

PRODUCTIVITY AND DIVERSITY IN A CROSS-FEEDING POPULATION OF ARTIFICIAL ORGANISMS

Philip Gerlee^{1,2} and Torbjörn Lundh^{3,4}

¹Center for Models of Life, Niels Bohr Institute, Blegdamsvej 17, 2100, Copenhagen, Denmark

²E-mail: gerlee@nbi.dk

³Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, SE-41296, Göteborg, Sweden

⁴E-mail: torbjorn.lundh@chalmers.se

Received August 18, 2009

Accepted March 17, 2010

Cross-feeding interactions are a common feature of many microbial systems, such as colonies of *Escherichia coli* grown on a single limiting resource, and microbial consortia cooperatively degrading complex compounds. We have studied this phenomenon from an abstract point of view by considering artificial organisms that metabolize binary strings from a shared environment. The organisms are represented as simple cellular automaton rules and the analog of energy in the system is an approximation of the Shannon entropy of the binary strings. Only organisms that increase the entropy of the transformed strings are allowed to replicate. This system exhibits a large degree of species diversity, which increases when the flow of binary strings into the system is reduced. Investigating the relation between ecosystem productivity and diversity we find that diversity is negatively correlated with biomass production and energy uptake, while it correlates positively with energy-uptake efficiency. By performing invasion experiments, we show that the source of diversity is negative frequency-dependent selection acting among the different species, and that some of these interactions are intransitive, another mechanism known to promote diversity.

KEY WORDS: Artificial life, cross-feeding, productivity, species diversity.

The origin of biodiversity has been a long standing problem in ecology (Hutchinson 1959), and the evolution and maintenance of diversity was long difficult to account for, especially in the light of the proposed competitive exclusion principle that states that several species competing for the same resources cannot co-exist. However recent advances have revealed several ways in which diversity can be maintained, for example it can be promoted by ecological mechanisms, such as competition (Tilman 1982; Czaran et al. 2002), mutualism (Bronstein 1994), and niche partitioning (Schluter 2000), and also by genetic mechanisms such as fitness trade-offs (Rainey et al. 2000) and adaptive decay (Ostrowski et al. 2007).

Related to these issues is the question of how species diversity influences ecosystem productivity (Waide et al. 1999).

Several experiments and theoretical models have been devised to resolve this issue, but many of the results have been inconclusive and even contradictory. Some models have predicted that diversity should peak at intermediate productivity (Kassen et al. 2000), and that this can occur either by spatial niche formation or frequency-dependent selection (Chow et al. 2004), whereas in other experiments diversity was observed to increase as a function of the productivity (Travisano and Rainey 2000). Yet another hypothesis states that species diversity might increase at low resource levels (Valentine 1971) and there is some support for this suggestion in aquatic ecosystems (Hessler and Sanders 1967; Fryer and Iles 1969).

These differing observations most likely follow from the fact that the relation between biodiversity and productivity is

dependent on the system under consideration and emerges as a result of the system-specific species interactions in the ecosystem. The relation also seems to depend on which scale it is studied; at continental scales, the general trend is a positive correlation, whereas the unimodal pattern seems to dominate studies performed on smaller scale (Waide et al. 1999). Another source of discrepancy stems from the definition of ecosystem productivity. Some researchers have taken it to be the rate of energy flow through the system (Chow et al. 2004), whereas others have equated it with the rate of primary production in the system (or some suitable proxy such as rainfall) (Odum 1959) or the amount of nutrients available for the organisms (Kassen et al. 2000).

One of the simplest ecological systems in which diversity emerges, and is stably maintained, is in populations of *Escherichia coli* growing in a homogeneous environment limited by a single resource, usually glucose. The diversity is facilitated by cross-feeding (syntrophy), where one strain partially degrades the limiting resource into a secondary metabolite that is then used by a second strain. This phenomenon was first observed by Helling et al. (1987) and has since been reported to occur in other systems such as methanogenic environments (Stams 1994), bacteria engaging in nitrification (Costa et al. 2006), and degradation of xenobiotic compounds (Dejonghe et al. 2003; Katsuyama et al. 2009). In the case of *E. coli* grown on glucose, the second strain is clearly subject to frequency-dependent selection (as it would starve in the absence of the primary degrader), however it has also been hypothesized that the first strain is dependent on the second one, as the secondary metabolite could be toxic at high concentrations (Pelz et al. 1999). This raises the question why cross-feeding emerges in the first place, and why a single strain that completely degrades the nutrient is not evolutionarily superior.

This question has been investigated by Pfeiffer and Bonhoeffer (2004) using a theoretical model, and their results showed that cross-feeding naturally emerges under the assumption that ATP production is maximized whereas the total concentrations of enzymes and intermediates are minimized. Further they showed that the evolution of cross-feeding depends on the dilution rate in the chemostat, and that a stable polymorphism is more likely to emerge at low dilution rates.

A different approach was taken by Doebeli (2002) who investigated the emergence of cross-feeding in the framework of adaptive dynamics. In this case, the conditions for evolutionary branching and the appearance of cross-feeding are that there is a trade-off between uptake efficiency of the primary and secondary metabolites, and that this trade-off function has a positive curvature. The model also makes the correct prediction that cross-feeding is less likely to occur in serial batch culture, in which the primary resource is not replenished (Rozen and Lenski 2000). This highlights the necessity of the secondary metabolite being present for an extended period of time for cross-feeding to evolve.

In this article, we present a more general model of the evolution of cross-feeding, which is not aimed at modeling a specific biological system, but rather extracts and models the general principles governing systems in which cross-feeding might emerge. This type of modeling approach is typically labeled as “artificial life” (Adami 1998), an interdisciplinary field that through simulation and emulation of living systems tries to extract the general principles governing living systems. A special class of artificial life systems, which have been used extensively in biological research (Adami 2006), are those which use self-replicating code as a means of *in silico* imitating the *in vivo* evolutionary process by considering an evolving population of digital organisms, which replicate and perform simple computational tasks. The pioneering system of this kind was Tierra, developed by T. Ray (1992), and today the most widely used platform is Avida, developed by researchers at Caltech (Wilke and Adami 2002).

The Avida system does not try to mimic the precise details of living organisms and evolution, but has nevertheless been very useful for understanding various aspects of biology and evolution, such as evolution of genetic architecture (Gerlee and Lundh 2008) and metabolic pathways (Gerlee et al. 2009), genetic interactions (Lenski et al. 1999), and the evolutionary origin of complex features (Lenski et al. 2003). One study in particular dealt with the relation between productivity and diversity in a microcosm of evolving digital organisms (Chow et al. 2004), and concluded that diversity peaks at intermediate productivity and that this occurs due to frequency-dependent selection promoted by the influx of mixed resources into the system.

A general drawback of the above-mentioned system is that the fitness function, which influences the replication rate of the digital organisms, consists of a list of computational tasks, which is *a priori* decided. Further, the interactions between the organisms are weak and only by explicitly implementing flow between computational resources do the interactions resemble those observed in real ecosystems (Yedid et al. 2009). To ameliorate this situation, we have devised a novel Artificial Life system, named *Urdar* (Urðarbrunnr is one of the three wells that lie beneath the world tree Yggdrasil in Norse mythology. The name means well of fate.) in which the fitness of an organism is defined in a more general sense and where interactions between organisms are at the core of the model. The fitness of the organisms in this model is directly related to their ability to extract energy from a common environment, and is more closely connected to the fundamental concept of energy that drives many ecological interactions. On the other hand, this system does not take self-replication directly into account and can therefore be seen as a complement to for example Avida that explicitly handles self-replication of the digital organisms.

The *Urdar* platform is more of an analogy than a classical model, because there are no existing groups of organisms being

realistically modeled, but on the other hand—it realizes an evolutionary process similar to the one occurring in nature. What can then be learned from such a system? We hope that this platform will be like a sand box, where concrete experiments can be carried out, investigating the evolution and dynamics of cross-feeding populations and also for analyzing more general questions about evolution. In this first investigation of *Urdar*, we have focused on how the resource supply of the ecosystem influences species diversity and productivity. The main conclusion is that low resource levels tend to increase the diversity, by implicitly increasing the variety of the food supply and thus opening up new niches in the environment. Although the efficiency of the ecosystem at harvesting energy increases at low resource levels, the overall productivity decreases, which demonstrates that efficiency and productivity are two distinct features of an ecosystem. Further we show, by performing invasion experiments, that the diversity in the system is caused by negative frequency-dependent selection and intransitive competition.

The Model

To motivate the suggested platform *Urdar*, let us imagine a situation in which a population of different strains of bacteria inhabit a petri dish continually supplied with a given nutrient. The bacteria only partially metabolize the nutrient, which is added at a certain rate, so other bacteria might extract energy from the “leftovers” of this successive degradation. Assume that this experiment is carried out for a long period of time, so that strains that do well will increase their share of the total population. Because we can imagine that different strains of bacteria have variations to their metabolism, we have that if a single species dominates the population, a certain type of leftovers will be abundant in the free pool of metabolites. Hence that would lead to higher number of offsprings of a species that is specialized on extracting energy from that kind of leftovers. Please note that the model we will present is not specific to bacteria, but could represent any ecosystem where resources are consecutively degraded by several species, creating a network of interdependence. The general structure of these interactions are the main goal of our study, and how they relate to ecosystem properties such as productivity.

