regardless of whether it is a driver or a passenger, it is almost instantly fixed in the population or completely removed from it. It is the action of the drift and selection chain that causes fixation or removal and favors driver mutations. Therefore, the probability of instantaneous fixation of a passenger mutant is usually smaller than that of a driver mutant. However, because the passengers may arise more often than drivers, the possibility of fixation of passenger mutant is not negligible (see Section 7 for details).

In summary, the limit model state-space is composed of pairs of non-negative integers; this is because each individual is fully characterized by such a pair, and the entire population is composed of identical individuals. At a time a new mutant arises, it is instantly fixed or removed from the population, with probabilities depending on its fitness, and so the population is again homogeneous. Such model seems to account for the influence of driver and passenger mutations. It is interesting that it clearly displays the non-monotonous dependence on the parameter d of passenger fitness (Lemma 9.1). Simulations in Figure 10 fully corroborate theoretical analysis.

In addition, simulations show a similar effect in the complete model, as depicted in consecutive panels of Figure 2. In the complete model, the non-monotonicity of the $e(d)$ corresponds to the balance between downward and upward trends of subsets of trajectories in Figure 2F. The balance is delicate: if the influx of deleterious passenger mutants is limited, drift and selection purge the mutants and population fitness keeps increasing. Only when the influx is sufficiently large, population fitness decreases in part of realization of the process.

We studied by simulation a range of special cases of fitness trajectories and pedigrees of clones originating from driver mutations. A theory of such clones in the Tug-of-war process is still missing. Simulations show how rich is the behavior of this process (Figures 3-9).

With all the reservations, present paper places McFarland's Tug of War model into the rigorous framework of Moran Model, which allows analyzing it using the well-developed toolbox of time-continuous Markov chains and theory of operator semigroups. Let us notice that our formulation is different from McFarland's original model as spelled out in [21, 22, 23]. The model there is a state-dependent branching process. To our best knowledge, these models were not rigorously explored. One of the subsequent papers from McFarland's group [24] explores experimentally the dependence of fitness on the rate of deleterious passenger mutations. However, the references to the mathematical model are only qualitative.

Supplement: The related semigroup of Markov operators

In Section 4, we have studied in detail the Markov chain related to the intensity matrix Q_S. Since in this chain only a finite number of states can be reached from any given starting point, existence of such chain is an elementary matter. Existence of the chain of mutations (i.e. that related to Q_M) is also clear, as this chain consists of two independent Poisson processes (one for drivers and one for passengers). The question we have never answered is whether there is a Markov chain related to the intensity matrix Q of (3.7).

This question is non-trivial because there are so-called *explosive* intensity matrices that are so 'poorly designed' that they do not determine the related Markov chain: additional rules need to be specified to describe the chain after the random time of *explosion* (see [4, 9, 25] and references given there). According to the theorem of Kato [16] (discussed e.g. in [1] Chapter 5, [2] pp. 334–338, [3] pp. 74-80, [4] Chapter 3 and [15] pp. 642–647; see also Section 4 in [13] for W. Feller's proof of this result), for any intensity matrix, whether explosive or not, there is a related minimal Markov chain which, however, is undefined after explosion.

Therefore, in this section we show that Q of (3.7) is non-explosive and our argument boils down to the statement that the sum of two intensity matrices, one of which is non-explosive and the other is bounded, is non-explosive. This statement is most naturally proved in the language of semigroups of Markov operators, as we will now explain. Such semigroups are analytical tools for treating Markov chains, and in the later chapters we will use them extensively.

The analysis involves the space $\ell^1 = \ell^1(\mathfrak{P})$ of functions $x : \mathfrak{P} \to \mathbb{R}$ which, because $\mathfrak{P}$ is a countable set, can be considered sequences $x = (\xi_\mathfrak{p})_{\mathfrak{p}\in\mathfrak{P}}$ where $\xi_\mathfrak{p} = x(\mathfrak{p})$ is the value of x at $\mathfrak{p} \in \mathfrak{P}$. Elements $(\xi_\mathfrak{p})_{\mathfrak{p}\in\mathfrak{P}}$ of ℓ^1 such that $\xi_\mathfrak{p} \ge 0, \mathfrak{p} \in \mathfrak{P}$ and $\sum_{\mathfrak{p}\in\mathfrak{P}} \xi_\mathfrak{p} = 1$ are termed *distributions*.

With each time-continuous Markov chain with values in $\mathfrak{P}$ one may associate the probabilities $p_{\mathfrak{p},\mathfrak{q}}(t)$ that the chain starting at a $\mathfrak{p} \in \mathfrak{P}$ will be at a $\mathfrak{q} \in \mathfrak{P}$ at time $t \ge 0$. These so-called transition probabilities are conveniently gathered in the matrices

$$P(t) = (p_{\mathfrak{p},\mathfrak{q}}(t))_{\mathfrak{p},\mathfrak{q}\in\mathfrak{P}},$$

which in turn may be identified (see [4], Chapter 2 for details) with the operators in ℓ^1 defined by the formula

$$P(t)x = x \cdot P(t), \qquad x \in \ell^1, t \ge 0,$$

where $x \cdot P(t)$ is the product of a row-vector x and the matrix $P(t)$. All $P(t)$'s are Markov operators in that they leave the set of densities invariant: if x is an initial distribution of the chain, then $P(t)x$ is its distribution at time $t \geq 0$. Moreover, the Markov property of the chain is expressed in the semigroup property:

$$P(t)P(s) = P(s+t), \qquad s,t \geq 0.$$

Under mild, natural assumptions on transition probabilities we also have

$$\lim_{t \to 0+} \|P(t)x - x\| = 0, \qquad x \in \ell^1.$$

These properties are summarized in the statement that $\{P(t), t \geq 0\}$ is a *strongly continuous semigroup of operators* in ℓ^1.

Thus, with each Markov chain we have the associated (uniquely determined) strongly continuous semigroup of Markov operators. Conversely, if all Markov chains with the same transition probabilities are identified, one may speak of *the* Markov chain related to a strongly continuous semigroup of Markov operators.

