

Evolutionary dynamics and highly optimized tolerance

Tong Zhou^a, J.M. Carlson^{a,*}, John Doyle^b

^aDepartment of Physics, University of California, Santa Barbara, CA 93106, USA

^bControl and Dynamical Systems, California Institute of Technology, Pasadena, CA 91125, USA

Received 30 June 2004; received in revised form 31 January 2005; accepted 18 March 2005

Available online 17 May 2005

Abstract

We develop a numerical model of a lattice community based on Highly Optimized Tolerance (HOT), which relates the evolution of complexity to robustness tradeoffs in an uncertain environment. With the model, we explore scenarios for evolution and extinction which are abstractions of processes which are commonly discussed in biological and ecological case studies. These include the effects of different habitats on the phenotypic traits of the organisms, the effects of different mutation rates on adaptation, fitness, and diversity, and competition between generalists and specialists. The model exhibits a wide variety of microevolutionary and macroevolutionary phenomena which can arise in organisms which are subject to random mutation, and selection based on fitness evaluated in a specific environment. Generalists arise in uniform habitats, where different disturbances occur with equal frequency, while specialists arise when the relative frequency of different disturbances is skewed. Fast mutators are seen to play a primary role in adaptation, while slow mutators preserve well-adapted configurations. When uniform and skewed habitats are coupled through migration of the organisms, we observe a primitive form of punctuated equilibrium. Rare events in the skewed habitat lead to extinction of the specialists, whereupon generalists invade from the uniform habitat, adapt to their new surroundings, ultimately leading their progeny to become vulnerable to extinction in a subsequent rare disturbance.

© 2005 Elsevier Ltd. All rights reserved.

Keywords: Complexity; Evolution; Robustness

1. Introduction

Biology and ecology are rich with examples illustrating the importance of environmental factors in shaping the traits of organisms (Bonner, 1988; Levin, 1999; Maynard Smith, 1989; Newman and Palmer, 2003; Dykhuizen and Davies, 1980; Thompson, 1994). Relatively simple organisms which reproduce rapidly, such as *E. coli*, have also provided an ideal setting to study the effects of variable mutation rates and changing environments in a controlled laboratory setting (Drake, 1991; Horst et al., 1999; Sniegowski et al., 1997, 2000; Taddei et al., 1997). The goal of the work presented here is to develop models to explore the kinds of fundamental processes which are frequently discussed in ecology and

evolutionary biology, but are difficult to abstract and generalize in that context because of the overwhelming complexity of case specific details. We presented a variant of the model in this paper previously (Zhou et al., 2002), along with initial studies of its basic micro- and macro-evolutionary characteristics, and a detailed discussion of the relevance of the lattice barrier model to biological robustness and evolution. This paper is devoted to a more systematic study of the model with more realistic assumptions about the “organisms” and their habitat. We focus on phenotypic convergence and the impact of the environment, the relative extinction vulnerability of generalists vs. specialists, and the role of mutation rate on adaptation and stability.

Our numerical model is motivated by Highly Optimized Tolerance (HOT) (Carlson and Doyle, 1999, 2000, 2002; Doyle and Carlson, 2000; Newman, 2000; Reynolds et al., 2002; Robert et al., 2001; Zhou and

*Corresponding author. Tel.: +1 805 893 5369.

E-mail address: carlson@physics.ucsb.edu (J.M. Carlson).

Carlson, 2000; Zhou et al., 2002), a mechanism for complexity which is based on robustness. HOT emphasizes the efficiency and robustness tradeoffs that dominate engineering design and, we claim, biological evolution, but when possible these tradeoffs are illustrated using the models and methods of statistical physics (Stauffer, 1985). The aim is to both make engineering concepts accessible in greatly simplified models, and sharpen the distinctions between the very different perspectives of engineering and physics. While the importance of robustness and resilience in biology and ecology has long been recognized (Holling, 1973; Savageau, 1976), it is receiving renewed attention. Studies include robustness *tradeoffs* as a mechanism that drives complexity in biology (Lauffenburger, 2000; Yi et al., 2000; Csete and Doyle, 2002, 2004; Stelling et al., 2004; El-Samad et al., 2005), and the relationship of robustness with evolvability (Kirschner and Gerhart, 1998; Caporale, 2003).

We have aimed to develop the simplest possible model that illustrates the link between robustness and evolved complexity, and allows us to explore the consequences of a variety of ecological scenarios. While the lattices represent a severe abstraction of biological organisms, this model has several advantages over both more and less abstract representations. Many models which aim to address fundamental issues associated with evolution and the underlying logic associated with Darwinian mechanisms focus on dynamical processes involving evolution on abstract fitness landscapes. Recent reviews of this work appear in Gavrilets (2004) and Weiss and Buchanan (2004). Other theoretical studies aimed at explaining extinction patterns and statistics in the fossil record (Raup, 1986) have focused on exclusively macroscopic definitions of fitness. In these models species lack genotypic and phenotypic notions of structure, heredity, and expression, and instead are represented in terms of scalar fitness variables obeying dynamical rules, whereby the system self-organizes to a critical point (SOC) or the “edge of chaos” (Bak and Sneppen, 1993; Drossel, 2001; Kauffman, 2000; Newman and Palmer, 2003; Sneppen et al., 1995; Sole et al., 1997). Such models lack the explicit connection between macroscopic natural selection based on fitness and microscopic random mutation of the genotype (configuration), and cannot evolve traits that are not put in by hand. In contrast, our model retains concrete and distinct, if simplified, notions of genotype, phenotype and fitness, along with non-trivial environmental coupling which ultimately influences the phenotypic traits of the organisms. We find that behavior of the model is surprisingly rich, with numerous meaningful analogies to observed phenomena in biology and ecology.