We set up such an experiment using artificial organisms or agents that are capable of successive degradation (transformation) of metabolites from which they extract energy used for self-maintenance and reproduction. More specifically, let R be a pool of resources (or metabolites) $\{r_i\}$ where each r_i is a binary string of length L , as for example $r_i = 00101, \dots, 01110$. Let A be the population $\{a_j\}$ of agents (or organisms), where each agent a_j is represented by a function that transforms binary strings into new binary strings, $a_j : R \rightarrow R$. We can view this mapping as a

“metabolic digestion” of the string being transformed. More precisely the agents in A transforms resource strings from R in the following way:

$$r_i^{\text{new}} = a_j(r_i^{\text{old}}).$$

Let now a positive function E on the binary strings in R represent the “energy state” of such a string. If the agent a_j is able to extract energy from the resource string r_i , we have that $E(r_i^{\text{new}}) < E(r_i^{\text{old}})$, and the amount of energy extracted is given by

$$\Delta E_j = E(r_i^{\text{old}}) - E(r_i^{\text{new}}).$$

The evolutionary dynamics are then introduced by a possible replication of the agent a_j to a daughter agent whenever $\Delta E_j > 0$. Replication in the current model is asexual and offsprings have just a single parent organism. The offspring is mutated with probability μ , and replaces another agent in the population, thus keeping the population size constant. The constant population size can be thought of as either being imposed by a space constraint, or by the carrying capacity of an additional nutrient required for biomass synthesis (assuming that the evolutionary dynamics related to this trait occurs on a much slower timescale). The probability for a reproduction to take place is an increasing function of ΔE_j with zero probability if $\Delta E_j \leq 0$. Hence a successful type of agent, is one which is able to effectively extract energy from the binary resource strings in R , and the content of R in turn depends on which agents constitute the population. To feed the system with energy, strings in the resource pool R are continually being replaced with new high-energetic strings at a rate γ , representing a flow of energy into the system. A schematic of the modeling framework is shown in Figure 1, which illustrates how binary strings are metabolized by the organisms and flow through the system.

The framework described so far is quite general, and we will in the following describe the particular choices we have made in the current study. First, the agents a_j are chosen to be elementary cellular automata (CA), one of the simplest notions of digital algorithms (see Appendix A and Fig. A1 for an example). The reason for that particular choice in *Urdar* is that such functions are well studied in the literature starting from the work of Wolfram (1983). They are simple, but when applied in a sequence show a surprisingly wide range of complexity. The second choice we made was using an approximated Shannon Entropy as the energy function E , which gives an estimate of the amount of disorder a binary string contains (see Appendix B), associating a low entropy (low level of disorder) with a high “energy” state of the string, that is, we set $E = 1 - s$. To motivate such a choice, one can see organismal metabolism as extraction of energy (usually in the form of ATP molecules) through the successive degradation of chemical compounds, but one other fundamental feature of metabolic

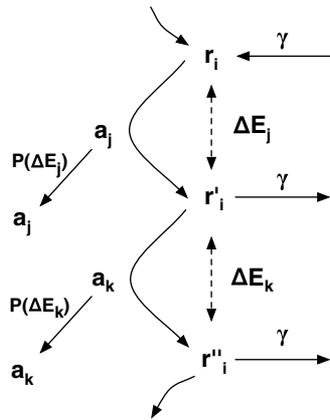


Figure 1. A schematic view of the model. The agents a in the model digest binary strings r by applying CA-rules, transforming r to r' . To each such metabolic step, we can associate a difference in energy ΔE (visualized with dotted lines). The reproduction of each agent depends on how much it can decrease the energy of the binary string and occurs with probability $P(\Delta E)$ (represented by the arrows on the left-hand side). The binary strings exist in a common pool which they enter (and leave) at a rate γ , as shown by the arrows on the right-hand side.

processes is that organisms maintain their degree of internal order by increasing the disorder (entropy) of the compounds being metabolized; or in the words of White et al. (1973):

“Thus the order of the foodstuffs is altered through oxidation, to maintain the high degree of order of the cell. The sum of such processes in the organism may be presumed to comprise a major fraction of the basal metabolic rate. If the supply of food or oxygen ceases, the tendency toward equilibrium is not counterbalanced, and the expected equilibria attained.”

This viewpoint is both common and well established.

“Thus the device by which an agent maintains stationary at a fairly high level of orderliness (=fairly low level of entropy) really consists in continually sucking orderliness from its environment.” (Schrödinger 1944)

One could of course make use of a more sophisticated “artificial chemistry” by assigning higher energy, and hence fitness, to an organism that is able to transform strings into certain patterns, instead of just increasing the entropy; but in our effort for simplicity and a more open-ended fitness function, we have chosen the current setup.

The probability for agent a_j to reproduce, as a function of the energy it extracts from a binary string, is set to

$$P(\Delta E) = \begin{cases} \frac{1 - \exp(-\Delta E/\beta)}{1 - \exp(-\beta)}, & \text{if } \Delta E > 0 \\ 0, & \text{if } \Delta E \leq 0, \end{cases} \quad (1)$$

where β is a positive parameter indicating the level of competitive pressure among the agents. In the limit of $\beta \rightarrow 0$ selection is weak as any $\Delta E > 0$ gives a probability of reproduction very close to unity, whereas for larger β selection is stronger as the magnitude of ΔE is more important for determining the value of $P(\Delta E)$ and hence the reproductive success of the organisms.

An example of applying CA-rules to binary strings is shown in Figure 2, where three rules are allowed to digest a string with a low entropy to binary strings with successively increasing entropy. This is the type of interactions we can expect in the model, in particular at low γ when the strings are replenished at a low rate. This figure also illustrates the fact that the CA-rules in general make small changes to the food string during digestion. In fact there is no CA-rule that can, in a single metabolic step, increase the entropy of a fairly ordered string to the maximum attainable entropy. This is similar to individual metabolic reactions in

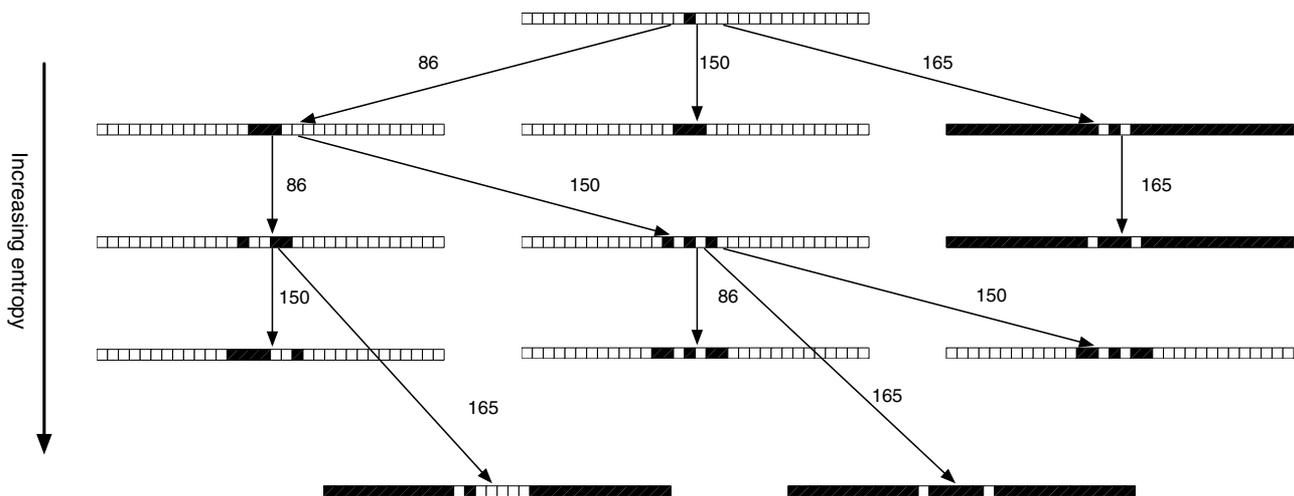


Figure 2. The transformation of binary food strings by three CA-rules. Only transformations that increase the entropy are shown and they have been truncated at a metabolic depth of four. The number of possible transformations is greater for the three rules together than for a single isolated rule suggesting the possible advantage of cross-feeding among the species in the model.

real organisms that generally only change the free energy of the metabolites a small amount, whereas the metabolism as a whole is responsible for the major difference in free energy between the nutrients taken up by the organism and the waste products being excreted. This fact also suggests that *Urdar* can be viewed as a model of the early stages of life on earth when the metabolic repertoire of organisms was much smaller and cross-feeding was possibly more prominent.

Note that in the current setup, the mapping between the genotype and phenotype of the agents is one-to-one, where the genotype corresponds to the integer value representing the rule (ranging from 0 to 255), and the phenotype simply is the action of the rule on the strings that are metabolized. All organisms implementing the same CA-rule are consequently referred to as belonging to the same species. Furthermore, there is no speciation occurring in the model because all the 256 different species already exist—it is just their frequencies that vary. In the current setup, we have chosen not to explicitly model self-replication, to keep things simple. In future extensions of the model both sex and self-replication can be included. However, the first planned amendment is to introduce a genome to add a separation between genotype and phenotype, which will introduce a more realistic fitness landscape with for example frequent neutral mutations.