There are two commonly used infinitesimal descriptions of strongly continuous semigroups of Markov operators in ℓ^1. First, (see e.g. [2, 12, 15]) a strongly continuous semigroup determines and is determined by its generator $\mathfrak{A}$, defined by

$$\mathfrak{A}x = \lim_{t \to 0} t^{-1}(P(t)x - x),$$

on the domain $\mathcal{D}(\mathfrak{A})$ composed of x such that the limit on the right-hand side exists. Second, as proved by Doob [10] (see [4, 14]) the limits, called intensities,

$$q_{\mathfrak{p},\mathfrak{p}} := \lim_{t \to 0} \frac{p_{\mathfrak{p},\mathfrak{p}}(t) - 1}{t}, \qquad \mathfrak{p} \in \mathfrak{P},$$
$$q_{\mathfrak{p},\mathfrak{q}} := \lim_{t \to 0} \frac{p_{\mathfrak{p},\mathfrak{q}}(t)}{t}, \qquad \mathfrak{p},\mathfrak{q} \in \mathfrak{P}, \mathfrak{p} \neq \mathfrak{q}$$

exist, and $q_{\mathfrak{p},\mathfrak{q}}, \mathfrak{p} \neq \mathfrak{q}$ are finite. However, even if all intensities are finite, knowing the entire intensity matrix $Q := (q_{\mathfrak{p},\mathfrak{q}})_{\mathfrak{p},\mathfrak{q} \in \mathfrak{P}}$ is not equivalent to knowing $\mathfrak{A}$. For, whereas $\mathfrak{A}$ contains the entire information on the semigroup $\{P(t), t \geq 0\}$, the matrix Q in general does not. Nevertheless, if Q is non-explosive, Q and $\mathfrak{A}$ may be somewhat identified: for a typical $x \in \mathcal{D}(\mathfrak{A})$, the product $x \cdot Q$ can be computed, and $\mathfrak{A}x$ turns out to be equal to this product. For explosive intensity matrices this is not the case; see e.g. the already cited [1, 4], and in particular Chapter

3 in [4]. In fact, for an explosive matrix there are many different Markov chains, many different semigroups of Markov operators and many different generators related to this matrix.

Coming back to the chain of interest, Section 4, and specifically Theorem 4.1 imply that there exists a Markov chain related to the intensity matrix Q_S of (3.4) and (3.6) which is well-defined for all $t \geq 0$. In particular, Q_S is non-explosive. This is because the related chain reaches an absorbing state by passing through a finite number of transient states. This rules out the possibility of explosion, since an exploding chain is passing through an infinite number of states in a finite time. Therefore, by Kato' Theorem (see the references earlier on), there is a unique strongly continuous semigroup of Markov operators $\{P_S(t), t \geq 0\}$ in ℓ^1 with the generator, say $\mathfrak{S}$, identified with Q_S.

The case of intensity matrix Q_M of (3.3) is simpler, because all its entries are bounded in absolute value by $N\lambda$, while a bound does not exits for the matrix (3.4)–(3.6). It follows that for *any* $x \in \ell^1$ the product $x \cdot Q_M$ may be computed and belongs to ℓ^1, where x is a row-vector, and the map

$$\ell^1 \ni x \mapsto \mathfrak{M}x := x \cdot Q_M$$

is bounded. Hence, Q_M may be identified with a bounded linear operator $\mathfrak{M}$ and the semigroup of Markov operators related to Q_M may be defined as the exponent of this operator:

$$P_M(t) = \sum_{n=0}^{\infty} \frac{t^n \mathfrak{M}^n}{n!}, \qquad t \geq 0.$$

(See e.g. [4] Section 2.3 for details.). Further, operator $\mathfrak{M}$ is the generator of semigroup $\{P_M(t), t \geq 0\}$.

Boundedness of the operator $\mathfrak{M}$ guarantees that the operator

$$\mathfrak{A} := \mathfrak{S} + \mathfrak{M}$$

is well-defined on $\mathcal{D}(\mathfrak{A}) := \mathcal{D}(\mathfrak{S})$ and, in view of the Phillips Perturbation Theorem (see e.g. [2, 4, 12, 15]), is a generator of a strongly continuous semigroup, say $\{P(t), t \geq 0\}$. On the other hand, using the Trotter Product Formula, which says (see the monographs cited above) that

$$P(t)x = \lim_{n\to\infty} \left[P_S\left(\frac{t}{n}\right) P_M\left(\frac{t}{n}\right)\right]^n x, \qquad t \geq 0, x \in \ell^1,$$

we check that this semigroup is composed of Markov operators, because so are $\{P_S(t), t \geq 0\}$ and $\{P_M(t), t \geq 0\}$. It can be argued that this semigroup describes the minimal Markov chain related to Q. But, by Kato's Theorem, if Q were explosive, this semigroup could not be composed of Markov operators. This shows that Q is non-explosive, and thus that the minimal chain is well-defined for all times $t \geq 0$. It is this minimal chain related to Q that models the evolution of our population under selection, drift and mutations. In other words, by *the* Markov chain related to Q we mean the unique minimal chain related to this matrix: since Q is non-explosive this chain is well-defined for all $t \geq 0$.

Data Availability

The authors affirm that all data necessary for confirming the conclusions of the article are present within the article, figures, and tables.

Acknowledgments

AB was supported by the National Science Centre (Poland) grant 2017/25/B/ST1/01804. MK and MKK were supported by the National Science Centre (Poland) grant 2018/29/B/ST7/02550.

References

[1] J. Banasiak and L. Arlotti. Perturbations of Positive Semigroups with Applications. Springer, 2006.

[2] A. Bobrowski. Functional Analysis for Probability and Stochastic Processes. An Introduction. Cambridge University Press, Cambridge, 2005.

[3] A. Bobrowski. Convergence of One-parameter Operator Semigroups. In Models of Mathematical Biology and Elsewhere, volume 30 of New Mathematical Monographs. Cambridge University Press, Cambridge, 2016.

[4] A. Bobrowski. Generators of Markov Chains. From a Walk in the Interior to a Dance on the Boundary. Cambridge University Press, Cambridge, 2020.

[5] A. Bobrowski and M. Kimmel. Asymptotic behavior of joint distributions of characteristics of a pair of randomly chosen individuals in discrete-time Fisher–Wright models with mutations and drift. Theor. Popul. Biol., 66(4):355–367, 2004.

[6] A. Bobrowski and M. Kimmel. An Operator Semigroup in Mathematical Genetics. Springer Briefs in Applied Sciences and Technology. Mathematical Methods. Springer, 2015.

[7] A. Bobrowski, M. Kimmel, R. Chakraborty, and O. Arino. A semigroup representation and asymptotic behavior of the Fisher–Wright–Moran coalescent. In "Handbook of Statistics 19: Stochastic Processes: Theory and Methods", Chapter 8, C.R. Rao and D.N. Shanbhag, eds. Elsevier Science, Amsterdam, 2001.

[8] A. Bobrowski, N. Wang, R. Chakraborty, and M. Kimmel. Non-homogeneous infinite sites model under demographic change: Mathematical description and asymptotic behavior of pairwise distributions. Math. Biosci., 175(2):83–115, 2002.

