# Evolution of species from Darwin theory: A simple model

M.A. Moret [a,b], H.B.B. Pereira [a,c,*], S.L. Monteiro [d], A.C. Galeão [d]

[a] Programa de Modelagem Computacional - SENAI - CIMATEC, 41650-010 Salvador, Bahia, Brazil
[b] Departamento de Física - UEFS 44031-460 Feira de Santana, Bahia, Brazil
[c] Departamento de Ciências Exatas - UEFS 44031-460 Feira de Santana, Bahia, Brazil
[d] Laboratório Nacional de Computação Científica, Petrópolis, RJ, Brazil

## ARTICLE INFO

## ABSTRACT

Evolution of species is a complex phenomenon. Some theoretical models take into account evolution of species, like the Bak–Sneppen model that obtain punctuated equilibrium from self-organized criticality and the Penna model for biological aging that consists in a bit-string model subjected to aging, reproduction and death. In this work we propose a simple model to study different scenarios used to simulate the evolution of species. This model is based on Darwin's ideas of evolution. The present findings show that punctuated equilibria and stasis seem to be obtained directly from the mutation, selection of parents and the genetic crossover, and are very close to the fossil data analysis.

© 2011 Elsevier B.V. Open access under the Elsevier OA license.

Biological evolution takes place in terms of an intermittent burst of activity separating relatively long periods of quiescence, rather than in a gradual manner [1,2]. Both theoretical and empirical studies of the last decades suggest that the complex pattern of selection imposed on geographically structured populations by heterogeneous environments and coevolution can paradoxically maintain stasis at the species level over long periods of time. By contrast, genetic mechanisms are not known to explain species-wide stasis [3–6]. In this sense, an established concept of biological evolution is the order–disorder conflict in biological evolution route to equilibrium. The Bak–Sneppen (BS) model [7] proposes the biological evolution as a self-organized criticality system [7–9]. On the other hand, the Penna model for biological aging [10–13] is based on Darwinian evolution with mutations and it is a representation of the Darwinian conflict. It has played a role similar to the Ising model for magnetic systems in the sense that it is a minimal model that retains the essentials of Darwinian dynamics. Like the Ising model, the Penna model uses binary variables to represent genes: 0 for ordinary genes and 1 for harmful ones. Originally, the Penna model focused on problems of biological aging. Furthermore, some applications to several different evolutionary problems substantially increased the scope of the Penna model.

Our purpose here is to provide a different procedure to simulate the evolution of species based on Darwinian evolution, using mutation, reproduction and crossover processes. The proposed model takes into account the "survival of the fittest" and some ideas present in Genetic Algorithms (GA). Two main differences can be observed in our model. The first one is the fact that we simulate biological evolution of species by using an artificial adaptiveness genetic code and we impose the selection of parent's genes from the parent's fitness (a mechanism like a natural selection). And, the second one is the fact that we do not use a binary representation of the genes. Instead we use a very simple taxonomy to simulate gene evolution. A commensurable taxonomy to evolution of species can be obtained from Mahalanobis distances [14]. Recently, temporal and environmental (spatial) patterns of morphological change in two species of Middle Devonian brachiopods [3] was shown. Eldredge et al. [3] found Mahalanobis values from canonical discriminant analysis of morphometric data show that this distance ranges to values close to the interval [0, 6].

---

* Corresponding author at: Programa de Modelagem Computacional - SENAI - CIMATEC, 41650-010 Salvador, Bahia, Brazil.
E-mail addresses: mamoret@gmail.com (M.A. Moret), hbbpereira@gmail.com (H.B.B. Pereira).

## 1. Description of the proposed model

Evolution of species is a complex phenomenon of biological interest. The simple procedure proposed suggests to us that species have different evolutionary stages and stasis is observed if the mutation process has a very low probability of happening.

In this paper we use some ideas based on Genetic Algorithms (GA's). Ref. [15] has proposed the GA [16] that is a stochastic procedure of search. This procedure is based on Darwin's ideas (i.e. it is based on mechanisms of natural selection and the response of the genetic code). The use of a codified population of points allows the GA to employ the ideas of adaptation and "crossover", which are essential in the biological processes. This type of algorithm normally uses a chain of bits to represent each individual of the population. Classical GA uses chains of bits in all its processes (i.e. mutation, reproduction and crossover). In the mutation process, some bits, randomly chosen, can be modified in all the individuals. This chain of bits emulates the individual's chromosome of the population in all GA. In the reproduction process, two individuals are successively chosen for crossover until a new generation in the population is formed.