IMPLEMENTATION

To conclude the model description, (an online version of the platform is available at: <http://www.math.chalmers.se/~torbjrn/Urdar/urdar.html>) let us sum up its main features. The dynamics, depicted schematically in Figure 1, during one update can be described in the following way: (1) Each agent in the population picks randomly a resource string r_j from the resource pool R and transform it accordingly to its CA-rule and then puts the transformed string back into the resource pool; (2) The efficiency of the “metabolic process” just occurred is evaluated by measuring the energy difference ΔE of the string before and after the “digestion/transformation.” This is done by drawing a random number x uniformly between 0 and 1, and if $P(\Delta E) > x$ the agent reproduces, replacing a randomly picked agent with a copy of itself; (3) With probability μ the offspring will be mutated uniformly to another CA-rule; and (4) To keep energy flowing into the system, after all agents have been updated, a fraction γ of the strings is replaced with high energy binary strings.

The replacement rate γ can be seen as a flow rate of energy into the system. If that rate is high, there will be less interaction through cross-feeding among the agents in A , as strings are flushed out at a high rate, but if, on the other hand, γ is set to zero, the whole process will slow down to a halt. The strings introduced into the system are random binary strings, however with a low entropy (high degree of order). The new strings are constructed by, at each position, adding a 1 with probability p_0 and a 0 with the

complementary probability $1 - p_0$. The Shannon entropy of such strings is given by

$$s_0 = p_0 \log_2 \frac{1}{p_0} + (1 - p_0) \log_2 \frac{1}{1 - p_0}, \quad (2)$$

where \log_2 is the logarithm with base 2, that is, $2^{\log_2 x} = x$. By setting p_0 close to unity, we can create strings that, although being random, have a low entropy. In order not to bias the resource pool with strings dominated by ones, at an equal rate we add strings that have the probabilities reversed, that is, are dominated by zeros instead. To ensure that the wide variety of strings flowing into the system does not bias the results, we also performed simulations in which a single fixed string with approximately the same s_0 was supplied as energy source, and these simulations showed qualitatively the same behavior, in particular a sustained species diversity. We also simulated the system with an inflow of strings based on a 01 and 011-patterns, with random bits flipped, and the results again showed similar dynamics and ecosystem properties to the case examined in this article.

THE PARAMETERS

One of the goals of this study is to monitor how the whole system behaves with respect to the various parameters introduced. For the sake of convenience for the reader, let us list the main parameters in the model.

- γ is the inflow rate of new high energetic binary resource strings into the pool R . After each update, that is, after all agents have digested a resource string, the probability for each resource string in the pool to be replaced by a new fresh one, keeping the total number of resource strings constant, is γ . Here, we will typically set $\gamma \in [0.003, 0.3]$.
- μ is the mutation probability in the reproduction, where an agent is uniformly changed to another of the 256 CA-rules. We will use $\mu = 0.01$ as a default value of the mutation rate.
- β is the level of selective pressure, as it determines the importance of ΔE in calculating the reproductive rate, see equation (1). The default value of in the current study is $\beta = 0.1$, giving an intermediate selection pressure.

The population size is set to $N_A = 1024$, and the number of binary strings in the resource pool is $N_R = 5N_A = 5120$. The size of the binary strings is set to $L = 100$, and level of order in the inflowing strings is $p_0 = 0.95$, which gives, through equation (2), an initial energy of $E_0 = 1 - s_0 \approx 0.8$. The initial condition of each simulation is a uniform distribution of species, that is, $1024/256 = 4$ organisms of each species, and a resource pool consisting of strings with the initial energy E_0 .

Results

The dynamics in *Urdar* are driven by the addition of new high-energy strings to the system. Without this flow, no organismal divisions can occur and the dynamics are frozen. The parameter γ , which controls the flow rate, thus corresponds to the rate of energy flow into the ecosystem. To investigate how the artificial ecology behaves in different resource regimes, we have therefore systematically varied γ over 2 orders of magnitude and measured several properties related to diversity, productivity, and species interactions.

SPECIES DIVERSITY

Figure 3 shows the species distribution as a function of the number of time steps for (A) high flow rate ($\gamma = 0.3$) and (B) low flow

rate ($\gamma = 0.003$), in two typical simulations. These plots reveal a qualitative difference in the dynamics at different flow rates, where the species distribution in the high flow regime is dominated by a single species, whereas at low flow rates the species diversity is larger. This difference can be quantified by looking at the Shannon index of the species distributions, defined as

$$H(t) = \sum_{i=0}^{255} f_i(t) \ln \frac{1}{f_i(t)},$$

where $f_i(t)$ is the fraction of species i at time t in the population. The Shannon index of the two simulations above are shown in Figure 3C and D. In both cases, the Shannon index starts at $H(0) = \ln 256 \approx 5.5$, as all species are equally abundant in the beginning of the simulation. Rapidly however the diversity drops due to

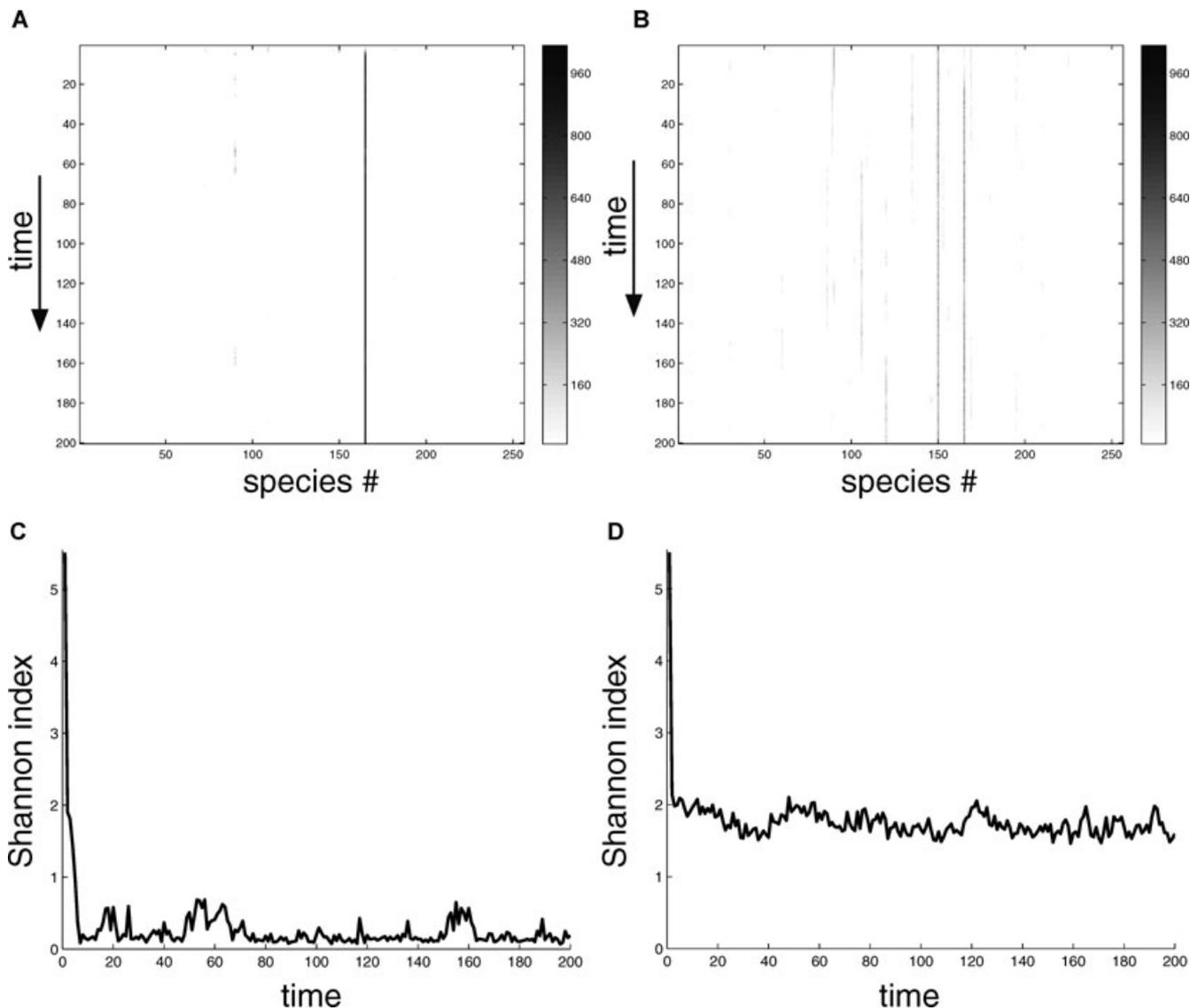


Figure 3. A time evolution of the species distribution in the ecosystem for (A) $\gamma = 0.3$ and (B) $\gamma = 0.003$. The two lower panels (C) and (D) show the time evolution of the Shannon diversity index for the simulations in (A) and (B), respectively.