[9] K. L. Chung. Lectures on boundary theory for Markov chains. With the cooperation of Paul-André Meyer. Annals of Mathematics Studies, No. 65. Princeton University Press, Princeton, N.J.; University of Tokyo Press, Tokyo, 1970.

[10] J. L. Doob. Stochastic Processes. Wiley, 1953.

[11] R. Durrett. Probability Models for DNA Sequence Evolution. Springer, New York, Second Edition, 2008.

[12] S. N. Ethier and T. G. Kurtz. Markov Processes. Characterization and Convergence. Wiley, New York, 1986.

[13] W. Feller. On boundaries and lateral conditions for the Kolmogorov differential equations. Ann. of Math. (2), 65:527–570, 1957.

[14] D. Freedman. Markov Chains. Holden-Day, Inc., 1971.

[15] E. Hille and R. S. Phillips. Functional Analysis and Semi-Groups. Amer. Math. Soc. Colloq. Publ. 31. Amer. Math. Soc., Providence, R. I., 1957.

[16] T. Kato. On the semi-groups generated by Kolmogoroff's differential equations. J. Math. Soc. Japan, 6(1):1–15, 03 1954.

[17] J. F. C. Kingman. Poisson Processes. The Clarendon Press, Oxford University Press, New York, 1993.

[18] T. G. Kurtz. Extensions of Trotter's operator semigroup approximation theorems. J. Functional Analysis, 3:354–375, 1969.

[19] T. G. Kurtz. A limit theorem for perturbed operator semigroups with applications to random evolutions. J. Functional Analysis, 12:55–67, 1973.

[20] T. G. Kurtz. Applications of an abstract perturbation theorem to ordinary differential equations. Houston J. Math., 3(1):67–82, 1977.

[21] C. D. McFarland. The role of deleterious passengers in cancer. PhD thesis, 2014.

[22] C. D. McFarland, K. S. Korolev, G. V. Kryukov, S. R. Sunyaev, and L. A. Mirny. Impact of deleterious passenger mutations on cancer progression. Proceedings of the National Academy of Sciences, 110(8):2910–2915, 2013.

[23] C. D. McFarland, L. A. Mirny, and K. S. Korolev. Tug-of-war between driver and passenger mutations in cancer and other adaptive processes. Proceedings of the National Academy of Sciences, 111(42):15138–15143, 2014.

[24] C. D. McFarland, J. A. Yaglom, J. W. Wojtkowiak, J. G. Scott, D. L. Morse, M. Y. Sherman, and L. A. Mirny. The damaging effect of passenger mutations on cancer progression. Cancer Research, 77(18):4763–4772, 2017.

[25] J. R. Norris. Markov Chains. Cambridge University Press, Cambridge, 1997.

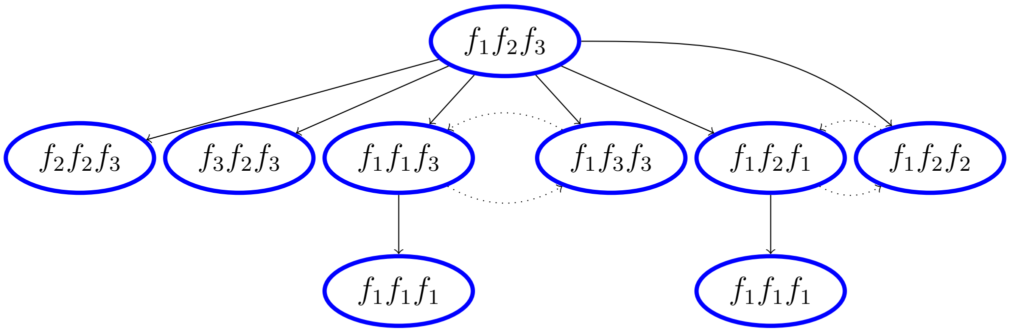

FIGURE 1. Calculating probability p_1 in the case $N = 3$. Dotted lines denote communication between events associated with (f_1, f_1, f_3) and (f_1, f_3, f_3), and (f_1, f_2, f_1) and (f_1, f_2, f_2).

