Genome heterogeneity drives the evolution of species

Mattia Miotto [1,2,*,†] and Lorenzo Monacelli [1,*,‡]

[1]*Department of Physics, Sapienza University of Rome, Rome, Italy*
[2]*Center for Life Nanoscience, Istituto Italiano di Tecnologia, Rome, Italy*

Most of the DNA that composes a complex organism is noncoding and defined as *junk*. Even the coding part is composed of genes that affect the phenotype differently. Therefore a random mutation has an effect on the specimen fitness that strongly depends on the DNA region where it occurs. Understanding how this heterogeneous composition influences the fitness evolution of individuals is hampered by the complexity of the problem and a clear picture is missing. Here we study a minimal model for the evolution of an ecosystem where two antagonist species struggle for survival on a lattice. Each specimen has a unique toy genome that codes for its phenotype. The gene pool of populations changes in time due to the effect of random mutations on genes (entropic force) and of interactions with the environment and between individuals (natural selection). We prove that the relevance of each gene in the manifestation of the phenotype is a key feature for evolution. In the presence of a uniform gene relevance, a mutational meltdown is observed. Natural selection acts to quench the ecosystem in a nonequilibrium state that slowly drifts, decreasing the fitness and leading to the extinction of the species. Conversely, if a specimen is provided with a heterogeneous gene relevance, natural selection wins against entropic forces and the species evolves, increasing its fitness. We finally show that heterogeneity together with spatial correlations are responsible for spontaneous sympatric speciation.

I. INTRODUCTION

Evolution by natural selection shaped the marvelous biodiversity we presently observe in nature. Starting with one (or few) living organisms, the tree of life progressively branched as living beings managed to survive in different environments through adaptations and speciations [1].

Both processes arise by a complex interplay between intrinsic forces (e.g., mutations) and extrinsic ones, provided by the interactions between the different components of the ecosystem [2]. This results in changes in behavior, morphology, and physiology (or combinations thereof) in the organisms.

Since its first formulation, the theory of evolution regarded natural selection, i.e., the effect of environment and interaction between inter- and intraspecies in selecting the organisms with maximum fitness, as the pivotal mechanism of evolution. Several more years were required to clearly define the "microscopic" role of genome mutations (alongside genetic drift, hitchhiking, etc.) as the other main component of the evolutionary process [3–6].

Quantifying the simultaneous effects of natural selection and genomic mutations is far from being an easy task [6–9].

Its understanding is, however, not only a fascinating theoretical challenge but carries important practical implications. In fact, a continuously increasing literature highlights the connection between the rules governing ecology to the one regarding cell populations [10–12]. Preeminent is the case of cancer cells, that are subject to both high entropic forces (fast mutation rates) and a strong natural selection, due to the selective pressure given by both competition with the immune system and the effect of anticancer therapies. Bacteria drug resistance is under much scrutiny, too, since the selection done by antibiotics influences the evolution of resilient traits. Experimentally, the evolution of genomes during speciation has been studied only recently with the availability of next-generation sequencing technologies [13–15]. In particular, recent findings highlighted how mutations on certain genes considerably enhance the speciation process [13]. Also, the increase of the genome length due to duplication errors (e.g., the presence of redundant genes/chromosomes) has been pointed as one mechanism for evolution/speciation [16].

Parallel to experiments, several theoretical models were developed to study ecosystem dynamics [17,18] and evolution, both at the molecular level [7,19] and at the population level, such as population-environment interaction [20–23] and species-species interactions [24–31].

In particular, agent-based models proved to be very efficient to include spatial information and features [32–36]. Here evolution is accounted for either at phenotype level, i.e., the phenotype of a species is randomly modified generation by generation according to a particular law [21,23,31], or at genotype level, i.e., an evolution law is assumed for a genome upon which the phenotype is computed [37–40]. More refined models have been considered accounting for dominant and

---

*The authors contributed equally to the present work.

†mattia.miotto@roma1.infn.it

‡lorenzo.monacelli@uniroma1.it

recessive alleles, sex recombination (crossover), or spatially resolved ecosystem dynamics [37,41,42].

This work aims to study how heterogeneity in gene relevance affects the evolution of an ecosystem. Heterogeneity in genome expression is known to play an important role in evolution: the expression level of the encoding gene strongly affects the evolution rate of proteins [43]. Here we focus on the effect of this heterogeneity on the overall fitness, and how this affects the dynamics of the whole ecosystem. In particular, we study the evolution of a minimal ecosystem on a lattice initially composed by one species of predators and one of preys. Each specimen possesses a toy genome composed of $3N$ genes that encode for three essential macrophenotypic features of the animal: (i) its capability of moving/hunting, (ii) its fertility, and (iii) the mean lifetime of the specimen (mortality).

The genome pool of the two prey-predator species can change in time due to random mutations at the level of the single genes (entropic force) and to predation/death events (selection force). Aiming at limiting the number of parameters and developing an as minimal as possible evolutionary model, we do not insert mating in the model, i.e., we have an asexual reproduction. Since mating is not taken into consideration, we will not look for speciation according to Mayr's biology speciation concept, where different species are separated if mating does not produce fertile offsprings [5]. Instead, we speak of speciation in terms of differences in phenotype distributions [44], where well-separated peaks can be interpreted as different species while the width of each peak accounts for the intraspecies differentiation.

Two different kinds of genomes are considered: (i) a uniform genome, where all the genes have the same impact on the phenotype, and (ii) a heterogeneous genome, where each gene has a different weight on the overall phenotype. We show that while in the first case entropy dominates, fostering the mutational meltdown [45–47], in the latter case natural selection allows the ecosystem to increase predator fitness. For each genome we also look for the emergence of spontaneous speciation.

II. MODEL

We consider a variant of the ECOLAT (ecosystem on lattice) model [35], where each site of a $L \times L$ square lattice can be occupied exclusively either by the environment or a prey or a predator. To use the same notation adopted in [35], we identify preys as fishes (f) and predators as sharks (s). At every discrete time step, fishes can move, breed, or remain still with probability p_f^m, p_f^f and $1 - p_f^m - p_f^f$. Sharks can move (p_s^m) or remain still ($1 - p_s^m$). Furthermore, predators eat preys whenever they step into a cell occupied by a prey. In this case, sharks can reproduce with probability p_s^f. If a shark does not eat a fish during its round, it can die for starvation with probability p_s^d. We assumed that fishes may only die murdered by a predator (i.e., $p_f^d = 0$). See Appendix A for more details on the simulation steps.

The set of three probabilities, $\{p_{(s,f)}^m, p_{(s,f)}^f, p_{(s,f)}^d\}$, constitutes the macrophenotype of each individual, dictating the rates with which the specimen carries out three essential tasks

of life. Each specimen has a genome that codes the macrophenotype of the individual. The macrophenotype is obtained as a weighted average over all the N genes (g_i) that codes for a particular feature:

$$p_y = \sum_{i=1}^{N} g_{yi} W_i, \quad y = \begin{smallmatrix}(m,f,d)\\(s,f)\end{smallmatrix}, \tag{1}$$

where W_i expresses the weight of the ith gene in the manifestation of the phenotype [see Fig. 1(a)].

The genome is subject to point mutations, i.e., Poisson-distributed random events in time occurring with a constant rate for each gene every time a new individual is born (reproduction is asexual). If a mutation event occurs on a gene, its value is reset to a random number between 0 and 1.

In our framework the outcome of a mutation does not depend on the value the gene possesses before the mutation, as is often the case in real mutation of a nucleic basis. However, since we are modeling effective genes directly related to the phenotype, this is a necessary requirement to simulate the effect of silent mutations. In fact, changes in the phenotypes are often due to the accumulation of mutations, which by themselves do not influence the phenotype.

