Rate of environmental variation impacts the predictability in evolution

Diego Cirne and Paulo R. A. Campos
Departamento de Física, Universidade Federal de Pernambuco, 50740-560 Recife-PE, Brazil

(Received 25 August 2022; accepted 9 December 2022; published 21 December 2022)

In the two last decades, we have improved our understanding of the adaptive evolution of natural populations under constant and stable environments. For instance, experimental methods from evolutionary biology have allowed us to explore the structure of fitness landscapes and survey how the landscape properties can constrain the adaptation process. However, understanding how environmental changes can affect adaptation remains challenging. Very little progress has been made with respect to time-varying fitness landscapes. Using the adaptive-walk approximation, we survey the evolutionary process of populations under a scenario of environmental variation. In particular, we investigate how the rate of environmental variation influences the predictability in evolution. We observe that the rate of environmental variation not only changes the duration of adaptive walks towards fitness peaks of the fitness landscape, but also affects the degree of repeatability of both outcomes and evolutionary paths. In general, slower environmental variation increases the predictability in evolution. The accessibility of endpoints is greatly influenced by the ecological dynamics. The dependence of these quantities on the genome size and number of traits is also addressed. To our knowledge, this contribution is the first to use the predictive approach to quantify and understand the impact of the speed of environmental variation on the degree of parallelism of the evolutionary process.

DOI: 10.1103/PhysRevE.106.064408

I. INTRODUCTION

The concept of fitness landscapes is widely used in evolutionary biology, combinatorial optimization, and the physics of disordered systems [1,2]. First proposed by Sewall Wright at the beginning of the nineteenth century [3], fitness landscapes establish a relationship between an individual's genotype and its reproductive success. The concept of the fitness landscape is analogous to that of energy landscapes in physics, relating possible states of a system and their corresponding energy levels [4]. While from a biological perspective, evolution is portrayed as a hill-climbing process toward higher peaks of the fitness landscape, physical systems are driven to states of low energy [5].

Experimental methods from evolutionary biology have allowed us to explore the structure of fitness landscapes in microbial species and address how the topography of those landscapes can constrain the adaptation process [6]. The typical experimental approach consists of characterizing the topography of empirical fitness landscapes by analyzing the interactions among a small subset of mutations and reconstructing all possible genotypes from the wild type to the evolved [7]. Genotypic fitness landscape models, those directly mapping from genotypes to fitness, have been widely used to explain experimental data [8,9]. One important class of genotypic fitness landscape models is the NK landscape model. Accordingly, the fitness of a given genotype configuration follows from the contribution of L loci. The contribution of each locus to fitness depends not only on the state of the focal locus, but also on the states of K other loci [10]. We can create a broad spectrum of scenarios by changing the epistatic parameter K. When $K = 0$, a smooth, single-peaked fitness landscape is obtained, whereas, in the other extreme $K = L - 1$, a completely uncorrelated fitness landscape is generated. The increase of K results in a larger number of local optima, induced by the presence of sign epistasis [11]. Indeed, the NK model can be seen as a biological version of the p-spin spin-glass model [12–14].

Another relevant class of fitness landscape models is referred to as phenotypic fitness landscapes [15]. Instead of mapping genetic states to fitness, phenotypic landscapes map a phenotypic domain to fitness. The most prominent instance of phenotypic landscapes is Fisher's geometric model (FGM), proposed by Fisher in the 1930s [16]. In the FGM, selection and mutation act on quantitative traits. A point in a multidimensional phenotypic space represents a phenotype in which each dimension corresponds to a trait. In a given environment, each trait has an optimal value, and the overall fitness decreases with the distance to the optimum phenotype [17]. In the model, variations resulting from mutations will alter trait values, and so the mutant offspring will lie at a different position in the phenotypic space. If such change moves the mutant offspring to a position closer to the location of the optimum phenotype, the mutation is said to be beneficial, otherwise, the mutation will have a detrimental effect on fitness [18]. The model predicts that deleterious mutations are more frequent, in agreement with the findings of experimental evolutionary biology [19,20], and those mutations are even more likely as one approaches the optimum phenotype. The number of traits, used as a proxy for phenotypic complexity, plays a role by affecting both the rate of beneficial mutations as well as the distribution of mutation effects [21–23].

 ©2022 American Physical Society

An important and useful extension of original Fisher's formulation is the one in which mutation effects on the phenotypes are assumed to be additive. So deviations from additivity on the genotype-fitness map are a direct consequence of the nonlinear mapping from phenotype to fitness [24]. Epistasis emerges due to this nonlinearity of the phenotype to fitness map and is particularly important around the optimum phenotype, where the curvature is larger. The level of epistasis and ruggedness of the fitness landscape are fundamental features of the process of adaptation [11,25,26], influencing its degree of repeatability and predictability [27–29].

FGM has also been successfully employed to interpret experimental data, such as the distribution of fitness effects of random mutations [23], distributions of epistasis [30,31], and dynamics of fitness growth in microbial populations [32], to mention just a few. A genotypic landscape under FGM has recently been proposed and used to study the properties of selected mutations from experimental data of bacterial populations [26]. The genotypic landscape under Fisher's model provides a map genotype → phenotype → fitness, in which an additional layer is considered relative to genotypic landscape models. The discrete nature of the genotypic space is responsible for the emergence of interesting features of the resulting fitness landscape. Despite the smoothness of the phenotypic landscape, the resulting fitness landscape displays fitness optima, and its number grows exponentially with the genome size, as commonly found in genotypic fitness landscapes [24].

The FGM has been a valuable tool in the study of the impact of environmental variations on many processes found in both evolutionary and ecological contexts [33–35], including those leading to ecological diversification and speciation events [36,37]. A common approach for the FGM in the study of environmental variation is to assume that the population or community adapts to a dynamic optimum phenotype [33,34]. The effect of the environment on reshaping the landscape has analogous to physical systems, such as the effect of temperature in changing the energy landscape associated with protein folding, RNA macromolecules, and amorphous solids [38–40]. The influence of the rate of environmental variation is a quite controversial issue, especially in the face of the debate about the role of climatic and ecological changes in shaping biodiversity [41]. Within an evolutionary perspective, a previous study shows that the rate of environmental variation strongly influences the distribution of adaptive substitutions [35]. On the other hand, another study demonstrated that, at least under the perspective of the FGM, the pattern of diversification and speciation is dependent not on the rate of environmental changes but on the net magnitude of those changes over a given time interval [37]. In the current contribution, we use the framework of the FGM equipped with a genetic basis to survey the consequences of environmental variations on the predictability of evolution. The limitations in predicting evolution mainly arise because of the random nature of processes such as point mutations and other sources of genetic variability and the strength of stochasticity in finite systems [42]. Our limited understanding of selection, i.e. the knowledge of traits or features that would need to be adjusted to increase organisms' adaptation in a given environment, also hinders our ability to predict [43,44]. Lastly, the environment itself is dynamic, and so environmental variables such as climatic conditions and predator abundance might fluctuate and affect selective pressure [43,45].

Through an adaptive-walk approximation, we address the role of environmental variations in shaping the evolutionary process. Measurements of predictability and path divergence, among others, quantify the degree of repeatability of the evolutionary process upon environmental variation. In this framework, one depicts the whole population as a single entity that travels through the fitness landscape toward higher fitness [46]. At each time step, the walker moves to one of its fitter single-mutation neighbors. The adaptive walk ends when the walker reaches a fitness peak [47]. The properties of adaptive walks are intrinsically linked to the topography of fitness landscapes. For instance, the distribution of walk lengths relates to the size distribution of the basins and the degree of the ruggedness of the fitness landscape [48–50]. The framework of adaptive walks has been largely used in several optimization problems, such as the study of searching strategies in complex networks [51], in the exploration of protein fitness landscapes [52], and other combinatorial optimization problems [53].

The paper is organized as follows. Section II A describes in detail the FGM, which here is provided with a genetic basis, and how environmental variation is included in the modeling. In Sec. II B we define the statistical measurements of our interest: predictability and mean path divergence. In Sec. III we present our simulation results. Finally, Sec. IV presents our conclusions.

