Time-dependent extinction rate and species abundance in a tangled-nature model of biological evolution

Matt Hall,[1] Kim Christensen,[2] Simone A. di Collobiano,[2] and Henrik Jeldtoft Jensen[1,*]
[1]*Department of Mathematics, Imperial College, 180 Queen's Gate, London SW7 2BZ, United Kingdom*
[2]*Blackett Laboratory, Imperial College, Prince Consort Road, London SW7 2BW, United Kingdom*
(Received 4 February 2002; revised manuscript received 18 March 2002; published 12 July 2002)

We present a model of evolutionary ecology consisting of a web of interacting individuals, a tangle-nature model. The reproduction rate of individuals characterized by their genome depends on the composition of the population in genotype space. Ecological features such as the taxonomy and the macroevolutionary mode of the dynamics are emergent properties. The macrodynamics exhibit intermittent two-mode switching with a gradually decreasing extinction rate. The generated ecologies become gradually better adapted as well as more complex in a collective sense. The form of the species abundance curve compares well with observed functional forms. The model's error threshold can be understood in terms of the characteristics of the two dynamical modes of the system.

DOI: 10.1103/PhysRevE.66.011904 PACS number(s): 87.23.Cc, 87.23.Kg, 05.65.+b

I. INTRODUCTION

The dynamics and organization of biological ecosystems is a fascinating example of complex interacting systems with many levels of emerging structure and time scales. Biological evolution creates intricate taxonomic hierarchies presumably as an effect of mutation, natural selection, and the ensuing adaptation. Taxonomic structures from the level of individuals through species and genera up to kingdoms are generated and vanish again in a never ending succession. Different strata in the hierarchy are described by very different time scales and with very different types of dynamics. At the level of individuals, fairly well defined characteristic lifetimes exist for each specific type (species) and the population dynamics can be considered smooth. This picture changes as one considers the system at the more coarse grained level of species and genera. The lifetime distribution of, e.g., genera is broad (see, e.g., Ref. [1]) and the dynamics is intermittent [2–5]. In the spirit of the traditional approach of statistical mechanics it is interesting to consider models, defined at a microscopic level, which are able to reproduce the large scale temporal and taxonomic structures.

In the present paper we consider a model of individuals identified solely by their genome. The model was introduced in Ref. [6] (where we presented a discussion of the qualitative features) under the name the tangled-nature (TaNa) model with an allusion to Darwin's notion of the *Tangled Bank* to stress the model's emphasis on ecological interactions. We combine ecology with evolution by considering interacting individuals that can multiply (sexually or asexually) subject, potentially, to mutations. The size of the total population fluctuates, the average being controlled by the amount of available resources. From these three minimal ingredients emerge segregation in genome space, to be interpreted as the appearance of species, and a complex intermittent dynamics, to be interpreted as extinction and creation events at the higher taxonomic levels. The entire taxonomic hierarchy is an emergent property of the dynamics at the microscopic level of individuals. We characterize the configurations generated in genotype space in terms of the species abundance curve, and find a good qualitative agreement with the functional form typically found for real ecosystems. The intermittent dynamics is characterized by the statistics of the duration of the quasistable epochs or in other words the waiting times between transitions. We find a broad distribution of durations and observe a gradual change of the average extinction rate. No stationary state is ever reached.

Related models

Many mathematical models of biological evolution have been developed according to the usual statistical mechanics agenda of generating the macroscopic complex behavior from simplistic microscopic definitions. An elegant review of this endeavor has recently been given by Drossel [7]. For reviews from a more biological point of view, see, e.g., the two excellent papers by Burt [8] and Leroi [9]. Here we limit ourselves to a discussion of similarities and differences between our model and related studies.

Let us first mention models that define the ecosystem in terms of individuals. Higgs and Derrida [10] studied speciation in a model consisting of a fixed number of individuals. Each individual is represented by a genome modeled as a string of zeros or ones, like in Eigen and co-workers' seminal work on quasispecies [11]. Higgs and Derrida demonstrated that a sexually reproducing population breaks up into distinct species when only individuals with a sufficiently similar genome sequence are allowed to produce offspring. This agrees with a large bulk of experimental work [12]. Gavrilets and co-workers [13–15] have made use of similar models generalized in particular to be able to study geographical and temporal aspects of speciation. These studies differ from ours in assuming a fixed population size and by defining a fitness function for pairs of individuals that is constant if the Hamming distance between the genomes of two individuals is

*Author to whom correspondence should be addressed. Email address: h.jensen@ic.ac.uk; URL:http://www.ma.ic.ac.uk/~hjjens/

small enough, and zero otherwise. Our model allows the total size of the population to fluctuate and the fitness of pairs of individuals (or in the asexual case single individuals) depends on the composition of the population at a given instant in time.

It is also important to mention the fitness landscape approach first pioneered by Wright [16,17], who considered gene frequencies, and was brought to the attention of the statistical mechanics community mainly through Kauffman's so-called NK model [18,19]. The main focus of the NK model, and of the later coevolutionary NKC model [19], is the study of epistatic interactions (the influence of one gene on another) by use of fitness functions. The main difference between our model and Kauffman's models is that the fitness of an individual in our system depends on the *frequencies* by which other locations in genotype space are occupied.

Taylor and Higgs [20] have studied pleiotropy and epistasis (the influence of one gene on several traits and the influence of one gene on another) in a model that combines and generalizes aspects of the Higgs-Derrida model with the epistatic interactions of Kauffman's models. Taylor and Higgs then derive a phenotypical fitness for the specific genotype. Kaneko and Yomo [21] have also studied models in which the difference between phenotype and genotype is accounted explicitly. In our model we make the drastic simplification not to distinguish between genotype and phenotype.

Other models consider species as the elementary building block; these models neglect the specifics of the dynamics arising from reproduction and mutations at the level of individuals. The simplest of these models is the Bak-Sneppen model [22]. The model aims to demonstrate that coevolutionary interactions are sufficient to produce intermittent dynamics that is then related to intermittency in the fossil record and to Eldredge and Gould's concept of punctuated equilibrium [2–5]. Each species is characterized by a single number between zero and one, the fitness, and the total number of species is kept constant. The model has interesting statistical properties but is difficult to relate to biological evolution.

Species level models of more detail than the Bak-Sneppen model have been formulated recently by McKane, Alonso, and Solé [23] and by Drossel, Higgs, and McKane [24]. The emphasis in these models is on predator-prey interactions and food webs and are generalizations of early work by May [25] and May and Anderson [26]. Our model is intended to include all types of interactions between individuals, e.g., antagonistic or collaborative relationships, in addition to predator-prey competitions. Another important difference is that we define our model at the level of individuals in order to be able to study the emergence of species, something not possible in a species based model.

Most models of biological evolution assume that the dynamics is in a statistically stationary state. One marked exception is the model considered by Sibani and co-workers [27–30]. This is an abstract species based model consisting of random walks in a rugged fitness landscape. The statistics of the jumps in this landscape are the same as the record statistics considered some time ago by Sibani and Littlewood [31]. The pace of the dynamics of the model gradually slows down as indicated by a logarithmically decreasing extinction rate. As we shall see below our individual based model also exhibits a decreasing average extinction rate, a property found to be consistent with analysis of the fossil record [1].