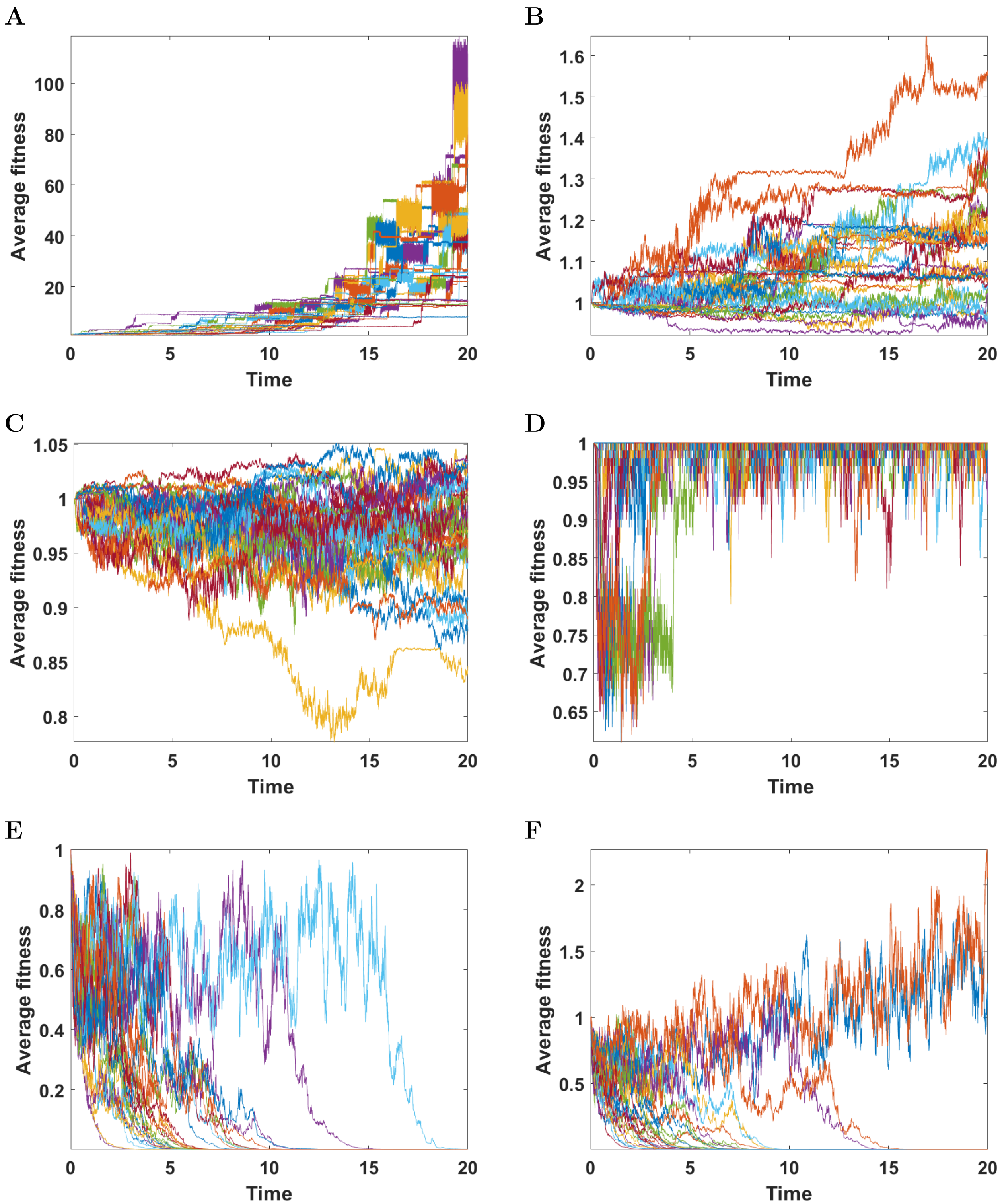

FIGURE 2. Average fitness of individuals. Results for 30 simulations with parameters: **A**: $s = 0.8$, $d = 0.05$, $p = 0.1$, $L = 5$, $N = 50$ (strong positive selection); **B**: $s = 0.1$, $d = 0.01111$, $p = 0.1$, $L = 5$, $N = 50$ (equilibrium); **C**: $s = 0.01$, $d = 0.05$, $p = 0.5$, $L = 5$, $N = 50$ (negative selection); **D**: $s = 0$, $d = 0.5$, $p = 0.5$, $L = 5$, $N = 50$ (lack of impact of driver mutations); **E**: $s = 0.01$, $d = 0.5$, $p = 0.1$, $L = 100$, $N = 20$; **F**: $s = 0.05$, $d = 0.5$, $p = 0.1$, $L = 100$, $N = 20$ (large mutation rate, passengers prevailing).

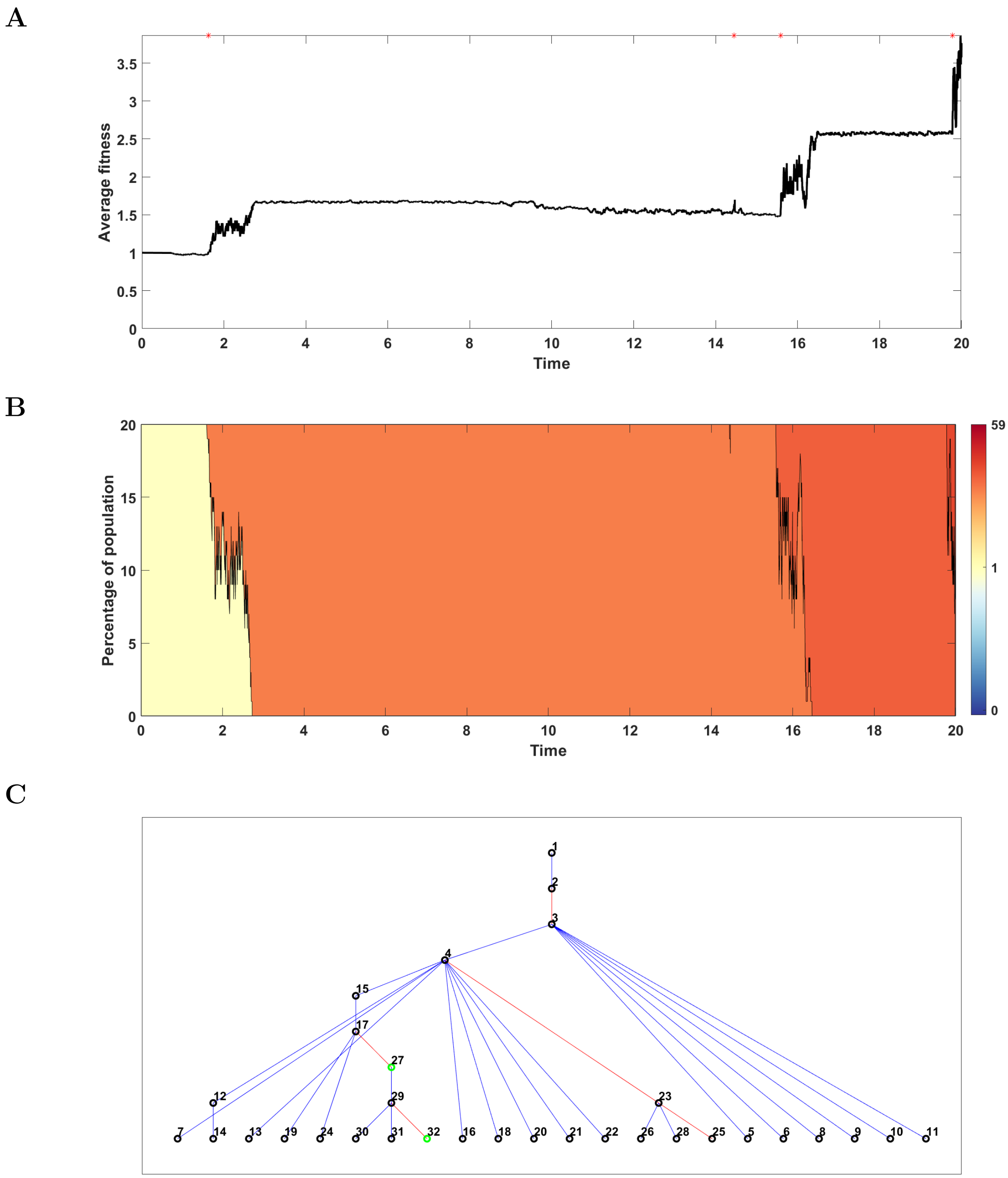

FIGURE 3. Results for one simulation on $N = 20$ individuals with parameters: $s = 0.8$, $d = 0.05$, $p = 0.1$, $L = 2$ (strong positive selection). **A**: Average fitness of population. Mutation events are marked with red (driver) and blue (passenger) asterisks; **B**: Time succession patterns of clones started with driver mutations colored according to fitness of given clone; **C**: Genealogies of the clones initiated by drivers and passengers. Lines between nodes represent driver (red) and passenger (blue) mutations. Green circles mark are clones alive at the end of simulation ($t = 20$).