II. MATERIALS AND METHODS

A. The model

In the strong-selection weak-mutation regime, the evolutionary dynamics of natural populations can be seen as performing an uphill climbing on a fitness landscape, in which beneficial mutations fix sequentially, up to reaching a local fitness peak. Moreover, because the time of appearance of new variants is much larger than the time to fixation of beneficial mutations in such a limit, the population can thus be depicted as a single entity (isogenic population) performing an adaptive walk on this fitness landscape. Once a population has reached a local fitness peak, adaptive dynamics may stall as further adaptation requires crossing a valley. Because of the discrete nature of the genotypic space, each genome is described as a binary sequence of size L, $S = (s^{(1)}, s^{(2)}, \ldots, s^{(L)})$ with $s^{(\alpha)} \in \{0, 1\}$, and so there are 2^L possible genotype configurations. The dichotomous nature of the locus state follows the usual design of experimental setups based on the absence (0) or presence (1) of point mutations [54]. At each time step of the dynamics, the walker moves to one of its immediate neighbors, i.e., those differing by a single digit, with higher fitness. Here we use the so-called natural adaptation walk [48,55,56], in which the chance of moving to a given neighbor is proportional to the fitness advantage conferred by that neighbor.

Regarding the fitness landscape, we consider the FGM but equipped with a genetic basis. To each genotype there is a corresponding phenotype with N traits. As standard, the phenotype vector $\vec{r}_S = (r_S^{(1)}, r_S^{(2)}, \ldots, r_S^{(N)})$ associated with an individual carrying genome S rules its level of adaptation.

The fitness landscape is simply built by ascribing a phenotype to genotype $S_0 = (0, 0, \ldots, 0)$, and then defining L displacement vectors, $\{\vec{\eta}_k\}$ with $k = 1, \ldots, L$, each one related to a mutation in locus k relative to the wildtype sequence $S_0 = (0, 0, \ldots, 0)$. In this manner, the phenotype of the antipode of S_0, which corresponds to the sequence $\bar{S}_0 = (1, 1, \ldots, 1)$, is simply $\vec{r}_{\bar{S}_0} = \vec{r}_{S_0} + \vec{\eta}_1 + \vec{\eta}_2 + \cdots + \vec{\eta}_L$, where $\vec{r}_{S_0}$ is the phenotype vector associated with S_0. Each component of each vector $\vec{\eta}_k$ is drawn from a Gaussian distribution of null mean and standard deviation σ, so that mutations are supposed to be isotropic and have no preferred direction, hence affecting all traits similarly [18]. Furthermore, notice that mutation in locus k always causes the same phenotypic displacement regardless the state of any other locus, once $\vec{\eta}_k$ is a constant vector. In sum, mutations are combined additively.

Within the framework of the FGM, the fitness of an individual with phenotype $\vec{r} = (r^{(1)}, r^{(2)}, \ldots, r^{(N)})$ is a function of the distance of its phenotype to the optimum phenotype, here denoted as $\vec{h}$. As usual, the fitness is estimated as

$$f(\vec{r}) = \exp\left(-\frac{1}{2\alpha^2}\sum_{\ell=1}^{N}(r^{(\ell)} - h^{(\ell)})^2\right), \tag{1}$$

where the parameter α determines the inverse of the strength of selection. Here we set $\alpha = 1$.

1. Moving optimum

As aforementioned, we are mainly interested in studying the properties of evolutionary trajectories performed by adaptive walks in a scenario of time-varying landscapes. In phenotypic terms, the total amount of change brought about by the event of environmental variation will end up the same. However, the rate at which such event takes place is a key parameter of our modeling. In all scenarios here studied, at time $t = 0$ the optimum phenotype $\vec{h}$ is set at $\vec{r}_{S_0}$, the corresponding phenotype of genotype S_0. According to Eq. 1, such initial condition corresponds to the original environment $\mathcal{F}_0$ (Notice that $\vec{r}_{S_0}$ plays no role in the dynamics). After τ time steps, the phenotypic optimum ends up on $\vec{r}_{\bar{S}_0}$, the phenotype of S_0's antipode, $\bar{S}_0$. Every time step, the optimum phenotype, while $\vec{r}_{\bar{S}_0}$ is not reached, will change by an amount $\vec{v}_h$,

$$\vec{v}_h = \frac{\vec{r}_{\bar{S}_0} - \vec{r}_{S_0}}{\tau}, \tag{2}$$

which defines the rate of change of the optimum phenotype in the transient time interval τ. Note that $\tau = 1$ resembles standard adaptive walk studies of static fitness landscapes, i.e., the optimum phenotype is instantaneously placed at the final location. Those variations of the optimum phenotype will reshape the fitness landscape, generating a family of fitness landscapes $\mathcal{F}_0 \to \mathcal{F}_1 \to \cdots \to \mathcal{F}_\tau$, where $\mathcal{F}_0$ denotes the fitness landscape in which the phenotypic optimum is placed at $\vec{r}_{S_0}$, whereas for $\mathcal{F}_\tau$ the phenotypic optimum is placed at $\vec{r}_{\bar{S}_0}$.

During the transient time $t \leqslant \tau$, the adaptive walker moves while the landscape is also being modified. Once the final fitness landscape $\mathcal{F}_\tau$ is reached, the adaptive walker will continue moving up to finding a fitness peak of $\mathcal{F}_\tau$. In short, the adaptive walk is definitely concluded once the transient time has elapsed and a local fitness peak of $\mathcal{F}_\tau$ has been reached. It may frequently happen, especially when the timescale of the adaptive process is shorter than that of ecological changes, that one temporarily reaches a given local fitness peak while the fitness landscape is still under environmental variation. If, by chance, the local fitness peak is no longer a local fitness peak owing to subsequent changes in the fitness landscape, we reestablish the dynamics of the adaptive walk. In Fig. 1 we present an illustration to help the understanding of the modeling, and in Table I a pseudocode of the implementation of the model.

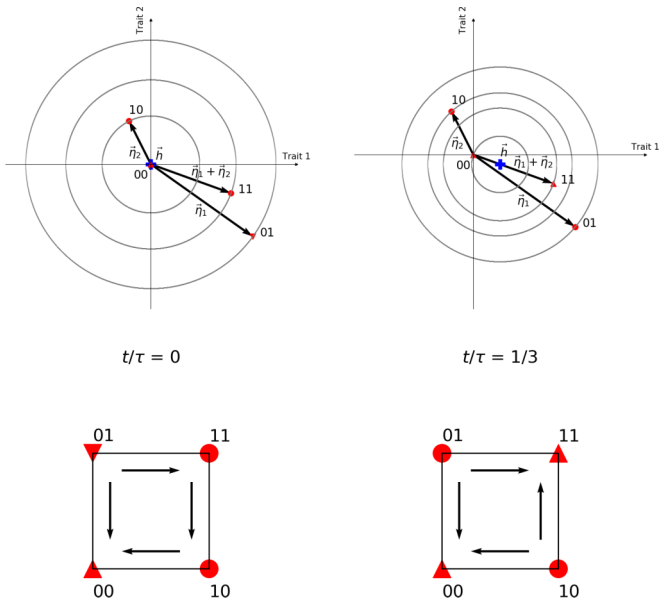

FIG. 1. Illustration of a hypothetical realization of the phenotype-fitness map. One has genome size $L = 2$, number of traits $N = 2$, and transient time $\tau = 3$. We show the illustration corresponding to two different times $t = 0$ and $t = 1$. The phenotype displacement vectors $\{\eta_k\}$ are also displayed. Phenotypes associated with all combinations of genotype are marked in red, and their distances to the optimum phenotype, which is represented by a blue plus sign, are evinced by concentric circles centered on the latter. Lower panels illustrate the corresponding genotype-fitness maps, with arrows indicating the direction of ascending fitness, down and up triangles denote minima and maxima local, respectively, and circles otherwise. As shown, the ecological dynamics entails the rearrangement of the fitness landscape. In this example, the fitness landscape displays a single local maximum at $t = 0$ and two fitness peaks at $t = 1$.

2. Implementing adaptive walks

In the adaptive walks, the whole population is visualized as a single entity traveling through a genotypic space. The walker always moves toward a domain of higher fitness. When a local optimum is achieved the process ends. There are different versions of adaptive walks. The one considered here, dubbed natural adaptive walks, chooses a fitter neighbor among all those beneficial variants with a probability that is proportional to the fitness advantage it confers [47,48]. Suppose that at a given time, the state of the population is described by genotype S whose fitness is f, and let us denote the set of its one-mutational step neighbor by $\{S_{\text{neigh}}\}$. For each genotype S_i in $\{S_{\text{neigh}}\}$, we calculate its selective effect as $s_i = f_i - f$,

TABLE I. Pseudocode for the adaptive walk with environmental variation.

Pseudocode

```
generation = 0
h⃗ = r⃗_{S_0}
walker = S_0
repeat
    h⃗ ← h⃗ + v⃗_h
    if (walker ≠ local maximum)
        adaptive step
    generation ← generation + 1
until (generation = τ)
repeat
    adaptive step
until (walker = local maximum)
```

where f_i is the fitness of S_i. Thus, the probability of moving to state S_i in the next step is given by

$$P_i(s_i) = \frac{s_i}{\sum_j s_j}, \tag{3}$$

where the sum in the denominator runs over all genotypes in $\{S_{\text{neigh}}\}$, in which $s > 0$. Note, however, that in our model the landscape is not static during a time interval τ, meaning that the fitness values f_i are themselves evolving. The process is still Markovian, but the transition probabilities are updated while the landscape is reshaped ($t \leqslant \tau$).