The procedure we are proposing is an iterative algorithm. The population is modified at each iteration and a new generation is determined from this iterative process. It is interesting to comment that individuals of a generation $t + 1$ are not necessarily "children" of the previous one ($t$). From Darwin's ideas, it is desirable that individuals of the population have different degrees of adaptation to the environment. Then, we propose an algorithm based on Darwinian ideas to simulate the evolution of species and we use a chain of values to represent the chromosome of each individual of the population. As commented by Holland [3], two simple organism present Mahalanobis distances in values close to [0, 6]. Thus, the values from 0 to 6 are used to specify the level of fitness, i.e. the evolutive genetic value in the "chromosome" of each individual.

In other words, we propose as chromosome a set of 32 alleles formed by the numbers (from 0 to 6) associated with each level of fitness. These values in this set range from the value zero that represents a totally unfit allele to six that represents complete fitness. We recall that if we use two bits (0 or 1) to represent unfitness (0) and fitness (1) the results follow the same behavior, but this adapted taxonomy allows a marked gap between individuals that present worse and better fitness.

In order to simulate the "survival of the fittest", essential in the Darwinian proposal, we define the fitness probability ($P_i$) of an individual ($i$) in any generation as:

$$P_i = \frac{Q_i}{\sum\limits_{j=1}^{M} Q_j} \tag{1}$$

where $M$ is the number of individuals, $Q_k = \sum_{i=1}^{N_a} v_i$ is the chromosome quality, $N_a$ is the number of alleles and $v_i$ is the value of each allele.

Under these assumptions, our simple genetic procedure can be summarized as follows:

(1) *Initial population*: A set of $M$ individuals is randomly generated to initiate the simulation;
(2) *Mutation process*: With predefined probability of mutation ($P_{Mut}$) randomly choose which individuals, $x_i(t)$ ($i = 1, \ldots, M$), will participate in the mutation process, then choose which genes will be changed;
(3) *Reproduction process*: Calculate the fitness probability (reproduction) for each individual and select a pair of individuals, $k$ ($P_k$) and $l$ ($P_l$). The choice of pair $k$ and $l$ is due to the probability of how close they are to the greatest value calculated in this generation, as previously proposed by Moret [17,18];
(4) *"Crossover" process*: Mix randomly the chains of "bits" of $k$ and $l$, in order to produce two new individual candidates $k^*$ and $l^*$. These two candidates will compete among themselves to be born;
(5) Repeat the processes of reproduction and crossover until $p = P_{Cross} \times M$ (where $P_{Cross}$ is a prescribed percent of $M$) individuals be obtained. $M - p$ individuals will be maintained live in the next generation. We randomly choose these $M - p$ individuals from their fitness probability obtained in the Step 3 (previous generation);
(6) Go back to the Step 2.

The mutation process is a stochastic process over the chromosome. Thus, mutation in the gene is a random process in a specific allele. Its integer number is modified to a randomly selected value in the interval [0, 6]. The reproduction process lets more fit individuals to take part in the crossover. And, "crossover" procedure guarantees the parent's genes are transferred to the next generation. Fig. 1 shows the crossover process used in the evolutive algorithm.

## 2. Scenarios, simulations and discussions

It is highlighted here that, to simulate the genetic evolution process in accordance to the Darwinian perspective, two important aspects must be taken into account: (i) an individual "parent" transmits its genetic characteristics and (ii) the measurement of the fitness consists of the sum of the chromosome "alleles".

