Emergence of species and punctuated equilibrium in the Tangle Nature model of biological evolution

Henrik Jeldtoft Jensen

Department of Mathematics, Imperial College London, South Kensington Campus,
London SW7 2AZ, UK

Abstract

We consider an individual-based model of evolution. Species emergence as segregation of the population on to discrete positions in a hyper-cubical genotype space. The model emphasises interaction between coexisting genotypes. Although the rate of the micro-dynamics, consisting of reproduction and killing of individuals is essentially constant, the macro-dynamics is punctuated. The macro-evolution consists of the creation and annihilation of configurations in genotype space equivalent to sequences of different ecosystems. We review the definition of the model and results concerning distribution lifetimes of individuals, genotypes and configurations. We discuss the non-stationarity of the model and how this agrees with the fossil record. Finally, we describe how the model is able to produce species abundance distributions of a form qualitatively in agreement with observation on real ecosystems.

1. Introduction

Per Bak was a great creator of metaphors and a great source of inspiration. For example, the sandpile metaphor of self-organised criticality has reached people far beyond the research community of physics. The simple model of co-evolution published

by Bak and Sneppen [1] in 1993 has inspired great numbers of physicist and, even more impressive, biologists. [1] I believe that this paper informed many people working in the field of statistical mechanics and complex systems that biological evolution contains some open theoretical questions which can be addressed by the modelling techniques used in statistical mechanics. It certainly was the Bak–Sneppen paper that brought to my attention issues such as Punctuated Equilibrium, the Red Queen effect and the distribution of lifetimes of taxonomic structures. The Tangle Nature model, reviewed here, was developed as a response. The work described below has progressed during the last 6 years in collaboration with P. Anderson, K. Christensen, S.A. di Collombiano and M. Hall. Per Bak was a person with distinct and articulated views, he was kind to let us know that he liked very much the Tangled Nature model and found the approach to be highly worthwhile. It is in this sense that the following review is to be read as an example of the impact of Per Bak's scientific work.

2. The tangled nature model

The philosophy of the Tangled Nature model is well described by Charles Darwin's wonderful closing words in "The Origin of Species":

> *It is interesting to contemplate an entangled bank, clothed with many plants of many kinds, with birds singing on the bushes, with various insects flitting about, and with worms crawling through the damp earth, and to reflect that these elaborately constructed forms, so different from each other, and dependent on each other in so complex a manner, have all been produced by laws acting around us* [4].

The Tangled Nature model of evolutionary ecology [5–8] is an attempt to model the evolution of the highly interconnected and interdependent net comprising ecosystems. In this respect the model is of the same spirit as the Bak–Sneppen model [1], however, the Bak–Sneppen model is a much more abstract model. The Bak–Sneppen model is defined at the level of species. Each species is described by a single randomly assigned fitness parameter and the individual species are created and annihilated while the total number of species remain fixed. The Tangled Nature model is an individual-based model and attempts to focus on the understanding of the laws emerging as one proceeds from the level of micro-dynamics, consisting of mutation-prone reproduction and death involving individuals, to the level of macro-dynamics, consisting of the creation and extinction of species and higher taxon.

The model investigates frequency-dependent selection in a population of individuals, or organisms, characterised by their position in genotype space. The population segregates into a set of distinct types. While the dynamics at the level of individuals is ticking along at a constant rate, the population dynamics at the level of "species",

---

[1] See e.g. Campbell's monumental introductory text [2] where he refers to the review by Bak et al. [3] as further reading to the Chapter *Tracing Phylogeny: Macroevolution, The Fossil Record, and Systematics*.

or more correctly genotypes, is strongly intermittent [5,6]. Long periods of quiescence are followed by hectic periods of activity. We interpret the quiet periods as established ecosystems and call these quasi-evolutionary stable strategies (q-ESS) with a reference to Maynard Smith's game theoretic concept: Evolutionary Stable Strategies [9]. The lifetime distribution at different "taxonomic levels" are power law like, with exponents consistent with the fossil record [5]. The model is for relevant sizes of the genomes always in a non-stationary state. Adaptation occurs in the model at a collective level and consists of a gradual (on average) increase in the stability of successive ecosystems (the quasi-stable configurations in genotype space) [6]. Depending on the connectivity of the interactions in genotype space the dynamics of the model is able to generate species abundances distributions which compares well with typical observations [8].