The paper is organized as follows. In the following section, we define the model in detail. In Sec. III we discuss the modes of the model's emergent dynamics. In Sec. IV we show how the configurations generated dynamically gradually become better adapted in a collective sense. Section V demonstrates that the ecologies generated in the model exhibit characteristics similar to those observed in real ecologies. Section VI contains an analysis of the error threshold. We briefly present in Sec. VII a scan of the behavior of the model for a range of the control parameters and in Sec. VIII we conclude and summarize.

II. DEFINITION OF MODEL

We describe here in detail the structure and dynamics of the tangled-nature model.

A. Interaction

We represent an individual by a vector $\mathbf{S}^\alpha = (S_1^\alpha, S_2^\alpha, \ldots, S_L^\alpha)$ in genotype space $\mathcal{S}$. This representation is frequently used; see, e.g., Refs. [10,11,13,19,32]. Here S_i^α may take the values ± 1, i.e., $\mathbf{S}^\alpha$ denotes one of the corners of the L-dimensional hypercube (in the present paper we use $L=20$). The coordinates S_i^α may be interpreted as genes with two alleles, or a string of either pyrimidines or purines. We think of genotype space $\mathcal{S}$ as containing all possible ways of combining the genomic building blocks into genome sequences. Many sequences may not correspond to viable organisms. Whether this is the case or not is for the evolutionary dynamics to determine. All possible sequences are made available for evolution to select from.

Individuals are labeled by greek letters $\alpha, \beta, \ldots = 1,2,\ldots,N(t)$. When we refer, without reference to a specific individual, to one of the 2^L positions in genome space, we use roman superscripts $\mathbf{S}^a, \mathbf{S}^b, \ldots$ with $a, b, \ldots = 1,2,\ldots,2^L$. Many different individuals $\mathbf{S}^\alpha, \mathbf{S}^\beta, \ldots$, may reside on the same position, say $\mathbf{S}^a$, in $\mathcal{S}$.

The ability of an individual α to reproduce is controlled by $H(\mathbf{S}^\alpha, t)$:

$$H(\mathbf{S}^\alpha, t) = \frac{1}{cN(t)} \sum_{\mathbf{S}\in\mathcal{S}} J(\mathbf{S}^\alpha, \mathbf{S}) n(\mathbf{S}, t) - \mu N(t), \qquad (1)$$

where c is a control parameter (see below), $N(t)$ is the total number of individuals at time t, the sum is over the 2^L locations $\mathbf{S}$ in $\mathcal{S}$ and $n(\mathbf{S}, t)$ is the occupancy of position $\mathbf{S}$. Two positions $\mathbf{S}^a$ and $\mathbf{S}^b$ in genome space are coupled with the fixed random strength $J^{ab} = J(\mathbf{S}^a, \mathbf{S}^b)$ that can be either positive, negative or zero. The coupling is nonzero with probability Θ (throughout the paper we use $\Theta = 0.25$), in which case we assume $J^{ab} \neq J^{ba}$ to be a deterministic but erratic function of the two positions $\mathbf{S}^a$ and $\mathbf{S}^b$. We have checked that the specific details of the form of the distribution of the nonzero values of the function $J(\mathbf{S}^a, \mathbf{S}^b)$ are irrelevant. We choose accordingly a form mainly determined by its numerical effi-

ciency. In the following subsection we describe the details of the specific procedure used. The distribution of the generated interaction strengths is shown in Fig. 3 below.

1. Generation of interaction matrix

The interaction between two locations in genotype space, $\mathbf{S}^a$ and $\mathbf{S}^b$ is generated as a product $J(\mathbf{S}^a,\mathbf{S}^b) = \Theta(\mathbf{S}^a,\mathbf{S}^b)I(\mathbf{S}^a,\mathbf{S}^b)$. The first factor $\Theta(\mathbf{S}^a,\mathbf{S}^b)$ is obtained by interpreting the sequences $\mathbf{S}^a$ and $\mathbf{S}^b$ as binary numbers (letting $-1 \mapsto 0$) and perform the XOR operation on the binary pair to obtain a new integer. This integer is used as an index in a lookup list to obtain either a 0 or 1 as the value of $\Theta(\mathbf{S}^a,\mathbf{S}^b)$. In case 1 is returned, the element of the $I(\mathbf{S}^a,\mathbf{S}^b)$ matrix is obtained in a similar way. This time, however, two arrays are needed. Each auxiliary array is of length 2^L and now the arrays contain uniformly distributed random numbers drawn from the interval $[-1,+1]$. The pair of arrays is necessary in order to reproduce the asymmetry of the $I(\mathbf{S}^a,\mathbf{S}^b)$ matrix. Two indices are generated from the $\mathbf{S}^a$ and $\mathbf{S}^b$. The first one via the same XOR operation is used to calculate the $\Theta(\mathbf{S}^a,\mathbf{S}^b)$ matrix element, whereas the second is simply the integer representing $\mathbf{S}^b$. The strength of interaction is taken to be the product of the members of each array at the appropriate location. This ensures that the elements of the matrices are nonsymmetric due to the second array index depending on the order of the operation. This procedure is numerically extremely efficient and deterministic, but has the side effect of generating a distribution of a slightly unusual form, see Fig. 3.

We stress that the coupling matrix $J(\mathbf{S}^a,\mathbf{S}^b)$ is meant to included *all* possible interactions between two individuals of a given genomic constitution. In our simplistic approach, a given genome is imagined to lead uniquely to a certain set of attributes (phenotype) of the individuals/organisms. The locations $\mathbf{S}^a$ and $\mathbf{S}^b$ represent blueprints for organisms that exist *in potentia*. The positions may very likely be unoccupied but, if we were to construct individuals according to the sequences $\mathbf{S}^a$ and $\mathbf{S}^b$ the two individuals would have some specific features. The relationship between an organism of *design* $\mathbf{S}^a$ and one of design $\mathbf{S}^b$ may be as predator and prey or parasitic, i.e., $J^{ab}>0$ and $J^{ba}<0$, but it can also be collaborative ($J^{ab}>0$ and $J^{ba}>0$) or antagonistic ($J^{ab}<0$ and $J^{ba}<0$), see Fig. 1. And certainly in some cases J^{ab} may represent less direct couplings, e.g., some animals may not eat trees, nevertheless they breath the oxygen produced by the rain forest. In order to emphasis co-evolutionary aspects we have *excluded* "self-interaction" among individuals located at the same position $\mathbf{S}$ in genome space, i.e., $J(\mathbf{S},\mathbf{S}) = \mathbf{0}$ for all $\mathbf{S} \in \mathcal{S}$. It is important to mention that including self-interactions of the same order of strength as the J couplings do not change the qualitative behavior of the model.