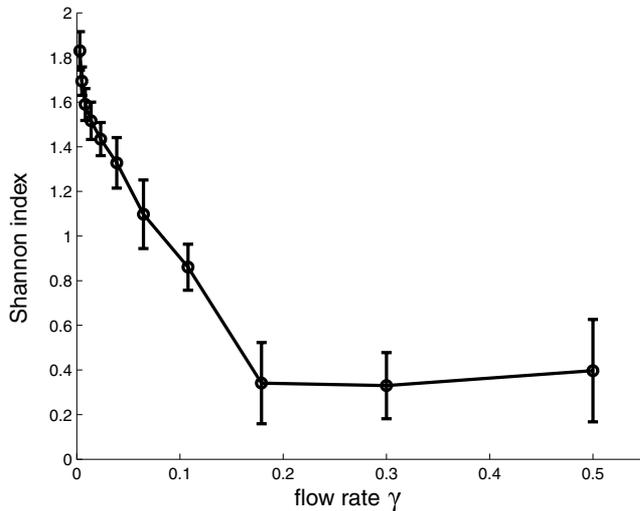


Figure 4. The Shannon diversity index of the species distribution as a function of the flow rate γ . Each datapoint was averaged over 20 simulations and the error bars represent one standard deviation.

selection among the species, and fluctuates around a mean value of 0.27 in the high flow case and 1.65 in the low flow regime, showing that the species diversity is significantly influenced by the flow rate (one-tailed binomial-test $N = 100$; $P < 10^{-31}$). In the low flow case, the system never settles in a steady state with regard to the species distribution, but it seems rather as the species are continually being replaced, a feature we will analyze in detail in the section on species interactions. On the other hand, in the high flow case the fluctuations in the Shannon index are due to mutations that spawn new species (in the available range 0–255), which are unable to coexist with, or outcompete the dominant species 92.

These results indicate that the diversity increases as the resource level in the system drops, and this trend was investigated systematically by measuring the time average of the Shannon index (excluding the first 20 time steps) for several values of γ in the range 0.003–0.3. The results are shown in Figure 4 and reveals that the diversity is a decreasing function of the flow and exhibits an approximately linear decrease with the flow rate γ , except for a saturation for high values of γ . The diversity in the system is also maintained, although at a slightly lower value, even if the mutation rate is set to zero during a simulation (data not shown).

This shows that at low flow rates the artificial ecosystem will tend to contain a higher number of species (at appreciable levels). This can partly be explained by the fact that at lower flow rates each food string stays in the system for a longer period of time; at $\gamma = 0.3$ the average number of successful metabolic reactions each strings has been through is approximately 1, whereas for $\gamma = 0.003$ it is approximately 6. This is in contrast with the average number of updates a string spends in the resource pool,

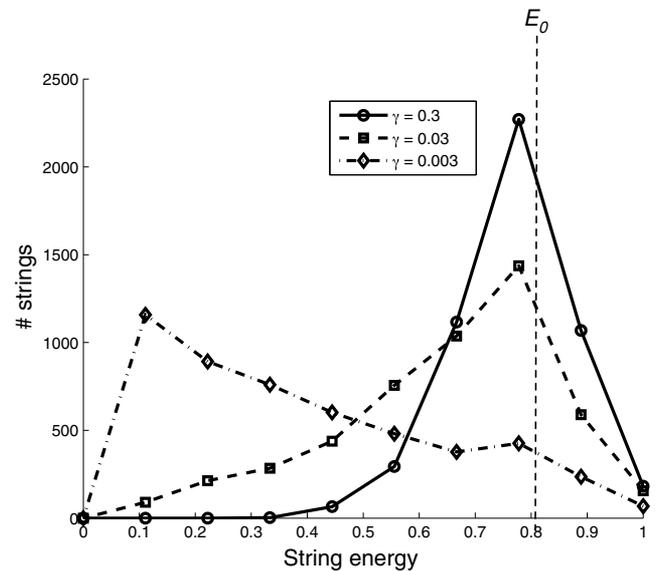


Figure 5. A histogram of the food string energy sampled at three different flow rates. E_0 corresponds to the mean energy of the inflowing strings.

which is given by $1/\gamma \approx 3$ and 300, considerably larger than the corresponding depth. The increase in metabolic depth leads to a larger diversity among the resources at lower flow rates, as illustrated in Figure 5, a histogram of the food string energy for three different flow rates. At high flow ($\gamma = 0.3$), the energy distribution is centered close to $E_0 \approx 0.2$, the initial energy of the strings, whereas for lower flow rates the distribution is skewed toward lower energies and is less peaked.

The organisms in the system interact with each other through the reuse of food strings. These interactions will be analyzed more closely in the section on species interactions, but first we will turn to the productivity in the ecosystem and investigate how it relates to the flow of energy in the system.

ECOSYSTEM PRODUCTIVITY

The productivity of an ecosystem can be defined in several ways and we will in this section investigate how the different measures apply to our artificial ecology. The most common measure is the rate of primary production by autotrophs usually measured in $[\text{g yr}^{-1} \text{m}^{-2}]$ (Waide et al. 1999). In our system, the amount of biomass (the number of organisms) is constant, but the rate of production can be compared to the reproduction rate ρ , defined as the number of divisions/update. Another way to look at productivity is to measure how much of the energy available to the ecosystem (sun light, chemical energy etc.) is being transformed into biomass. The equivalent in the artificial ecology is the difference in energy between the strings flowing into the system and the strings being removed, as this reflects the amount of energy that has been “transformed” into biomass. More precisely,

we calculate the total difference in energy between the strings being removed from the system and the strings being introduced in each update,

$$\mathcal{E} = \sum_k (E_0 - E(r_k^{\text{out}})), \quad (3)$$

where the sum runs over all strings being replaced (on average γN_R number of strings), and $E_0 \approx 0.8$ is the energy of the input strings. This measure, in the units of bits/update, we term the rate of energy uptake in the ecosystem, because it, in our analogy, corresponds to the amount of energy extracted per unit time by the population from its environment. The efficiency of this process is defined as

$$\eta = \frac{\mathcal{E}}{\gamma N_R E_0}, \quad (4)$$

where γN_R is the average number of strings being replaced each update. Because the energy of a completely degraded string equals zero, the maximum energy difference for each replaced string equals E_0 . Hence, the maximum value of \mathcal{E} is $\gamma N_R E_0$. From this follows that the efficiency lies between 0 and 1, and is a dimensionless quantity.

The influence of the flow rate γ on these measures related to productivity was analyzed for γ in the range 0.003 to 0.3 and the results can be seen in Figure 6. Figure 6A shows that the reproduction rate ρ increases linearly with $\log \gamma$, meaning that increasing the flow with a factor 100 only increases the reproduction rate by a factor 4. The energy uptake rate \mathcal{E} also shows a similar dependence on γ . When γ is increased from 0.003 to 0.3, \mathcal{E} increases from 5 to 35 bits/update. On the other hand, the efficiency η exhibits an inverse relationship with γ compared to the other measures, instead decreasing linearly with $\log \gamma$.

The similar behavior in reproduction rate and energy uptake rate is to be expected, as what drives organismal division is precisely decrease in string energy. For each successful string manipulation, the probability of division is given by equation (1), which is an increasing function of the energy change. This means that we are lead to the conclusion that productivity increases as a function of the flow. These facts taken together with the observation that the diversity is a decreasing function of the flow ($H \approx -1.4\gamma + 0.3$ for $0.02 < \gamma < 0.18$, see Fig. 4) means that the correlation between productivity and diversity in *Urdar* is negative. The energy-uptake efficiency on the other hand decreases with the flow rate. Although the flow of resources is increased by a factor 100, the reproduction rate is only increased by a factor 4, which shows that the ecosystem as a whole uses the common resource 25 times more efficiently in the low-flow regime.

The flow of energy into the system does not only depend on the flow rate, but also on the energy E_0 of the strings flowing into the system. The total energy flux is in fact given by the product