A key abstraction which emerges in this and previous studies of HOT is the development of efficient barriers which limit the damage associated with cascading failure

events. We discussed the barrier abstraction, which is much more versatile than it may appear, in the context of specific biological examples in Zhou et al. (2002). Barriers can be viewed somewhat literally as physical barriers such as armor and membranes, but also more abstractly as the result of regulatory mechanisms that prevent cascading failures in the state space dynamics of a system. In this paper, we have modified the model so that two key abstractions—niches and mutation rates—are represented more realistically. First, we replace the variable definition of fitness, with a variable rate of mutation. Now our lattices can be fast or slow mutators, and we investigate the consequences which arise when the mutation rate itself is subject to mutation. Second, we replace the designated niches with distinct habitats which are characterized by different disturbance patterns. We investigate the consequences which arise when organisms can migrate between these new habitats. We also develop new techniques for analyzing evolutionary patterns in the model, specifically the family tree.

The rest of this paper is dedicated to first defining the model and then investigating the effects of coupling two habitats, one in which the disturbances are uniform and the other skewed. The variable mutation rates add to the richness of the dynamics. Slow mutators dominate during periods of stasis, but fast mutators accelerate adaptation after extinction events in the skewed habitat.

2. Population dynamics and evolving lattices

Our model consists of two habitats, each of which contains a population of organisms. There is an upper bound of 1000 on the total population S in each habitat, and we begin our simulations with 1000 randomly generated lattices per habitat. Competition between organisms occurs within each habitat and is associated with competition for space in communities of bounded size. Each individual organism is represented by a 16×16 square percolation lattice. Each site on each lattice is either occupied or vacant, and the binary list of occupied and unoccupied sites is a lattice’s “genotype.” Contiguous nearest neighbor occupied sites define connected clusters, with the clusters and their locations defining a lattice’s “phenotype.” The connection between genotype and phenotype here is transparent, but not entirely trivial, in that the feature of a lattice that will be crucial for fitness is the location and size of connected clusters.

The effect of environmental perturbations is modeled by “sparks”, chosen from a probability distribution $P(i, j)$. This distribution is the defining characteristic of the habitat. A spark impacts each lattice in a habitat at the same site (i, j) . If on a given lattice (i, j) is vacant then nothing happens. However, if (i, j) is occupied the entire connected cluster of occupied sites containing the

original sparked (i, j) site “burns” and is thus lost. In the forest fire models that use lattices (Drossel, 2001; Carlson and Doyle, 2002) these sparks were interpreted as initiating a fire, which burned a connected cluster until stopped by a barrier of vacant sites. Here the interpretation is even more abstract, and is intended to model an external perturbation from the environment on a lattice, which here is interpreted as an individual organism. The most literal interpretation is that sites then correspond to distinct internal components of the organism that are lost in a cascading failure throughout a connected cluster following a perturbation (spark) in the cluster. The failure is only stopped by barriers, and the total damage an organism suffers is the size of the loss of a connected cluster. Nevertheless, we will continue to use the term “spark” and “burn” to describe the initiation and propagation of damage in the lattice. There are many less literal interpretations of barriers and failure that connect directly to cellular biochemistry, and occur in the state space of biological networks, cells, and organisms. Zhou et al. (2002) discusses additional interpretations of this model.

The time line for evolution is defined in terms of “generations,” and the evolutionary dynamics is defined by the following sequence at each generation: (1) each parent lattices “divides” by being copied into two offspring, with some probability of mutation, determined by their per site mutation rate, (2) with some small probability the remaining offspring may migrate to the other habitat and compete there, (3) all offspring lattices in the same habitat are subjected to the same spark, drawn from the $P(i, j)$ for the habitat, which determines fitness Y , (4) depending on Y , some offspring may “die” immediately and are discarded, (5) all surviving offspring, including invaders, then compete for space also based on fitness Y to become parents for the next generation with the rest being discarded. The process then repeats. Note that “parents” survive to the next generation if and only if they reproduce and divide, so this is asexual reproduction by simple division. In the absence of habitat limits and deaths, the population would double each generation. To complete the model, we must define Y , the mechanism for immediate death, the mechanism for migration, and the mechanism of competition for space. We will also allow for mutation of the mutation rate.

Mutation corresponds to inverting the occupation of a site upon division, from vacant to occupied or from occupied to vacant. A site has a certain probability to mutate if itself or one of its neighbor is void. In most of our simulations we have chosen one of two different values for this probability. An organism can have mutation probability set to 0.01, (fast mutators), or to 0.0001, (slow mutators). The absolute values of mutation rates here are not intended to model any specific biochemical rates, but merely to create two types of

organisms with very different rates. We also allow the mutation rate itself to mutate between the characteristic fast and slow values. The mechanism we use is to include an extra “bit” or gene in addition to our lattice, which indicates the mutation rate. The parent lattice’s rate is passed on to any offspring, subject to mutation from the fast to the slow rate or vice versa, with a small probability of 0.005. This loosely represents the kind of variable mutation rates which have been studied extensively in bacteria populations (Drake, 1991; Horst et al., 1999; Sniegowski et al., 1997, 2000; Taddei et al., 1997).

The fitness is the yield $Y = \rho - l$ of a lattice, namely the remaining density computed after one spark (drawn from $P(i, j)$) impacts the lattice. Here ρ is the initial density, and l is the loss due to the spark. The average fitness could also be defined in a habitat as $\langle Y \rangle$ averaged over the given $P(i, j)$. The actual expressed fitness depends on the particular spark history experienced by a lineage and thus has a meaning on every time scale, but we will use fitness to mean the particular Y based on one spark in a given generation. Thus computing yield and fitness is a stochastic process: a single spark impacts a single lattice, which mimics fitness evaluated on a noisy landscape with a short time horizon, typically sampling the most common events according to the $P(i, j)$. This definition of yield creates a tradeoff.