The width of the distribution of couplings in the first term in Eq. (1) is determined by the parameter c. The $J(\mathbf{S}^a,\mathbf{S}^b)$ are all distributed between -1 and 1 and therefore $J(\mathbf{S}^a,\mathbf{S}^b)/c$ can assume values between $-1/c$ and $1/c$. A very inhomogeneous population, in which different types of individuals can influence each other in very different ways, corresponds to a broad range of possible couplings and, hence, to a small value of c.

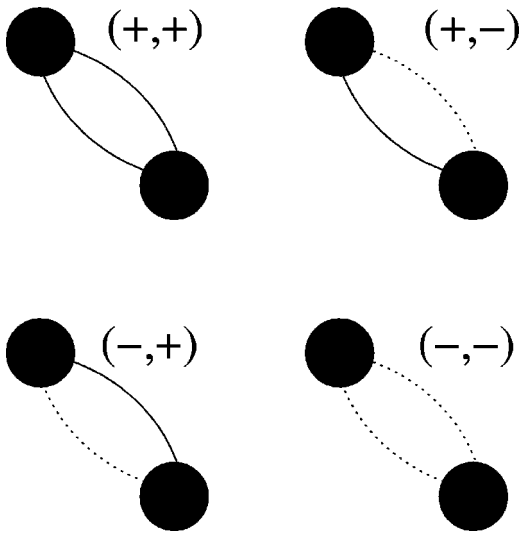

FIG. 1. Examples of possible realizations of the couplings J^{ab} between different positions $\mathbf{S}^a$ and $\mathbf{S}^b$ in genotype space representing collaborative $(+,+)$, antagonistic $(-,-)$ or predator prey $(+,-)$, and $(-,+)$ relationships.

The conditions of the physical environment are simplistically described by the term $\mu N(t)$ in Eq. (1), where μ determines the average sustainable total population size. That is, the total carrying capacity of the environment. An increase in μ corresponds to harsher physical conditions. This is a simplification, though one should remember that what is often considered as the physical conditions, e.g., temperature or oxygen density, is to a degree determined by the activity of other organisms and is therefore really a part of the biotic conditions. Consider, for example, the environment experienced by the bacterial flora in the intestines. Here one type of bacteria live very much in an environment strongly influenced by the presence of other types of bacteria. In this sense some fluctuations in the environment may be thought of as included in the coupling matrix $J(\mathbf{S}^a,\mathbf{S}^b)$.

B. Reproduction, mutations, and annihilation

Asexual reproduction consists of one individual being replaced by two copies. Successful reproduction occurs for individuals $\mathbf{S}^\alpha$ with a probability per time unit given by

$$p_{off}(\mathbf{S}^\alpha,t) = \frac{\exp[H(\mathbf{S}^\alpha,t)]}{1+\exp[H(\mathbf{S}^\alpha,t)]} \in [0,1]. \tag{2}$$

In the case of sexual reproduction an individual $\mathbf{S}^\alpha$ is picked at random and paired with another randomly chosen individual $\mathbf{S}^\beta$ with Hamming distance $d = \frac{1}{2}\Sigma_{i=1}^{L}|S_i^\alpha - S_i^\beta| \leqslant d_{max}$ (allowing at most d_{max} pairs of genes to differ). The pair produces an offspring γ with probability $\sqrt{p_{off}(\mathbf{S}^\alpha,t)p_{off}(\mathbf{S}^\beta,t)}$, where S_i^γ is chosen at random from one of the two parent genes, either S_i^α or S_i^β. For $d_{max} \geqslant 1$ this procedure may be thought of as being similar to recom-

bination. The maximum separation criterion has been studied by several authors, see, e.g., Refs. [10,13].

We allow for mutations in the following way: with probability p_{mut} per gene we perform a change of sign $S_i^\gamma \rightarrow -S_i^\gamma$, during the reproduction process.

For simplicity, an individual is removed from the system with a constant probability p_{kill} per time step (we use $p_{kill}=0.2$). This procedure is implemented both for asexual and sexually reproducing individuals. Darwinian evolution assumes that reproduction rates are of prime importance. The total number of offspring produced by an individual will of course depend on p_{off} as well as on the duration of the period during which the individual is reproductively active, which on the other hand depends on p_{kill} since p_{kill} will affect the life expectancy of a particular individual. It is important to mention that our main interest is concerned with long time behavior at the level of the entire ecology, e.g., the lifetimes of species (or types) rather than the lifetimes of individuals. Given this perspective, the assumption of a constant p_{kill} common to all individuals, is not expected to be a serious limitation. We simply let the genotype dependence of p_{off} to account for the combined reproductive efficiency caused by an individual's reproduction rate and lifetime expectancy.

A time step consists of *one* annihilation attempt followed by *one* reproduction attempt. One generation consists of $N(t)/p_{kill}$ time steps, which is the average time taken to kill all currently living individuals.

Initially we place $N(0)=500$ individuals at randomly chosen positions. The results are independent of initial conditions. We obtain the same results if all individuals are located at the same position initially.

The present paper's main focus is on the asexual mode of reproduction and results presented are for asexual individuals except otherwise stated.

III. DYNAMICAL STABILITY

Neglecting fluctuations in the occupancy $n(\mathbf{S},t)$, the above dynamics is described by the following set of equations (one equation for each position in the genotype space):

$$n(\mathbf{S},t+1)=n(\mathbf{S},t)+\{p_{off}(\mathbf{S},t)[2(1-p_{mut})^L-1] - p_{kill}\}\frac{n(\mathbf{S},t)}{N(t)}+2p_{mut}(1-p_{mut})^{L-1} \times \sum_{\langle \mathbf{S}',\mathbf{S}\rangle} p_{off}(\mathbf{S}',t)\frac{n(\mathbf{S}',t)}{N(t)}, \quad (3)$$

where the sum is over the nearest neighbors of $\mathbf{S}$. Stationary solutions require the system to find configurations in genotype space for which all positions satisfy the demand that either $n(\mathbf{S},t)=0$ or if $n(\mathbf{S},t)\neq 0$ [neglecting the mutational back flow represented by the last term in Eq. (3)], we must have

$$p_{off}=\frac{p_{kill}}{2(1-p_{mut})^L-1}\equiv p_{q-ESS}. \quad (4)$$

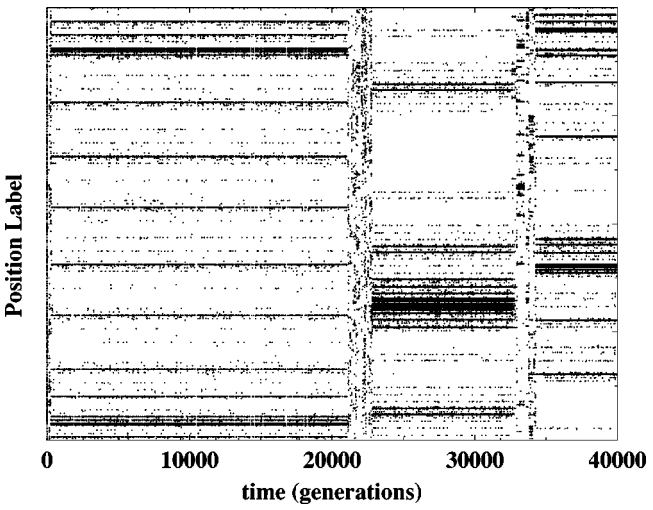

FIG. 2. The occupation in genotype space plotted as a function of generation time. The genotypes are enumerated in an arbitrary manner. If a position is occupied at a given moment in time, a dot is placed at the corresponding number along the *y* axis at that instant in time. Parameters are $c=0.5$, $\mu=0.005$, and $Lp_{mut}=0.25$.

