

A Cellular Automata Model for Adaptive Sympatric Speciation

Samira El Yacoubi and Sebastien Gourbière

MEPS/ASD - University of Perpignan
52, Paul Alduy Avenue, 66860 Perpignan, Cedex, France
{yacoubi, gourbier}@univ-perp.fr
<http://www.univ-perp.fr/see/rch/cef/index.html>

Abstract. The emergence of new species is one of the trickiest issues of evolutionary biology. We propose a cellular automata model to investigate the possibility that speciation proceeds in sympatry, focusing on the importance of the structure of the landscape on the likelihood of speciation. The conditions for speciation are shown to be limited whatever the landscape being considered, although habitat structure best favours the emergence of new species.

1 Introduction

Understanding the origin and maintenance of diversity is a fundamental problem in biology. Ecology and genetics usually focus on the maintenance of diversity addressing essential theoretical issues and trying to answer major questions of applied biology related to the harmful impact of human activities on the current biodiversity. More romantically, evolution is also concerned by the appearance of new forms of life on earth.

The emergence of new species is among the most controversial topics in evolution, whose origin is usually tracked back to Darwins seminal *The Origin of species* (1859). Several definitions of species have been provided (see [3], Ch. 1 for review), all fitting parts of the complexity of the process. Surely, the most used definition of species is related to the biological species concept; *species are groups of interbreeding natural populations that are reproductively isolated from other such groups* [18]. Starting from this definition, theoreticians investigate speciation as the emergence of two sets of individuals from one and the evolution of reproductive isolation between them.

Different scenarii have been proposed to account for such a diversification process. The most approved one is allopatric speciation, in which pools of individuals of a single initial population get geographically isolated because of an external event (like mountains formation). Sub-populations then differentiate, simply because they evolve separately. Sympatric speciation, the emergence of two species within a set of individuals living and reproducing in the same place, is much more tricky and has long been seen as a purely and unlikely theoretical hypothesis. The difficulty with sympatric speciation is that the continual mixing

due to reproduction between individuals of the two incipient species (referred to as gene flow by evolutionary biologists) opposes the differentiation process.

Adaptive sympatric speciation provides an appealing answer to this issue. The basic idea is that, if splitting of the population is due to adaptation of the two pools of individuals to different ecological resources, then every single individual (whatever the pool it belongs to) has an advantage to mate with a partner using the same resource as it. This is simply because mating with an individual exploiting another resource eventually leads to produce unfit offsprings, which are unable to live on either resource. Also intuitive, this verbal argument calls for quantitative investigations under the many different ecological and genetic situations observable in the diversity of biological organisms.

In this contribution, we use cellular automata as a simple formalism to investigate adaptive sympatric speciation. Some work on sympatric speciation has been done in the past in the framework of CA's [1]. We focus in this paper on the importance of habitat structure on the probability of speciation by habitat specialization and evolution of assortative mating through a one-locus two-allele model (see [3], Ch. 4).

2 The Proposed Cellular Automata Model

2.1 Cellular Automata Approach

Cellular automata (CA) are a class of spatially and temporally discrete mathematical systems characterized by local interaction and an inherently parallel form of evolution. First introduced by von Neumann in the early 1950s to act as simple models of self-reproduction in biological systems, CA are considered as models for complex systems in computability theory, mathematics, and theoretical biology [4,15,16,22,24,25]. In addition to these theoretical aspects of CA, there have been numerous applications to physics, biology, chemistry, biochemistry among other disciplines. The studied phenomena include fluid and chemical turbulence, plant growth, ecological theory, DNA evolution, propagation of infectious diseases, urban social dynamics and forest fires. CA have also been used as discrete versions of partial differential equations in one or more spatial variables [2,8,9,12,23].

A cellular automaton consists generally of a regular array of identically programmed units called "cells" which interact with their neighbours subject to a finite set of prescribed rules for local transitions. Each cell is characterized by a particular state taken in a discrete set of values. Time progresses in discrete steps. The state of a cell at time $t + 1$ is a function only of its own state and of the states of its neighbours at time t .

In a mathematical formalism, a cellular automaton is defined as the quadruple $\mathcal{A} = (\mathcal{L}, \mathcal{S}, N, f)$, where \mathcal{L} is a regular lattice which consists of a periodic paving of a d -dimensional space domain, \mathcal{S} is a discrete state set, N is the neighbourhood of size n defined by the mapping

$$\begin{aligned} N : \mathcal{L} &\longrightarrow \mathcal{L}^n \\ c &\longrightarrow N(c) = \{c_1, c_2, \dots, c_n\} \end{aligned} \quad (1)$$

and f is a function which specifies the transition rule defined by

$$\begin{aligned} f : \mathcal{S}^n &\longrightarrow \mathcal{S} \\ s_t(N(c)) &\longrightarrow s_{t+1}(c) \end{aligned} \tag{2}$$

where $s_{t+1}(c)$ is the state of the cell c at time $t + 1$.

The dynamics definition is augmented with initial and boundary conditions which depend on the considered application.

2.2 Biological Background

We considered a haploid population with discrete and non-overlapping generations living in a landscape including two types of habitat. Life-cycle involves a viability selection stage followed by reproduction. During the viability selection stage, each individual survives according to its adaptation to the habitat it is living in. All the surviving individuals are allowed to reproduce. Individuals first enter a mating pool according to some preferences they have and then mate randomly within the pool. Offsprings are laid close to the place occupied by their mother before mating. The genetic underlying ecological adaptation and mating pool preferences are as simple as possible. One ecological locus with two alleles, A and a , provides adaptation to the first and second type of habitat, respectively. One mating locus with two alleles, B and b , provides preferences for a first and a second mating pool, respectively.