A spark may immediately kill a lattice organism if Y is too small. We choose this death probability to be $\max(0, 1 - 2Y^2)$. The motivation for this choice is that we want $Y = 0$ to give certain death, sufficiently large Y to give no death, with the probability of death decreasing with increasing Y . Note that surviving the spark with probability $\max(0, 1 - 2Y^2)$ only guarantees that the offspring then compete for space, not that they are guaranteed to have offspring themselves in the next round. We also want it to be possible for random lattices to survive instant death. Since the percolation density for these lattices is approximately .59, small losses are suffered in random lattices below this threshold (Stauffer, 1985). Thus random lattices of $\rho \approx 1/2$ have a death probability of roughly $1/2$. This will on average allow a large population of random lattices to survive at a constant size, so lattices can evolve from completely unstructured random lattices, and no organizational mechanisms are built into the model. Only competition between organisms would drive evolution away from random. The exact form of $\max(0, 1 - 2Y^2)$ is otherwise chosen for simplicity, and provided the general features are preserved, all results are qualitatively similar.

If a lattice survives the initial spark, it may migrate from one habitat to another with small probability 0.005, creating a random diffusion between habitats. Finally, all the remaining offspring, including invaders, compete for one of the total of 1000 spaces in its habitat

based solely on its Y ranking. If there are less than 1000 candidates, all become parents and divide and mutate. If there are more than 1000, the 1000 lattices with the highest Y values become parents, and the rest are discarded.

A simpler version of this model was first introduced in Zhou et al. (2002), where we stressed the importance of the time-scale of evaluating fitness (i.e. comparisons between Y and $\langle Y \rangle$). In addition to a more comprehensive investigation of all phases of the evolutionary process, including effects of variability in the habitat and migration, the version of the model we define here includes several significant technical refinements over our previous study. The major changes are as follows. In the original model, mutation can occur at any site on the lattice and is totally random, in accordance with the simplest notion of random mutation in biology. However, after the initial transient, we find this is extremely inefficient in advancing the structural design during the later stages in the evolution. In our model, too many offspring create a computational burden when computing the connected clusters. Therefore, in this updated version of the model, mutation can only occur at a void site or one of its neighbors. Since highly evolved lattices have high densities, and consist of compact regions of unit density, separated by linear barriers, creating isolated voids in the middle of large connected clusters was the most probable mutation in our previous model. However, this process simply produces noise, with no significant evolutionary change. All the mutations in the new model are concentrated in moving the barrier patterns in a lattice. Comparing the two algorithms for mutation, this modification of focused mutation significantly increases the numerical speed of our investigations, without introducing significant qualitative changes in the patterns which emerge. Our main motivation for focused mutation is computational overhead, so this is clearly an “engineering” feature of our model. However, there is growing evidence that real organisms have a seemingly endless variety of mechanisms that appear to focus mutations in ways that are very far from random, and appear to significantly skew mutations both to be much larger and much more beneficial than random point mutations. Thus our focused mutation is primarily for convenience, but is not without biological analogues (Kirchner and Gerhart, 1998; Caporale, 2003)

Another important change is associated with our treatment of variable mutation rates. In our original model, the effects of variable mutation rates were represented heuristically by including organisms with different definitions of fitness and a scheme based on effectively different time scales for fitness evaluation and mutation. However, in the more realistic treatment included here, we fix a single time axis for reproduction, disturbances, and competition, and obtain different

macroscopic adaptation rates for different organisms as a result of different microscopic mutation rates.

Finally, here we also introduce a different treatment of the death rate and the niches. These two issues are coupled because they both relate to the probability of extinction, which terminates the simulation. In our previous model, an organism died when its fitness fell below a fixed value, taken to be 0.4. As a result a single rare event could kill a highly evolved community. We introduced protective niches of size 50 for each species, which preserved organisms regardless of the death criterion, enabling the system to continue evolving forever by preventing extinction by hand. In this version the protective niches are removed, and we allow extinction of lineages. However, the likelihood of extinction depends on properties of the habitat, which have different environmental characteristics. Lineages which adapt to common events in environments with skewed $P(i,j)$ are most at risk of extinction, whereas lineages adapted to uniform $P(i,j)$ rarely go extinct. Since death is defined in terms of a probability, given by $\max(0, 1 - 2Y^2)$, it is very unlikely that a single rare event will kill every organism even in the skewed habitat. Rather, complete extinction requires multiple rare events. The two habitats allow organisms to invade one habitat from the other, which sustains non-zero populations of the system.

3. Fast and slow mutators compete in coupled habitats

We begin by describing the most complex scenario with the richest features, with fast and slow mutators in populations with a migration between uniform and skewed habitats. We have also thoroughly explored all the simpler scenarios that lead naturally to this picture. The results are consistent, so we will focus on the richest scenario, but additional details are summarized in Zhou et al. (2005). There are two habitats, each specified entirely by $P(i,j)$, one uniform and one skewed. A uniform habitat is characterized by a constant $P(i,j) = P$. This represents a homogeneous environmental influence, where each disturbance has the same frequency. For a skewed habitat $P(i,j)$ is non-uniform, and instead exhibits some regions where sparks are common and others where sparks are rare. Note that there is no a priori variation in the “intensity” of the sparks, and the size of the resulting damage is due entirely to the phenotype, the connected clusters, of the organism. Even this minimal setting of disturbances which occur with different frequencies has been shown to lead to large fluctuations and heavy-tailed distributions of losses for both deliberately designed and highly evolved lattices. For our numerical results we take $P(i,j) \sim \exp(-(i+j)/\xi)$, where ξ is 1/6 of the lattice size. Thus the most probable disturbances occur in the upper left hand

corner of the lattice, and the likelihood falls off exponentially with distance departing from that point. Our results, including sensitivity to changes in the $P(i,j)$, are qualitatively independent of the specific form of $P(i,j)$, aside from the fact that it is skewed.

After an initial transient period, the system reaches a roughly “steady state” which has large fluctuations in populations but in which the organisms have highly organized and characteristic configurations. A sample snapshot of a set of configurations is illustrated in Fig. 1, where for each figure the gray scale represents an average over species (fast and slow mutators) and habitat. The organisms in the uniform habitat have neat cross barriers separating the lattice into four nearly equal-sized regions. The exact shape and number of regions can vary between runs, but are all qualitatively similar. A configuration like this is so well adapted to the uniform habitat that it is very unlikely for the population to ever become extinct. The patterns associated with the competing fast and slow mutators are similar (the fast mutator patterns are slightly noisier, i.e. the population is more diverse), which reflects the fact that the mutation rate may mutate within a lineage. As we will see, the commonality of the patterns is due to a common single, ancient ancestor.