Moreover, in order to tune the average value of the new mutated gene, we choose the new value $g_i = x$ according to the following probability distribution:

$$q(x) = \left(\frac{1}{\langle x \rangle} - 1 \right)(1-x)^{\frac{1}{\langle x \rangle} - 2}, \tag{2}$$

where $\langle x \rangle$ is a tunable parameter that determines the average value of the gene after the mutation [see Fig. 1(b)]. Such kinds of asymmetric distributions are widely used for phenotype modeling, e.g., in the contest of bacterial growth rates, where slow rates dominate over fast ones [21,22,48]. In this way deleterious mutations can be enforced to be more common than favorable ones. This is quite reasonable, as a random mutation in the coding DNA produces a random change in the amino acid chain of a protein, which is far more likely to produce an unfolded structure than a more functional one [49,50].

Finally, in this work we will consider two weight distributions [see Fig. 1(c)]:

(1) A uniform distribution, where all genes equally contribute to manifest the phenotype (discussed in Sec. III A),

$$W_i = \frac{1}{N}. \tag{3}$$

(2) A power-law distribution allowing for the presence of preeminent genes, as discussed in Sec. III B,

$$W_i \propto i^{-\alpha}. \tag{4}$$

To assess the outcome of evolution, we define the fitness measured as the average size of the population. Such a definition, profoundly linked with the usual fitness measurements (growth rate or the number of nephews per individual), is enforced by the finite carrying capacity dictated by the lattice. Additional details on the simulations and the link to the source code to reproduce our results are provided in Appendixes A and B.
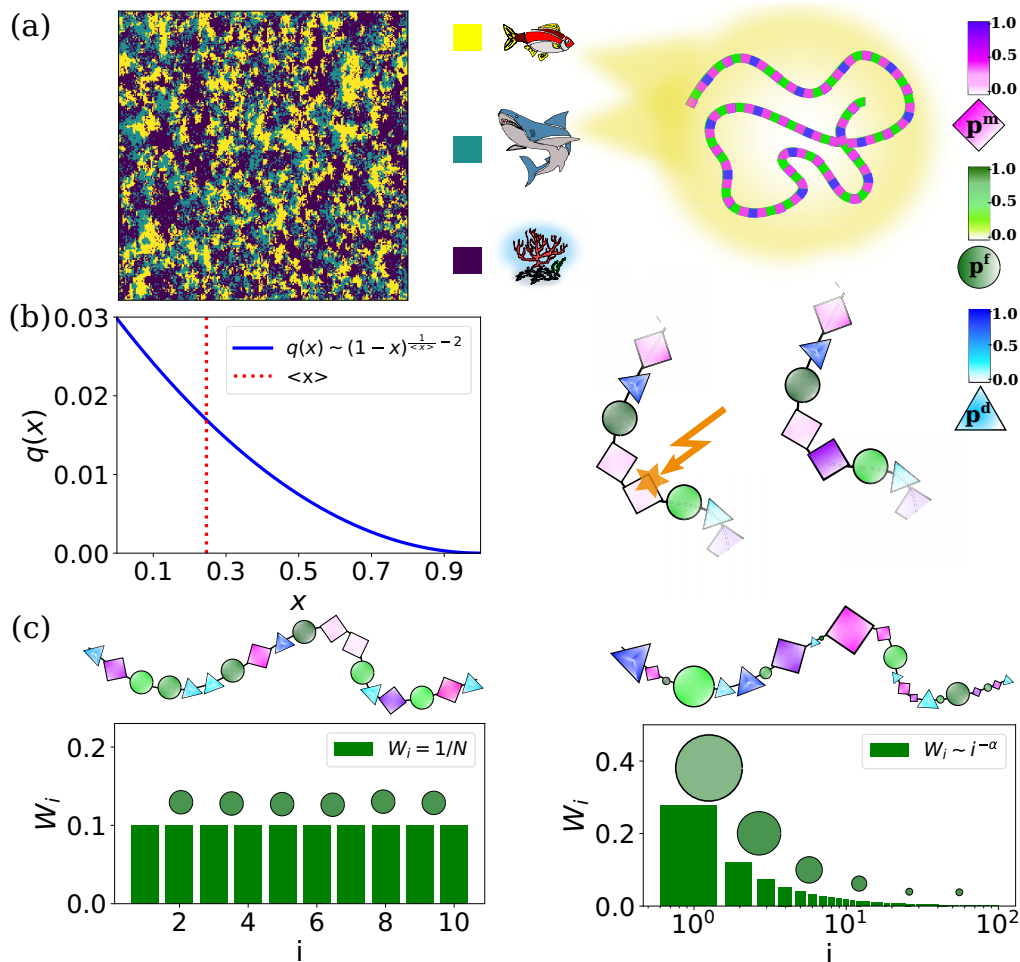
FIG. 1. (a) Snapshot of an EVOLAT configuration in the metastable steady-state regime. Fishes are colored in yellow, sharks in blue, while purple cells represent the environment. Each animal on the lattice possesses a toy genome composed of $3N$ genes. Each gene is represented by a real number between 0 and 1 that concurs in manifesting three macrophenotypes associated with the animals. Animal mobility is represented by the probability of moving (p^m), their fertility by the breeding probability (p^f), and their lifetime by the death probability (p^d). (b) Mutation events can happen with a rate μ with equal probability on each gene. If one gene is subject to a mutation, a new value between 0 and 1 is extracted from the underline phenotype distribution $q(x)$, whose asymmetry accounts for the fact that mutations tend to be deleterious for the organisms. (c) The phenotype of the individual is obtained as the weighted average over all the N independent genes encoding for it. In the present work we assume two possible kinds of weights with a uniform mean (all genes equally concur to the phenotype) or a power-law weighted mean (some genes influence the phenotype more than others).

III. RESULTS

A. Uniform genome

To begin with, we investigated the model behavior in the presence of a uniform genome where each phenotype is determined as the uniform average over all the genes associated with that phenotype [as given by (1) with (3)]. After preparing the ecosystem in a nonequilibrium steady state (NESS) with fixed initial phenotypes, we turned on mutations and followed the evolution of prey's and predator's traits over long periods. To prevent predators to acquire infinite lifetimes, we strongly favored deleterious mutations in the shark death rate (entropic force). This is achieved by tuning the $\langle x \rangle$ parameter in (2). In fact, if the average phenotype is above (or below) $\langle x \rangle$, random mutations will try to restore it around $\langle x \rangle$.

Figure 2(a) shows the dynamic of the predator mortality distribution for various initial conditions (marked by different

colors). In all simulations, sharks have a well-defined phenotype at each time step, so no speciation is observed. All distributions converge toward a unique final distribution with a mean close to the average value chosen for the entropic force (dashed line in the figure), even with different timescales. Note that the final state is metastable since the presence of a finite death rate for sharks introduces a non-null probability of fluctuation in the shark number, leading to either prey of predator extinction.

All the observed metastable states are indistinguishable, so no memory on the starting parameters is preserved. The time required to reach the final state diverges following quite remarkably a power law:

$$\tau\left(p_{s\,0}^d\right) = A p_{s\,0}^{d-\gamma}, \qquad \lim_{N\to\infty} \gamma = 2.57, \qquad (5)$$