We have simulated different scenarios. In these simulations, for each initial population, the procedure was executed using some different initial parameters (32 alleles randomly obtained; alleles varying from 0 to 6; and population from 2 to 500 individuals). It was observed that when the number of individuals is greater than 150 there is no significant changing in the evolution behavior, showing non-dependence on the initial population to reach the "final" generation. For this reason,
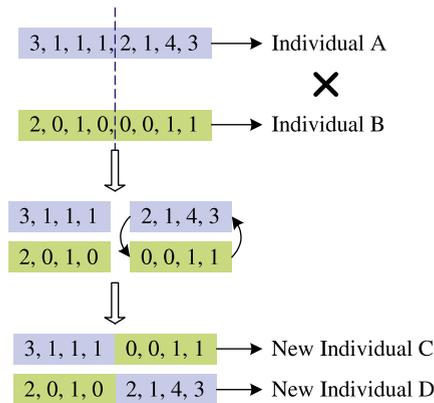
**Fig. 1.** Crossover process where the alleles are given by $v_i = 0, 1, \ldots, 6$. We recall that the new individual will be $D$ $(Q_D > Q_C)$.
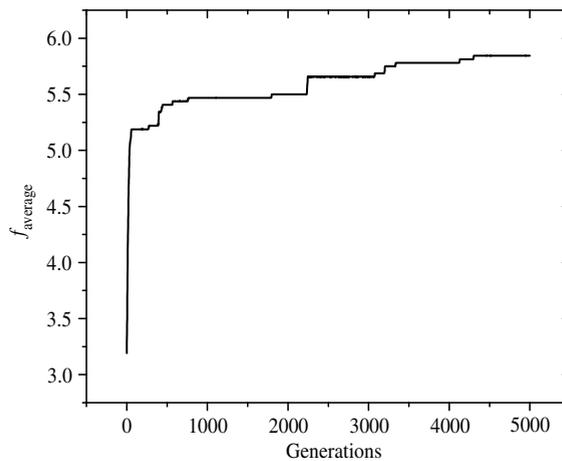


**Fig. 2.** Evolution of individual fitness ($f_{average}$) as function of number of generations for the average fitness to $P_{Cross} = 60\%$ of Crossover and $P_{Mut} = 0.005\%$ of mutation.

all results that will be presented were obtained from a population of 500 individuals. In order to follow the evolution process we propose as fitness:

$$f_i = \frac{Q_i}{N_a}. \tag{2}$$

On the extreme situations, for example, $P_{Cross} = 1\%$ $P_{Mut} = 1\%$ or on the other one, $P_{Cross} = 100\%$ and $P_{Mut} = 10\%$ the final population does not present stationary equilibria.

To some intermediate values of $P_{Cross}$ local equilibria can be obtained. Fig. 2 depicts the evolution of value of fitness when the value of $P_{Cross} = 60\%$ and $P_{Mut} = 0.005\%$ are kept fixed.

Fig. 2 shows that punctuated equilibrium occurs in different time of evolution of species. This behavior occurs if the number of mutation is small. Thus, a very small percentage of individuals is involved in genetical improvement of the population as whole. We recall that Fig. 2 has the same shape observed theoretically [7] or by "lineage through time" [23,24]. Thus, Fig. 2 presents exactly the same behavior observed by Reznick and Ricklefs [24] in their recent review entitled "Darwin's bridge between microevolution and macroevolution".

Several studies have confirmed the widespread occurrence of lineage stasis and punctuation by cladogenesis in the fossil record [1,2,19–21]. The stable lineages detected in fossil data sets appear to be, for the most part, species that were abundant, widespread and participants in many local ecosystems.

Recently the hypothesis that the power law of the lifetime distribution could be simply a consequence of the branching-like structure of single population dynamics rather than an effect of the interactions among different species [22] was suggested. Our proposal is analyze the adaptation and speciation of evolution of specie. We note that the rate of fitness of new lineages is initially high and then levels off in last steps of evolution. From the proposed approach stasis and punctuations are observed in the evolution (Fig. 2) by using only the Darwinian perspective. It is worthwhile to connect our approach with a model for a phylogenetic tree for North American wood warblers based on more than 9 kb of mitochondrial and nuclear

intron DNA sequence [23,24]. The "lineage through time" plot [24] derived from the phylogeny presents the same shape that we obtain (Fig. 2).

## 3. Concluding remarks

In summary, we proposed a simple model to analyze the evolution of species. Despite of its simplicity this model leads to consistent evolution processes. We observed that an evolution process just based on mutation will be a random-like process. On the other hand, an evolution process for which only the fitness and crossover take place will lead to a final generation with genetic material of the individuals present in the initial population. For these extreme situations, punctuated equilibria are improbable to take place. Therefore, a competition between mutation, reproduction and crossover must exist as long as a real evolution of species is focused. The proposed model attains these requirements. Furthermore, the mutation process seems to be the most important feature to differentiate species. We recall that the probability of a mutation process must be very small. Thus, in our simulations with 32 alleles and population with 500 individuals, only one or two alleles mutate in each generation in order for stasis to occur; consequently, punctuated equilibria must occur in different times of evolution.