B. Quantifying the degree of repeatability in evolution

As defined above, our model presents stochasticity on two levels. First, the generation of a family $\{\mathcal{F}_0, \mathcal{F}_1, \ldots, \mathcal{F}_\tau\}$ of fitness landscapes relies on assigning L displacement vectors $\{\vec{\eta}_k\}$. The adaptive walker steps account for the second. We perform Monte Carlo simulations on both levels. For a given family $\{\mathcal{F}_0, \mathcal{F}_1, \ldots, \mathcal{F}_\tau\}$, after many independent trials, one has an ensemble of evolutionary trajectories to proceed with the statistical analysis. We consider different measurements to quantify the degree of repeatability of the evolutionary process. The first one is the predictability with respect to the evolutionary trajectories (first defined and applied by Ref. [57]), hereafter called Path Predictability, and calculated as

$$P_2^{\text{path}} = \sum_{\{q\}} O^2(q), \tag{4}$$

where $O(q)$ is the probability of sampling the trajectory q, i.e, its relative frequency over trials in the ensemble $\{q\}$ of evolutionary trajectories generated for a given time-varying fitness landscape. This latter condition means that in the calculation of P_2^{path}, as given by Eq. (4), we hold the same genotype-phenotype-fitness mapping when producing the ensemble of trajectories. Besides, the rate of change of the optimum phenotype is the same (i.e., the same time τ). $1/P_2^{\text{path}}$ is an estimate of the number of effective pathways exploited by the adaptive walker [58]. Similarly, we may ask about the probability that two randomly chosen paths terminate at the same local optimum after the transient time τ has elapsed. In this case we are dealing with the ensemble of endpoints instead of paths yet in the very same fashion of Eq. (4).

A limitation of P_2^{path} is that it does not account for similarities among evolutionary pathways. Two trajectories that slightly differ equally contribute to P_2^{path} as two quite different and divergent paths. A preferable quantity is the mean path divergence to assess the similarity between evolutionary pathways [59]. The mean path divergence depends on the pairwise divergence $d(q_\alpha, q_\beta)$ between two fixed and arbitrary paths q_α and q_β, and estimated as follows: For each sequence S_α belonging to q_α, we look for the minimum Hamming distance between S_α and all sequences in q_β, $h(S_\alpha, q_\beta)$. Only the minimum distance is stored. The same process is performed on the opposite direction, switching the roles of q_α and q_β. The divergence between the two evolutionary paths, q_α and q_β, is finally taken as the mean value of those shortest Hamming distances [60,61],

$$d(q_\alpha, q_\beta) = \frac{1}{n_\alpha + n_\beta}\left(\sum_{S_\alpha \in q_\alpha} h(S_\alpha, q_\beta) + \sum_{S_\beta \in q_\beta} h(S_\beta, q_\alpha)\right), \tag{5}$$

where n_α (n_β) stands for the length of trajectory q_α (q_β). The above definition for the divergence between two evolutionary paths ensures that $d(q_\alpha, q_\beta)$ is symmetric. The mean path divergence, estimated from an ensemble of evolutionary trajectories $\{q\}$, is then calculated as

$$\bar{d} = \sum_{q_\alpha \in \{q\}} O(q_\alpha) \sum_{q_\beta \in \{q\}} O(q_\beta)\, d(q_\alpha, q_\beta). \tag{6}$$

Here we emphasize that, unlike P_2^{path}, $\bar{d}$ makes sense when defined for an ensemble of paths sharing the same initial and final sequences. Therefore, the definition of mean path divergence, as given in Eq. (6), corresponds to the subset of evolutionary paths that terminated at the same endpoint. Then the average of the $\bar{d}$'s weighted by the accessibility of the endpoints is taken, which we denote $\bar{D}$.

III. RESULTS

A. The mean walk length

One of the fundamental quantities in the study of properties of fitness landscapes is the number of steps taken during the adaptive walks up to meeting a local fitness peak [55], which corresponds to the number of substitutions in the genome. The mean walk length is dependent not only on the dynamics chosen but also on the level of correlation of the fitness landscape [48,55]. For example, for completely random and uncorrelated fitness landscapes and upon greedy adaptation, in which the fittest among the one-mutational neighboring sites is always chosen, the mean number of steps is just $\bar{L} = e - 1$ [47,55], where e denotes the Euler's constant. We are mainly interested in checking the role of the key parameters, such as the rate at which the environmental variation is implemented and the number of traits on the evolutionary trajectories produced by adaptive walks.

In Fig. 2 we observe that the rate of the environmental variation, $1/\tau$, influences the dynamics. The dependence on the genome size and number of traits is also explored in the

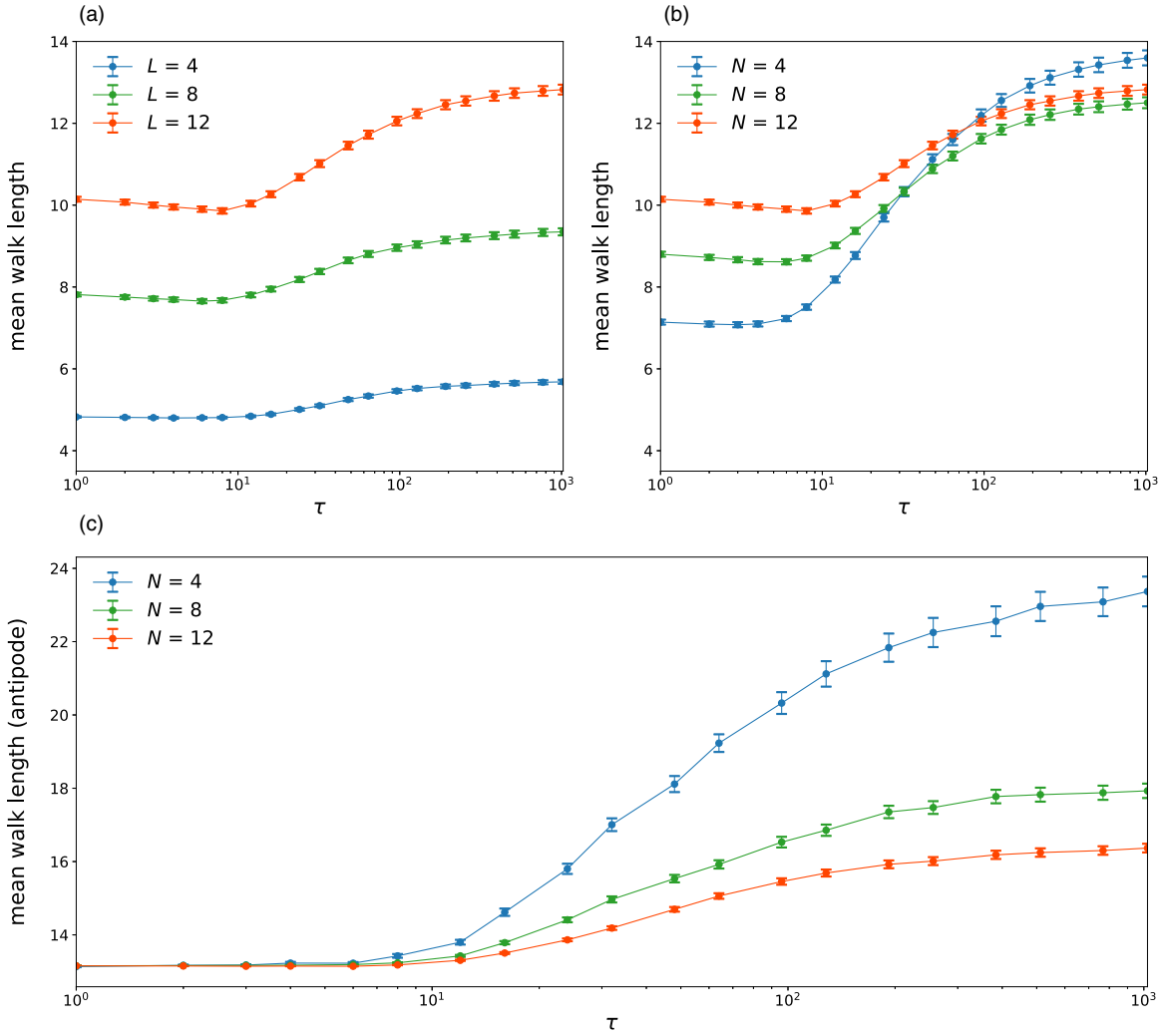

FIG. 2. Mean walk length as a function of the time interval τ of environmental changes. In panel (a), the number of traits is kept constant, $N = 12$, while distinct curves correspond to different genome size L. In panel (b), the genome size is constant, $L = 12$, while distinct curves correspond to different number of traits N. In both plots, $\sigma_{\text{mut}} = 0.05$. In panel (c), we show the mean walk length of the subset of adaptive walks that ended up at the antipode $\bar{S}_0$. The error bars are standard error of the mean over 1000 independent fitness landscapes.