where τ is the transient time before the metastable steady state is reached, $p_{s\,0}^d$ is the initial value of the predator death
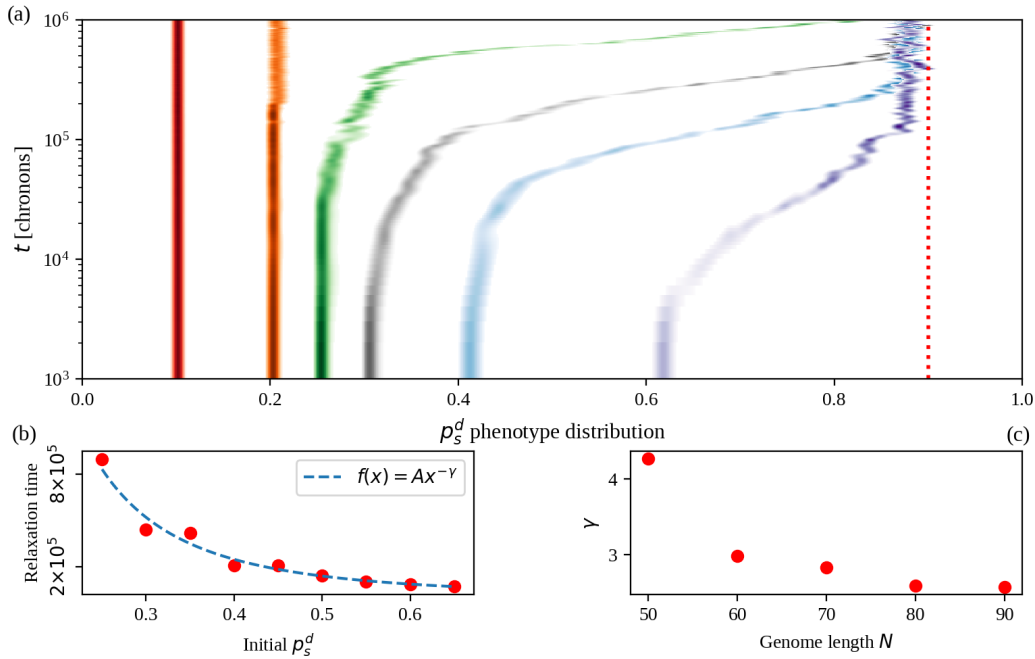
FIG. 2. (a) Time evolution of shark mortality distribution. The dashed vertical line is the stationary state for predator growing in a breeding farm (no natural selection). The simulations were performed starting from all values of p_{sd} with a spacing of 0.05. For presentation reason, only a subset of the simulation are reported with different colors. (b) Relaxation time as a function of the initial phenotype. The relaxation times below $p_{sd} = 0.25$ are not shown, as those simulations did not reach the final metastable state in the time of 10^6 steps. As seen, a power law perfectly fits the relaxation time. The predator relaxation time diverges as a power law in a way that resembles the dynamic of glassy systems. (c) Exponent of the power law as a function of the genome length. It converges to a value of about 2.57. Simulation in panels (a) and (b) are performed with $N = 90$.

rate, and N is the number of genes that determine the features of each individual [see Fig. 2(b)]. We repeated the simulations by varying the genome size N from 50 to 90. All the simulations show the same divergence of the relaxation time, perfectly fitted by the power law of Eq. (4) with a γ exponent that depends on the genome size, but converges quickly after $N = 80$, as shown in Fig. 2(c). The divergence of the transient time as p_s^d becomes smaller is similar to a system close to a glassy phase [51] where the distance between the initial and final phenotypes plays the role of the difference between the quenched and equilibrium temperatures in a typical quenching experiment. Such behavior provides great insight into the evolutionary dynamics. This quenching is a result of natural selection that freezes the system in a nonequilibrium state for long times. In fact, phenotypes subject to the sole effect of entropic forces would converge to the final metastable steady state exponentially; the dynamics of the average p_s^d of individuals subject only to random mutation without natural selection converges to the entropic value [the dashed red line in Fig. 2(a)] as

$$p_s^d(t) = \left[p_s^d(0) - \langle sd \rangle \right] e^{-p_{\mathrm{mut}}t} + \langle sd \rangle , \qquad (6)$$

where p_{mut} is the mutation rate and $\langle sd \rangle$ is the average entropic value for the shark death rate (see Appendix C for the derivation). So, in the absence of natural selection, the typical time to reach $\langle sd \rangle$ is always the mutation rate $1/p_{\mathrm{mut}}$, independent of the initial value. This analysis unveils the role played by natural selection, which acts stronger the more the individuals are fitted, quenching the dynamics.

The death rate of predators is strongly related to their fitness. In particular, we define the fitness of a species at a given time as the average number of individuals in the lattice. This accounts for the finite carrying capacity of the system. We cannot use the growth rate, as is usually done, since we are studying a system with a finite carrying capacity. On sufficiently long timescales with respect to the typical oscillation of the system [35] but small with respect to the typical mutation rates, the mean number of individuals is constant and the mean growth rate is zero. In Fig. 3 we report the correlation between the fitness and the death rate, computed from all the configurations extracted in Fig. 2.

Furthermore, we see how the ecosystem and predators, in particular, react to negative mutations. In fact, at long times, shark fitness decreases due to the entropic force of the mutations that increases the death rate. If a population has a phenotype distribution whose mean is far from the mean entropy value [$\langle x \rangle$ in (2)], a mutation of the genome will tend to produce an individual with lower fitness than the rest of the population. The lower the fitness the faster this individual will be suppressed by natural selection.

If the fitness worsens only slightly (the mutation is very soft), the specimen will reproduce and that deleterious mutation will remain inside the genome pool. On long timescales, the accumulations of these very soft negative mutations slowly drive the whole population toward the entropic limit. This phenomenon is observed in real ecosystems and is known as mutational meltdown [45–47], which may cause the extinction of the population. The meltdown we observe originates by the even impact genes have on the phenotype. In general, to
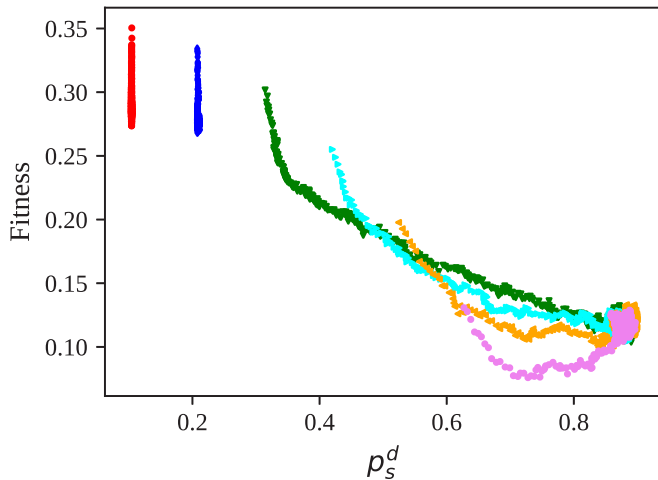
FIG. 3. Correlation between the fitness of predators and their death rate for the simulations in Fig. 2. Colors distinguish different starting values for the shark death rate. The Pearson correlation is –0.98. The residual quadratic error on the fitness from the linear fit is 0.01. Therefore the shark death rate is a very good estimator for the fitness of the predators.

prevent the mutational meltdown, a huge population is required. With a huge population, the genome variability improves the effect of natural selection, preventing the meltdown. We discuss how it scales with the system-carrying capacity in Appendix D. In particular, we show that in our model the timescale τ increases sublogarithmically with the number of individuals (i.e., the dimension of the lattice).

Figure 2 clearly shows how evolution is not able to improve the fitness of predators. In fact, although natural selection is freezing the dynamics (strongly reducing the effect of genetic drift), if we wait a sufficient time, sharks will finally increase their death rate, therefore reducing their fitness. This behavior is opposite to what is commonly observed in nature, where the combined effect of mutations and natural selection leads to an overall improvement of the fitness on long times.

This finding highlights the deleterious role of entropy in the evolution of the species. We will show in the next section that the capability to improve the fitness can be recovered even in the presence of a strong entropic force if genome variability is considered. This feature alone will be able both to allow predators to evolve and to provide a mechanism for spontaneous speciation.