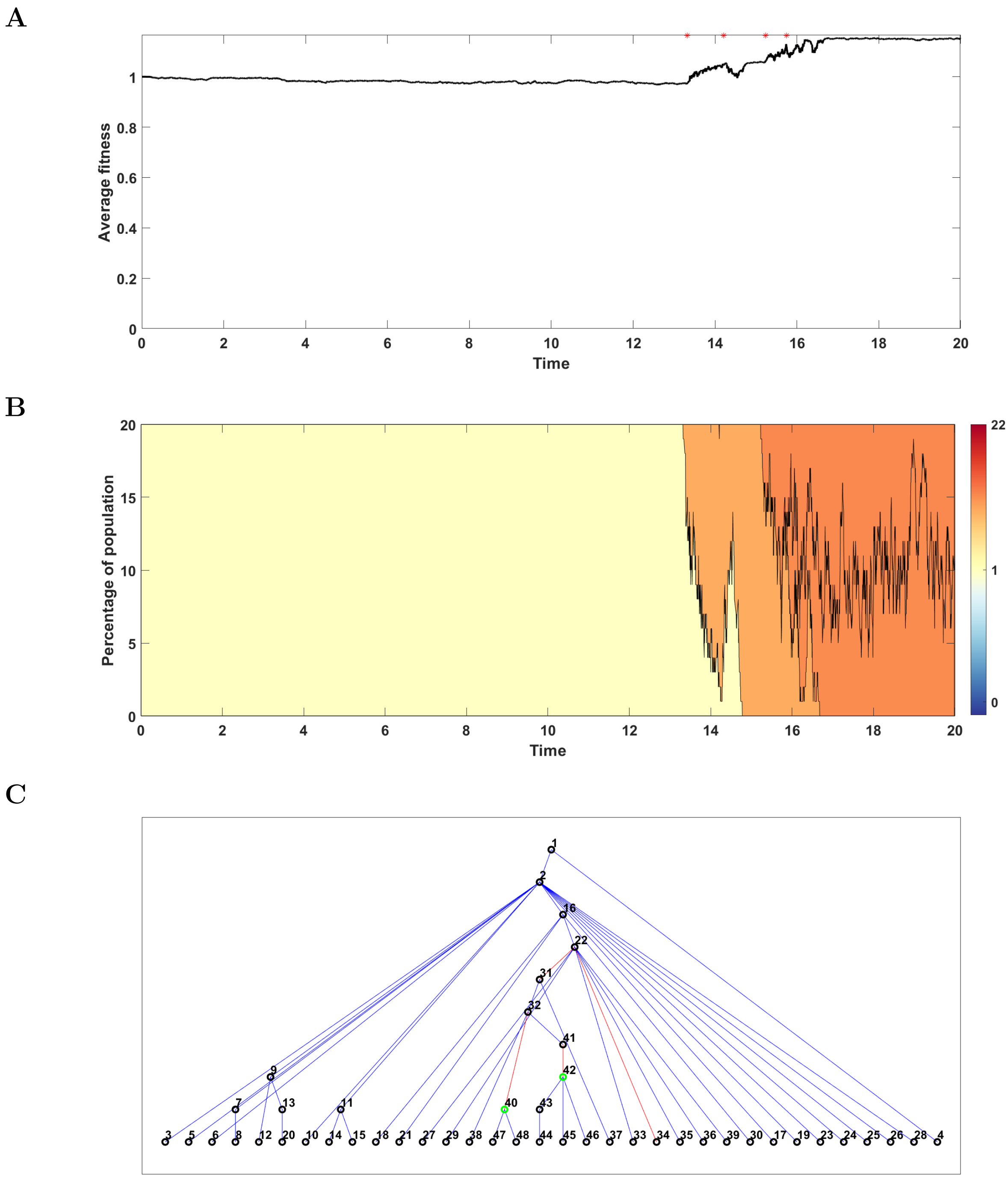

FIGURE 4. Results for one simulation on $N = 20$ individuals with parameters: $s = 0.1$, $d = 0.01111$, $p = 0.1$, $L = 2$ (equilibrium). **A**: Average fitness of population. Mutation events are marked with red (driver) and blue (passenger) asterisks; **B**: Time succession patterns of clones started with driver mutations colored according to fitness of given clone; **C**: Genealogies of the clones initiated by drivers and passengers. Lines between nodes represent driver (red) and passenger (blue) mutations. Green circles mark clones alive at the end of simulation ($t = 20$).