In general, as expected, the slow mutators are very stable, and preserve high fitness configurations, while

the fast mutators adapt rapidly, and are much more effective in producing rapid evolutionary change. Each of these strengths has a special role to play in coupled habitats. In particular, after the system has evolved past the initial transient phase, the slow mutators dominate the uniform habitat, because they preserve a successful body plan in an environment where extinction is not a risk. However, after an extinction in the skewed habitat, the slow mutating invaders from the uniform habitat, will sometimes transform to fast mutators, in order to adapt more rapidly to the new surroundings (while there is an advantage to rapid adaptation, fast mutation can also lead to undesirable mutations, and thus in some cases slow mutators will adapt more successfully). If fast mutators adapt most successfully, there is now an evolutionary advantage for their lineage to transform again to slow mutators, in order to preserve the adapted phenotype. However, in the skewed habitat the lattices are always racing against the clock, because of the risk of extinction which sets in once they successfully adapt. A large extinction event typically leads to another invasion. We illustrate a representative time evolution of this kind of pattern in more detail below.

The organisms in the skewed habitat have patterns which are very similar to those in the uniform habitat. This is due to a common ancestor, and repeated cycles of extinction in the skewed habitat, followed by invasion from the uniform habitat. The differences in the skewed habitat is that invasion is followed by the adaptive loss of barriers in the lower right corner, which gives higher yield for the more common sparks. The average configurations in Fig. 1 illustrate that in the skewed habitat, the fast mutators have more variable patterns, with noisier boundaries compared to the slow mutators.

HOT arises whenever fitness is maximized using some optimization strategy. In this case, the optimization is performed by mutation and selection and leads to non-generic, highly structured states with yields well in excess of the corresponding randomly generated configurations. The price paid for the high yields can be that the resulting HOT configurations can be extremely fragile to rare events as defined by a given $P(i,j)$, or to changes in the $P(i,j)$ relative to what was assumed in the design or was part of the evolutionary history. An example of the former is that lineages eventually go extinct in the skewed habitat when hit by several rare sparks in near succession in time. An example of the latter is that lattices adapted to the skewed habitat never successfully invade the uniform habitat after the initial transient.

Fig. 2 illustrates the time evolution of the population sizes in the two habitats. The graph is complicated, but it illustrates the typical large-scale trends. We will zoom in on important features below. The uniform habitat is represented by the green (slow mutators) and yellow (fast mutators) curves. The total population size is 1000, and at later stages the habitat is clearly dominated by

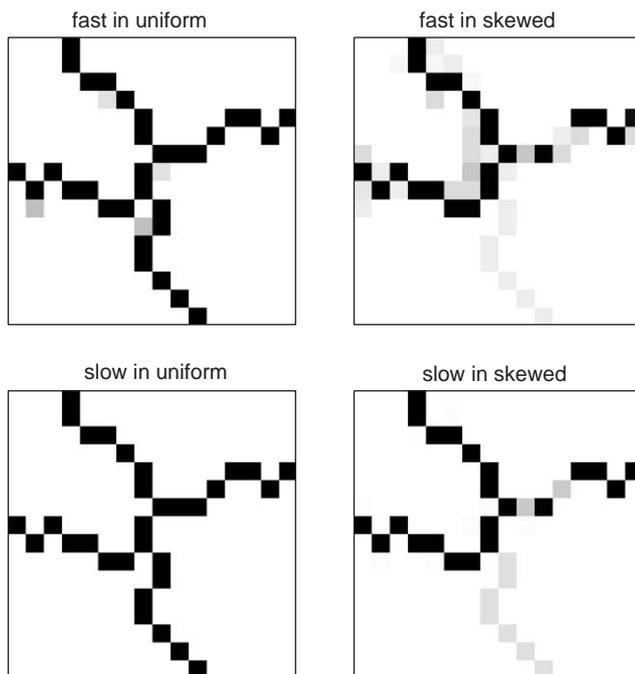


Fig. 1. Sample snapshots of the average configurations of fast and slow mutators in the uniform and skewed habitat in the steady state. Black indicates void sites and white indicates occupied. The gray scale ranges from white (occupied for every individual), to black (void for every individual) for the community average and indicates the average occupation of the site.

the slow mutators (green), which preserve the high fitness configuration (Fig. 1). During the initial transient there is a significant population of fast mutators, but

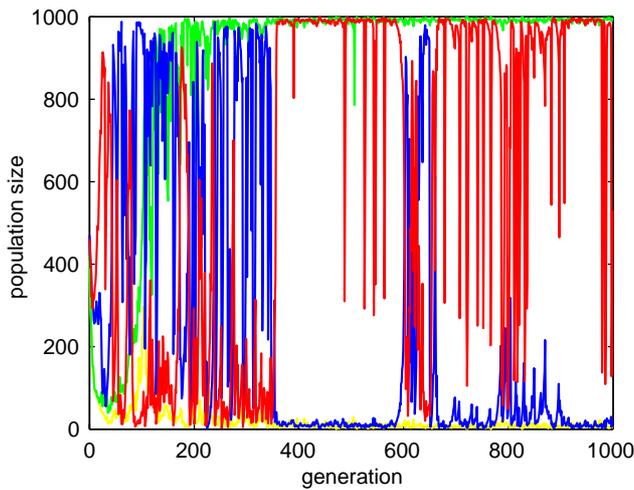


Fig. 2. Population dynamics for coupled habitats, containing fast and slow mutators. The green line indicates slow mutators in the uniform habitat. The yellow line indicates fast mutators in the uniform habitat. The red line indicates slow mutators in the skewed habitat. The blue line indicates fast mutators in the uniform habitat.

after generation 200 or so, very few remain. The skewed habitat is much more complicated. There is obvious competition between the slow mutators (red line) and the fast mutators (blue line). It is still generally true that slow mutators are better at preserving configurations while fast mutators are better at exploring change. In many cases, when there is need for change, like just after the moment of invasion of uniform habitat organisms into the skewed habitat, fast mutators excel first in losing extra barriers to adapt to the new environment. But after finding better performing patterns, they become slow mutators, which take over from their fast siblings since they are better at preserving the adapted configuration. The process repeats when rare events lead again to extinction.