B. Heterogeneous genome

To include the effect of genome heterogeneity in the simulation, i.e., the uneven impact that different genes have on the phenotype, we inserted a gene-dependent weight [Eq. (4)] in Eq. (1). Figure 4 depicts the time evolution of the distribution of the shark's mortality starting from different initial values for two different exponents of (4), namely, $\alpha = 1$ and $\alpha = 2$. The evolution is observed with both choices of exponents. Predators improve their expected lifetime and fitness (see the correlation in Fig. 3) much above the final metastable state obtained in Fig. 2. The improvement of the shark lifetime is a general feature, as discussed in Appendix B, where we report
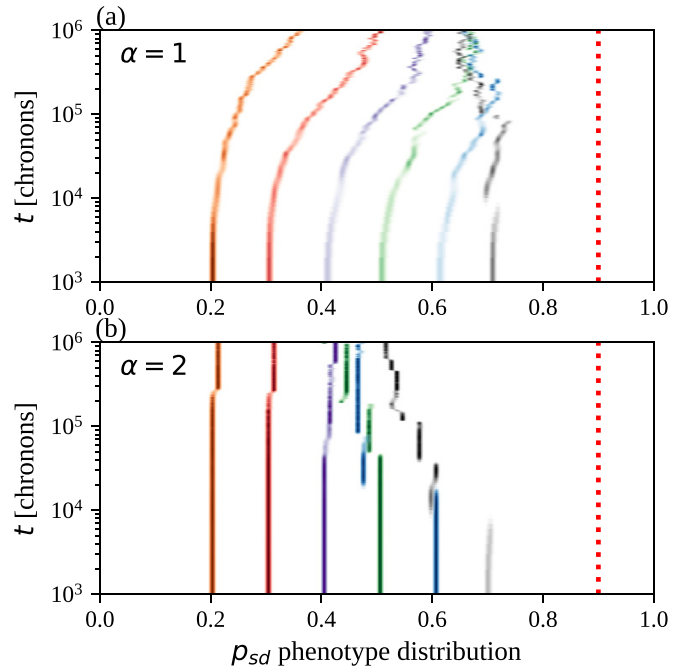

FIG. 4. (a) Evolution of the shark death-probability distribution as a function of time in the presence of a heterogeneous genome. The power-law exponent α [see (4)] is equal to 1. Heterogeneity allows the species to evolve to higher average lifetime, while this does not occur in the presence of a uniform genome (Fig. 2). Different colors correspond to different starting values. (b) Same as in (a) but with $\alpha = 2$. Both simulations exhibit a qualitative different behavior if compared to a uniform genome. The less fitted individual can improve their fitness far above the entropic value.

several simulations performed with the same parameters and all showing the same qualitative behavior. Such behavior is in strong contrast to what happens with the uniform genome, as we show in Fig. 2.

Positive evolution occurs thanks to the combination of two effects: (i) the phenotype population is normally quenched (see Sec. III A) and (ii) genes exert a different role on the phenotype manifestation. The quenching allows the quasispecies to survive in an out-of-equilibrium situation where common random negative mutations usually kill the individuals. The species can survive long enough that a very rare, positive mutation occurs in an important gene (with a big W_i), causing a discontinuity in the phenotype distribution and a sudden evolution of the population. This speciation creates two kinds of predators with a quantitative different phenotype. The quasispecies which is more fitted to the environment very quickly gets fixed while the less fitted alleles vanish from the gene pool of the population.

Such a rare, discontinuous event is not possible in a uniform genome, where each gene has only a moderate impact on the overall phenotype and a massive number of positive mutations would be required to obtain the same shift (that is so rare an event that it never occurs in practice). Conversely, heterogeneity allows the positive discontinuity in the phenotype to depend on the mutation of a relatively small number of genes, which is much more likely to occur. The resulting dynamics is something that never reaches a metastable state, as the
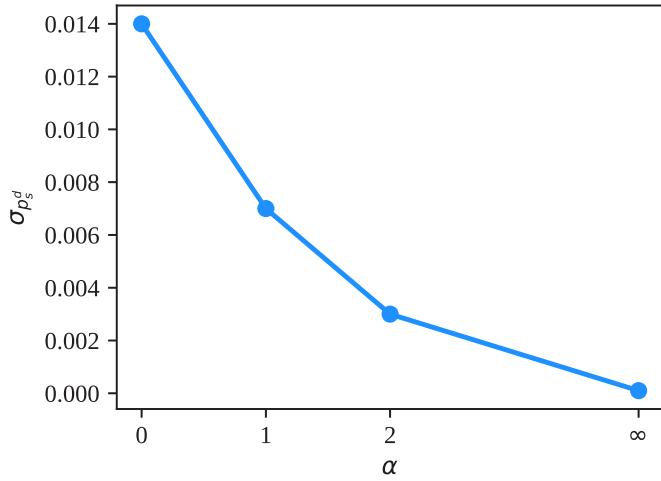
FIG. 5. Root square of the variance of the shark death probability as a function of the genome heterogeneity, measured thought the α exponent.

TABLE I. Pearson spatial correlations of intra- and interspecies and conditional probabilities. f_1 refers to fishes with p_f^f lower than 0.83, while f_2 refers to those with p_f^f higher than 0.83.

Corr[s, s]	0.418 ± 0.001	$P(s|s)$	0.590 ± 0.001
Corr[f_1, f_1]	0.575 ± 0.002	$P(f_1|f_1)$	0.641 ± 0.002
Corr[f_2, f_2]	0.594 ± 0.003	$P(f_2|f_2)$	0.632 ± 0.002
Corr[f_1, f_2]	-0.113 ± 0.002	$P(f_1|f_2)$	0.024 ± 0.001
Corr[s, f_1]	-0.142 ± 0.003	$P(s|f_1)$	0.144 ± 0.001
Corr[s, f_2]	-0.105 ± 0.003	$P(s|f_2)$	0.144 ± 0.001

more fitted the population becomes, the more quenched is the dynamics. This reflects the intrinsic nonequilibrium character of evolving ecosystems. Notably, increasing the heterogeneity reduces the differentiation between individuals, which can be measured by the width of the phenotype distributions (as shown in Fig. 5).

Heterogeneity has another important effect on preys phenotype distribution. In Fig. 6 we show the prey breeding rate as it evolves as a function of time. After a while, the prey population splits into two well-separated traits. In particular, the two species have a different reproduction rate (p_f^f) of 0.78

and 0.90, respectively. To assess whether the two populations could be considered as different species, apart from their different fertility, we computed the spatial Pearson coefficient:

$$\text{Corr}[i, j] = \frac{P(i \cap j) - P(i)P(j)}{\sqrt{P(i \cap i) - P^2(i)}\sqrt{P(j \cap j) - P^2(j)}}, \quad (7)$$

where $P(i)$ is the probability of finding species i in any site of the lattice, and $P(i \cap j)$ is the probability of finding i and j in near-neighbor sites. We further evaluate the probability of finding a species a in a near-neighbor site given that species b occupies a certain site, $P(a|b)$:

$$P(a|b) = \frac{P(a \cap b)}{P(b)}. \quad (8)$$

As one can see from Table I, intraspecies correlation coefficients are positive while different species show a negative correlation. Indeed, the anticorrelation exhibited by the two different species of preys is comparable to the anticorrelation found between preys and predators.
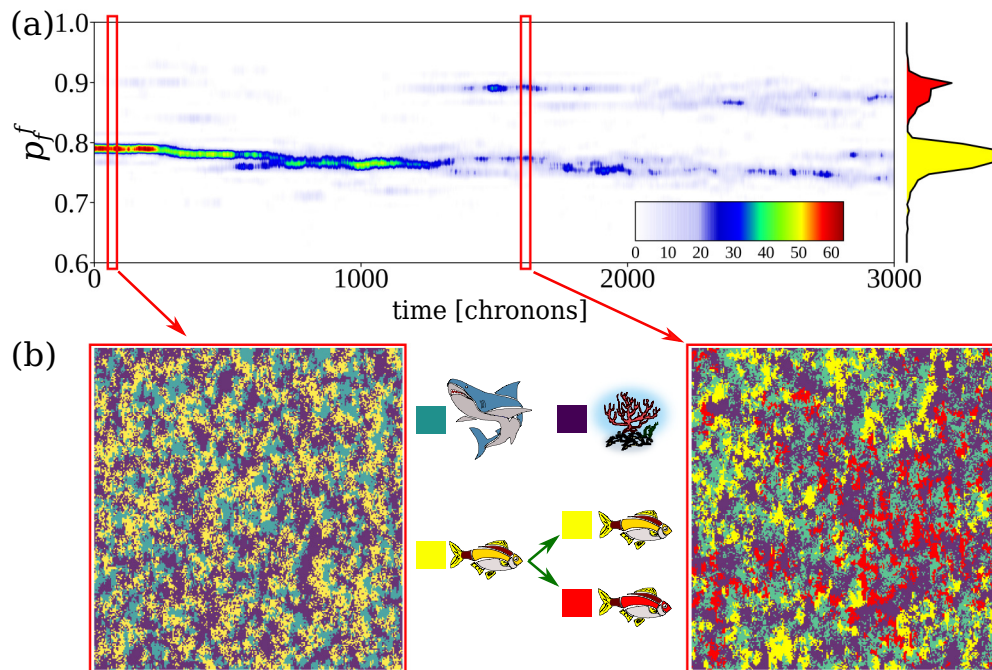