A

B

C

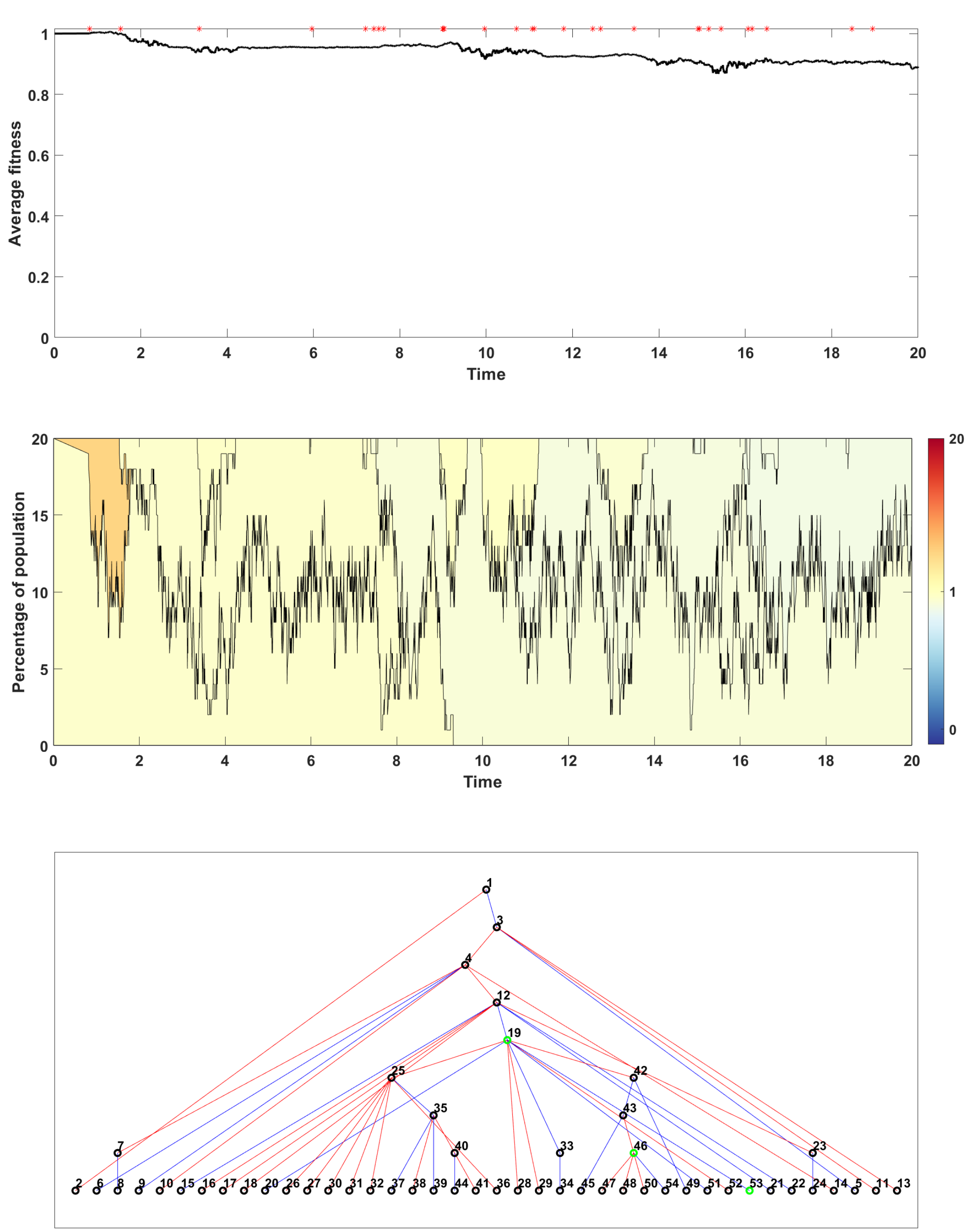

FIGURE 5. Results for one simulation on $N = 20$ individuals with parameters: $s = 0.01$, $d = 0.05$, $p = 0.5$, $L = 2$ (negative selection). **A**: Average fitness of population. Mutation events are marked with red (driver) and blue (passenger) asterisks; **B**: Time succession patterns of clones started with driver mutations colored according to fitness of given clone; **C**: Genealogies of the clones initiated by drivers and passengers. Lines between nodes represent driver (red) and passenger (blue) mutations. Green circles mark clones alive at the end of simulation ($t = 20$).

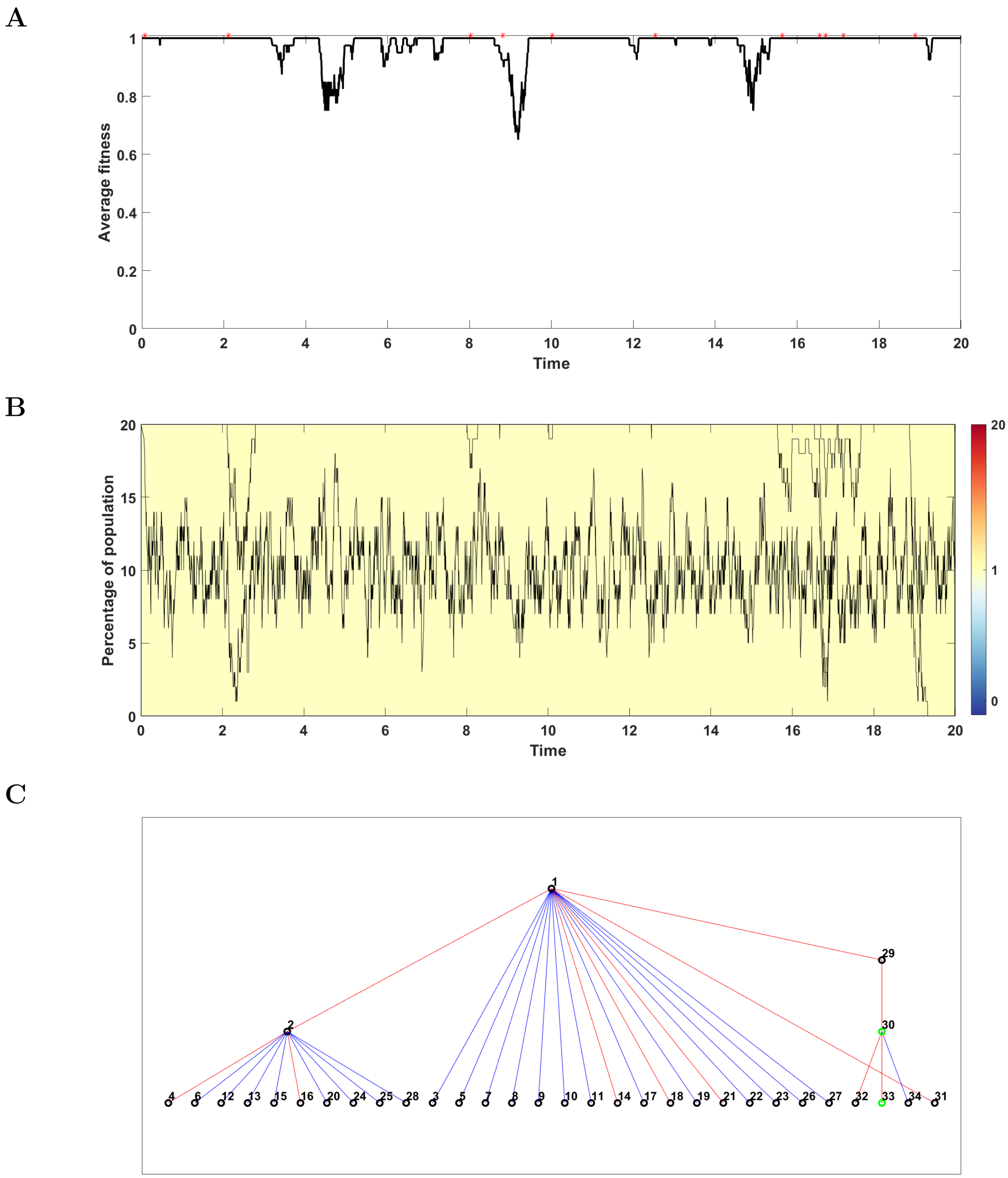

FIGURE 6. Results for one simulation on $N = 20$ individuals with parameters: $s = 0$, $d = 0.5$, $p = 0.5$, $L = 2$ (lack of impact of driver mutations). **A**: Average fitness of population. Mutation events are marked with red (driver) and blue (passenger) asterisks; **B**: Time succession patterns of clones started with driver mutations colored according to fitness of given clone; **C**: Genealogies of the clones initiated by drivers and passengers. Lines between nodes represent driver (red) and passenger (blue) mutations. Green circles mark clones alive at the end of simulation ($t = 20$).