The fitness $p_{off}(\mathbf{S}^a,t)$ of individuals at a position $\mathbf{S}^a$ depends on the occupancy $n(\mathbf{S}^b,t)$ of all the sites $\mathbf{S}^b$ with which site $\mathbf{S}^a$ is connected through couplings J^{ab}. Accordingly, a small perturbation in the occupancy at one position may be able to disturb the balance in Eq. (4) between $p_{off}(\mathbf{S},t)$, p_{kill}, and p_{mut} on connected sites. In this way an imbalance at one site can spread as a chain reaction through the system, possibly causing a global reconfiguration of the occupancy in genotype space.

We show in Fig. 2 the occupancy in genotype space plotted as a function of time for asexual reproduction. Periods of stable configurations are separated by fast transitions. We have called the stable periods as ''quasievolutionary stable strategies'' (qESS) since they are reminiscent of the evolutionary stable strategies (ESS) introduced by Maynard Smith [33]. The transition periods between the qESS are characterized by a rapidly changing occupation in genotype space. We call these periods for *hectic periods* to emphasize the hectic rearrangement of the occupation of positions in genotype space, i.e, $n(\mathbf{S},t)$ changes very rapidly as a function of t during the hectic phases in contrast to the situation during the qESS.

It is interesting to investigate just how stable the qESS are. We have done this by applying different types of perturbations in the qESS. The result is that the qESS are very stable against global perturbations, such as a brief or a lasting increase in control parameters μ,c or p_{kill}. Changes of up to 50% in these parameters, either permanently or for a period of 100 generations, only effect the total population size and is typically not able to kick the population out of its present qESS configuration in genotype space. In contrast, a similar perturbation of the mutation rate easily destabilizes the qESS configuration.

We stress that the segregation (or speciation) to be discussed below is an effect of different couplings between different positions $\mathbf{S}^a$ and $\mathbf{S}^b$. When we assume $J(\mathbf{S}^a,\mathbf{S}^b)=J_0$

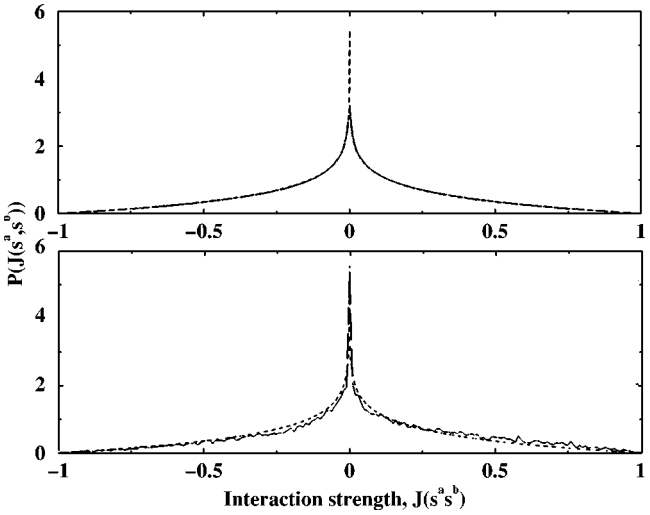

FIG. 3. The distribution from which the values of the couplings $J(\mathbf{S}^a,\mathbf{S}^b)$ are drawn at the start of the simulations (dashed curve) together with the probability density function of the couplings between occupied sites (solid curve) during the hectic periods (top panel) and during the qESS (bottom panel). Parameters are $c=0.01$, $\mu=0.01$, and $Lp_{mut}=0.2$.

independent of $\mathbf{S}^a$ and $\mathbf{S}^b$, the population is not concentrated around a subset of the positions in genotype space, instead the population is smeared out through the space in a diffuse manner. Self-interaction, however, can cause segregation in a rather trivial way. Namely, if we include a distribution of $J(\mathbf{S},\mathbf{S})$ values, segregation may occur even in the case where all interaction terms assume the same value: $J(\mathbf{S}^a,\mathbf{S}^b)=J_0$ for $\mathbf{S}^a\neq\mathbf{S}^b$. However, this type of selection of configurations in genotype space is not very interesting since for the sites to become occupied is determined by the arbitrarily assigned self-interactions $J(\mathbf{S},\mathbf{S})$ and not by the collective dynamical adaptation at play when $J(\mathbf{S},\mathbf{S})=0$ and $J(\mathbf{S}^a,\mathbf{S}^b)$ assumes a distribution of different J values. In reality one will expect the selection of species to be caused by a mixture of self-interaction and interaction between different species. To decide which one is dominant might be difficult and will certainly be system specific.

There is a significant difference between the distribution of active couplings, $p_{act}[J(\mathbf{S}^a,\mathbf{S}^b)]$, in the qESS and the distribution during the hectic transitions. We show in Fig. 3 the distribution from which the $J_{bare}(\mathbf{S}^a,\mathbf{S}^b)$ are sampled together with the distribution of couplings between occupied sites after a large number of generations. During the hectic phases there is clearly no noticeable difference between the ''bare'' distribution of the $J(\mathbf{S}^a,\mathbf{S}^b)$ and the distribution of active couplings, i.e., couplings between occupied positions. During the qESS we observe a slight bias towards positive J values of the active couplings. This slight shift towards more positive couplings will, according to Eqs. (1) and (2), lead to an increased reproduction rate during the qESS. The manifestation of this difference between the hectic periods and the qESS is illustrated in Fig. 4. We see that the distribution of H values in the qESS during both modes of the dynamics contains a narrow peak. In the qESS, the peak in $p(H)$ is separated from a strongly negative band of support. The values of

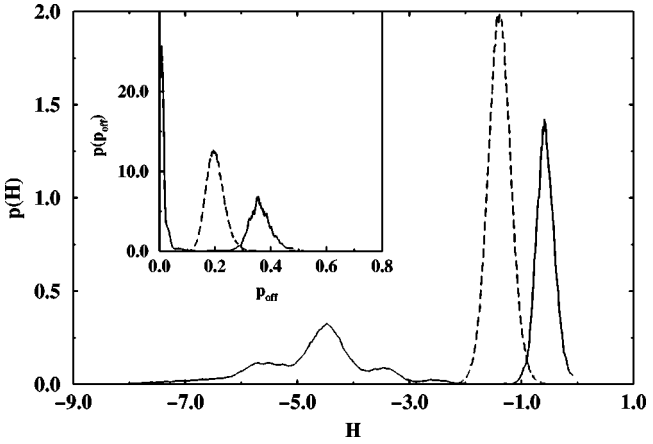

FIG. 4. The probability density function for the weight function H (main frame) and reproduction rates p_{off} (inset) during the hectic transitions (dashed curve) and in the qESS (solid curve). Parameters are $c=0.08$, $\mu=0.005$, and $Lp_{mut}=0.25$.