FIG. 6. (a) Time evolution of the distribution of fish filiation probability. At time zero, the population of fishes is prepared with a filiation rate of 0.8. After about 1500 chronons, the population splits into two species with a well-separated trait. (b) Snapshots of two EVOLAT configurations during the evolution. In the left snapshot, only a species of fish is observed. In the right-side one, two fish species have formed. The spatial distribution of sharks differs in the proximity of the two species, showing that sharks are adopting different hunting strategies.

Predators do not see any difference between the two kinds of preys, as we can see from the conditional probabilities of finding a shark next to a prey $[P(s|f_1) = P(s|f_2)]$. On the other hand, a small but significant difference is present between the near-neighbor correlations $(C[f_1, f_1] \neq C[f_2, f_2])$, measuring the spatial organization of the two prey species.

Note that the presence of the spatial organization of the animals is a very important feature for sympatric speciation [52], since no physical barriers are present in the system.

In a mean-field scenario, this coexistence between preys with different reproduction rates would be impossible, as it would lead to the supremacy of the prey with the higher reproduction rate.

IV. DISCUSSION

We simulate the evolution of a minimal prey-predator system. Each species is characterized by a toy genome through which three macrophenotypes manifest. Providing each specimen with N genes for each phenotype, we showed that the gene relevance [i.e., the weight of each gene according to Eq. (1) in coding for phenotypes is a key feature for evolution].

To survive, an organism must be robust to deleterious mutations. A convenient choice, the organism can opt for could be to rely on many genes for the manifestation on one phenotype. If the information for the phenotype is evenly distributed in several genes, then a damage to some gene has the minimum impact on the resulting phenotype. As a counterpart, we showed that this condition favors the accumulation of detrimental mutations that lead to the mutational meltdown. Conversely, packing all the phenotype information in few genes enables abrupt variations of phenotype. This prevents the mutational meltdown, at the cost of reducing the differentiation between individuals and consequently the adaptability of the species (as shown in Fig. 5). In this case a drastic change in the environment would provoke a sudden extinction of the species.

Relying on a heterogeneous distribution of information in many genes assures both broad differentiation of a huge genome and the possibility to have astonishing positive mutations that drive the evolution and prevent the mutational meltdown. EVOLAT, albeit representing a minimal model of an ecosystem, reproduces a rich variety of scenarios upon varying a single parameter. In fact, by tuning the α exponent in Eq. (4), i.e., assigning different weights to genes associated with the same phenotype, the system exhibits different behaviors. Indeed, a uniform distribution of weights ($\alpha = 0$) leads to a progressive reduction of the fitness due to the accumulation of detrimental mutations (see Fig. 2). On the opposite side, if only one gene encodes for all the traits ($\alpha = \infty$), the mutational meltdown is prevented as species can improve their fitness. However, in this regime there is no differentiation inside the same population, exposing the species to the threat of sudden environmental changes. Life lies in between, where the high impact of a few genes prevents the mutational meltdown, while the bulk of the remaining genes guarantees differentiation (see Fig. 5).

Notably, fishes provided with a heterogeneous genome respond to the fluctuation of the environment and to the natural selection operated by sharks by a sympatric speciation. Spontaneous speciation, and particularly a sympatric one, is rarely reproduced [37] by models that manage to observe differentiation due to Mendelian inheritance [53]. According to our simulations, two key ingredients are important to observe the emergence of sympatric speciations: the heterogeneity in gene relevance and the explicit spatial extension of the simulation. Moreover, our minimal model provides a theoretical framework to deal with the everlasting debate about the physical feasibility of evolution.

The idea that life had evolved by the combined action of mutations in the genome and natural selection of the most fitted individuals is accepted by a vast majority of scientists. The skeptics argue that random mutations on the genome should progressively increase the disorder (entropy) and consequently be incompatible with life. Indeed, this happens only in the case of a uniform genome, where mutations lead the phenotypes toward the entropic values. This argument is therefore based on the wrong assumption that all the genes equally contribute to the phenotype.

In our simple model, each part of the genetic sequence influences the phenotype with different weights. This is a coarse-grain representation of the underlying biological mechanism through which the phenotype is manifested. In nature, only 2%–3% of DNA is "coding" (it can be translated into mRNA and then into proteins). Several studies [54,55] have shown that also the noncoding DNA (often referred to as *junk*) can affect the phenotype. This is a mechanism for strong heterogeneity in how the phenotype manifests from the underlying genetic sequence.

Overall, in our opinion the most limiting choices we made were not to include sexual reproduction, correlations in gene expression, and modifications in the food chain (a prey cannot become a predator). This limits the sources of variation between individuals only to the effect of mutations. On the other hand, the reduced number of parameters allows one to look for general features of evolution.

In conclusion, we proved that heterogeneity in gene relevance is a key feature to prevent a mutational meltdown in a species. Moreover, we showed evidence that spatial correlations are fundamental to account for sympatric speciation. These findings contribute to disentangling how genomes change to create new species and provide a step forward in understanding the mechanisms of evolution.

The source code to run the simulations to reproduce the calculations can be found in the GitHub repository under the name of EvoLat [56].

APPENDIX A: DETAILS ON THE COMPUTATIONAL METHODS

In the following, we report the details on how the simulation is performed.

The starting parameters are

(i) The initial phenotype for preys and predators.

(ii) The average phenotype without natural selection ($\langle x \rangle$).

(iii) The size of the lattice.

(iv) The mutation rate for each gene.

(v) The number of genes for each phenotypic feature.

(vi) The initial number of preys and predators in the lattice.

All the specimens are then prepared with a genome in which all genes are equal to the target phenotypic feature as provided in the input. Then we present the workflow of each chronon in the simulation:

(i) A random cell in the lattice is extracted.

(ii) If the cell is empty, the step is over

(iii) We compute the phenotype of the selected individual from its genome.

(iv) We extract and perform the move for the selected individual.

This cycle is repeated a number of times equal to the number of sites in the lattice to compose a single unit of time. The move of each individual depends if it is a prey or a predator. For preys,

(i) We extract if the prey moves, breeds, or dies with probability p_f^m, p_f^f, or p_f^d.

(ii) If the prey moves or breeds, we extract randomly a near-neighbor lattice site. If it is empty we move the prey in that site; otherwise the step is over.

(iii) If the prey dies, it is removed from the lattice.

For the predator,

(i) We extract if the predator moves with probability p_s^m.

(ii) If the predator moves, we extract randomly a near-neighbor lattice site. If it does not contain another predator, it is moved in that cell.

(iii) If the predator is moved in a cell contained by a prey, it can breed with probability p_s^f.

(iv) If the predator breeds, two offspring are located in the new position and the old position of the prey.

(v) If the predator did not eat any pray in this step, it can die with probability p_s^d.

The breeding process is common for both preys and predators.

(i) Two offspring are located in the current position of the individual and in the empty cell from which the last move is performed.

(ii) Each new individual has a genome that is copied from the parent.

(iii) For each gene of the new individual, a random mutation can occur with probability chosen in the input.

(iv) If the gene mutates, its value is replaced with the one extracted from the distribution $q_x(x)$, as discussed in the main text.