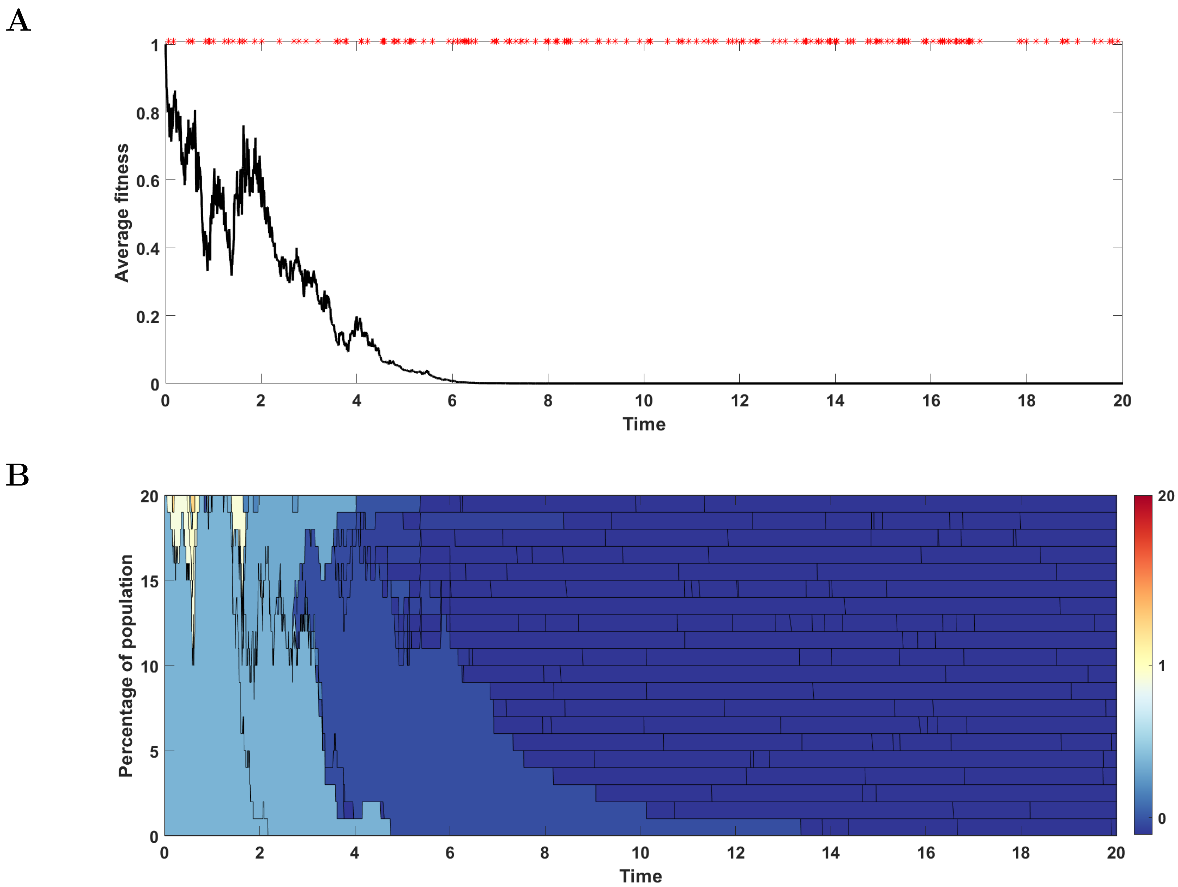

FIGURE 7. Results for one simulation on $N = 20$ individuals with parameters: $s = 0.01$, $d = 0.5$, $p = 0.1$, $L = 100$ (large mutation rate, passengers prevailing). **A**: Average fitness of population. Driver mutation events are marked with red asterisks; **B**: Time succession patterns of clones started with driver mutations colored according to fitness of given clone.