3. Model definition

The model is easily specified mathematically. Consider a time-dependent number of individuals labelled by $\alpha = 1, 2, \ldots, N(t)$. Each individual is characterised by which corner of the L-dimensional hypercube it belongs to, i.e., by the vector $\mathbf{S}^{\alpha} = (S_1^{\alpha}, S_2^{\alpha}, \ldots, S_L^{\alpha})$ with $S_i^{\alpha} = \pm 1$. In each time step, an individual is chosen at random and removed from the system with the fixed probability p_{kill} equal for all individuals. During each time step, an individual α may reproduce with the time and type-dependent probability $p_{off}/N(t)$ give by $p_{off}(\mathbf{S}^{\alpha}, t) = 1/(1 + \exp^{-1}[H(\mathbf{S}^{\alpha}, t)])$, where $H(\mathbf{S}^{\alpha}, t)$ is a weight function calculated in the following way:

$$H(\mathbf{S}^{\alpha}, t) = \frac{1}{cN(t)} \sum_{\mathbf{S} \neq \mathbf{S}^{\alpha}} J(\mathbf{S}^{\alpha}, \mathbf{S}) n(\mathbf{S}, t) - \mu N(t) \ . \tag{1}$$

Here μ determines the carrying capacity and $n(\mathbf{S}, t)$ denotes the number of individuals of genotype $\mathbf{S}$. A genotype $\mathbf{S}^{\alpha}$ is with probability Θ coupled to another genotype $\mathbf{S}$. The coupling $J(\mathbf{S}^{\alpha}, \mathbf{S}) \in [-1, 1]$ is asymmetric. The parameter c is a constant that effectively scales the width of the distribution of the couplings. We will discuss the couplings below. Though both sexual and asexual reproduction can be defined in a natural way we will here concentrate on asexual reproduction. In this case, $p_{off}(\mathbf{S}^{\alpha}, t)$ determines the probability for the following action: first replace individual α by two copies $\mathbf{S}^{\beta} = \mathbf{S}^{\alpha}$ and $\mathbf{S}^{\gamma} = \mathbf{S}^{\alpha}$, then remove the original $\mathbf{S}^{\alpha}$. After the faithful copies have been produced mutations are introduced by changing the sign of the components $i = 1, 2, \ldots, L$ of the offspring $S_i^{\delta} \mapsto -S_i^{\delta}$, for $\delta = \beta, \gamma$, with probability p_{mut}. The total probability that a mutation occurs in one of the two offspring is accordingly $2L p_{mut}$. Similar models have be considered e.g. in Refs. [10–12].

Though several interpretations are possible, the model was developed to contribute to bridging the gap between evolution at the level of individuals and at higher ecological levels. From this perspective the vector $\mathbf{S}^{\alpha}$ is to be interpreted as the genotype or even the "genome" of individual α. The individual components of the vector can be thought of in several ways. In the most microscopic interpretation S_i^{α} labels the nucleotides (though, of course, we should allow S_i^{α} to assume four values in that case). At a more coarse grained level S_i^{α} could correspond to different alleles of a gene (again more than

two values would in general be needed) or we can think of S_i^α representing a group of genes.

There are 2^L different possible genotypes in the model, some of these corresponds to sequences of genes or nucleotides which do not correspond to any viable organism. By far the majority of randomly sequenced combinations of genes will not produce a genome that can support a living organism. We focus on co-evolution and eco-logical interaction between organisms. Accordingly we neglect self-interactions, i.e., $J(\mathbf{S}^\alpha, \mathbf{S}^\alpha) = 0$. This makes it difficult for organisms to live/reproduce in isolation. The coupling $J(\mathbf{S}^\alpha, \mathbf{S}^\beta)$ represents, in the widest sense, interactions between organisms. Trophic relations are only one, though important, possibility. The J-coupling between two organisms can also represent much more indirect effects, e.g. the production of a nutrient by β which is consumed by α—something like the relationship between a tree and me: the tree produces oxygen and I breathe it. We cannot determine in detail all possible couplings between all possible types of organisms, nor are we interested in the specific details, hence we simply draw the coupling strengths $J(\mathbf{S}^\alpha, \mathbf{S}^\beta)$ from a convenient distribution. I think an analogy with the periodic table of elements is helpful. Let us consider all the different reactions (or interactions) between a certain element, say Selenium, and the other elements taken one by one: Hydrogen, Helium,... Unnihexium, and even more as we include theoretically calculate reactions with the as yet not produced higher elements. These couplings, or reactions, vary in nature as we scan through the periodic table. Of course, this is all very systematic and deterministic, nevertheless the overall picture is one of a fair range of possibilities, which in a rough way may be represented by a random collection of possibilities.