plot. Small values of τ, corresponding to abrupt changes in the environment, are associated with shorter adaptive walks. Therefore, there is an increased chance that the walker more rapidly gets trapped in a local optimum of the fitness landscape. This observation occurs even though at the end of the process of environmental change, the fitness landscape ends up the same ($\mathcal{F}_\tau$). When τ is very small, the dynamics of ecological changes is faster than that of the adaptation process. The adaptive walks take longer as the transient time τ is augmented. The mean walk length faces a steep growth in the range $\tau \in [10, 100]$. Note that the onset of this domain coincides with the timescale of the adaptation to a constant environment, which is usually of the order of 10, as one can infer when τ is small. In this limit, the dynamics of adaptation is the primary mechanism determining the duration of the adaptive walks. Therefore, the steep increase of the mean walk length with τ might be related to a transition in which the ecological dynamics start to influence the duration of adaptive walks.

As expected, increased genome size results in longer adaptive walks, a fact not seen, for instance, in uncorrelated fitness landscapes [47,55]. The reason underlying this observation is that in the FGM sign epistasis becomes prevalent around the optimum phenotype [24,26]. Thus, we expect that the local optima of the fitness landscape will also settle in this domain. After the transient time τ has elapsed, the optimum phenotype will be settled on the antipode of S_0, $\bar{S}_0 = (1, 1, \ldots, 1)$. Therefore, genotypic local optima of the landscape will likely not be so dissimilar to $\bar{S}_0$. Thus, mutations might accumulate and drive the population towards the genotypic domain around $\bar{S}_0$. In Fig. 2(b), the genome size is kept constant at $L = 12$, while the distinct curves denote different numbers of traits. The dependence on the number of traits is less trivial. While for small τ, in which the timescale of the ecological dynamics is shorter than that of the evolutionary dynamics, the mean walk length rises with N. Furthermore, at intermediate and large τ we observe that the mean walk length becomes considerably large for small N. Nonetheless, the scenario seems

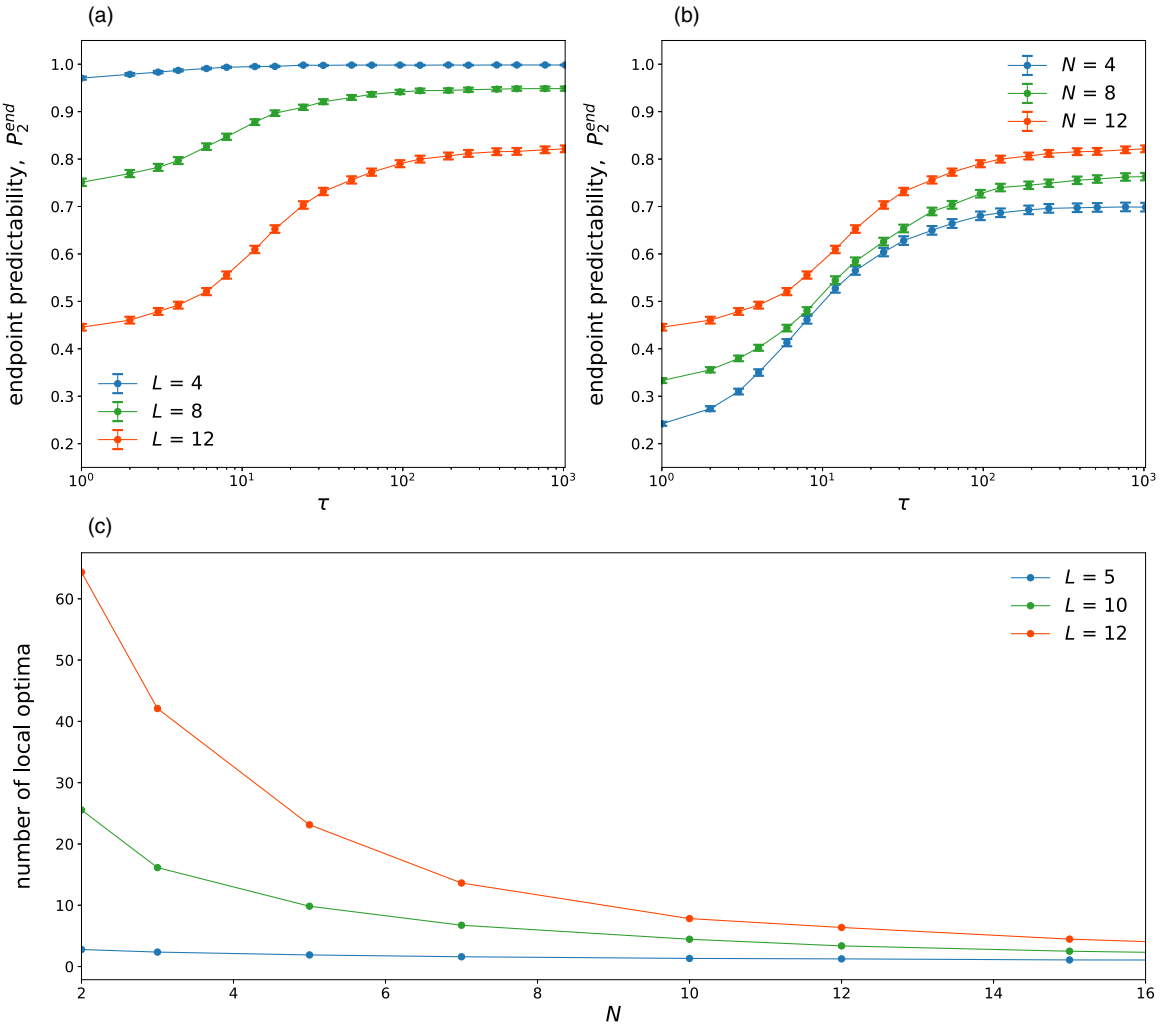

FIG. 3. Endpoint predictability, P_2^{end}, as a function of the time interval τ of environmental changes (upper panels). In panel (a), the number of traits is kept constant, $L = 12$, while distinct curves correspond to different genome size N. In panel (b), the genome size is kept constant, $N = 12$, while distinct curves correspond to different number of traits L. In panel (c), we plot the number of local optima of the fitness landscape at the end of ecological dynamics, which corresponds to $\mathcal{F}_\tau$, with the number of traits N. The genome sizes are indicated in the legends. The error bars are standard error of the mean over 1000 independent fitness landscapes.

much simpler when we only look at the mean walk length of the subset of trajectories that ended up at $\bar{S}_0$, as illustrated in Fig. 2(c). Under this restriction, a smaller number of traits means longer adaptive walks. This behavior can be understood in the light of classical studies of elementary properties of the FGM. We have learned that the fraction of beneficial mutations is approximately equal to $\frac{1}{2}erfc(\frac{x}{\sqrt{2}})$, with $x = \frac{r}{2d}\sqrt{N}$ and $erfc$ is the complementary error function [18,21,62]. Here d is the phenotypic distance to the optimum and r is the magnitude of the mutation vector. Thus, the fraction of beneficial mutations decreases with both the number of traits and the magnitude of the mutation. In this way, the decrease of the mean walk length towards $\bar{S}_0$ with the number of traits is probably a direct consequence of the reduced availability of beneficial mutations with increased N, turning their route less erratic.

B. Predictability

Predictability quantifies the level of uncertainty or entropy of the evolutionary dynamics. The larger the predictability, the lower the entropy or level of uncertainty concerning the variable of interest. It is important to highlight that sequence sizes L used in our simulations are consistent with those used in empirical studies of predictability. Indeed, empirical studies deal with a subset of the whole genome. For instance, in the seminal work of Weinreich *et al.* a sequence of size $L = 5$ is used to make predictions about predictability [63].