To explore the mechanisms responsible for these fluctuations in more detail, four different episodes in the evolutionary dynamics, which occur within two qualitatively similar time windows, corresponding to the left and right columns in Fig. 3. The top row is an expanded graph of the population sizes from Fig. 2, using the same color code. In both cases, the slow mutators (green) dominate the uniform habitat, while the situation is more complicated in the skewed habitat. The remaining

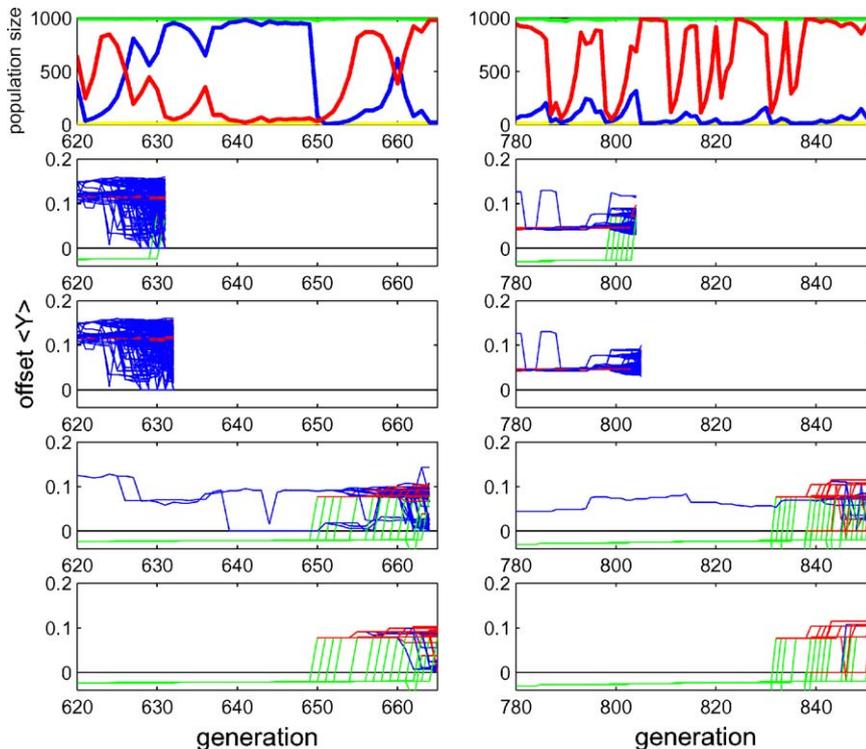


Fig. 3. The family tree for two time windows of the simulation shown in Fig. 2. The color coding is same as that in Fig. 2. The left and right column correspond to two different time windows, both of which illustrate, first, the pattern of failed invasion (second and third rows), and, second, extinction and successful invasion (fourth and fifth rows). The top row is a blown up image of the corresponding time window from Fig. 2. The remaining rows are trees tracing backwards from current generation lattices of the skewed habitat only, from starting points, which differ by one generation for the second and third rows, and, similarly, the fourth and fifth rows. The top (red/slow and blue/fast) portion illustrates the skewed habitat, where the value of $\langle Y \rangle$ is displaced by 0.6 on the vertical axis, and the bottom (green/slow) illustrates the uniform habitat, where the value of $\langle Y \rangle$ is displaced by 0.7 on the vertical axis. In the third row, we see no lines tracing back to parent lattices in the uniform habitat. In the bottom row, we see that all members of the current generation in the skewed habitat trace back to parent lattices in the uniform habitat.

rows are family trees, which are useful in understanding the detailed mechanisms of mutation and invasion responsible for the large fluctuations in Figs. 2 and 3. The second and third rows of Fig. 3 illustrate two examples of failed invasions, one around generation 630, and the other around generation 805. In contrast, the fourth and fifth rows illustrate examples of successful invasions from the uniform habitat to the skewed habitat.

4. Analyzing family trees

Tree structures are familiar tools for analysis in evolutionary biology. Most commonly, phylogenetic trees begin with a common ancestor and depict the emergence of new species as a forward branching structure. In these cases it is only the density and topology of the branching structure which is significant, and all known species are included. A complete phylogenetic tree recording all progeny from some initial starting point can be constructed for the model, but is relatively uninformative. Such a tree would have loops, but not because of cross breeding or symbiosis, but rather because mutations from different parent lattices could produce identical offspring configurations, given our limited genetic code. Fast mutators branch more rapidly than slow mutators, but in either case the overall structure is sufficiently dense and complicated that it does not help us visualize the underlying evolutionary patterns and processes. Instead, we focus on a subset of the full tree, and add meaning to the vertical and horizontal axes on which the tree structure is graphed. We refer to our construction as the “family tree.”

We begin with a designated current generation, and trace backwards through the lineage, arriving eventually at a common ancestor. The horizontal axis of our tree represents the generation, preserving a record of the time line for evolution in the family tree. Since each individual has one parent, this can be done unambiguously. On the other hand, while each parent has two offspring, not all offspring are included in the family tree, only those whose lineage survives until the current generation. As a result an undivided branch of our tree does not signify the absence of mutation, but rather indicates that tracing back from the current generation, there is an interval over which a parent who has progeny in the current generation produces only one offspring who has progeny in the current generation, who also produces one offspring who has progeny in the current generation, and so on. A branch splits in the family only when two of a common parent's offspring both have progeny in the current generation.