In all the performed simulations, an initial thermalization time of 1000 chronons is used to allow the ecosystem to reach its nonequilibrium, quasisteady state in the absence of mutations [35], which occurs in a timescale much higher than the typical evolution of species (usually in three or four generations). Then we study the out-of-equilibrium dynamics induced by the evolution of the phenotype.
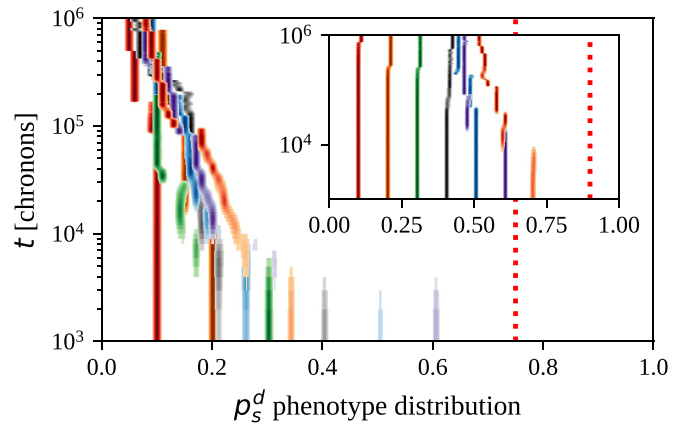

FIG. 7. Evolution of the shark death-probability distribution as a function of time in the presence of a heterogeneous genome. Heterogeneity allows the species to evolve to a higher average lifetime, while this does not occur in the presence of a uniform genome. Different colors correspond to different starting values. The less fitted individual can improve their fitness far above the entropic value. The dashed vertical red line represents the entropic force value of 0.9 [$\langle x \rangle$ in Eq. (C2)]. The inset show dynamics obtained by a different value (0.75) of the entropic force.

APPENDIX B: SIMULATION PARAMETERS

All the simulations were started with the following parameters (if not differently specified): $p_f^f = 0.2$, $p_f^m = 0.8$, $p_f^d = 0$, $p_s^m = 0.7$, $p_s^f = 1$, $p_s^d = 0.6$. The entropic forces [Eq. (2)] were $\langle x \rangle = 0.5$ for all the parameters, except for the predator death rate, which we set to $\langle p_s^d \rangle = 0.9$. We used a lattice size of $L = 256 \times 256$ with a mutation rate of 0.016 per gene per chronon. Each simulation is initialized by a random configuration fully occupied by preys and predators in equal ratio. Then we wait for 1000 chronons to reach the nonequilibrium metastable steady state of the prey-predator dynamics. Then the simulation consists of 10^6 chronons, saving snapshots, and the state of the system once every 1000 chronons. For the data in Fig. 4 and Table I of the main paper, we extracted snapshots every 40 chronons, while we computed the correlation coefficients averaging every 400 chronons to avoid sampling correlated configurations.

To test the consistency of our results, we also performed the simulation by changing the initial phenotype, genome length, and entropic forces, but no significant difference has been observed. As an example, in Fig. 7 we reported the shark death-probability distribution as a function of time in the presence of a heterogeneous genome for different values of the initial phenotype and of the entropic force. In all cases the distributions tend to drift away from the entropic force values (dotted lines). The observed behaviors were robust to statistical fluctuations. In fact, different simulations, starting with the same set of parameters, exhibit the same trend (see Fig. 8).
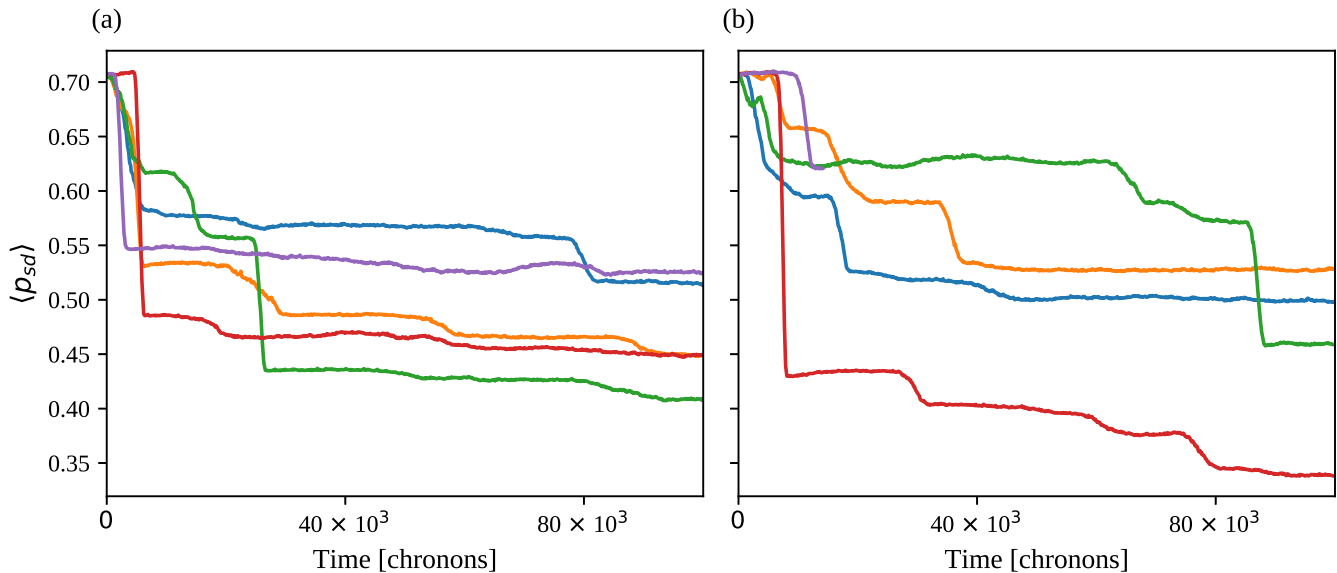
FIG. 8. (a) Evolution of the average shark death rate. Simulations performed with an heterogeneous genome $\alpha = 2$, all with the same initial values of p_s^d. (b) Same as in (a) but showing other independent realizations of the system evolution.

APPENDIX C: ENTROPIC FORCES IN FARMS

Here we compute the dynamical evolution of a predator population in a "farm," i.e., supplying the predator with an infinite source of food, allowing them to reproduce without the effect of natural selection. In this case each gene of the new generation will have a probability of a mutation given by the mutation rate. Therefore the master equation for the genome of the phenotype x is

$$g_{x_i}(t + \delta) = p_{\text{mut}} \int_0^1 dy\, y f_x(y) + (1 - p_{\text{mut}}) g_{x_i}(t), \quad \text{(C1)}$$

where $f_x(y)$ is the probability distribution of the entropic forces that set the average mutations. We can define the average value of the entropic force as

$$\langle x \rangle = \int_0^1 dy\, y f_x(y). \quad \text{(C2)}$$

So if we start with

$$g_{x_i}(0) = p_x, \quad \text{(C3)}$$

in the limit $\delta \to 0$ we get

$$\frac{d}{dt} g_{xi} = p_{\text{mut}} \left(\langle x \rangle - g_{x_i} \right). \quad \text{(C4)}$$

That is a standard first-order differential equation, whose solution is

$$g_{x_i}(t) = (p_x - \langle x \rangle) e^{-p_{\text{mut}} t} + \langle x \rangle. \quad \text{(C5)}$$

The phenotype is the generic weighted average of the genome:

$$p_x(t) = \sum_{i=1}^N w_i g_{x_i}(t). \quad \text{(C6)}$$

Since each gene evolves independently in the absence of natural selection, we have

$$\langle p_x \rangle (t) = (p_x - \langle x \rangle) e^{-p_{\text{mut}} t} + \langle x \rangle. \quad \text{(C7)}$$

Therefore the phenotype converges to the expected value of the entropic forces after a typical time of $1/p_{\text{mut}}$ independent of the particular initial state p_x.