1. Predictability with respect to the endpoints

We start our analysis of predictability with respect to the outcomes of the adaptive walks, i.e., the endpoints of the evolutionary trajectories. As we know, the endpoints are local optima of the fitness landscape. In general, they are not equally accessible, and some of them can even be unreachable by dynamics. In Fig. 3 the dependence of the predictability with respect to the endpoints, P_2^{end}, on the transient time τ is shown. Its relation with the genome size, L, and the number of traits, N, is also explored. We see that the rate of environmental variation, $1/\tau$, strongly affect the outcome

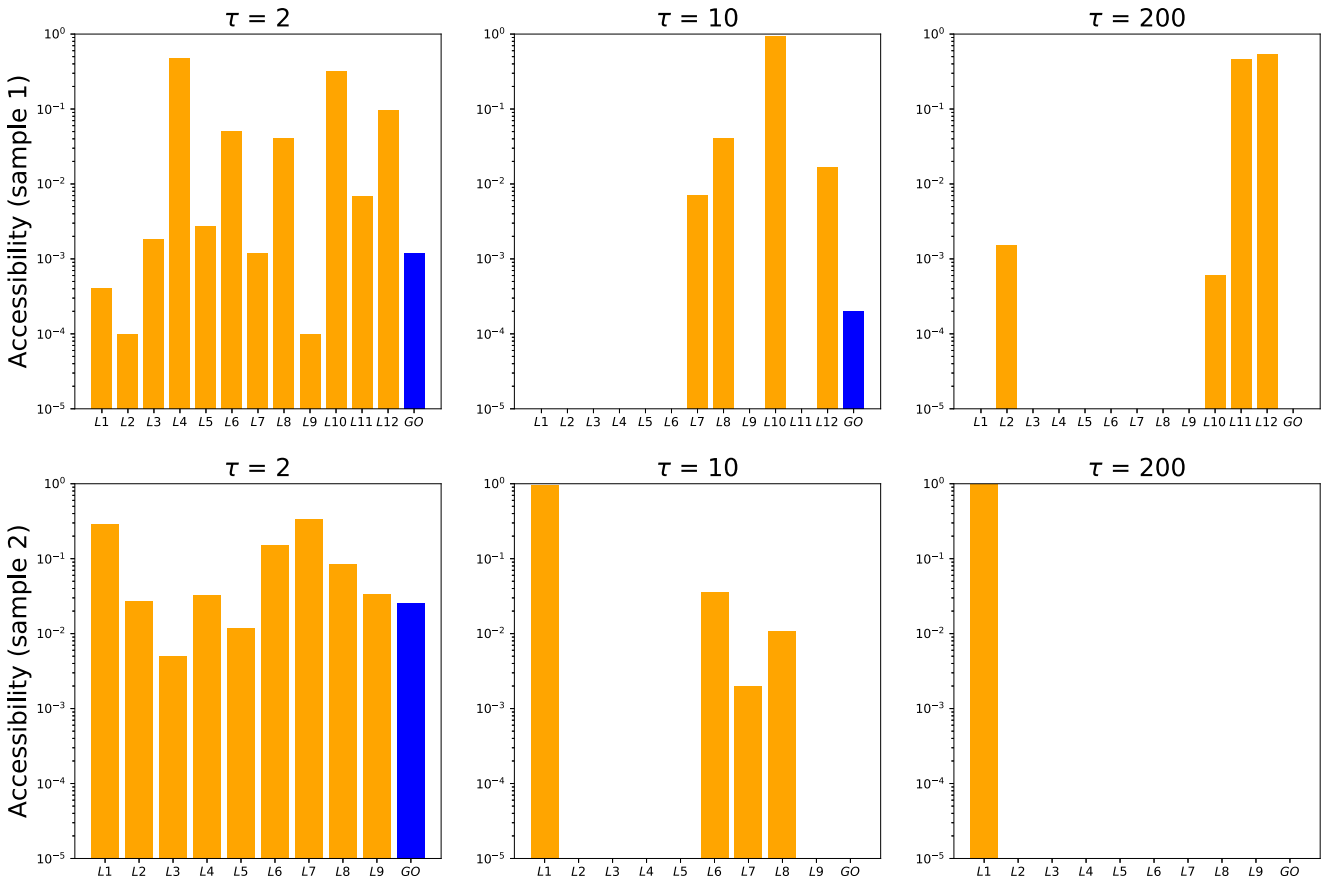

FIG. 4. Accessibility of the endpoints. Results are shown for two independent samples of the fitness landscape, and measures are made for three distinct values of τ, as indicated in the subtitle of the plots. The parameter values are $L = 12$, $N = 12$ and $\sigma_{\text{mut}} = 0.05$. The local optima are in the crescent ordering of fitness from left to right. The blue column denotes the global optimum.

of adaptive walks. Smoother environmental variations lead to more predictable outcomes. The relation of the predictability with L and N can be understood when looking at how the number of local maxima of the fitness landscape scales with both quantities. In Fig. 3(c) we see that while the number of local optima grows with the genome size L, as expected, it shrinks with the number of traits N. A smoother landscape is expected to be more predictable concerning the outcomes, whereas the endpoints of adaptation may be less predictable on rugged landscapes [64], though it remains to understand the relation of the predictability with τ. At this point, a more microscopic analysis is needed. In Fig. 4 we show the accessibility of the endpoints, the fraction of the adaptive walks that terminated at a given local optimum, for two distinct samples of fitness landscapes. For each sample of the fitness landscape, we simulated 10 000 adaptive walks. The local optima are in ascending order of fitness from left to right. The blue column denotes the global optimum. A total of 13 local optima were accessed for the former landscape when $\tau = 2$, but just four of them were visited when $\tau = 200$. The second sample displays a similar qualitative scenario. These results demonstrate that the rate at which the environmental change is introduced is a key mechanism in determining the outcome of the evolutionary process. The ecological dynamics is, in fact, reshaping and altering the attraction basin of the endpoints.

Alternatively, the entropy $S = -\sum_i p_i \ln p_i$ is exhibited in Fig. 5, where the sum is taken over all endpoints and p_i is the probability of the walk ending up at endpoint i, i.e., its accessibility. The results are in complete agreement with those observed for the predictability. Just note that higher predictability means lower entropy.

2. *Predictability with respect to the evolutionary pathways*

Now we turn our attention to the evolutionary trajectories from the wild-type sequence $S_0 = (0, 0, \ldots, 0)$ to the local optima of the fitness landscape. The dependence of the predictability with respect to the evolutionary pathways, P_2^{path}, on τ is shown in Fig. 6. Except for small N, the predictability behaves as a monotonic increasing function of τ. In Fig. 6(b), we analyze the predictability of the evolutionary paths towards $\bar{S}_0 = (1, 1, \ldots, 1)$. Once again, the behavior of the predictability for small N is distinguished from those observed for a larger number of traits N, and helps to explain the pattern shown in the left panel.

3. *Mean path divergence*

As pointed out before, the mean path divergence provides valuable insight into the evolutionary dynamics, allowing us to assess the similarity between evolutionary pathways and giving us an idea of the domain in the genotypic space

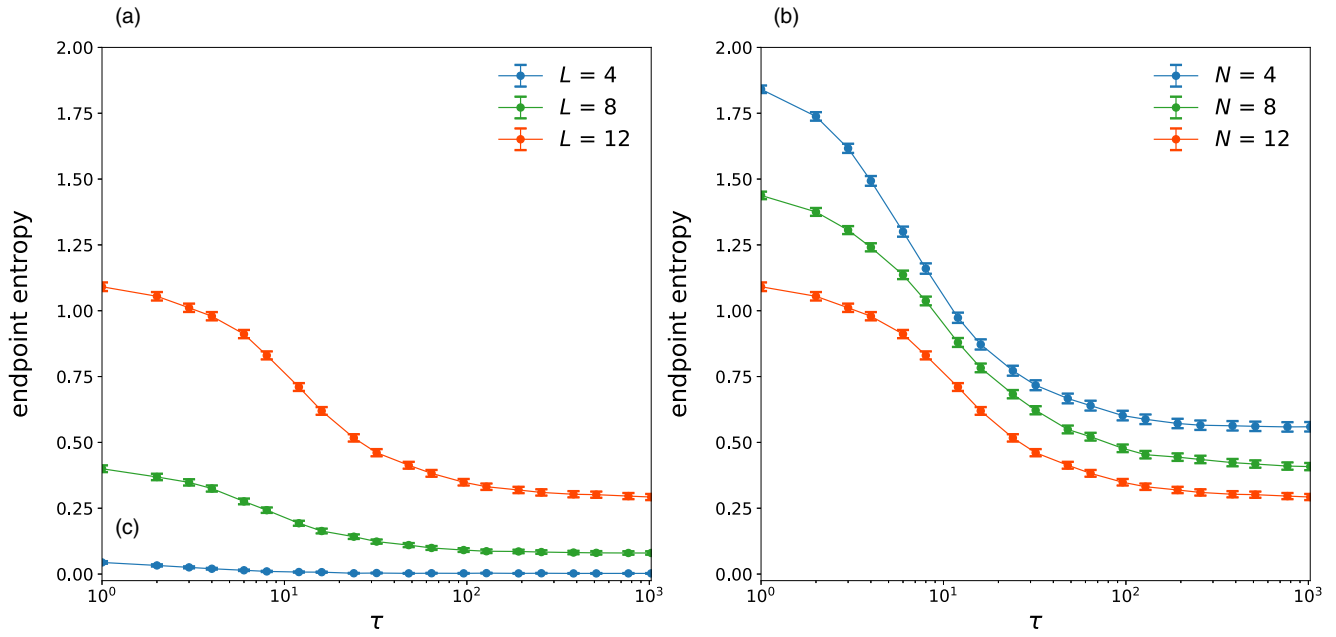

FIG. 5. Entropy of endpoints, $S = -\sum_i p_i \ln p_i$, as function of the transient time τ. In panel (a), the number of traits is kept constant, $N = 12$, while distinct curves correspond to different genome sizes L. In panel (b), the genome size is kept constant, $L = 12$, while distinct curves correspond to different number of traits N. The error bars are standard error of the mean over 1000 independent fitness landscapes.