It is important to stress that though the couplings $J(\mathbf{S}^\alpha, \mathbf{S}^\beta)$ are assigned at random they are not changing in time. The effect of a type $\mathbf{S}^\beta$ organism on a type $\mathbf{S}^\alpha$ organism is *always* represented by the same number $J(\mathbf{S}^\alpha, \mathbf{S}^\beta)$. This obviously does not imply that the viability of an organism of type $\mathbf{S}^\alpha$ will be time independent. The interesting question in the Tangled Nature model is how the "fitness" or viability of a given type of organisms changes with the overall configuration in genotype space. Although the couplings between different positions in genotype space is "hardwired" from the onset, this does not mean we have predetermined the type of genotype configurations that will evolve. On the contrary, amongst all the possible ways of distributing a population in genotype space, the micro-dynamics of reproduction, mutation and annihilation of individuals will have to find ways to occupy sets of genotype positions which form networks of couplings that can produce a balance between killing and reproduction, i.e., configurations consisting of genotype positions which, at least intermittently, allow the balance $p(\mathbf{S}^\alpha, t) \simeq p_{kill}$. We will see that this can typically be done in many different, more or less adequate, ways.

4. Speciation, intermittency and ecosystems

We now describe briefly how macroscopic structures emerge from the microscopic dynamics, much more detail can be found in Refs. [5–8]. Fig. 2 depicts the occupancy in genotype space (size $2^L \simeq 10^6$ for $L = 20$) during the first 5000 time step and
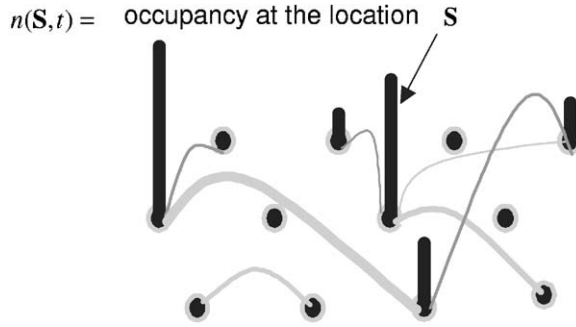
Fig. 1. Sketch of genotype space. The columns indicate the size of the population of a particular genotype. The lines connecting genotype positions indicate the variety of strengths of the couplings $J(\mathbf{S}^{\alpha}, \mathbf{S}^{\beta})$. The position indicated by the arrow is linked to three other positions of which only two are occupied.

the corresponding variation in the population size and diversity. The entire population was initially placed at one particular position $\mathbf{S}_0$. All the positions in genotype space have been labelled in an arbitrary way up along the y-axis and a dot is placed at the label of each occupied position at each time step. For the first 2400 time steps the entire population remains concentrated at $\mathbf{S}_0$. This happens because the size of the population is too big: in Eq. (1) the only non-zero term is the $-\mu N(t)$ (since $n(\mathbf{S}) = 0$ for $\mathbf{S} \neq \mathbf{S}_0$) and therefore $p_{off} \simeq 0$. As the effect of p_{kill} gradually reduces $N(t)$, the reproduction probability $p_{off}(\mathbf{S}_0, t)$ gradually increases and reproduction kicks in. When this happens, mutations occur leading to other genotype positions becoming occupied, hence the many parallel lines in Fig. 1 after time step 2400. The increased diversity allows the first term in Eq. (1) to become non-zero, this can increase p_{off} further leading to a growth in the population size as seen in Fig. 2. It is possible to identify events of speciation in the model. In Fig. 2 the original genotype splits
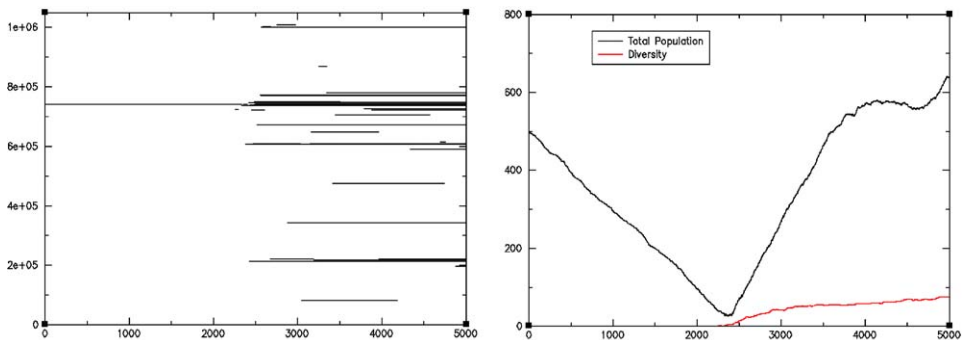