History of individual lineages revels prolonged intervals with little or no change (equilibrium or stasis) interspersed with intervals of rapid change (punctuation) that are associated with origin of new species [1]. Our results show that stasis and punctuation can be obtained from natural selection, because selection of pairs $i, j$ (reproduction process) for crossover depend on the fitness of $(f_i, f_j)$. If individuals increase their fitness, due the mutation process, they are responsible for the punctuations in different times of evolution. On the other hand, the fossil records reveal that some lineages rapidly diversify into new species whereas others decline. Combining the proposed evolution model with the Penna model [10] we can simulate the reason some species decline.

Finally, we observe that existence of small mutations, selection of pairs and crossover seem to be the keys that cause species-wide stasis. Mutations can make the entire population improves itself, even if these mutations increase the fitness of a small number of individuals. Furthermore, the self organized criticality (SOC) behavior observed in BS model [7] seems to be due the small number of individuals better fitted in the population. This small number of individuals behaves like an attractor to the evolution process. Thus, these results seem to represent a phylogenetically refined SOC.

## Acknowledgments

## References

[1] S.J. Gould, N. Eldredge, Paleobiology 3 (1977) 115.
[2] S.J. Gould, N. Eldredge, Nature 366 (1993) 223.
[3] N. Eldredge, J.N. Thompson, P.M. Brakefield, S. Gavrilets, D. Jablonski, J.B.C. Jackson, R.E. Lenski, B.S. Lieberman, M.A. McPeek, W. Miller, Paleobiology 31 (2005) 133.
[4] B.S. Lieberman, E.S. Vrba, Paleobiology 31 (2005) 113.
[5] G. Hunt, Proc. Natl. Acad. Sci. USA 104 (2007) 18404.
[6] L.J. Revell, L.J. Harmon, D.C. Collar, Syst. Biol. 57 (2008) 591.
[7] P. Bak, K. Sneppen, Phys. Rev. Lett. 71 (1993) 4083.
[8] M. Paczuski, S. Maslov, P. Bak, Phys. Rev. E 53 (1996) 414.
[9] S. Maslov, Phys. Rev. Lett. 77 (1996) 1182.
[10] T.J.P. Penna, J. Stat. Phys. 78 (1995) 1629.
[11] T.J.P. Penna, S. Moss de Oliveira, D. Stauffer, Phys. Rev. E 52 (1995) R3309.
[12] D. Stauffer, P.M.C. De Oliviera, S.M. De Oliveira, T.J.P. Penna, J.S.S. Martins, Anais da Academia Brasileira de Ciencias 73 (2001) 15.
[13] T.J.P. Penna, D. Stauffer, Int. J. Mod. Phys. C 6 (1995) 233.
[14] P.C. Mahalanobis, Proc. Nat. Inst. Sci. India 2 (1936) 49.
[15] J. Holland, Adaptation in Natural and Artificial Systems, U. Mich. Press, Ann Arbor, 1975.
[16] D.A. Goldberg, Genetic Algorithm in Search, Optimization, and Machine Learning, Addison-Wesley, Reading, 1989.
[17] M.A. Moret, P.M. Bich, F.M.C. Vieira, Phys. Rev. E 57 (1998) R2535.
[18] M.A. Moret, P.G. Pascutti, P.M. Bisch, M.S.P. Mundim, K.C. Mundim, Phys. A 363 (2006) 260.
[19] J.B.C. Jackson, A.H. Cheetham, Trends Ecol. Evol. 14 (1999) 72.
[20] M.J. Benton, P.N. Pearson, Trends Ecol. Evol. 16 (2001) 405.
[21] B.S. Lieberman, C.E. Brett, N. Eldredge, Paleobiology 21 (1995) 15.
[22] S. Pigolotti, A. Flammini, M. Marsili, A. Maritan, Proc. Natl. Acad. Sci. USA 102 (2005) 15747.
[23] D.L. Rabosky, I.J. Lovette, Proc. R. Soc. Lond. B 275 (2008) 2363.
[24] D.N. Reznick, R.E. Ricklefs, Nature 457 (2009) 837.