H in this band are so negative that the corresponding $p_{off}(H)$ are negligible (see the inset in Fig. 4). Genotype positions corresponding to this band consist of unfit positions next to highly occupied and very fit positions. The reason these positions are occupied at all is that they are supplied by mutations occurring on the neighboring fit positions. The conclusion of these considerations is that the dynamics during the qESS as well as during the hectic periods are controlled by the reproduction of individuals with H values in the two respective peaks of $p(H)$.

The location of the peaks of $p(H)$ is determined in the following way. During the hectic periods the occupation of positions in genotype space is highly unstable and $n(\mathbf{S},t+1)$ is only related to $n(\mathbf{S},t)$ in an erratic way, the balance equation (3) is never fulfilled for nonzero $n(\mathbf{S},t)=0$. The only constraint on p_{off} during the hectic periods is accordingly that the total population remains constant on average, which implies that on average $p_{off}=p_{kill}$. This explains why in Fig. 4 the peak in $p(H)$ during the hectic periods corresponds to a peak in $p(p_{off})$ centered at $p_{kill}=0.2$. The situation is different during the qESS. Here the occupation of the selected positions in genotype space remains approximately constant and Eq. (4) applies. Substituting the relevant values $p_{kill}=0.2$, $p_{mut}=0.0125$, and $L=20$ into Eq. (4) produces $p_{off}=0.36$, which explains the position of the peak in $p(p_{off})$ during the qESS.

IV. TIME-DEPENDENT AVERAGES

For simplicity we concentrate again on the asexual model in this section. For large genome length L the system is always in a transient. The time needed to reach the stationary state increases exponentially with L and is therefore unreachable for any biologically relevant values of L.

A. Increasing q-ESS durations

The gradual change in the statistical measures of the model is seen directly as a slow increase with time of the average duration of the qESS. To demonstrate this we show

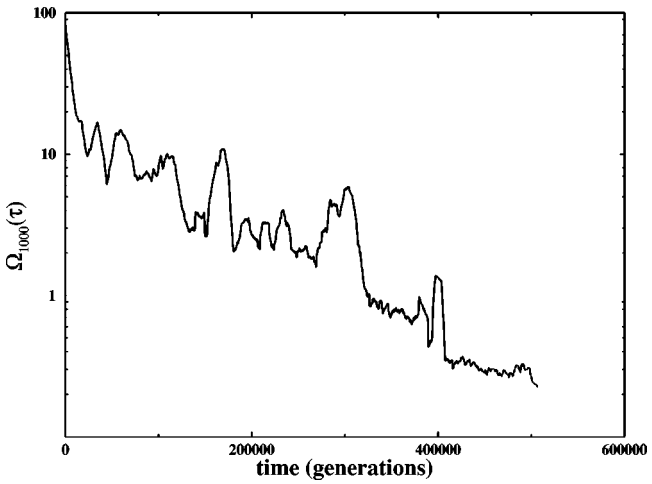

FIG. 5. The average number of transitions during a window of size $T=1000$ generations as a function of generation time. Parameters are $c=0.01$, $\mu=0.01$, and $Lp_{mut}=0.2$. The average is over 400 realizations.

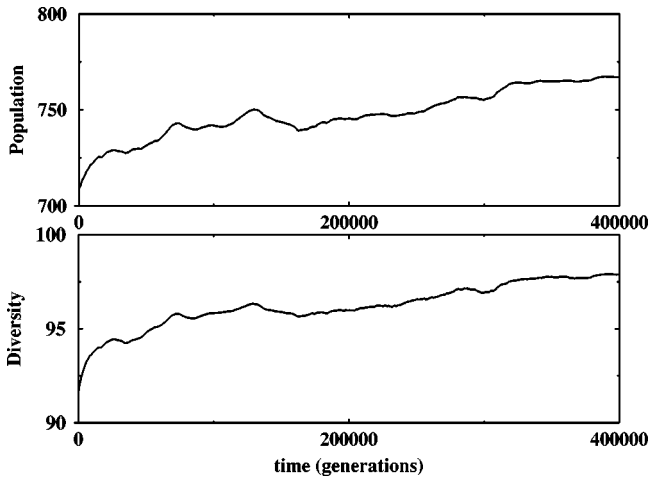

FIG. 6. The ensemble averaged total population (top) and diversity (bottom) as function of generation time for the same ensemble as in Fig. 5.

in Fig. 5 the average number of transitions $\Omega_T(t)$ between qESS within a time window of fixed size T as a function of time t measured in number of generations. We choose T sufficiently big to get reasonble statistics on the number of transitions during T generations and keep T small compared with the total number of generations simulated to be able to fit in a large number of T windows. The result does not depend qualitatively on the choice of T. It is clear that $\Omega_T(t)$ decreases with increasing t, however, it is very difficult to obtain sufficient statistics to be able to determine the functional dependence of $\Omega_T(t)$ on t, though a very slow exponential t dependence is suggested in Fig. 5. Despite these sampling difficulties, it is evident that the duration of the qESS, on average, increases with time. This corresponds to a decrease in the extinction rate, consistent with analysis of the fossil record [1].

B. Increasing population size, diversity, and complexity

The gradual growth of the duration of the stable qESS epochs indicates that the dynamics of the system is able to produce more stable or better adapted configurations in genotype space. It is difficult to test quantitatively the stability of the qESS with respect to perturbations. That the population is distributed in an increasingly more efficient manner in genotype space can be seen directly from the increase in the total population size $N(t)$ averaged over an ensemble of different realizations of the stochastic elements of the dynamics. Figure 6 contains the average total population $\langle N(t)\rangle$ together with the ensemble average of the diversity $\langle D(t)\rangle$, where $D(t)$ is defined as the number of different occupied positions $\mathbf{S}$ in genotype space at time t. The average diversity also increases with time.

Let us briefly consider how the total population size can increase. We saw in Sec. III that essentially p_{off} is narrowly distributed either about p_{kill} or, as in the qESS, about the value p_{q-ESS} in Eq. (4). The increase in $N(t)$ is therefore not an effect of a gradual increase in p_{off}. Simulations indeed confirm that the average offspring probability always remains constant over the entire run. In biological observations and experiments, for reasons of uniqueness, the reproduction rate is identified as *fitness*. In this sense the fitness of the individuals remains, on average, constant in the TaNa model, as presumably is also the case in biological macroevolution; though the microbial experiments by Lenski and Travisano [34] demonstrate that the reproductive fitness can increase as a result of adaptation in microevolution.

The increase in the average population size $\langle N(t)\rangle$ observed in the TaNa model is caused by the system's ability to generate configurations that increase the interaction term in the weight function $H(\mathbf{S},t)$ defined in Eq. (1). When the first term increases, the second term $\mu N(t)$ in Eq. (1) can increase as well, while the total $H(\mathbf{S},t)$ remains, on average, fixed.