explored by the adaptive walks. As already mentioned before, in place of measuring the mean path divergence, the initial and endpoints of the ensemble of trajectories must be the same [59,60]. We will show results for the mean path divergence over the set of endpoints, $\bar{D}$, corresponding to an average weighted by the accessibility of those endpoints, and additionally $D_{\bar{S}_0}$, which is $\bar{d}$ for the subset of paths concluded at the antipode of S_0, $\bar{S}_0$ (see Fig. 7). In all scenarios, we see that the mean path divergence is a decreasing function of the transient time τ, suggesting that smoother variation of the environmental causes evolutionary pathways to remain more condensed (less erratic and diffuse) over the genotypic space. Nicely, the mean path divergence also displays an unequivocal behavior concerning its dependence on the number of traits,

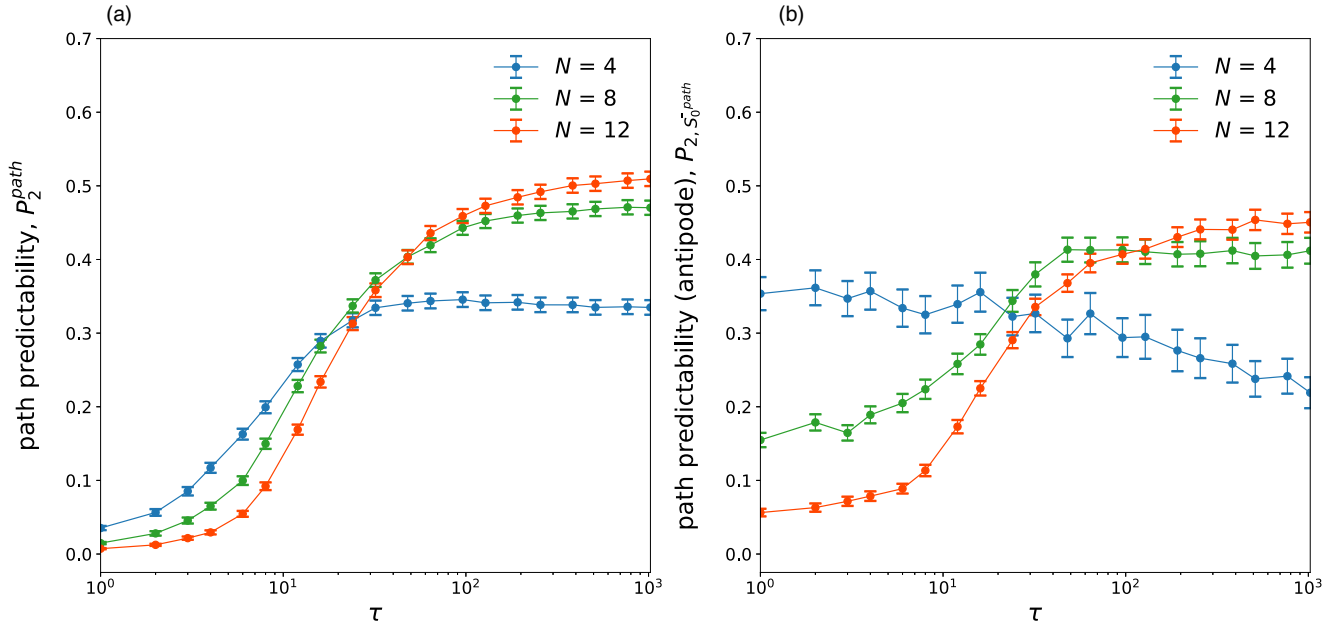

FIG. 6. Predictability with respect to the evolutionary paths. In panel (a), P_2^{path} corresponds to an average over all endpoints, whereas panel (b) shows results for the predictability for the a single endpoint, $\bar{S}_0$. In the panels, the genome size is $L = 12$, and distinct values for the number of traits, N, are considered, as indicated in the legends. The error bars are standard error of the mean over 1000 independent fitness landscapes.

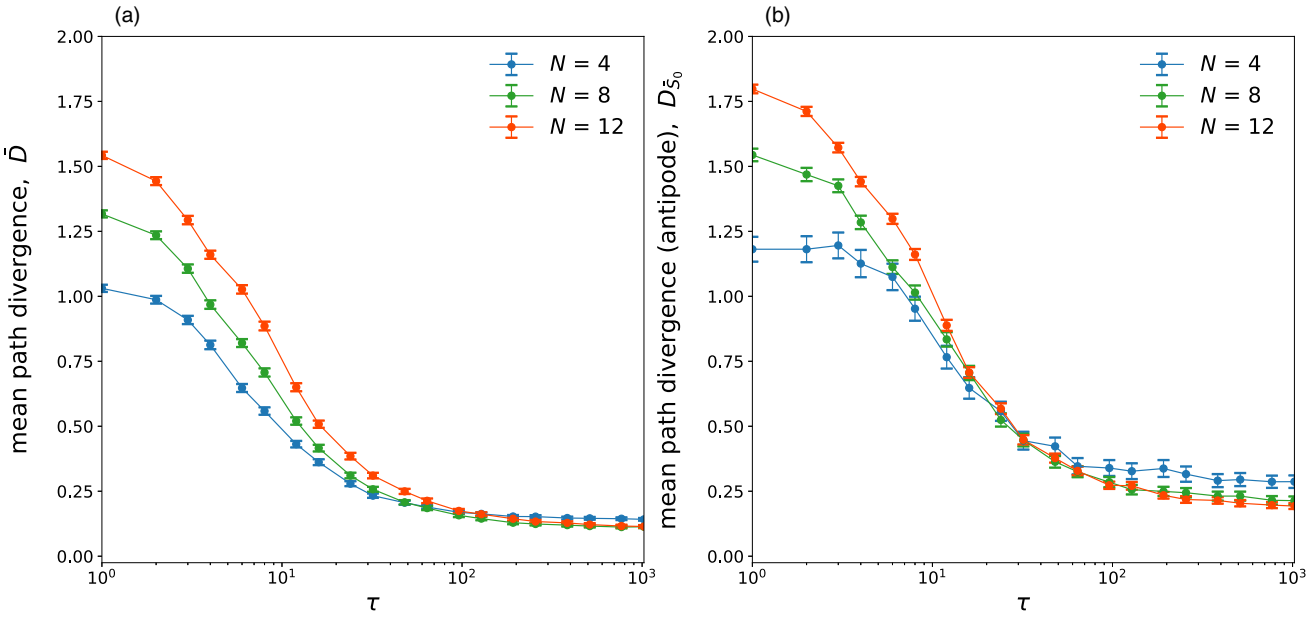

FIG. 7. Mean path divergence vs the transient time τ of environmental changes. In panel (a), the measure is taken over all endpoints, whereas in panel (b), the measure refers to the collection of evolutionary paths towards the antipode $\bar{S}_0$. The parameter values are genome size $L = 12$, standard deviation for the phenotypic effect of mutations, $\sigma_{\text{mut}} = 0.05$. The number of traits is indicated in the legends. The error bars are standard error of the mean over 1000 independent fitness landscapes.

with the mean path divergence growing with the number of traits. As τ becomes large, the influence of the number of traits ceases, and the curves nearly collapse.