We also include a long-range measure of fitness on the vertical axis of our family tree, which is the average yield $\langle Y \rangle$ based on the complete $P(i,j)$. That is, $\langle Y \rangle = \rho - \langle l \rangle$

where ρ is the initial density, and $\langle l \rangle$ is the average loss computed with respect to the full distribution of sparks. We could have plotted the instantaneous fitness, on which the death and selection criterion of a given generation are based. However, this measure is noisy due to the stochastic selection of a single spark from the distribution. As a result, an offspring lattice which is completely identical to the parent typically has different instantaneous fitness. While $\langle Y \rangle$ does not uniquely identify the lattice, identical lattices have identical values of $\langle Y \rangle$, which facilitates our ability to identify patterns and branches in the family tree. For example, sharp drops in $\langle Y \rangle$ are indications that barriers in rare spark region have been lost, perhaps enhancing fitness in the short run but risking large losses to rare sparks.

For the dynamics in Fig. 3 we will focus initially on the bottom row, which illustrate two similarly successful invasions from the uniform habitat to the skewed habitat. In each case we choose a particular moment, corresponding to generation 665 and 852 and trace back the parentage of the lattices which exist at that time, which in this case involves invasions starting around 650 and 830. As seen in the top row, between generation 620 and 665, for the most part fast mutators (blue line) dominate the skewed habitat. Around generation 650 this population in the skewed habitat drops sharply, due to a rare spark. The fourth row shows family trees one generation before that in the bottom row, and we can see that the lineage that survived the near extinction at 650 has high average yield, meaning it preserved barriers in the lower right corner. Because of this the fast/skewed population recovers temporarily, but adapts further and is killed entirely by a series of rare sparks by 665. In the meantime, invaders from the uniform habitat begin taking over the skewed habitat, which corresponds in the bottom row to all family lines from the current generation in the skewed habitat tracing back to parent lattices from the uniform habitat.

A similar episode occurs in the time window between generation 780 and 852. In this period, the skewed habitat is primarily dominated by slow mutators (red line). However, the relatively frequent large drops in their population suggest that many lineages have lost all barriers in the rare spark region, and have just two connected clusters, as seen in Fig. 1. Around generation 830 a near extinction occurs and invaders from the uniform habitat begin taking over the skewed habitat, similar to the situation near generation 650 above. The fourth row again shows the family tree one generation before the “extinction” event where we see the coexistence of recent invaders and descendants of an ancestor with much longer persistence in the skewed habitat, and who had a relatively high $\langle Y \rangle$. This cycle of extinction and invasion is repeated throughout the record.

Not all invasions are successful, and must generally be facilitated by extinction events. In both cases shown in

the second and third rows, family trees (second row) contain lineages which trace back to parents from the uniform habitat, which are extinguished one generation later (the third row). In both cases, the total population size (top row, red plus blue total) of the skewed habitat remains relatively stable, at or near the maximum of 1000 (top row) during the invasion. The invaders are not killed immediately by sparks, but are out-competed by lineages who are already adapted to the skewed habitat.

The long-term pattern of adaptation and barrier loss which occurs when invading lattices from the uniform habitat take over the skewed habitat is best illustrated by extending the family tree even further back in time. In Fig. 4 we illustrate a family tree which spans much of the period in the two time windows illustrated in Fig. 3, as well as the interval in between. The individuals in the skewed habitat around generation 800 (before the second successful invasion) are mostly offsprings of the invaders from the uniform habitat which took over the skewed habitat in the successful invasion which occurred around generation 650. This is illustrated in Fig. 4. The invaders carry with them the configurations which evolved in the uniform habitat, which are not optimal for the skewed habitat. This configuration is adapted gradually in the skewed habitat, which leads to the slow, but steady rise in the long-term fitness $\langle Y \rangle$, evaluated with respect to the full $P(i, j)$. In this case, this process is carried out by slow mutators (red curve), which dominate the skewed habitat population during this time (Fig. 2). However, around generation 780, a large change occurs, which corresponds to the loss of a barrier, and consequently a big drop in $\langle Y \rangle$ (Fig. 4). Most of the time rare hits are irrelevant, so $\langle Y \rangle$

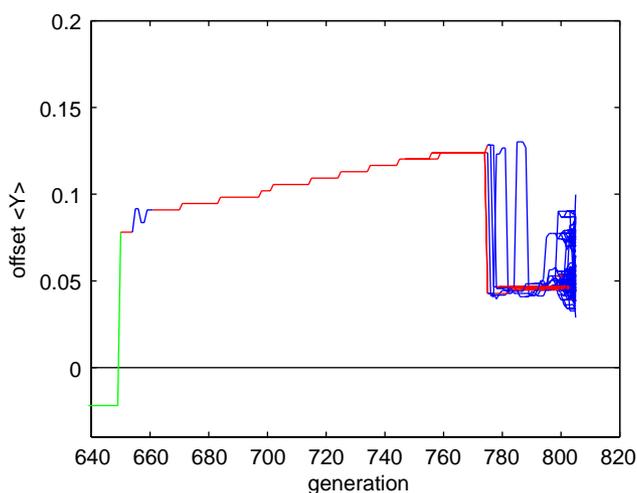


Fig. 4. Extended family tree illustrating the interval between the first successful invasion, and the second in Fig. 3. When lattices from the uniform habitat take over the skewed habitat, they adapt over time, leading to a gradual increase in long term fitness $\langle Y \rangle$. However, $\langle Y \rangle$ falls dramatically when a barrier is lost. This adaptation is beneficial for common events, but fatal for rare events.

underestimates the short-term fitness based on the common events. For common events, the barrier loss is beneficial, since it results in increased density. However, in the long term, this change leads to increased vulnerability, and ultimately to the extinction in the skewed habitat, which is followed by the second invasion.