APPENDIX D: MUTATIONAL MELTDOWN

The mutational meltdown is an evolutionary process where deleterious mutations accumulate in time progressively decreasing the fitness of the population until extinction. We observe an analogous process in Fig. 2. This process is known to be affected by the carrying capacity. To assess this we simulated two lattices with different L^2 number of sites ($L_1 = 256$ and $L_2 = 512$).

Looking at the relaxation times, we can spot a (small) dependence with the size (Fig. 9). The higher the carrying capacity, the slower the genomic meltdown. This is in complete accord to what we found in the literature [45–47].

To get an intuitive idea of how such an effect should scale with the size, let us consider a population of $N + 1$ individuals and suppose that each animal possesses a certain fitness extracted from a distribution. To simplify, let us identify the fitness with the growth rate λ of the individual so that the distribution of the fitness is the distribution of growth rate $p(\lambda)$. Let us make a further simplification that higher fitness corresponds to higher growth rates. In this situation, the population evolves under the effect of mutations and natural selection. Mutations tend to lower the fitness of each individual (i.e., reduce its growth rates). In particular, an accumulation of deleterious mutations must result in a decrease of the population fitness, so the growth rate distribution will drift toward smaller values of growth rate. On the other hand, natural selection favors animals with a high growth rate. In particular, the higher the growth rate of an individual from
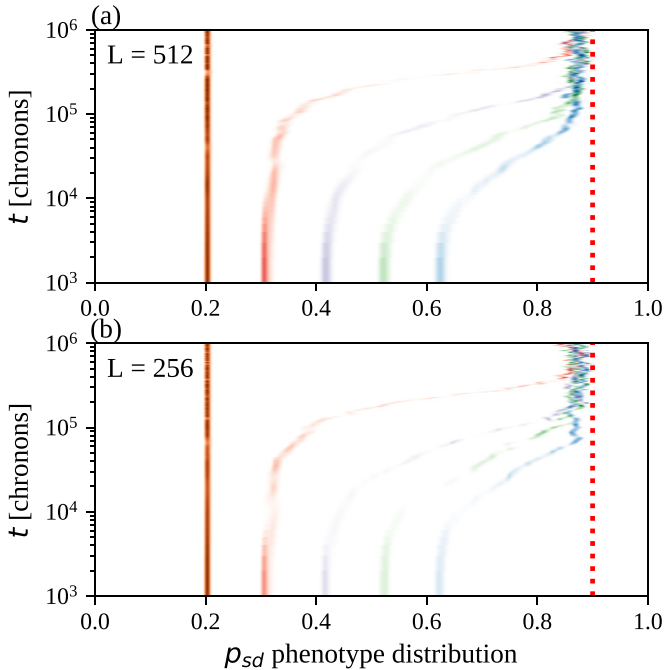
FIG. 9. (a) Time evolution of predator death-rate phenotype after quenching. Different colors correspond to different simulations performed starting from different values of p_{sd}, while the lattice size is set to $L = 512$. The predator relaxation time diverges as a power law in a way that resembles the dynamic of glassy systems. The dashed vertical line is the stationary state for predator growing in a breeding farm (no natural selection). (b) Same as in (a) but with $L = 256$.

the population mean growth rate, the more the progeny of this animal must be selected: the animal with the highest growth rate rapidly dominates, pushing the distribution toward high values of growth rates.

To quantity the robustness of the population to the meltdown, we can ask how the distance between the best-fitted
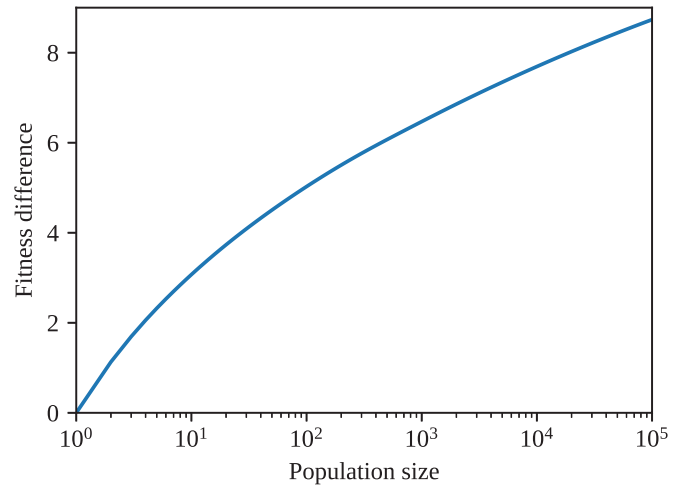

FIG. 10. Average difference between the most fitted and the mean individual as a function of the population size.

individual with respect to the mean varies with the size of the population, i.e., $\langle \lambda_M \rangle_N - \langle \lambda \rangle$. If the population is composed by $N + 1$ individuals and one of them has fitness λ_M, the probability that λ_M is the highest fitness is given by

$$P(\lambda < \lambda_M, N) = \left(\int_{-\infty}^{\lambda_M} p(\lambda) \right)^N. \tag{D1}$$

In turn, the probability of λ_M is obtained as

$$\tilde{P}(\lambda_M) = \frac{d}{d\lambda_M} P(\lambda < \lambda_M, N) = NP(\lambda < \lambda_M, N - 1)p(\lambda_M). \tag{D2}$$

In Fig. 10 we show that the distance between the best-fitted individual with respect to the mean varies with the size of the population, i.e., $\langle \lambda_M \rangle_N - \langle \lambda \rangle$, for a normal growth rate distribution. Indeed, as the number of samplings increases, so does the probability of sampling extreme values.

[1] D. J. Howard, A. P. D. B. D. J. Howard, S. H. Berlocher, and A. P. D. E. S. H. Berlocher, *Endless Forms: Species and Speciation* (Oxford University Press, Oxford, UK, 1998).

[2] O. F. Cook, Factors of species formation, Science **23**, 506 (1906).

[3] E. Freese and A. Yoshida, The role of mutations in evolution, in *Evolving Genes and Proteins* (Elsevier, New York, 1965), pp. 341–355.

[4] M. Kimura, *The Neutral Theory of Molecular Evolution* (Cambridge University Press, Cambridge, UK, 1983).

[5] E. Mayr, *What Makes Biology Unique?: Considerations on the Autonomy of a Scientific Discipline* (Cambridge University Press, Cambridge, UK, 2004).

[6] M. Nei, The new mutation theory of phenotypic evolution, Proc. Natl. Acad. Sci. **104**, 12235 (2007).

[7] M. Eigen, Self-organization of matter and the evolution of biological macromolecules, Die Naturwissenschaften **58**, 465 (1971).

[8] S. Forcelloni and A. Giansanti, Evolutionary forces and codon bias in different flavors of intrinsic disorder in the human proteome, J. Molec. Evol. **88**, 164 (2020).

[9] R. Hershberg, Mutation—The engine of evolution: Studying mutation and its role in the evolution of bacteria, Cold Spring Harbor Perspect. Biol. **7**, a018077 (2015).

[10] K. S. Korolev, J. B. Xavier, and J. Gore, Turning ecology and evolution against cancer, Nat. Rev. Cancer **14**, 371 (2014).

[11] N. McGranahan and C. Swanton, Clonal heterogeneity and tumor evolution: Past, present, and the future, Cell **168**, 613 (2017).

[12] Q. Zhang and R. H. Austin, Physics of cancer: The impact of heterogeneity, Annu. Rev. Condens. Matter Phys. **3**, 363 (2012).

[13] C. Ryan Campbell, J. W. Poelstra, and A. D. Yoder, What is speciation genomics? The roles of ecology, gene flow, and genomic architecture in the formation of species, Biol. J. Linn. Soc. **124**, 561 (2018).

[14] D. Cavalieri, J. P. Townsend, and D. L. Hartl, Manifold anomalies in gene expression in a vineyard isolate of *Saccharomyces cerevisiae* revealed by DNA microarray analysis, Proc. Natl. Acad. Sci. USA **97**, 12369 (2000).