The increase in the interaction term of $H(\mathbf{S},t)$ is achieved in several ways. First, the population is spread out onto an increasing number $D(t)$ of different genotypes, as seen in Fig. 6. Moreover, the evolutionary dynamics tends to produce occupied sites that are interacting with an increased number of other occupied sites, i.e., the number of nonzero terms $J(\mathbf{S}^\alpha,\mathbf{S})n(\mathbf{S})$ in $H(\mathbf{S}^\alpha,t)$ in Eq. (1) grows as the system produces configurations that are able to benefit better from the possible mutual interactions represented by $J(\mathbf{S}^\alpha,\mathbf{S})$. The distribution of active interaction links is shown at an early and a much later time in Fig. 7. We have not been able to resolve a shift with time in the distribution of the values of the active interaction strengths $J(\mathbf{S}^a,\mathbf{S}^b)$.

The increase in the diversity and the number of active links connected to an occupied position can be interpreted as an increase in the complexity of the configurations produced by the evolutionary dynamics. Selection and adaptation operate at the level of the entire configuration in genotype space rather than at the level of individual genotypes. This highlights that the biological concept of fitness makes most sense when considered as a *collective property of an ecology*, rather than an observable characteristic of the individual species or individual members of a population.

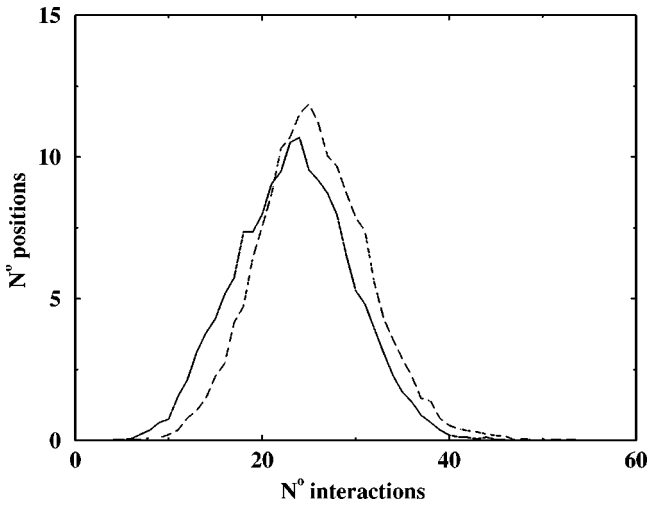

FIG. 7. The number of occupied positions in genotype space with a given number of active links connected to other occupied position in genotype space. The solid line is after 500 generations, the dashed line is recorded after 10^5 generations. Parameters are $c=0.01$, $\mu=0.01$, and $L_{Pmut}=0.02$.

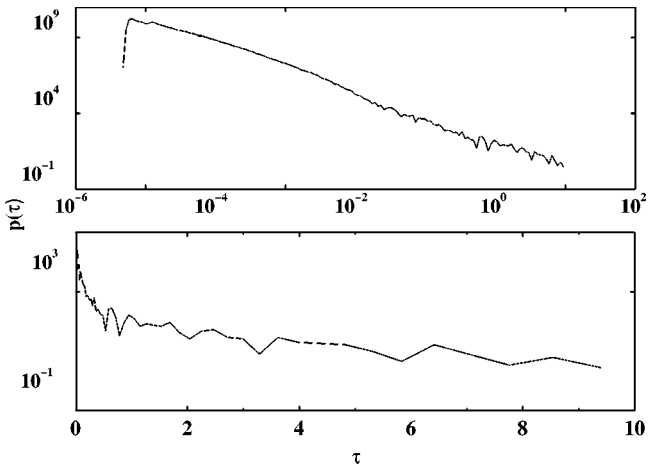

FIG. 8. The probability density $p(\tau)$ of the logarithmic waiting time τ. The top graph is a double logarithmic plot of $p(\tau)$ exhibiting a power law behavior in the region of small τ values. The bottom plot is a linear-log plot of the same data. Here one sees that the behavior of $p(\tau)$ for large values of τ is consistent with a slow exponential decay. Parameters are as for Fig. 7.

C. Record statistics

The observation in the preceding section that the dynamics of the TaNa model leads to an increase in a number of measurable quantities, taken together with the intermittent nature of the dynamics, suggests that the transitions between consecutive qESS epochs correspond to record transitions. One can imagine that some characteristic measures of the collective level of adaptation of the configurations generated in genotype space achieves an ever increasing value as the system undergoes a transition from one qESS to the next.

Sibani and co-workers [27–31] have studied record dynamics and shown that the probability for n records in a sequence of t independently drawn random numbers is Poisson distributed on a logarithmic time scale, or equivalently, that the logarithm of the ratio of the time between the kth and the $(k-1)$th record, $\tau_k=\ln(t_k/t_{k-1})$, is exponentially distributed, $P(\tau>x)=\exp(-\lambda x)$. Sibani and co-workers [27–30] have also demonstrated the relevance of record statistics to the dynamics of the Kauffman NK model [18,19].

Accordingly, it is interesting to investigate if the time dependence of the statistics observed in the TaNa model exhibits signs of record statistics. To do this, we study the distribution of the variable $\tau_k=\ln(t_k/t_{k-1})$, where t_k denotes the time at which the kth transition between consecutive qESS epochs occurs. We show in Fig. 8 that τ_k is exponentially distributed for large values and algebraically distributed for small values of τ_k.

The exponential tail in Fig. 8 suggests that the transition times in the TaNa model follow record statistics in the region of large τ values, corresponding to the regime of long qESS durations. The algebraic form of $p(\tau)$ for small τ values indicates significant correlations for transitions that occurred in rapid succession. The question is which quantity evolves according to record dynamics. We have so far not been able to identify a variable of the system, which jumps monotonously to ever higher values at the transition times. In the evolution models studied by Sibani *et al.* [27–30] the fitness increases through consecutive records. As mentioned above in the TaNa model the reproductive fitness p_{off} remains, on average, constant. The increase in the average duration of the qESS (see Fig. 5) suggests that the stability of the configurations in genotype space gradually increases. To explore this, one should study the temporal behavior of the eigenvalue spectrum of the stability matrix of the effective evolution equations in Eq. (3). We expect that the number of unstable directions, on average, decreases with time, though for a given realization fluctuations probably prevent a strictly monotonous behavior.