Fig. 2. The left frame shows the occupancy in genotype space during the first few thousand time steps. A dot is placed along the y-axis at the label of occupied genotypes. The total number of possible genotypes is about 10^6. The right frame contains the population size and the diversity as a function of time. Diversity is here defined as the number of occupied positions in genotype space.
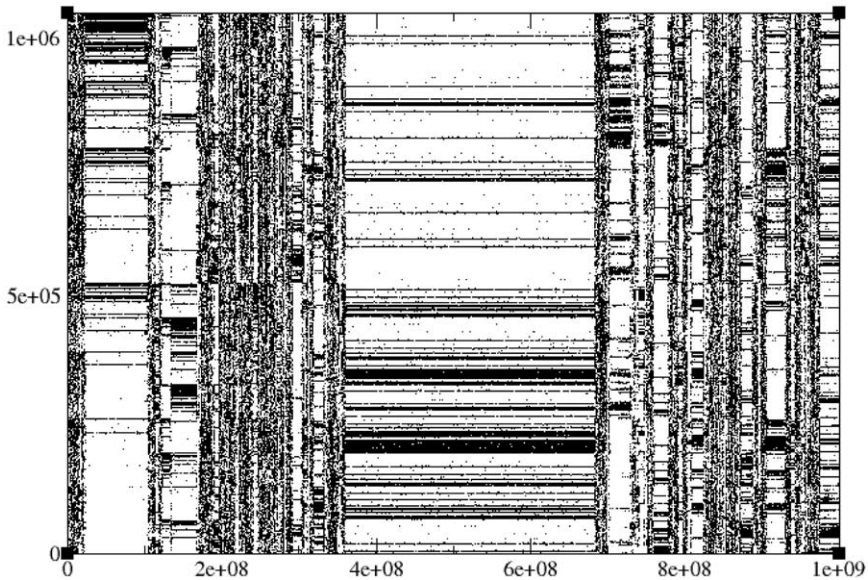
Fig. 3. The continuous micro-dynamics gives rise to an intermittent macro-dynamics consisting of long periods where the occupancy in genotype space changes little separated by brief hectic periods of extinction and creation. The number of time steps (10^9 in total) is along the *x*-axis.

into several other genotypes represented by the parallel lines emerging after time step 2400. Equally interesting is it that only a relatively small number (though much larger than one) of distinct genotypes are present simultaneously. The population does *not* smear out in a diffusive way across genotype space, instead the dynamics is, as seen in Fig. 3 intermittent. During the long q-ESS periods the occupancy is essentially concentrated on a small set of positions. Short line segments in Fig. 3 indicate that other types may pop up for shorter intervals during a q-ESS but there is clearly a dominant set of persistent types. A particular q-ESS comes very abruptly to an end when extinction events cascades through the system in the cause of a brief time period of hectic annihilation and creation. These hectic periods appear as very sharp regions in Fig. 3.

The quiet q-ESS periods can be thought of as particular configurations of the ecosystem. We can investigate the "species abundance distribution" (SAD) for these ecosystems, in our case this is simply the abundances of different types. The SAD are found to evolve slowly with time [8]. The functional form of the emerging SAD depends on the level of connectedness in genotype space. Except for very small values of the connectivity parameter Θ we find the SAD to evolve towards an approximate log-normal form [5,8] similar to what is frequently observed in real systems.