[15] B. Jesse Shapiro, J. Friedman, O. X. Cordero, S. P. Preheim, S. C. Timberlake, G. Szabó, M. F. Polz, and E. J. Alm, Population genomics of early events in the ecological differentiation of bacteria, Science **336**, 48 (2012).

[16] J. Pellicer, O. Hidalgo, S. Dodsworth, and I. J. Leitch, Genome size diversity and its impact on the evolution of land plants, Genes **9**, 88 (2018).

[17] A. J. Lotka, Analytical note on certain rhythmic relations in organic systems, Proc. Natl. Acad. Sci. USA **6**, 410 (1920).

[18] V. Volterra, Variations and fluctuations of the number of individuals in animal species living together, in *Animal Ecology*, edited by R. N. Chapman (McGraw-Hill, New York, 1931), pp. 409–448.

[19] D. Bajić, J. C. C. Vila, Z. D. Blount, and A. Sánchez, On the deformability of an empirical fitness landscape by microbial evolution, Proc. Natl. Acad. Sci. USA **115**, 11286 (2018).

[20] I. Bena, M. Droz, J. Szwabiński, and A. Pękalski, Complex population dynamics as a competition between multiple-time-scale phenomena, Phys. Rev. E **76**, 011908 (2007).

[21] A. De Martino, T. Gueudré, and M. Miotto, Exploration-exploitation tradeoffs dictate the optimal distributions of phenotypes for populations subject to fitness fluctuations, Phys. Rev. E **99**, 012417 (2019).

[22] D. De Martino, F. Capuani, and A. De Martino, Growth against entropy in bacterial metabolism: The phenotypic trade-off behind empirical growth rate distributions in *E. coli*, Phys. Biol. **13**, 036005 (2016).

[23] E. Kussell, Phenotypic diversity, population growth, and information in fluctuating environments, Science **309**, 2075 (2005).

[24] G. M. Abernethy, M. McCartney, and D. H. Glass, The interaction between predator strategy and prey competition in a pair of multi-predator multi-prey lattices, Commun. Nonlinear Sci. Numer. Simul. **56**, 9 (2018).

[25] M. Barbier, J.-F. Arnoldi, G. Bunin, and M. Loreau, Generic assembly patterns in complex ecological communities, Proc. Natl. Acad. Sci. USA **115**, 2156 (2018).

[26] M. He, Q.-H. Pan, and S. Wang, Final state of ecosystem containing grass, sheep and wolves with aging, Int. J. Mod. Phys. C **16**, 177 (2005).

[27] M. He, H. Ruan, and C. Yu, A predator prey model based on the fully parallel cellular automata, Int. J. Mod. Phys. C **14**, 1237 (2003).

[28] A. Pekalski and M. Droz, Self-organized packs selection in predator-prey ecosystems, Phys. Rev. E **73**, 021913 (2006).

[29] A. Pérez-Escudero, J. Friedman, and J. Gore, Preferential interactions promote blind cooperation and informed defection, Proc. Natl. Acad. Sci. **113**, 13995 (2016).

[30] S. M. J. Portalier, G. F. Fussmann, M. Loreau, and M. Cherif, The mechanics of predator–prey interactions: First principles of physics predict predator–prey size ratios, Funct. Ecol. **33**, 323 (2018).

[31] N. J. Savill and P. Hogeweg, Spatially induced speciation prevents extinction: The evolution of dispersal distance in oscillatory predator-prey models, Proc. R. Soc. London, Ser. B **265**, 25 (1998).

[32] T. Antal, M. Droz, A. Lipowski, and G. Ódor, Critical behavior of a lattice prey-predator model, Phys. Rev. E **64**, 036118 (2001).

[33] Dewdney, Sharks and fish wage an ecological war on the toroidal planet Wa-Tor, Sci. Am. **251**, 14 (1984).

[34] U. Dobramysl, M. Mobilia, M. Pleimling, and U. C. Täuber, Stochastic population dynamics in spatially extended predator-prey systems, J. Phys. A: Math. Theor. **51**, 063001 (2018).

[35] M. Miotto and L. Monacelli, Entropy evaluation sheds light on ecosystem complexity, Phys. Rev. E **98**, 042402 (2018).

[36] M. Mobilia, I. T. Georgiev, and U. C. Täuber, Fluctuations and correlations in lattice models for predator-prey interaction, Phys. Rev. E **73**, 040903(R) (2006).

[37] J. Friedman, E. J. Alm, and B. J. Shapiro, Sympatric speciation: When is it possible in bacteria? PLoS ONE **8**, e53539 (2013).

[38] P. A. P. Moran, Global stability of genetic systems governed by mutation and selection, Math. Proc. Cambridge Philos. Soc. **80**, 331 (1976).

[39] T. J. P. Penna, A bit-string model for biological aging, J. Stat. Phys. **78**, 1629 (1995).

[40] J. Swetina and P. Schuster, Self-replication with errors, Biophys. Chem. **16**, 329 (1982).

[41] J. Dalmau, Asymptotic behavior of Eigen's quasispecies model, Bull. Math. Biol. **80**, 1689 (2018).

[42] A. O. Sousa, Sympatric speciation in an age-structured population living on a lattice, Eur. Phys. J. B **39**, 521 (2004).

[43] D. A. Drummond, J. D. Bloom, C. Adami, C. O. Wilke, and F. H. Arnold, Why highly expressed proteins evolve slowly, Proc. Natl. Acad. Sci. USA **102**, 14338 (2005).

[44] J. Garcia-Bernardo and M. J. Dunlop, Phenotypic diversity using bimodal and unimodal expression of stress response proteins, Biophys. J. **110**, 2278 (2016).

[45] A. Grande-Perez, E. Lazaro, P. Lowenstein, E. Domingo, and S. C. Manrubia, Suppression of viral infectivity through lethal defection, Proc. Natl. Acad. Sci. U.S.A. **102**, 4448 (2005).

[46] M. Lynch, J. Conery, and R. Burger, Mutation accumulation and the extinction of small populations, Am. Nat. **146**, 489 (1995).

[47] T. Ohta. Slightly deleterious mutant substitutions in evolution, Nature (London) **246**, 96 (1973).

[48] D. De Martino, F. Capuani, and A. De Martino, Quantifying the entropic cost of cellular growth control, Phys. Rev. E **96**, 010401 (2017).

[49] S. Cocco, C. Feinauer, M. Figliuzzi, R. Monasson, and M. Weigt, Inverse statistical physics of protein sequences: A key issues review, Rep. Prog. Phys. **81**, 032601 (2018).

[50] M. Miotto, P. P. Olimpieri, L. Di Rienzo, F. Ambrosetti, P. Corsi, R. Lepore, G. G. Tartaglia, and E. Milanetti, Insights on protein thermal stability: A graph representation of molecular interactions, Bioinformatics **35**, 2569 (2019).

[51] W. Kob, F. Sciortino, and P. Tartaglia, Aging as dynamics in configuration space, Europhys. Lett. (EPL) **49**, 590 (2000).

[52] B. M. Fitzpatrick, J. A. Fordyce, and S. Gavrilets, What, if anything, is sympatric speciation? J. Evol. Biol. **21**, 1452 (2008).

[53] R. Aguilée, A. Lambert, and D. Claessen, Ecological speciation in dynamic landscapes, J. Evol. Biol. **24**, 2663 (2011).

[54] C. Biémont and C. Vieira, Junk DNA as an evolutionary force, Nature (London) **443**, 521 (2006).

[55] S. Prabhakar, A. Visel, J. A. Akiyama, M. Shoukry, K. D. Lewis, A. Holt, I. Plajzer-Frick, H. Morrison, D. R. FitzPatrick, V. Afzal, L. A. Pennacchio, E. M. Rubin, and J. P. Noonan, Human-specific gain of function in a developmental enhancer, Science **321**, 1346 (2008).

[56] https://github.com/mesonepigreco/EvoLat.