5. Interpretations and conclusions

In this paper, we have aimed to develop a simple, yet illustrative model for biological evolution, habitat specialization, and extinction vulnerability. The model retains a transparent connection between genotype, phenotype, environment, and fitness. Despite its extreme abstraction, our model provides a clear mechanism whereby microscopic processes of mutation and selection at the organism level (microevolution) are linked through competition and interaction with an uncertain environment, to large-scale “macroevolutionary” patterns associated with the population as a whole. The model captures what we believe is the essential mechanism that drives the evolution of complexity in organisms and ecosystems: the emergence of highly ordered robustness/resilience architectures from randomness through evolution and natural selection (Zhou et al., 2002).

In developing the model, we have focused on primitive representations and have sacrificed the rich behavior and complexity of fitness, loss, and the formation of robustness architectures (and networks) in real organisms and ecosystems. These topics would naturally complement and expand on the theoretical framework we have introduced here, and may most naturally be developed in the context of specific case studies. Topics of interest include genetic transfer and recombination and interactions beyond direct competition, which could lead to food chains, various types of symbiosis (Margulis, 1998), predation, and parasitism. Factors such as environmental change, introduction of exotic species, and stoichiometric constraints (Sterner and Elser, 2002) could also naturally be implemented in case-specific scenarios.

Our model is extremely abstract yet has a number of features that are strikingly reminiscent of real biology. Our lattice model exhibits both genotypic divergence and phenotypic convergence in response to the environment. In all variants of our model, organisms consistently cycle through similar phenotypes, but even very similar phenotypes, in the sense of having similar cluster sizes and barrier locations can have genotypes that differ substantially in their exact defining lattice configurations. Conversely, small changes in the genotype can break barriers, producing large changes in phenotype. Nevertheless, the process of purposeless mutation and selection

in our model, like biology, creates the *impression* of a clear direction in evolution (Sniegowski and Lenski, 1995), with results very similar to what would arise from purposeful engineering design.

The initial transient phase was not discussed in detail but it illustrates a relatively rapid take over of the habitat by the lineage of one dominant random lattice. Similar processes are widespread in biology and ecology. One well-studied case involves evolution of *Daphnia* in aquatic environments where amictic parthenogenesis (clonal reproduction) dominates during favorable growth conditions, leading to nearly complete genetic homogeneity of the population (De Meester, 1996). In our case, specialization for common events in the skewed habitat leads to extinction vulnerability to rare events. Similar processes involving specialization and extinction vulnerability are widely discussed (and debated) in ecology and paleontology (McKinney, 1997).

Long-term evolution in the uniform habitat typically leads to the evolution of stable, “generalist” patterns, consisting of a roughly uniform grid of barriers, which is roughly optimal in their habitat. The generalists essentially never suffer large extinction events. When the habitats are coupled, through migration of the organisms, the generalists are safe but suboptimal in habitats with skewed distributions. However, when the specialists inevitably suffer total extinction due to rare sparks, generalists that have migrated into the skewed habitat flourish. The generalists eventually evolve into specialists, and the cycle repeats (Dykhuizen and Davies, 1980; Thompson, 1994). The variable mutation rate, which is itself subject to mutation and selection, sometimes leads to an apparent division of labor, in which the fast mutators dominate in the adaptation phase, but are later replaced by slow mutators that effectively preserve favorable, high fitness, evolved patterns. Similar processes, and the role of variable mutation rates, are of great interest in studies of bacterial evolution (Drake, 1991; Horst et al., 1999; Sniegowski et al., 1997, 2000; Taddei et al., 1997). In these and other studies, a key constraint on rapid mutation rates is imposed by the inevitable tradeoff between rapid adaptation and deleterious genetic errors. This tension arises in our model as well, and detailed, quantitative investigation of this issue would be a natural next extension of our model.

Our basic framework is amenable to specific case studies involving primitive, rapidly evolving organisms in “microbial ecosystems,” where populations are subject to a range of stresses at different frequencies. Candidates include bacterial systems (Storz and Hengge-Aronis, 2000) as well as the zooplankton *Daphnia* and its algal prey. Both systems have the advantage that they can be studied by a variety of means, in both natural and laboratory settings, and using a range of methodologies, probing the system

from microscopic (DNA) scales to macroscopic consequences (turbidity). For problems involving a broad range of habitats and/or large populations of distinct cooperative or competing organisms, replacing the explicit lattice representation by the more abstract, and less computationally intense probability-resource-loss (PLR) HOT representation (Doyle and Carlson, 2000) may be natural. In other cases, increasing the fidelity of the model to better represent degrees of freedom which are available to individual organisms will be useful in uncovering the evolutionary advantage of networks and architectures responsible for the robustness and resilience of complex organisms.

Acknowledgements

This work was supported by the David and Lucile Packard Foundation, NSF Grant No. DMR-9813752, the James S. McDonnell Foundation, EPRI/DoD through the Program on Interactive Complex Networks, and the Institute for Collaborative Biotechnologies through grant DAAD19-03-D-0004 from the U.S. Army Research Office.