Predictability and mean path divergence are usually negatively correlated, although recent statistical analyses have revealed that single realizations of a fitness landscape is likely to incorporate idiosyncratic features or misleading correlations that disappear on examination of an ensemble of landscapes [56]. Therefore, the results displayed by predictability and mean path divergence for intermediate and large number of traits look consistent concerning their dependence on the transient time τ. The anomalous behavior observed when looking at the predictability for small N is likely to be a result of the susceptibility to the existence of a collection of evolutionary paths that are different, but genetically quite similar, differing in one or a few sequences. This certainly leads to spurious effects turning difficult to unveil in a clear way its relation with the rate of environmental variation.

4. Quasistatic approximation

In the quasistatic approximation, one ensures that the ecological timescale of environmental changes is slower than the timescale of adaptation. Here this is ensured by assuming that while the walker is moving through the fitness landscape, the latter remains static. In this case, changes in the landscape are allowed only after the adaptive walker has attained a local maximum of the fitness landscape. Variations in the fitness landscape occur precisely in the same way as described by Eq. (2), with the difference that now τ refers not to a transient time but the number of moves performed by the optimum phenotype before reaching the phenotype of $\bar{S}_0$, $\vec{r}_{\bar{S}_0}$.

In Fig. 8 we compare the results for the mean walk length and endpoint predictability of the quasistatic approximation with those of the original formulation. For $\tau = 1$, the two approaches are equivalent, corresponding to standard adaptive walks in static landscapes. The lower panel plots the mean walk length. In fact, the results for the quasistatic approximation provide an upper bound for the walk lengths. This occurs because, under the quasistatic, the walker will face all possible new arrangement of local fitness peaks. In contrast, in the original formulation, some of those fitness peaks have a short lifetime compared to the time required to reach a new fitness peak under a dynamic scenario. Interestingly, we also observe that the endpoint predictability is higher under the quasistatic scheme for small τ. The same qualitative scenario holds for the path predictability. At least concerning the degree of repeatability of the evolutionary process, when $\tau \approx 50$, the two formulations tend to become equivalent, and thereby the changes in the fitness landscape over time the walker spends from moving towards a fitness peak are not enough to change the distribution of those fitness landscapes over the genotypic space.

IV. CONCLUSIONS

In the past years, we have considerably enhanced our understanding of adaptation in constant and stable environments, both theoretically and empirically [65–67]. However, an understanding of how environmental changes can affect adaptation remains challenging. Environmental variation triggers organisms' responses, which may improve their fitness by modifying their traits. Phenotypic plasticity, the ability of an organism to express different phenotypes from the same genome in response to stimuli or inputs from the environment,

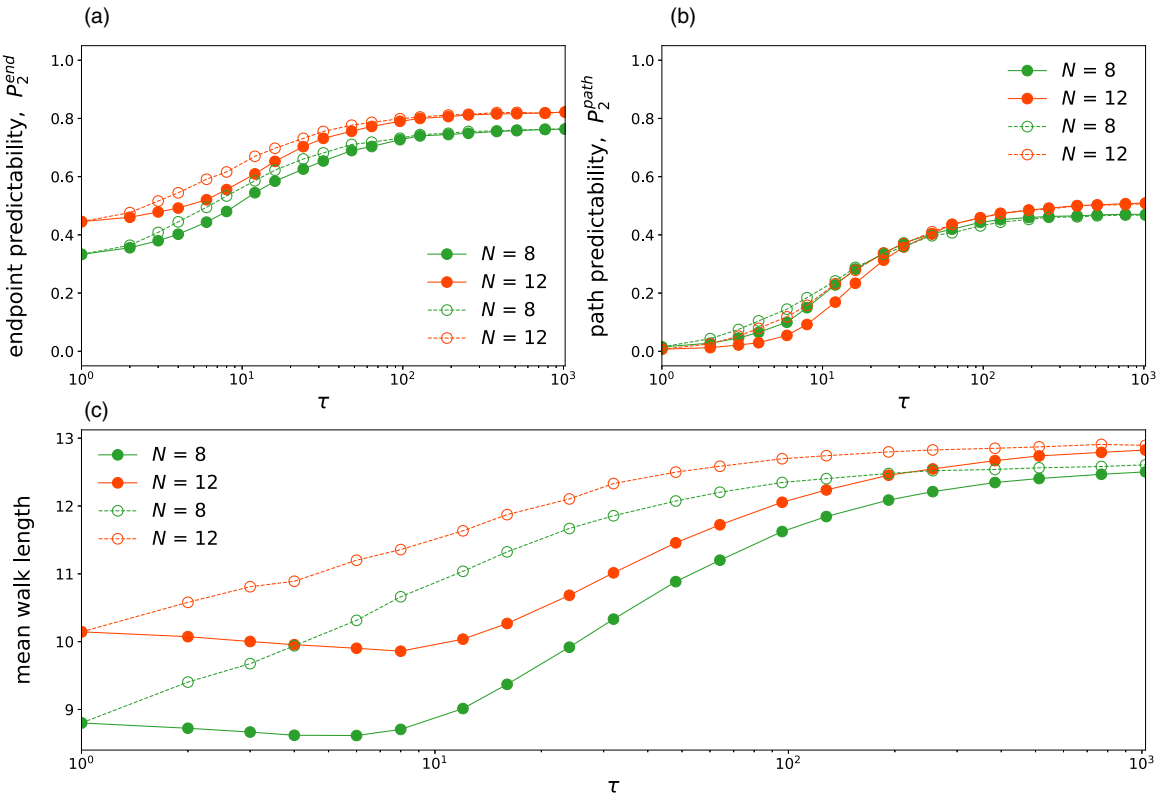

FIG. 8. Quasistatic approximation: endpoint predictability (a), path predictability (b), and mean walk and length (c) vs τ. It is important to highlight the for the quasistatic approximation τ represents the number of steps of the optimum phenotype up to reaching final state. Here we compare the results of the original formulation (filled symbols) with those of the quasistatic approximation (empty symbols). The parameter values are genome size $L = 12$, standard deviation for the phenotypic effect of mutations, $\sigma_{\text{mut}} = 0.05$. The number of traits are indicated in the legends.

is an important mechanism of adaptation, being ubiquitous [68,69]. Alternatively, evolutionary adaptation shaped by natural selection acting on genetic variation can also account for such responses [70,71].

To date, only a few experimental studies have been concerned with the role of environmental changes in microbial populations, in which the evolution of niche width and the maintenance of diversity is addressed [65,72]. In a recent study, Boyer *et al.* outlined an experimental setup to investigate the dynamics of evolution of yeast populations during adaptation to a changing environment and follow the role of the rate of switching between environmental conditions [71]. Their findings showed that different switching dynamics could select for different phenotypic and genotypic outcomes.

Our present study also aims to better characterize the dynamics of the evolution of populations under a scenario of environmental variation. It is important to emphasize that as opposed to the environmental dynamics used in Ref. [71], in our formulation, the switching between the two ecological conditions ensues from a series of temporary intermediate states. In particular, we look for quantification of the adaptive process itself by measuring not only the degree of repeatability of outcomes but also evolutionary trajectories. Using an adaptive walk approximation, expected to hold in a scenario of strong-selection weak mutation, we have shown that the rate of environmental variation influences the number of substitutions that occur during adaptive walks to a locally optimal genotype. In general, slowly changing environments leads to a greater number of substitutions.

According to our simulation results, abrupt environmental variations lead to more unpredictable outcomes. The increase of the predictability with the transient time τ can be understood as a drastic reduction of accessibility of some of the locally optimal genotypes. The rise of predictability with τ is notably steeper when the timescale of ecological variation surpasses that of the timescale of adaptation. The results reveal the role of the dynamics of environmental changes in shaping the attraction basin of the locally optimal genotypes. This occurs concomitantly with the constraint of the evolutionary paths towards those local optima of the fitness landscape, as clearly established by measures of the predictability with respect to the evolutionary pathways and mean path divergence.

A fundamental issue is understanding how predictable and parallel evolution response is across replicated environments [73]. To our knowledge, the current contribution is the first to use the predictive approach to quantify and understand the role of the speed of environmental variation on the degree of repeatability of the evolutionary process. This problem is not only an important matter from an evolutionary perspective but also closely related to the

theory of speciation, as the degree of parallelism of the evolutionary process can be used as a proxy for the likelihood of population divergence and subsequent reproductive isolation as a consequence of temporary or permanent geographic barriers [74]. The present results highlight the impact of the speed of environmental change on the process, which we believe sheds new light on future directions for the ecological theory of speciation and evolutionary biology.

ACKNOWLEDGMENTS

P.R.A.C. is partially supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). P.R.A.C. also acknowledges financial support from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) Project No. 0041/2022 and Universidade Federal de Pernambuco Edital Produção Qualificada. D.C. is supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) under Grant Number 302569/2018-9.