V. SPECIES ABUNDANCE FUNCTION

The fundamental quantity to describe an ecology is the species abundance function [35]. The species abundance $W(\rho)$ is the ratio W of species that contains a ratio ρ of the total population. It is well known that the a general precise concept of species is difficult to define, see, e.g., Chap. 15 in Ref. [36] for a list of seven different definitions of species. When one is dealing with asexual individuals, as we are doing here, it is probably best to define species from the point of view of genomic similarity. Hence, we use the term species to denote individual positions in genotype space. Ideally, one would perhaps define species as local regions in genotype space, unfortunately our present maximal system size (about 10^6 individuals and $L=20$) is too small for such a coarse grained definition. Our model remains a qualitative one. Using this definition of species we plot in Fig. 9 the species abundance function for the TaNa model during a qESS. A large number of positions are occupied by a small number of individuals; the occupancy of these positions is never established for extended periods during the qESS. The robust species contain a reasonable number of individuals and are distributed according to the broad peak. The peak can be fitted by a log-normal curve in a way similar to ob-

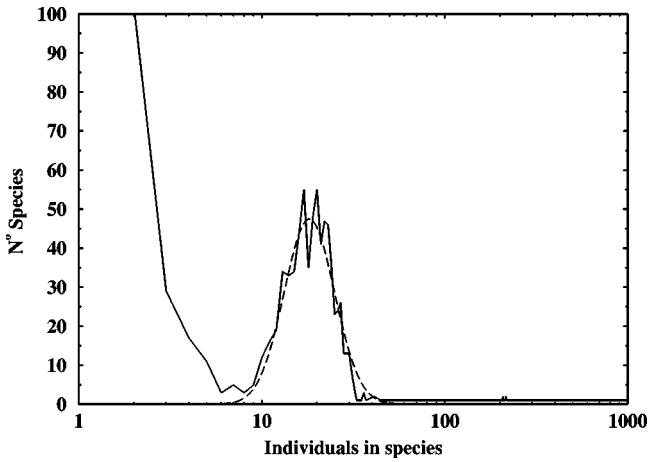

FIG. 9. The species abundance distribution. The peak in the distribution is compared with the log-normal form (dashed curve). Parameters are $c=0.01$, $\mu=5\times10^{-5}$, and $Lp_{mut}=0.2$.

served species abundance functions, see, e.g., Ref. [35]. We note that comparable species abundance functions are found in the predator-prey model studied by McKane, Alonso, and Solé [23].

VI. THE ERROR THRESHOLD

At sufficiently large mutation rates p_{mut}, offsprings are so different from their parents that the occupation in genotype space rapidly moves from one position to the next. When this happens, it becomes impossible to establish the qESS seen in Fig. 2 and consequently the entire simulation consists of one hectic period. The change from the behavior depicted in Fig. 2, where the hectic periods are of much shorter duration than the qESS, to the behavior where the qESS are absent, occurs over a very narrow region of p_{mut} values. Considering first large values of p_{mut} we gradually decrease p_{mut} in the simulations and we identify the threshold value, p_{th} of p_{mut} at which qESS are observed as the error threshold [7,11]. In Fig. 10 we plot the simulated value of p_{th} for different values of the width parameter c.

We can estimate the c dependence of p_{th} by the following argument. From Fig. 4 we know that the distribution of p_{off} in the hectic periods is centered about p_{kill} and in the qESS is centered about p_{q-ESS} defined in Eq. (4). Changing the parameter c will change the width of the distribution of the p_{off} values [see Eqs. (1) and (2)]. It will be possible to establish qESS in between the hectic periods if $p_{kill}+\sigma_p \geqslant p_{q-ESS}$, where σ_p is the half width of the peak in the distribution of p_{off} in the hectic periods. We translate this argument to the distribution of the H values and obtain the following estimate for p_{th}:

$$p_{th}=1-2^{-1/L}[(1-p_{kill})e^{-\alpha/c}+1+p_{kill}]^{1/L}. \qquad (5)$$

Here we have assumed that the width σ_H of the peak in the distribution of H values will be given by $\sigma_H=\alpha/c$ [see Eq. (1)] in which case α is a measure of the standard deviation of the factor in Eq. (1) multiplying $1/c$. We have used $\alpha=0.07$ to fit the simulation data in Fig. 10. This value is somewhat larger but of the right order of magnitude as the corresponding quantity measured during the simulation.

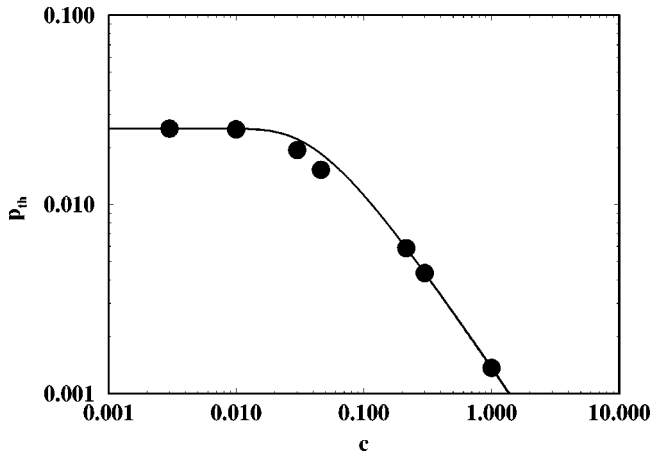

FIG. 10. The loss of q-ESS occurs for mutation rates above the circles. For comparison, the theoretically predicted error threshold $p_{mut}^{th}(c)$ is shown for $\alpha=0.07$ (see main text). The carrying capacity parameter is $\mu=0.005$.

VII. PARAMETER DEPENDENCE

For completeness, we present here the dependence on the parameters c and μ in the weight function H defined in Eq. (1).

We show in Figs. 11 and 12 the averaged occupation measured as the ratio between the average number of individuals and the average number of occupied positions in genotype space for purely asexual and sexually reproducing populations, respectively. As expected, the system is able to support the largest populations in the region of small μ parameter and broad distribution of coupling strength, i.e., small values

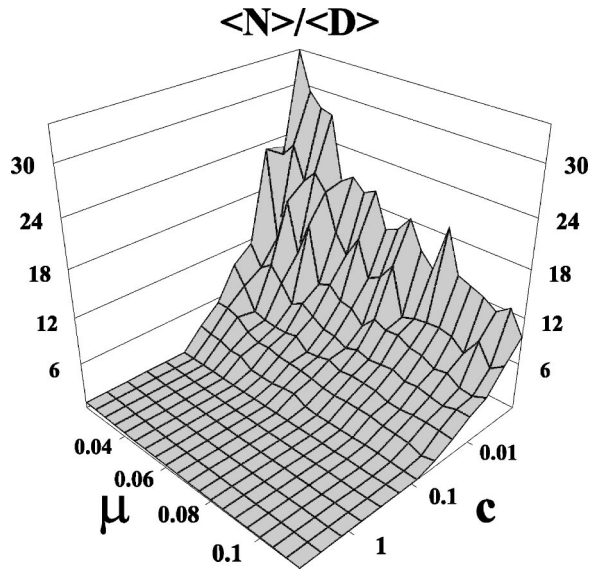

FIG. 11. The ratio between the average size of the population and the average diversity as function of the width parameter c and the physical environment parameter μ. The data are for a system with asexual reproduction with $Lp_{mut}=0.2$.