As successive ecosystems, or q-ESS periods, replace each other the distribution of coupling strengths between occupied genotypes become skewed towards positive
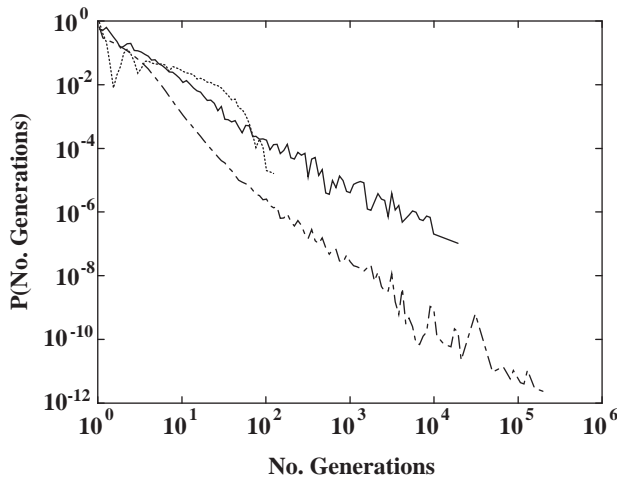
Fig. 4. Log–log plot of the distributions of durations of q-ESS (solid line), transition periods (dotted line) and lifetimes of occupied locations in genome space (dot–dashed line). We observe power-law-like behaviour in both the q-ESS and the lifetimes of genome space locations, but the transition periods exhibit an abrupt cut-off at much shorter times. The curve representing the distribution of lifetimes of individual genotype positions extends further than that of the q-ESS, indicating that locations may remain occupied from one q-ESS to another, surviving the transition.

couplings [8]. It is also very interesting to mention that the system is in a non-stationary state as indicated by a decrease in the average number of transitions (per time unit) between q-ESS [6]. Accordingly, the average extinction rate decreases gradually with time, a phenomena reported to be observed in the fossil record [13].

The distribution of durations is shown in Fig. 4 to be power law like with exponents roughly around -2 in agreement with observations on the fossil record [13]. The effect of increasing the mutation level can be studied [7] and an error threshold can be defined and the model related to Eigen's quasi-species model [10].

The power spectrum of the long-time dynamics have been investigated in a slightly simplified version of the Tangled Nature model by Rikvold and Zia [14]. The authors found that the extinction dynamics is characterised by approximate $1/f$ power spectra. It was suggested by Bak and collaborators [15] that the fossil record exhibit this behaviour. Although, the suggestion by Bak et al. was strongly challenged by Kirchner and Weil [16] on the basis of the incompleteness of the fossil record analysed by Bak et al. it is still of interest to develop an understanding of what to expect from theoretical models.

At the moment, we investigate the nature of the transitions between different q-ESS. We investigate the effect of including a limiting resource and will include spatial aspects in the Tangled Nature model. We expect to be able to make comparison with experimental observations on microbial communities in soil.

Acknowledgements

References

[1] P. Bak, K. Sneppen, Phys. Rev. Lett. 71 (1993) 4083.

[2] N.A. Campbell, Biology, The Benjamin/Cummings Publishing Company, Princeton, 1996.

[3] P. Bak, H. Flyvbjerg, K. Sneppen, New Sci. 71 (1994) 4083.

[4] C. Darwin, The Origin of Species by Means of Natural Selection, John Murray, more recent edition by Penguing Books 1968, London, 1859.

[5] K. Christensen, S.A. di Collobiano, M. Hall, H.J. Jensen, J. Theor. Biol. 216 (2002) 73.

[6] M. Hall, K. Christensen, S.A. Collombiano, H.J. Jensen, Phys. Rev. E 66 (2002) 011904.

[7] S.A. di Collobiano, K. Christensen, H.J. Jensen, J. Phys. A 36 (2003) 883.

[8] P. Anderson, H.J. Jensen, arXiv.org/cond-mat/030714, preprint, J. Theor. Biol., accepted for publication.

[9] J.M. Smith, Evolution and the Theory of Games, Cambridge University Press, Cambridge, 1982.

[10] M. Eigen, J. McCaskill, P. Schuster, J. Phys. Chem. 92 (1988) 6881.

[11] P.G. Higgs, B. Derrida, J. Mol. Evol. 35 (1992) 454.

[12] S. Kauffman, The Origins of Order, Oxford University Press, Oxford, 1993.

[13] M.E.J. Newman, P. Sibani, Proc. R. Soc. London B 266 (1999) 1593.

[14] P.A. Rikvold, R.K.P. Zia, Phys. Rev. E 68 (2003) 031913.

[15] R.V. Solé, S.C. Manrubia, M. Benton, P. Bak, Nature 388 (1997) 764.

[16] J.W. Kirchner, A. Weil, Nature 395 (1998) 337.