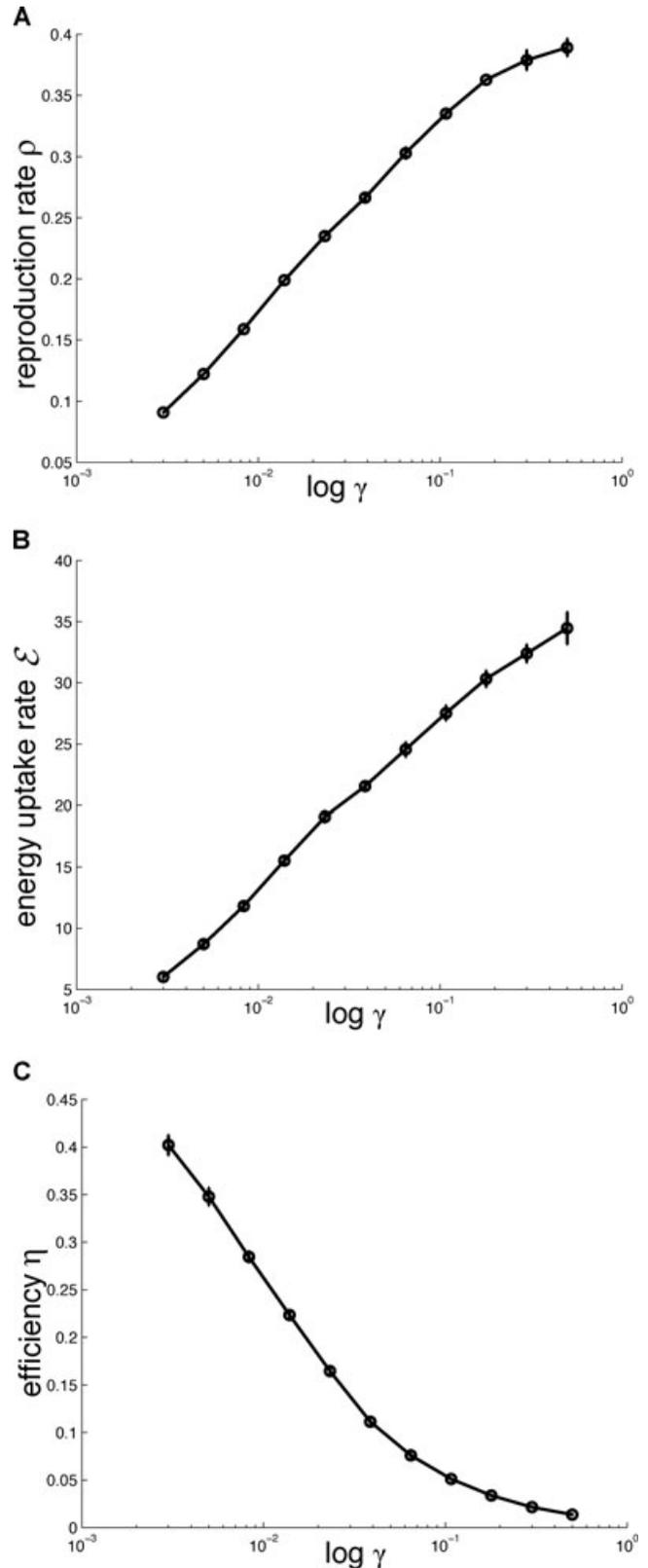


Figure 6. Three different measures related to productivity in the system. (A) the reproduction rate ρ , that is, the number of divisions per update, which corresponds to biomass growth, (B) The energy uptake rate \mathcal{E} , that is, the energy difference between outflow and inflow, and (C) the efficiency of the energy uptake η .

of γ and the amount of available energy in each string, which on average, is E_0 . A systematic study of how the diversity and the ecosystem properties previously introduced depend on the total energy flux showed that the diversity, productivity, and energy uptake are essentially constant along curves of constant energy flux. This means that if we vary the flow rate γ we can compensate for this by setting the initial energy equal to $E_0 = K/\gamma$ for some constant K . The ecosystem efficiency, on the other hand, shows a maximum at low flow rates and high available energy.

The choice of E_0 also affects the amount of energy extractable from each string, and peaks at $E_0 \approx 0.5$. This is due to the fact that a certain degree of disorder facilitates the functioning of the CA-rules in increasing the entropy (i.e., extracting energy) of the strings and is therefore specific to our choice of CA-rules as organisms. This effect is similar for all species, implying that for a given E_0 the species are affected in a similar way, which means that the influence on the ecosystem dynamics is small (except in an altered reproduction rate for all species).

It is also worth mentioning that because we have used disorder as a proxy for energy we have a conservation of the amount of order/disorder in the system, in analogy with energy conservation in real-world systems. This occurs because at the same time as a string transformation increases the entropy of the resource pool, the subsequent organismal reproduction reduces the entropy of the agent population. In a sense, we can view this as a transfer of order from the resource pool to the population. This hypothesis can most easily be tested by, during a simulation, dumping the entire resource pool and agent population into separate data files, and measure the rate of compression that can be achieved with the standard zip algorithm (Ornstein and Weiss 1993). What this shows is that over a range of flow rates the compression ratio (compressed/original size) of the resource pool C_R and of the population of agents C_A are approximately equal, $C_R \approx C_A$ (data not shown). Now, a low value of C_R corresponds to a low entropy resource pool from which large amounts of order can be extracted, and we therefore define the amount of available order as $\Omega_R = 1 - C_R$. We can now describe the conservation of order in the entire system as $\Omega_R + C_A \approx 1$. Or in other words, the amount of available order present in the resource pool plus the amount of disorder in the population is approximately constant.

SPECIES INTERACTIONS

As we have seen so far both the species composition and the ecosystem productivity are affected by the flow of energy into the system. What also has become clear is that species interactions play an important part in the dynamics of the system. The species composition that emerges at low flow rates in the ecosystem is more efficient at extracting energy from its environment suggesting a high degree of cooperativity in interactions between the species.

One of the advantages of a artificial ecology is that we can with ease manipulate the initial conditions and parameters of the model. We have made use of this flexibility to analyze species interactions with invasion experiments (Rainey and Travisano 1998). These experiments probe the reproduction rate of a species in the presence of only one other competing species, and can reveal if the two species can coexist or if one drives the other to extinction. The initial condition for each experiment was a 9:1 ratio in abundance between the two species (total population size was as before $N_A = 1024$) and the mutation rate was set to $\mu = 0$ throughout the entire experiment. An example of such a simulation can be seen in Figure 7, which shows the species abundance as a function of time when species 129 is paired of against species 145. The insets show the pattern the respective rules give rise to when iteratively applied to a string with an isolated zero and the rest ones (i.e., an illustration of the phenotypes of the species). These two species can clearly coexist, and do so in a 1:4 ratio (as opposed to the 9:1 initial condition). This balance is maintained by negative frequency-dependent selection, where the two species are better at extracting energy from strings that have previously been metabolized by the other species than the species itself.

These pairwise competition experiments were systematically performed for all $256 \times 256 = 65,536$ possible pairs (under flow rate $\gamma = 0.03$) and each simulation lasted 2000 updates or until only one species remained, and the abundance of both species was recorded. From these data, we can define a matrix C , where entry c_{ij} = the frequency of species i when the initial ratio between $i:j$ was 9:1. An excerpt of this matrix (row/column 90–166) is visualized in Figure 8, where black and white correspond to complete extinction of species i and j , respectively, and any shade in between signifies coexistence. Note that on the diagonal we have that $c_{ii} = 1$. This figure reveals that coexistence is a rather frequent feature of the interactions between the species in the model, and throws some light on the previously described high degree of diversity observed in the system. The species that can coexists are precisely those who are better at extracting energy from the strings that previously have been metabolized by the other species, than they are at recycling their own metabolites. The relative efficiency of these two processes then determines the equilibrium frequency between the species.

The interaction matrix in most cases satisfies $c_{ij} + c_{ji} = 1$, which means that the equilibrium concentration of the species is independent of the initial condition, but there are some interesting exceptions from this rule. First we have the anti-diagonal of the matrix where $c_{ij} + c_{ji} \approx 2$, and this is due to the underlying symmetry of the cellular automaton rules. The pairs on the antidiagonal are in fact rules that are inverses of each other when viewed in a binary representation. For example rule $145 = 10010001_2$ and its antidiagonal partner is rule $255 - 145 = 110 = 01101110_2$.

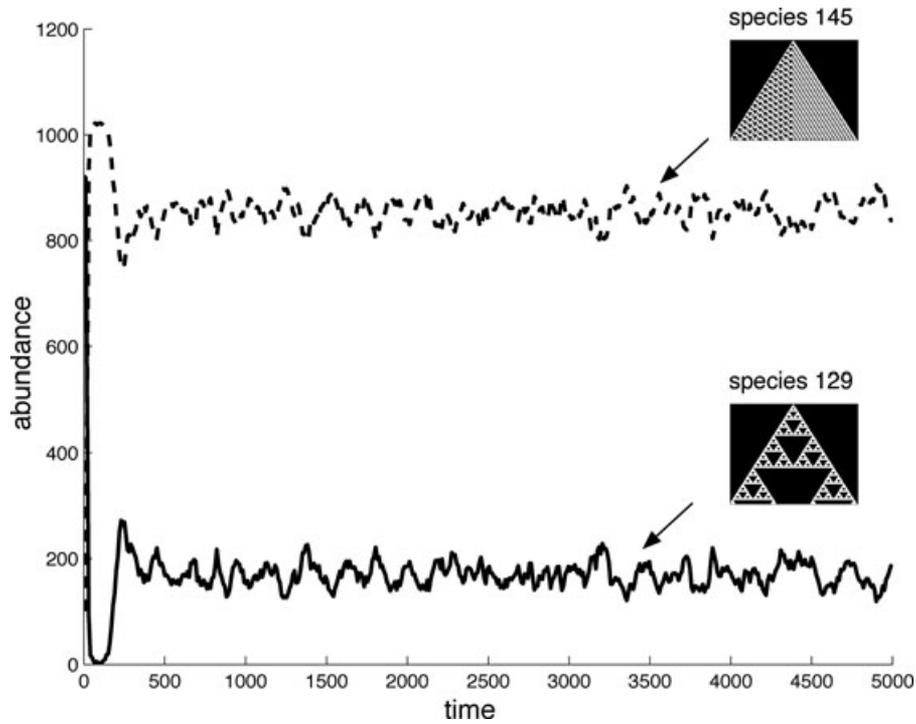


Figure 7. Pairwise competition experiment between species 129 and 145. At the beginning of the simulation species 129 dominates at a ratio 9:1, but almost instantly it declines in abundance and almost goes extinct. At this low abundance, it has a selective advantage and increases to a frequency that is maintained throughout the simulation.