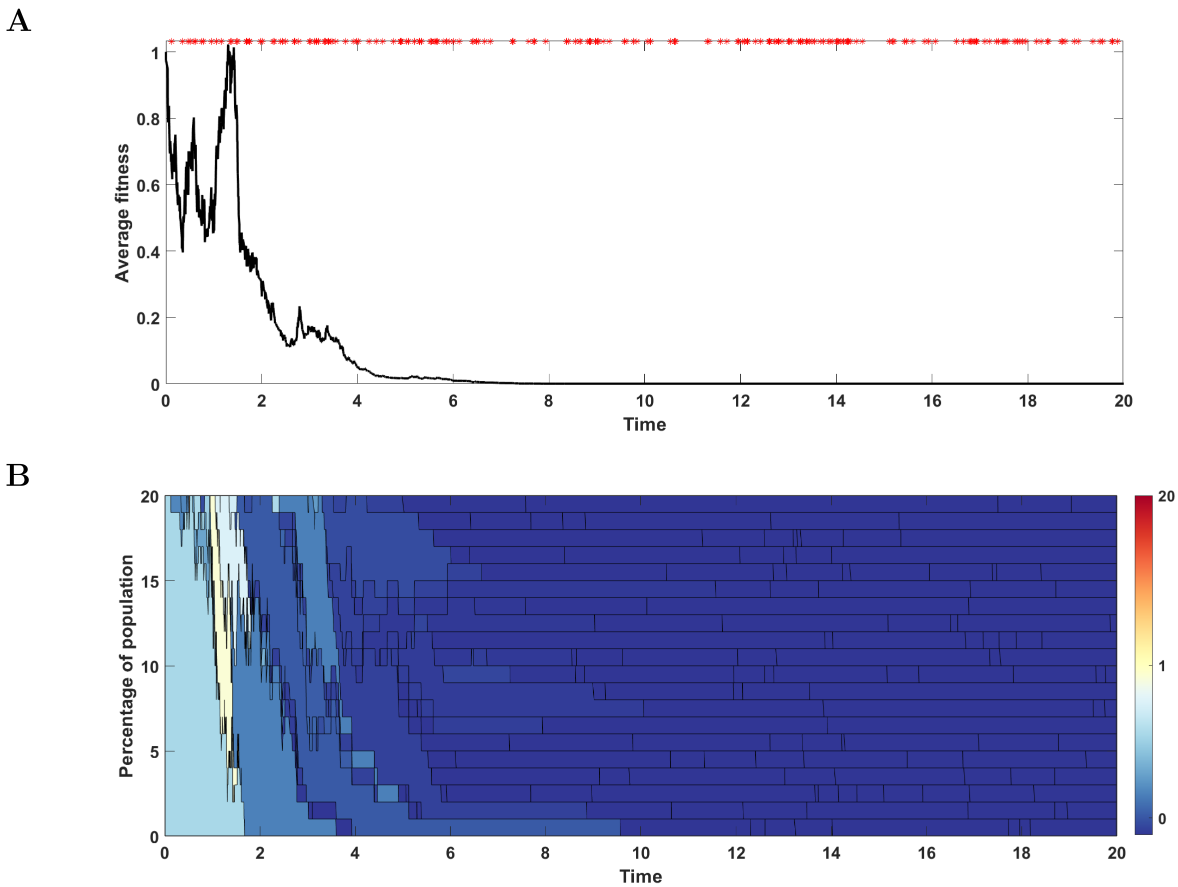

FIGURE 8. Results for one simulation on $N = 20$ individuals with parameters: $s = 0.05$, $d = 0.5$, $p = 0.1$, $L = 100$ (large mutation rate, passengers prevailing, case with decreasing fitness). **A**: Average fitness of population. Driver mutation events are marked with red asterisks; **B**: Time succession patterns of clones started with driver mutations colored according to fitness of given clone.

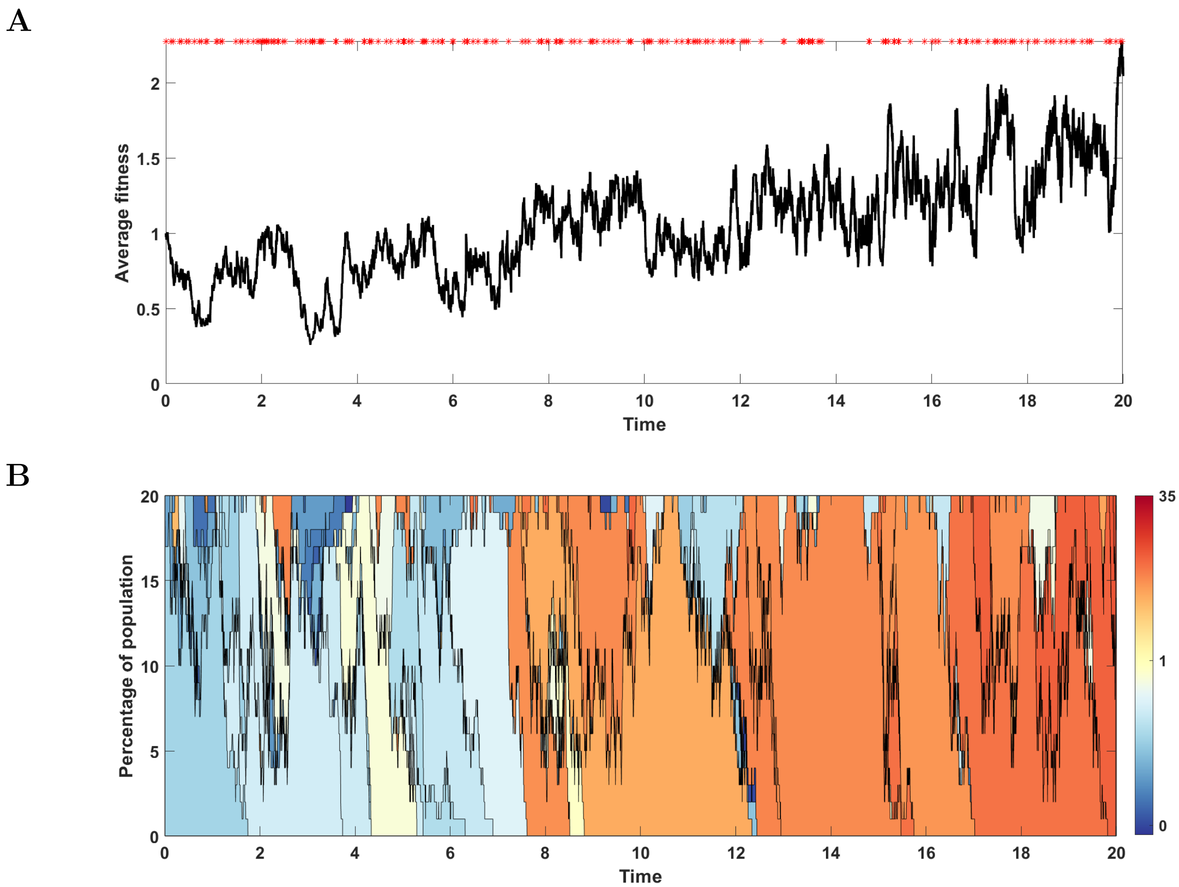

FIGURE 9. Results for one simulation on $N = 20$ individuals with parameters: $s = 0.05$, $d = 0.5$, $p = 0.1$, $L = 100$ (large mutation rate, passengers prevailing, case with increasing fitness). **A**: Average fitness of population. Driver mutation events are marked with red asterisks; **B**: Time succession patterns of clones started with driver mutations colored according to fitness of given clone.

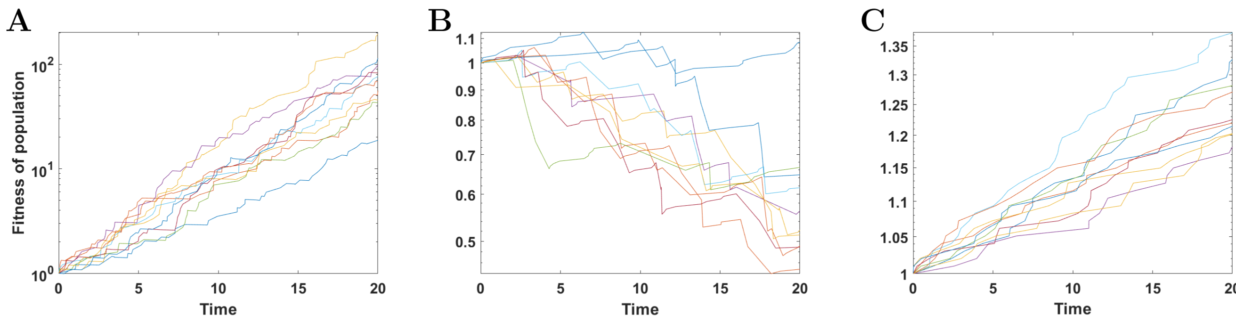

FIGURE 10. Results for 10 simulations of reduced process with $N = 20$ and remaining parameters: **A**: $s = 0.1$, $d = 0.01$, $p = 0.5$ ($sp > dq$); **B**: $s = 0.01$, $d = 0.1$, $p = 0.5$ ($sp < dq$); **C**: $s = 0.01$, $d = 0.5$, $p = 0.5$ ($sp < dq$). In all cases $L = 2$.)