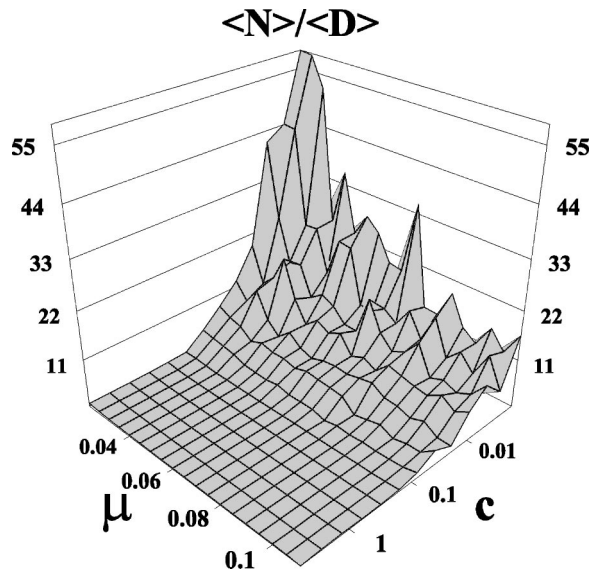

FIG. 12. The same data and parameters as in Fig. 11 except that the data in this figure is for a sexually reproducing population.

of c. The sexual reproduction is most sensitive to a decrease in the carrying capacity (increase in μ) or a decrease in the width of the range of possible $J(\mathbf{S}^a,\mathbf{S}^b)$ couplings (increase in c).

VIII. DISCUSSION AND CONCLUSION

The tangled-nature model may be considered as a mathematical framework for the study of evolutionary ecology. The dynamics of the model is defined at the level of individuals, either as asexual or as sexually reproducing individuals. All ecological structures in the model arise through emergence. The model is able to generate many of the observed features of biological evolution starting from a basic implementation of the key assumptions in the Darwinian evolution paradigm.

The density of individuals in genotype space segregates corresponding to the emergence of distinct species. The interaction between individuals gives rise to a jerky or intermittent macrodynamics, in which quasistable configurations (the qESS) in genotype space are abruptly replaced by new quasistable configurations. This mode of operation can be compared with the intermittent behavior observed in the fossil record and emphasized by Gould and Eldredge as the term ''punctuated equilibrium'' [4]. The TaNa model is always in a transient where the configurations generated as a result of adaptation to the coevolutionary selective pressure gradually produce configurations or ecologies in genotype space, which collectively exhibit a higher degree of adaptation, in the sense that the average lifetime of these qESS increases slowly. This behavior compares well with the observation that the fossil record indicates a decrease in the extinction rate [1]. The increase in the lifetime of the qESS is associated with an increase in the complexity (species diversity and the number of active interactions) of successive configurations. This gradual time dependence together with the intermittent nature of the dynamics suggest that some characteristic of the evolving ecosystem might be undergoing record statistics in the sense of Sibani and co-workers [27–30]. So far we have unfortunately not been able to identify the appropriate effective variable that moves through the records but we expect this variable to be related to the stability matrix of the effective dynamical equation.

The species abundance distribution generated by the TaNa model encourages future studies of larger populations with longer genome sequences. This will enable a hierarchical study of the taxonomic organization of the generated ecologies. Using distance criteria in genotype space one can study the clustering of individuals into species, of species into genera, etc. Future studies will also examine the phylogenetic structures in detail, especially during the radiation of species encountered in the transition periods between qESS. Using longer genome sequences and a more smoothly varying weight function, we expect the TaNa model to be able to illuminate the evolutionary competition between sexual and asexual reproductions.

ACKNOWLEDGMENTS

It is a pleasure to acknowledge very helpful discussions with Paolo Sibani. M.H. and S.A.C. were financially supported by EPSRC and K.C. gratefully acknowledges the financial support of EPSRC through Grant Nos. GR/R44683/01 and GR/L95267/01. We thank Paul Anderson for reading the manuscript.

[1] M.E.J. Newman and P. Sibani, Proc. R. Soc. London, Ser. B **266**, 1593 (1999).
[2] N. Eldredge, Evolution (Lawrence, Kans.) **25**, 156 (1971).
[3] S.J. Gould, Paleobiology **3**, 135 (1977).
[4] S. Gould and N. Eldredge, Paleobiology **3**, 114 (1977).
[5] N. Eldredge and S. Gould, Nature (London) **332**, 211 (1988).
[6] K. Christensen, M. Hall, A. di Collobiano, and H.J. Jensen, (2002), J. Theor. Biol. (to be published), e-print cond-mat/0104116.
[7] B. Drossel, Adv. Phys. **50**, 209 (2001).
[8] A. Burt, Evolution (Lawrence, Kans.) **54**, 337 (2000).
[9] A. Leroi, Evol. Devol. **2**, 67 (2000).
[10] P.G. Higgs and B. Derrida, J. Mol. Evol. **35**, 454 (1992).
[11] M. Eigen, J. McCaskill, and P. Schuster, J. Phys. Chem. **92**, 6881 (1988).
[12] E.E. Hoster, Evolution **47**, 1637 (1993).
[13] S. Gavrilets, Am. Nat. **154**, 1 (1999).
[14] S. Gavrilets, H. Li, and M.D. Vose, Proc. R. Soc. London, Ser. B **265**, 1483 (1998).
[15] S. Gavrilets, H. Li, and M.D. Vose, Evolution (Lawrence, Kans.) **54**, 1126 (2000).
[16] S. Wright, Proc. 6. Int. Congress of Genetics **1**, 356 (1932).
[17] S. Wright, Am. Nat. **131**, 115 (1988).
[18] S.A. Kauffman and S. Levine, J. Theor. Biol. **128**, 11 (1987).
[19] S. Kauffman, *The Origins of Order* (Oxford University Press, Oxford, 1993).

[20] C.F. Taylor and P.G. Higgs, J. Theor. Biol. **203**, 419 (2000).
[21] K. Kaneko and T. Yomo, Proc. R. Soc. London, Ser. B **267**, 2367 (2000).
[22] P. Bak and K. Sneppen, Phys. Rev. Lett. **71**, 4083 (1993).
[23] A. McKane, D. Alonso, and R.V. Solé, Phys. Rev. E **62**, 8466 (2000).
[24] B. Drossel, P.G. Higgs, and A.J. McKane, J. Theor. Biol. **208**, 91 (2001).
[25] R.M. May, Nature (London) **238**, 413 (1972).
[26] R.M. May and R.M. Anderson, Proc. R. Soc. London, Ser. B **219**, 281 (1983).
[27] P. Sibani, M.R. Schmidt, and P. Alstrøm, Phys. Rev. Lett. **75**, 2055 (1995).
[28] P. Sibani, Phys. Rev. Lett. **79**, 1413 (1997).
[29] P. Sibani and M. Brandt, Int. J. Mod. Phys. B **12**, 361 (1998).
[30] P. Sibani and A. Pedersen, Europhys. Lett. **48**, 346 (1999).
[31] P. Sibani and P. Littlewood, Phys. Rev. Lett. **71**, 1482 (1993).
[32] H. Wagner, E. Baake, and T. Gerische, J. Stat. Phys. **92**, 1017 (1998).
[33] J. Maynard Smith, *Evolution and the Theory of Games* (Cambridge University Press, Cambridge, England, 1982).
[34] R.E. Lenski and M. Travisano, Proc. Natl. Acad. Sci. U.S.A. **91**, 6808 (1994).
[35] S.P. Hubbell, *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton University Press, Princeton, 2001).
[36] D.J. Futuyma, *Evolutionary Biology* (Sinauer Associates, Inc., Sunderland, 1998).