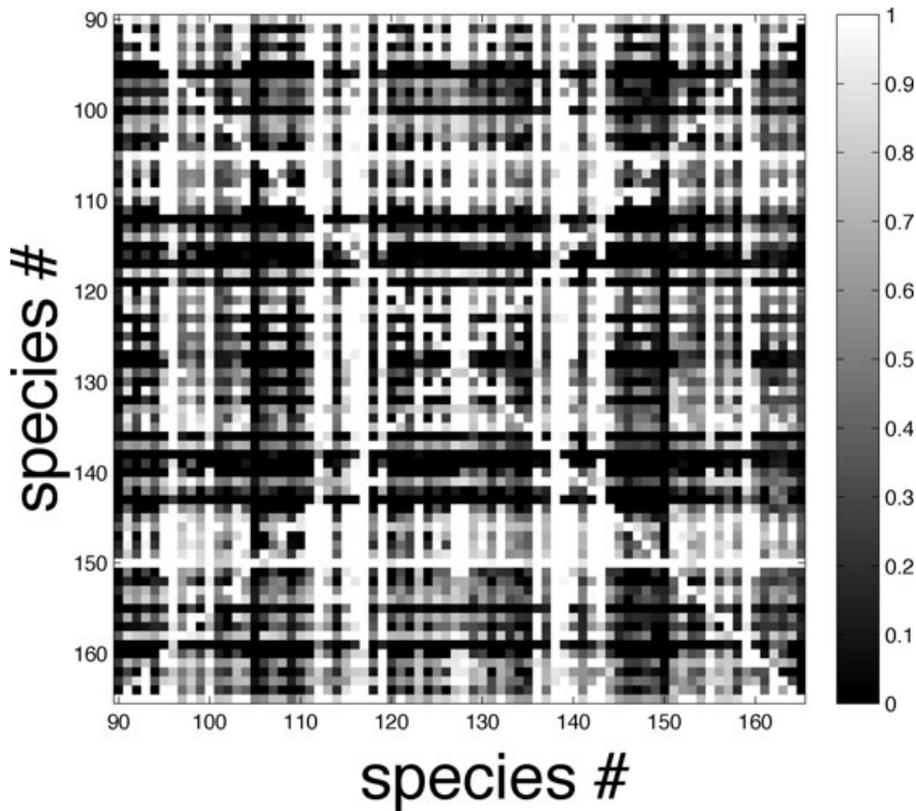


Figure 8. An excerpt of the interaction matrix constructed by pairwise competition experiments where c_{ij} = the frequency of species i when the initial ratio was 9:1 between $i:j$. For example the outcome of the experiment depicted in Figure 7 can be seen at row 129 and column 145, where the matrix entry equals 0.2.

When these rules are applied to a generic binary string the output strings they yield are inverses of each other, which by symmetry of the entropy function imply that they have the same entropy. This means that the two rules, when competing in isolation, are neutral and the only evolutionary force acting on the system is random drift. The consequence of this is that the dominant rule is more likely to win and therefore we observe $c_{ij} \approx c_{ji} \approx 1$ (or visually a white line) on the antidiagonal. Note that this does not imply that the two species are identical in their competition with other rules, and this has some important consequences for the dynamics of the model.

Second we have the cases in which $1 < c_{ij} + c_{ji} < 2$, which indicates that the initial condition in fact influences the equilibrium concentration. Upon further inspection, we found that the dynamics of these pairwise interactions contains two stable fixed-points, as opposed to one that is the case in all other interactions. Typically the only fixed-point lies either, in the case of coexistence, in the interior of the phase space at $(c, 1 - c)$, for the equilibrium concentration c , which satisfies $0 < c < 1$, or in the case of dominance at $(0,1)$. In the above-mentioned cases both an interior and a boundary fixed-point are present, and this implies that the dynamics can converge either to coexistence or dominance depending on the initial frequencies of the species.

For the species interactions that conform to one fixed-point dynamics, it is useful to introduce the notion of a dominating species, by which we mean that species i dominates j if the equilibrium concentration of i is greater than j in a pairwise competition experiment independent of the initial condition. This implies that $c_{ij} > 0.5$ and $c_{ji} < 0.5$, and we denote this by $i > j$. This definition allows us to compare the interactions between different pairs of species and see how they relate. For example, consider the neutral pair (114,141). These two are neutral in competition with each other, however, in competition with species 118, we have that $118 > 114$ and $141 > 118$. Essentially, this means that there is no hierarchical way to order these three species. This is even clearer in certain triples that do not contain any neutral pairs, as for example (126,134,141). For these species, we have $141 > 126$, $126 > 134$, and $134 > 141$, which means that they form an intransitive chain similar to the Rock–Paper–Scissors game ($R > S$, $S > P$, and $P > R$). These type of species interactions are also found in nature and are known to promote species diversity (Kerr et al. 2002; Laird and Schamp 2009).

To analyze how well the invasion matrix can describe the dynamics of the full model, we quantified the coexistence and invasion patterns in the following way: first the time-dependent species distribution was transformed into a binary form by considering a species to be “present” if its frequency was at least 10%. From these data, we calculated how likely a species A is to coexist with another species B as the conditional probability $Pr(B|A)$ of finding species B given that A is present. This mea-

sure can however be misleading as a rare species is more likely to show a large degree of coexistence, and we therefore formed a symmetric coexistence measure by taking the geometric average of $Pr(B|A)$ and $Pr(A|B)$. This measure is defined for any two pairs of species, which means that the result can be summarized in a matrix, and when we compare this coexistence matrix, calculated from 100 simulations, with the invasion matrix, we see that most cases of coexistence in the full model are explained by the fact that the two species equilibrate at nonzero frequencies in the invasion experiment. There are a few interesting cases that deviate from this tendency, for example, rule pair (109,150) shows considerable coexistence, but $c(150, 109) = 1$ suggesting complete dominance. To explain this observation one needs to invoke higher order interactions that are not captured by the pairwise invasion experiment.

The invasion patterns in the full model were quantified by observing the species configuration preceding the emergence of a given species A , and comparing it with the species composition during the time the species was present in the ecosystem. The species that existed prior to the emergence, but not at the later time, we label as having been invaded by species A . For each species, we calculated the probability that it is invaded by any other species in the model, and this shows that the invasion dynamics in the model can be well accounted for by the invasion experiments, although we again find a few exceptional cases that are due to higher order interactions. For example species 109 was found to be likely to invade 105, but in the pairwise invasion experiments we on the contrary found that species 105 outcompetes 109.

Discussion

The interactions occurring in a real ecosystem are enormously more complex than the rules that govern the interactions in the artificial ecosystem *Urdar*, yet the system exhibits a wide range of dynamical behavior. Most notable is the emergence of coexistence, maintained through negative frequency-dependent selection, which in turn is facilitated by cross-feeding among the species in the ecosystem. This is reminiscent of the dynamics observed in microbial communities (Helling et al. 1987; Rozen and Lenski 2000) and suggests that the model is able to capture the general principles governing these systems. The importance of cross-feeding increases when the inflow γ of new strings to the system is reduced. This occurs because each food string remains in the system longer and thus has a larger chance of being metabolized by several organisms. This makes the interactions between the organisms in the ecosystem stronger and consequently increases the importance of cross-feeding relations. The species compositions that emerge during starved growth conditions are therefore those which have strong mutualistic interactions.

The fact that at low flow rates give rise to a more diverse ecosystem might seem contradictory, but is actually in accordance with the results of Pfeiffer and Bonhoeffer (2004). Their model of the evolution of cross-feeding predicts that a second bacterial strain is more likely to emerge if the dilution rate in the chemostat is low. Only then does the intermediate metabolite exist in sufficiently high concentration to sustain a second bacterial strain. Although we do not speak of concentration of metabolites in *Urdar*, this corresponds to the fact that when, at low flow rates, the metabolites remain in the system longer, the resource pool R is more diverse (Fig. 5) and therefore can support a larger diversity of species, in the agent pool A .

In other words, the correlation between metabolic depth, or alternatively the diversity of the resource pool, and the number of species present in the artificial ecology suggests that the competitive-exclusion principle (Hardin 1960) also to some extent applies in *Urdar*. What we mean by this is that when γ is small, the system contains a larger number of distinct resources and it can therefore sustain a larger number of species, compared to a high flow rate when essentially only one type of resource exists and this only sustains a single dominating species.

Another factor that might influence the diversity is the fact that the system is subject to different levels of perturbations at different flow rates. At high flow rates, the system is less sensitive to the actions of a single organism, whereas at low flow the transformation of a single string will have a larger effect as it remains longer in the system. It has been hypothesized that intermediate perturbations can promote diversity (Grime 1973), however disentangling this effect from the stronger influence of cross-feeding in *Urdar* requires further investigation and is beyond the scope of this article.

The most salient feature of the model is the large degree of diversity it exhibits without a priori implementing frequency-dependent selection or specific niches that the species occupy. Instead these features emerge through the species interactions themselves. The artificial organisms and the binary food strings both inhabit a well-stirred environment, which means that the niches are formed not as physical locations, but rather as a given composition of resources. The diversity observed in these homogeneous growth conditions are in contrast with results from experiments performed with *P. fluorescens* (Kassen et al. 2000), where only heterogenous (nonstirred) environment gave rise to microbial diversity. This difference is most likely due to the fact that the above experimental system does not allow for cross-feeding and instead relies on selection of specialized types and niche competition as a means of generating diversity. On the other hand, it was shown in a recent paper by Saxer et al. (2009) that spatial homogeneity also can have a positive impact on diversity. In particular, they showed that populations of *E. coli* exhibited higher diversity in a well-mixed unstructured environment com-

pared to structured growth conditions, and hypothesize that the responsible mechanism is diffusion of nutrients between different sublcones, which is increased in the well-stirred environment. This points to the importance of system-specific interactions when determining the relation between different ecosystem properties, and the variety of mechanisms that can promote diversity in real ecosystems.