References

- Bak, P., Sneppen, K., 1993. Punctuated equilibrium and criticality in a simple-model of evolution. *Phys. Rev. Lett.* 71, 4083–4086.
- Bonner, J.T., 1988. *The Evolution of Complexity by Means of Natural Selection*. Princeton University Press, Princeton, New Jersey.
- Caporale, L.H., 2003. Natural selection and the emergence of a mutation phenotype: an update of the evolutionary synthesis considering mechanisms that affect genome variation. *Annu. Rev. Microbiol.* 57, 467–485.
- Carlson, J.M., Doyle, J., 1999. Highly optimized tolerance: a mechanism for power laws in designed systems. *Phys. Rev. E* 60, 1412.
- Carlson, J.M., Doyle, J., 2000. Highly optimized tolerance: robustness and design in complex systems. *Phys. Rev. Lett.* 84, 2529.
- Carlson, J.M., Doyle, J., 2002. Complexity and robustness. *Proc. Natl Acad. Sci.* 99, 2538.
- Csete, M., Doyle, J., 2002. Reverse engineering of biological complexity. *Science* 295, 2002.
- Csete, M., Doyle, J., 2004. Bowties, metabolism, and disease. *Trends Biotechnol.* 22, 446–450.
- De Meester, 1996. Local genetic differentiation and adaptation in freshwater zooplankton populations: patterns and processes. *Ecoscience* 3, 385–399.
- Doyle, J., Carlson, J.M., 2000. Power laws, highly optimized tolerance, and generalized source coding. *Phys. Rev. Lett.* 84, 5656.
- Drake, J.W., 1991. A constant rate of spontaneous mutation in DNA-based microbes. *Proc. Natl Acad. Sci.* 88, 7160–7164.
- Drossel, B., 2001. Biological evolution and statistical physics. *Adv. Phys.* 50, 209–295.
- Dykhuizen, D., Davies, M., 1980. An experimental model: bacterial specialists and generalists competing in chemostats. *Ecology* 61, 1213–1227.
- El-Samad, H., Kurata, H., Doyle, J.C., Gross, C.A., Khammash, M., 2005. Surviving heat shock: control strategies for robustness and performance. *Proc. Natl Acad. Sci.* 102, 2736–2741.

- Gavrilets, S., 2004. *Fitness Landscapes and the Origin of Species*. Princeton University Press, Princeton, New Jersey.
- Holling, C.S., 1973. Resilience and stability of ecological systems. *Ann. Rev. Ecol. Syst.* 4, 1–23.
- Horst, J.-P., Wu, T.-h., Marinus, M.G., 1999. *E. Coli* mutator genes. *Trends Microbiol.* 7, 29–36.
- Kauffman, S.A., 2000. *Investigations*. Oxford University Press, Oxford, New York.
- Kirschner, M., Gerhart, J., 1998. Evolvability. *Proc. Natl Acad. Sci.* 95, 8420–8427.
- Lauffenburger, D.A., 2000. Cell signaling pathways as control modules: complexity for simplicity? *Proc. Natl Acad. Sci.* 97, 5031–5033.
- Levin, S., 1999. *Fragile Dominion*. Perseus Books, Reading, MA.
- Margulis, L., 1998. *Symbiotic Planet: A New View of Evolution*. Perseus Books, New York.
- Maynard Smith, J., 1989. The causes of extinction. *Philos. Trans. R. Soc. B* 325, 241–252.
- McKinney, M.L., 1997. Extinction vulnerability and selectivity: combining ecological and paleontological views. *Annu. Rev. Ecol. Syst.* 28, 495–516.
- Newman, M., 2000. Applied mathematics and the power of design. *Nature* 405, 412.
- Newman, M.E.J., Palmer, R.G., 2003. *Modeling Extinction*. Oxford University Press, Oxford.
- Raup, D.M., 1986. Biological extinction in Earth history. *Science* 231, 1528–1533.
- Reynolds, D., Carlson, J.M., Doyle, J., 2002. Design degrees of freedom and mechanisms for complexity. *Phys. Rev. E* 66, 16108.
- Robert, C., Carlson, J.M., Doyle, J., 2001. Highly optimized tolerance in epidemic models incorporating local optimization and regrowth. *Phys. Rev. E* 63, 056122.
- Savageau, M.A., 1976. *Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology*. Addison-Wesley Pub. Co., Reading, MA.
- Sneppen, K., Bak, P., Flyvbjerg, H., Jensen, M.H., 1995. Evolution as a self-organized critical phenomenon. *Natl Acad. Sci. USA* 92, 5209–5213.
- Sniegowski, P.D., Lenski, R.E., 1995. Mutation and adaptation: the directed mutation controversy in evolutionary perspective. *Ann. Rev. Ecol. Syst.* 26, 553–578.
- Sniegowski, P.D., Gerrish, P.J., Lenski, R.E., 1997. Evolution of high mutation rates in experimental populations of *E. coli*. *Nature* 387, 703–705.
- Sniegowski, P.D., Gerrish, P.J., Johnson, T., Shaver, A., 2000. The evolution of mutation rates: separating causes from consequences. *BioEssays* 22, 1057–1066.
- Sole, R. V., Manrubia, S. C., Benton, M., Bak, P., 1997. Self-similarity of extinction statistics in the fossil record. *Nature* 388, 764–767.
- Stauffer, D., 1985. *Introduction to Percolation Theory*. Taylor, London.
- Stelling, J., Sauer, U., Szallasi, Z., Doyle, F.J., Doyle, J., 2004. Robustness of cellular functions. *Cell* 118, 675–685.
- Sterner, R., Elser, J., 2002. *Ecological Stoichiometry: The Biology of Elements from Molecules to the Biosphere*. Princeton University Press, Princeton, NJ.
- Storz, G., Hengge-Aronis, R., 2000. *Bacterial Stress Responses*. American Society for Microbiology, Washington, DC.
- Taddei, F., Radman, M., Maynard Smith, J., Toupance, B., Gouyon, P.H., Godelle, B., 1997. Role of mutator alleles in adaptive evolution. *Nature* 387, 700.
- Thompson, J.N., 1994. *The Coevolutionary Process*. University of Chicago Press, Chicago.
- Weiss, K.M., Buchanan, A.V., 2004. *Genetics and the Logic of Evolution*. Wiley, Hoboken, New Jersey.
- Yi, T.M., Huang, Y., Simon, M.I., Doyle, J., 2000. Robust perfect adaptation in bacterial chemotaxis through integral feedback control. *Proc. Natl Acad. Sci.* 97, 4649.
- Zhou, T., Carlson, J.M., 2000. Dynamics and changing environments in highly optimized tolerance. *Phys. Rev. E* 62, 3197.
- Zhou, T., Carlson, J.M., Doyle, J., 2002. Mutation, specialization, and hypersensitivity in highly optimized tolerance. *Proc. Natl Acad. Sci.* 99, 2049.
- Zhou, T., Carlson, J.M., Doyle, J., 2005. Evolutionary dynamics and highly optimized tolerance: extended version with supplementary details, Report.