[1] P. F. Stadler, in *Biological Evolution and Statistical Physics* (Springer, Berlin, Germany, 2002), pp. 183–204.
[2] D. Wales, *Energy Landscapes: Applications to Clusters, Biomolecules and Glasses* (Cambridge University Press, 2003).
[3] S. Wright, Genetics **16**, 97 (1931).
[4] B. W. B. Shires and C. J. Pickard, Phys. Rev. X **11**, 041026 (2021).
[5] S. Nowak and J. Krug, J. Stat. Mech.: Theory Exp. (2015) P06014.
[6] F. Blanquart and T. Bataillon, Genetics **203**, 847 (2016).
[7] J. De Visser, S. F. Elena, I. Fragata, and S. Matuszewski, Heredity **121**, 401 (2018).
[8] I. G. Szendro, M. F. Schenk, J. Franke, J. Krug, and J. A. G. De Visser, J. Stat. Mech.: Theory Exp. (2013) P01005.
[9] W. Rowe, M. Platt, D. C. Wedge, P. J. Day, D. B. Kell, and J. Knowles, J. R. Soc. Interface **7**, 397 (2010).
[10] S. Kauffman and S. Levin, J. Theor. Biol. **128**, 11 (1987).
[11] F. J. Poelwijk, S. Tănase-Nicola, D. J. Kiviet, and S. J. Tans, J. Theor. Biol. **272**, 141 (2011).
[12] V. M. de Oliveira and J. Fontanari, J. Phys. A: Math. Gen. **32**, 2285 (1999).
[13] V. M. de Oliveira and J. F. Fontanari, Phys. Rev. Lett. **85**, 4984 (2000).
[14] V. M. de Oliveira and J. F. Fontanari, Phys. Rev. Lett. **89**, 148101 (2002).
[15] H. A. Orr, Nat. Rev. Genet. **6**, 119 (2005).
[16] R. A. Fisher, *The Genetical Theory of Natural Selection* (Clarendon Press, Oxford, 1930).
[17] D. Waxman, J. Theor. Biol. **241**, 887 (2006).
[18] O. Tenaillon, Annu. Rev. Ecol. Evol. Syst. **45**, 179 (2014).
[19] L. Chao, Nature (London) **348**, 454 (1990).
[20] T. T. Kibota and M. Lynch, Nature (London) **381**, 694 (1996).
[21] D. L. Hartl and C. H. Taubes, J. Theor. Biol. **182**, 303 (1996).
[22] D. Waxman and J. R. Peck, Science **279**, 1210 (1998).
[23] G. Martin and T. Lenormand, Evolution **60**, 893 (2006).
[24] S. Hwang, S.-C. Park, and J. Krug, Genetics **206**, 1049 (2017).
[25] D. M. Weinreich, R. A. Watson, and L. Chao, Evolution **59**, 1165 (2005).
[26] F. Blanquart, G. Achaz, T. Bataillon, and O. Tenaillon, Evolution **68**, 3537 (2014).
[27] N. Colegrave and A. Buckling, Bioessays **27**, 1167 (2005).
[28] M. L. Salverda, E. Dellus, F. A. Gorter, A. J. Debets, J. Van Der Oost, R. F. Hoekstra, D. S. Tawfik, and J. A. G. de Visser, PLoS Genet. **7**, e1001321 (2011).
[29] L.-M. Chevin, G. Decorzent, and T. Lenormand, Evolution **68**, 1244 (2014).
[30] G. Martin, S. F. Elena, and T. Lenormand, Nat. Genet. **39**, 555 (2007).
[31] D. R. Rokyta, P. Joyce, S. B. Caudle, C. Miller, C. J. Beisel, and H. A. Wichman, PLoS Genet. **7**, e1002075 (2011).
[32] L. Perfeito, A. Sousa, T. Bataillon, and I. Gordo, Evolution **68**, 150 (2014).
[33] R. Bürger and M. Lynch, Evolution **49**, 151 (1995).
[34] I. Gordo and P. R. Campos, Biol. Lett. **9**, 20120239 (2013).
[35] S. Matuszewski, J. Hermisson, and M. Kopp, Evolution **68**, 2571 (2014).
[36] M. Amicone and I. Gordo, Evolution **75**, 2641 (2021).
[37] O. Freitas, S. B. Araujo, and P. R. Campos, Ecol. Model. **468**, 109958 (2022).
[38] W. Guo, S. Lampoudi, and J.-E. Shea, Proteins: Struct Func. Bioinfo. **55**, 395 (2004).
[39] P. Shah and M. A. Gilchrist, PLoS ONE **5**, e11308 (2010).
[40] C. Liu and Y. Fan, Phys. Rev. Lett. **127**, 215502 (2021).
[41] F. Botta, D. Dahl-Jensen, C. Rahbek, A. Svensson, and D. Nogués-Bravo, Curr. Biol. **29**, R1045 (2019).
[42] S. J. Gould, *The Structure of Evolutionary Theory* (Harvard University Press, 2002).
[43] P. Nosil, S. M. Flaxman, J. L. Feder, and Z. Gompert, Nat. Commun. **11**, 5592 (2020).
[44] A. Mas, Y. Lagadeuc, and P. Vandenkoornhuyse, iScience **23**, 101736 (2020).
[45] A. McGaughran, K. Morgan, and R. J. Sommer, PLoS ONE **9**, e87317 (2014).
[46] J. H. Gillespie, Evolution **38**, 1116 (1984).
[47] P. R. A. Campos and F. G. B. Moreira, Phys. Rev. E **71**, 061921 (2005).
[48] J. de Lima Filho, F. Moreira, P. Campos, and V. M. De Oliveira, J. Stat. Mech.: Theory Exp. (2012) P02014.
[49] S.-C. Park, I. G. Szendro, J. Neidhart, and J. Krug, Phys. Rev. E **91**, 042707 (2015).
[50] J. Neidhart and J. Krug, Phys. Rev. Lett. **107**, 178102 (2011).
[51] L. Prignano, Y. Moreno, and A. Díaz-Guilera, Phys. Rev. E **86**, 066116 (2012).
[52] P. A. Romero and F. H. Arnold, Nat. Rev. Mol. Cell Biol. **10**, 866 (2009).
[53] A. Liefooghe, F. Daolio, S. Verel, B. Derbel, H. Aguirre, and K. Tanaka, IEEE Trans. Evol. Comput. **24**, 1063 (2019).
[54] J. Franke, A. Klözer, J. A. G. de Visser, and J. Krug, PLoS Comput Biol **7**, e1002134 (2011).
[55] H. A. Orr, J. Theor. Biol. **220**, 241 (2003).
[56] S. M. Reia and P. R. Campos, R. Soc. Open Sci. **7**, 192118 (2020).

[57] S. W. Roy, PLoS ONE **4**, e4500 (2009).
[58] J. A. G. De Visser and J. Krug, Nat. Rev. Genet. **15**, 480 (2014).
[59] A. E. Lobkovsky, Y. I. Wolf, and E. V. Koonin, PLoS Comput Biol **7**, e1002302 (2011).
[60] M. Manhart and A. V. Morozov, in *First-Passage Phenomena and Their Applications* (World Scientific, 2014), pp. 416–446.
[61] A. E. Lobkovsky and E. V. Koonin, Front. Genet. **3**, 246 (2012).
[62] Y. Ram and L. Hadany, Theor. Pop. Biol. **99**, 1 (2015).
[63] D. M. Weinreich, N. F. Delaney, M. A. DePristo, and D. L. Hartl, Science **312**, 111 (2006).
[64] D. L. Hartl, Curr. Opin. Microbiol. **21**, 51 (2014).
[65] C. Bleuven and C. R. Landry, Proc. R. Soc. B. **283**, 20161458 (2016).
[66] B. H. Good, M. J. McDonald, J. E. Barrick, R. E. Lenski, and M. M. Desai, Nature (London) **551**, 45 (2017).
[67] V. S. Cooper, mSphere **3**, e00121-18 (2018).
[68] A. A. Agrawal, Science **294**, 321 (2001).
[69] V. Oostra, M. Saastamoinen, B. J. Zwaan, and C. W. Wheat, Nat. Commun. **9**, 1005 (2018).
[70] M. Shimada, Y. Ishii, and H. Shibao, Pop. Ecol. **52**, 5 (2010).
[71] S. Boyer, L. Hérissant, and G. Sherlock, PLoS Genet. **17**, e1009314 (2021).
[72] J.-N. Jasmin and R. Kassen, Proc. R. Soc. B **274**, 2761 (2007).
[73] R. B. Langerhans, J. Hered. **109**, 59 (2018).
[74] R. Yamaguchi and S. P. Otto, Evolution **74**, 1603 (2020).