Most of the cases of coexistence and patterns of invasion are attributable to the first-order interactions captured by the invasion matrix calculated from pairwise invasion experiments. The coexistence is explained by negative frequency-dependent selection, whereas the invasion patterns are caused by neutral and intransitive competition between the species. Intransitive interactions are known to give rise to oscillatory or even chaotic behavior (Huisman and Weissing 1999), and allows for a larger number of species to coexist on a fixed number of limiting resources. The intermittent dynamics (apart from the dominating species) observed in the low flow (Fig. 3B) are due to this, as intransitive competition among species can lead to sequential (and sometimes cyclic) replacement of species, instead of dominance by the most fit species. This suggests that we have, at least, three mechanisms in *Urdar*, which promote species diversity, frequency-dependent selection, neutral competition, and intransitive interactions, all being facilitated by the cross-feeding the organisms engage in.

The results obtained from comparing productivity and species diversity in *Urdar* show that the reproduction rate ρ and the energy uptake \mathcal{E} both correlate negatively with biodiversity. At low flow rates, we observe a high species diversity and a low productivity, whereas a high flow rate gives a low diversity and a high productivity. This observation deviates from the dominant view that the two are bimodally related, but agrees for example with the hypothesis stated in (Valentine 1971) and the studies by Fryer and Iles (1969) and Hessler and Sanders (1967). The energy-uptake efficiency η on the other hand is highest at low flow rates, showing that the ecosystem as a whole uses the common resources most efficiently when the flow rate is small. This coincides with a high species diversity suggesting a connection between ecosystem efficiency and biodiversity.

When comparing our results with productivity measures that have been used in the literature, we found that there are three ways to define productivity in *Urdar* and coincide with measures from the literature. The total reproduction rate ρ , defined as the number of divisions per update, can be compared to biomass growth in real ecosystems, which is a common measure of productivity (Waide et al. 1999). The amount of available nutrients in an ecosystem, a productivity measure used by Kassen et al. (2000), corresponds to the average energy of the food strings. This quantity is given by $\sum_i (E_0 - E_i)$ which approximately equals \mathcal{E} (as the strings being removed from the system are chosen randomly).

Chow et al. (2004) used the rate of energy flow into the ecosystem as a measure of productivity and this corresponds in *Urdar* (for a fixed E_0) to the flow rate γ . These quantities clearly measure different properties of the system, still they all correlate negatively with the diversity, suggesting a certain degree of coherence to them.

Outlook

Even though the model presented in this article is a gross simplification of a real ecosystem, one can argue that it is too complex and therefore loses much of its explanatory power. We would like to argue that to study the complexity of a phenomena, such as species diversity and its relation to ecosystem properties, one needs to devise a model that allows for such complexity. An objection to such an approach could be that the dynamics of the model are so complicated that nothing is actually learnt from studying it. In one sense this is true, but with a artificial system one has the possibility to investigate any observed phenomenon in absolute detail. The dynamics can be broken down into finest level of detail and from there the phenomenon can be explained. There are so to speak no hidden variables or unknown influences in the system. The biological insights gained from studying other Artificial Life-platforms such as *Avida* is evidence of the success of such a complex modeling approach.

The experiments presented in this article opens up for many new questions and ideas that are beyond the scope of this article. One question is for example to what degree the entries of the competition matrix can explain the dynamics of the full model, that is, how well can the dynamics of the model be approximated by a system of differential equations, which takes its parameters from the competition matrix C . This would show how well pairwise species interactions can explain the observed diversity, or if one needs to use higher order interactions (triplets, quadruplets etc.) to explain the dynamics, as was suggested by the observed discrepancy between the invasion experiments and the full-fledged simulations. One could also investigate the dynamics from another point of view by making use of the metabolic history of all food strings (i.e., the list of species each string has been metabolized by). This makes it possible to map out which species engage in cross-feeding, and from this information generate a network of ecological interactions. Another possibility is to examine to which extent the process of evolution maximizes productivity from an ecosystem point of view, that is, how well does the evolved species composition do compared to an optimal species composition, which maximizes productivity (for a given flow rate). Further, the model could also be extended to include features present in real biological systems, such as a distinction between the genotype and phenotype of the organisms and a spatial dimension that would impact the nature of the species interactions.

LITERATURE CITED

- Adami, C. 1998. Introduction to Artificial Life. Springer, New York.
- . 2006. Digital genetics: unravelling the genetic basis of evolution. *Nat. Rev. Genet.* 7:109–118.
- Bronstein, J. 1994. Our current understanding of mutualism. *Q. Rev. Biol.* 69:31–51.
- Chow, S. S., C. O. Wilke, C. Ofria, R. E. Lenski, and C. Adami. 2004. Adaptive radiation from resource competition in digital organisms. *Science* 305:84–86.
- Costa, E., J. Perez, and J.-U. Krefl. 2006. Why is metabolic labour divided in nitrification? *Trends Microbiol.* 14:213–219.
- Czaran, T., R. Hoekstra, and L. Page. 2002. Chemical warfare between microbes promotes biodiversity. *Proc. Natl. Acad. Sci. USA* 99:786–790.
- Dejonghe, W., E. Berteloot, J. Goris, N. Boon, K. Crul, S. Maertens, M. Höfte, P. D. Vos, W. Verstraete, and E. Top. 2003. Synergistic degradation of linuron by a bacterial consortium and isolation of a single linuron-degrading variovorax strain. *Appl Environ Microbiol* 69:1532–1541.
- Deutsch, A., and S. Dormann. 2005. Cellular automaton modeling of biological pattern formation. Birkhäuser, Boston.
- Doebeli, M. 2002. A model for the evolutionary dynamics of cross-feeding polymorphisms in microorganisms. *Pop. Ecol.* 44:59–70.
- Fryer, G., and T. Iles. 1969. Alternative routes to evolutionary success as exhibited by african cichlid fishes of the genus *tilapia* and the species flocks of the great lakes. *Evolution* 23:359–369.
- Gardner, M. 1970. Mathematical games: the fantastic combinations of john conway's new solitaire game "life". *Scientific Am.* 223:120–123.
- Gerlee, P., and T. Lundh. 2008. The emergence of overlapping scale-free genetic architecture in digital organisms. *Artificial Life* 14:265–275.
- Gerlee, P., T. Lundh, B. Zhang, and A. R. A. Anderson. 2009. Gene divergence and pathway duplication in the metabolic network of yeast and digital organisms. *J R Soc Interface* 6:1233–1245.
- Grime, J. P. 1973. Competitive exclusion in herbaceous vegetation. *Nature* 242:344–347.
- Hardin, G. 1960. The competitive exclusion principle. *Science* 131:1292–1297.
- Helling, R. B., C. N. Vargas, and J. Adams. 1987. Evolution of *escherichia coli* during growth in a constant environment. *Genetics* 116:349–358.
- Hessler, R., and H. Sanders. 1967. Faunal diversity in the deep-sea. *Deep-Sea Research* 14:65–79.
- Huisman, J., and F. J. Weissing. 1999. Biodiversity of plankton by species oscillations and chaos. *Nature* 402:407–410.
- Hutchinson, G. 1959. Homage to *santa rosalia* or why are there so many kinds of animals? *Am. Nat.* 93:145–159.
- Kassen, R., A. Buckling, G. Bell, and P. B. Rainey. 2000. Diversity peaks at intermediate productivity in a laboratory microcosm. *Nature* 406:508–512.
- Katsuyama, C., S. Nakaoka, Y. Takeuchi, K. Tago, M. Hayatsu, and K. Kato. 2009. Complementary cooperation between two syntrophic bacteria in pesticide degradation. *J. Theor. Biol.* 256:644–654.
- Kerr, B., M. A. Riley, M. W. Feldman, and B. J. M. Bohannan. 2002. Local dispersal promotes biodiversity in a real-life game of rock-paper-scissors. *Nature* 418:171–174.
- Laird, R. A., and B. S. Schamp. 2009. Species coexistence, intransitivity, and topological variation in competitive tournaments. *J. Theor. Biol.* 256:90–95.
- Lenski, R., C. Ofria, T. Collier, and C. Adami. 1999. Genome complexity, robustness and genetic interactions in digital organisms. *Nature* 400:661–664.
- Lenski, R., C. Ofria, R. Pennock, and C. Adami. 2003. The evolutionary origin of complex features. *Nature* 423:139–144.

- Lestienne, R. 1998. The creative power of chance. Univ. of Illinois Press, Illinois.
- Lindgren, K. 2003. Information theory for complex systems (lecture notes). Chalmers.
- Odum, E. 1959. Fundamentals of Ecology. Saunders, Philadelphia.
- Ornstein, D. S., and B. Weiss. 1993. Entropy and data compression schemes. *IEEE Tran. Inform. Theory* 39:78–83.
- Ostrowski, E. A., C. Ofria, and R. E. Lenski. 2007. Ecological specialization and adaptive decay in digital organisms. *Am Nat* 169:E1–E20.
- Pelz, O., M. Tesar, R. Wittich, E. R. B. Moore, K. N. Timmis, and W. R. Abraham. 1999. Towards elucidation of microbial community metabolic pathways: unravelling the network of carbon sharing in a pollutant-degrading bacterial consortium by immunocapture and isotopic ratio mass spectrometry. *Environ. Microbiol.* 1:167–174.
- Pfeiffer, T., and S. Bonhoeffer. 2004. Evolution of cross-feeding in microbial populations. *Am Nat* 163:E126–E135.
- Rainey, P., and M. Travisano. 1998. Adaptive radiation in a heterogeneous environment. *Nature* 394:69–72.
- Rainey, P. B., A. Buckling, R. Kassen, and M. Travisano. 2000. The emergence and maintenance of diversity: insights from experimental bacterial populations. *Trends Ecol. Evol.* 15:243–247.
- Ray, T. 1992. An approach to the synthesis of life. *in* C. Langton, C. Taylor, J. Farmer, and S. Rasmussen, eds. *Artificial Life II*, Pp. 371–408. Addison-Wesley, Redwood City, CA.
- Rozen, D., and R. Lenski. 2000. Long-term experimental evolution in *Escherichia coli*. viii. dynamics of a balanced polymorphism. *Am. Nat.* 155:24–35.
- Saxer, G., M. Doebeli, and M. Travisano. 2009. Spatial structure leads to ecological breakdown and loss of diversity. *Proc. Biol. Sci. Lond. B* 276:2065–2070.
- Schluter, D. 2000. The ecology of adaptive radiation. Oxford Press, Oxford.
- Schrödinger, E. 1944. What is Life? Cambridge Univ. Press, Cambridge.
- Shannon, C. E. 1948. A mathematical theory of communication. *The Bell System Technical Journal* 27.
- Stams, A. J. 1994. Metabolic interactions between anaerobic bacteria in methanogenic environments. *Antonie Van Leeuwenhoek* 66:271–294.
- Tilman, D. 1982. Resource competition and community structure. Princeton Univ. Press, Princeton.
- Travisano, M., and P. B. Rainey. 2000. Studies of adaptive radiation using model microbial systems. *Am. Nat.* 156:S35–S44.
- Ulam, S. 1962. On some mathematical problems connected with patterns of growth of figures. *Essays on Cellular Automata*, Univ. of Illinois.
- Valentine, J. 1971. Resource supply and species diversity patterns. *Lethaia* 4:51–61.
- von Neumann, J., and A. Burks. 1966. Theory of self-reproducing automata. Univ. of Illinois Press, Champaign.
- Waide, R., M. Willig, C. Steiner, M. G., L. Gough, S. Dodson, G. Juday, and R. Parmenter. 1999. The relationship between productivity and species richness. *Annu. Rev. Ecol. Syst.* 30:257–300.
- White, A., P. Handler, and S. E. L. 1973. Principles of biochemistry. 5th ed. New York, McGraw-Hill.
- Wilke, C., and C. Adami. 2002. The biology of digital organisms. *Trends Ecol. Evol.* 17:528–532.
- Wolfram, S. 1983. Statistical mechanics of cellular automata. *Rev. Modern Phys.* 55:601–644.
- Yedid, G., C. Ofria, and R. Lenski. 2009. Selective press extinctions, but not random pulse extinctions, cause delayed ecological recovery in communities of digital organisms. *Am Nat* 173:E139–E154.

Associate Editor: M. Doebeli

Appendix A

CELLULAR AUTOMATA

Cellular automata (CA) are a versatile tool in the field of Artificial Life and in biological modeling in general, mostly used for modeling spatiotemporal phenomena such as pattern formation (Deutsch and Dormann 2005) and interacting particle systems, for example, artificial chemistries (Adami 1998). We aim to take advantage of this common knowledge of CA in the proposed platform.

The history of CA goes back to the work of Ulam (1962) and von Neumann and Burks (1966), and in the 70s, Conway's Game of Life (Gardner 1970) which gave rise to much activity as it with simple rules could display life-like behavior such as self-reproduction and propagation. The study of CA was further extended by Wolfram (1983), who analyzed and classified the behavior of the simplest type of CA, namely one-dimensional, two-state, nearest neighbor rules. The main conclusion of that endeavor was that simple local rules of these CA can give rise to large-scale complex behavior. These are the type CA-rules we have employed in *Urdar*.

Cellular automata are dynamical systems discrete in both time and space. In the case of one-dimensional elementary CA, we can think of them as rules for changing the states of cells in a one-dimensional array having two possible states (typically 1 and 0). Each cell in such a configuration is updated in parallel as a function of its own current state and its two neighboring cells on the left and right (Fig. A1). This means that the CA-rule determines the state of each cell at time $t + 1$ as a function of triplets of cells at t . As we have two states per cell, this means that there are $2^3 = 8$ different triplets (Fig. A1), and each of these can map the middle cell to either 1 or 0, giving a total of $2^8 = 256$ different elementary CA-rules. Each of these are present in the simulation, and represent a species with a unique way of manipulating/metabolizing the binary food strings.

Appendix B

ENTROPY

The degree of disorder is best characterized by the information-theoretic entropy introduced by Shannon (1948). On a side note, it is said that von Neumann not only invented CAs, but at a seminar suggested to Shannon that he should call his new measure entropy because “no one really understands [its meaning], by using this word you will have an advantage over your adversary in any debate” (p. 105, Lestienne 1998). This entropy measures the degree of irregularity in a given structure, and does so by comparing the frequency of different substructures. In our case of binary strings, this is done by measuring the frequency of substrings of different lengths, and by then calculating how information about

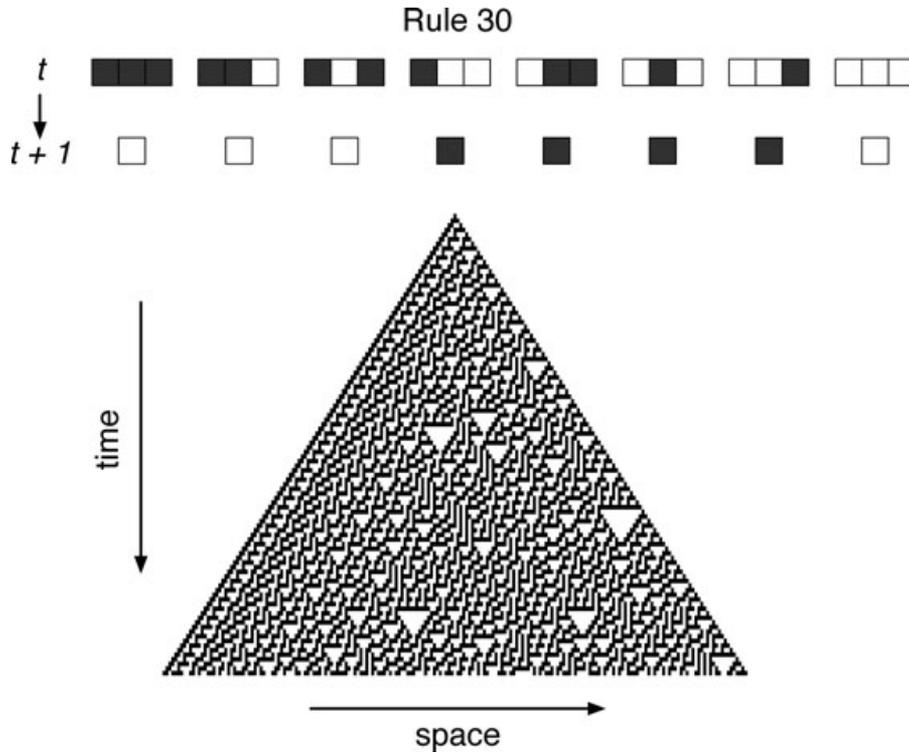


Figure A1. Rule table and realization of elementary CA-rule 130. The rule table shows how the CA-rule transforms the binary string by specifying the value of each bit at $t + 1$ as a function of triplets of bits at time t . The bottom panel shows multiple iterations of rule 30, when the initial condition (top row) consists of a single 1 surrounded by 0's. This would be the outcome of a single agent of species 30 repeatedly digesting the same binary string.

the content of the string increases as longer substrings are taken into account. In a fully regular string with alternating 1's and 0's ($r_1 = \dots 01010101 \dots$) no information is gained when looking at substrings longer than 2, that is, the structure is completely regular above this length scale, and this implies that the Shannon entropy of the string is $s = 0$ bits. On the other hand, for a completely random string, where each element is 1 or 0 with probability 1/2, no predictions about the structure can be made no matter how long substrings we consider, and this implies that the entropy is maximal and equal to $s = \log_2 2 = 1$ bit. The Shannon entropy can also be interpreted as the minimal average bits per symbol needed to transmit the “message” contained in the string.

More formally the Shannon entropy s is measured by calculating the block entropy of size m

$$S_m = \sum_{\sigma_m} p(\sigma_m) \log_2 \frac{1}{p(\sigma_m)}, \tag{A1}$$

where the sum is over all possible substrings σ_m of length m , and p the probability of finding the substring σ_m in the string. By then taking the difference $\Delta S_m = S_{m+1} - S_m$, the Shannon entropy is defined as the limit

$$s = \lim_{m \rightarrow \infty} \Delta S_m. \tag{A2}$$

This definition is not applicable in practice as it requires infinitely long strings, and therefore one has to approximate the true entropy with ΔS_m for some m . We have for computational reasons chosen $m = 2$, which still gives a reasonable approximation of the disorder in the binary strings. We are also using limited sized blocks for the entropy, not only for computational convenience, but also to avoid the well-known fact that when the block size m goes to infinity, no simple CA can increase the entropy of such an infinite binary string (Lindgren 2003).