In this background, we are interested in the joint evolution of habitat specialization and reproductive isolation. Speciation will be achieved if the two groups of individual adapted to each of the habitat get reproductively isolated by choosing different pools of mating. That is if the genetic structure of the population evolves towards one of the two following sets of genotypes $\{AB, ab\}$ or $\{Ab, aB\}$. This scenario, where individuals tend to mate assortatively with respect to genes non-involved in ecological adaptation, has been called *Assortative mating genes* by Maynard-Smith [17], who recognized it as the most plausible scenario for species to emerge in sympatry. It has received much attention leading to several modelling attempts (see [3,6,11] for reviews), none of them considering space explicitly as will be done using our cellular automata model.

2.3 Model Description

To model the spatiotemporal changes in the population, we propose a model based on a two-dimensional cellular automaton. It is defined on a square lattice where a cell can be either empty or occupied by a single individual which is assumed to carry one of the two ecological alleles (A or a) and one of the two mating alleles (B or b). We then consider four categories of individuals : AB, Ab, aB, ab which will be associated with the state values represented by a couple (x, y) where $x \in \{A, a\}$ and $y \in \{B, b\}$. A cell is given by its coordinates (i, j) in the square lattice \mathcal{L} and its state at time t is denoted by $s_t(i, j)$ which takes values in the discrete state set $\mathcal{S} = \{A, a\} \times \{B, b\} \cup \{*\}$, where $*$ represents an empty cell. The cell (i, j) also has a characteristic of habitat denoted by $c_{i,j} \in \{H_1, H_2\}$ corresponding to the two types of habitat.

Each transition step is divided on three processes : survival which depends on the ecological allele A or a and the site occupation H_1 or H_2 , mating depending on the second allele B or b and a local dispersal process of offsprings.

Survival is based on the following transition rules. An individual (A, \cdot) survives with probability s_A^1 if it lives in habitat H_1 and with probability s_A^2 if it lives in H_2 . An individual (a, \cdot) survives with probability s_a^1 if it lives in H_1 and with probability s_a^2 if it lives in H_2 , with $s_A^1 > s_a^2$ and $s_a^1 < s_a^2$.

For simplicity, we restrict ourselves throughout this paper to the case :

$$s_A^1 = s_a^2 = s \text{ and } s_A^2 = s_a^1 = 1 - s \text{ with } s \geq 0.5.$$

Let us denote by $\tilde{s}_t(i, j)$ the intermediate state of cell (i, j) after the survival phase which is applied only to occupied cells. For each cell (i, j) such that $s_t(i, j) \neq *$, the survival step is expressed by :

$$\tilde{s}_t(i, j) = \begin{cases} s_t(i, j) & \text{with probability} \\ * & \text{otherwise} \end{cases} \begin{cases} s_A^1 & \text{if } c_{i,j} = H_1 \text{ and } s_t(i, j) = (A, \cdot) \\ s_A^2 & \text{if } c_{i,j} = H_2 \text{ and } s_t(i, j) = (A, \cdot) \\ s_a^1 & \text{if } c_{i,j} = H_1 \text{ and } s_t(i, j) = (a, \cdot) \\ s_a^2 & \text{if } c_{i,j} = H_2 \text{ and } s_t(i, j) = (a, \cdot) \end{cases} \quad (3)$$

Mating. We assume that individuals mate randomly within two pools they join with some preferences defined by the mating allele, B or b , they carry. It is important to note that mating process is independent on the spatial localization of individuals, which conforms to Gavrillets definition of sympatric speciation [11]. Let S_1 and S_2 be the mating pools, we denote by p ($p \geq 0.5$) the preference probability of (\cdot, b) to be in S_1 and (\cdot, B) in S_2 .

Let $S_i^{(x,y)}$ be the set of individuals xy belonging to S_i , $i = 1, 2$. We have then $S_i = S_i^{(a,b)} \cup S_i^{(a,B)} \cup S_i^{(A,b)} \cup S_i^{(A,B)}$ and we can calculate the different mating probabilities as :

$$\begin{aligned} p_1(a, b) &= \frac{p}{|S_1|} (|S_1^{a,b}| + \frac{1}{2}|S_1^{a,B}| + \frac{1}{2}|S_1^{A,b}| + \frac{1}{4}|S_1^{A,B}|) + \\ &\quad \frac{1-p}{|S_2|} (|S_2^{a,b}| + \frac{1}{2}|S_2^{a,B}| + \frac{1}{2}|S_2^{A,b}| + \frac{1}{4}|S_2^{A,B}|) \\ p_2(a, b) &= \frac{p}{2|S_1|} (|S_1^{A,b}| + \frac{1}{2}|S_1^{A,B}|) + \frac{1-p}{2|S_2|} (|S_2^{a,b}| + \frac{1}{2}|S_2^{A,B}|) \\ p_3(a, b) &= \frac{p}{2|S_1|} (|S_1^{a,B}| + \frac{1}{2}|S_1^{A,B}|) + \frac{1-p}{2|S_2|} (|S_2^{a,B}| + \frac{1}{2}|S_2^{A,B}|) \\ p_4(a, b) &= \frac{p}{4|S_1|} |S_1^{A,B}| + \frac{1-p}{4|S_1|} |S_2^{A,B}| \end{aligned} \quad (4)$$

In the same way, we obtain the probabilities $p_i(a, B)$, $p_i(A, B)$ and $p_i(A, b)$, for $i = 1..4$. They represent all the mating probabilities in a set S_i , $i = 1, 2$, with a given category of individuals ab , aB , AB or Ab . Each individual xy has a probability $p_1(x, y) + p_2(x, y) + p_3(x, y) + p_4(x, y)$ of mating and can produce one of the fourth categories :

$$\begin{aligned}
 (a, b) &\longrightarrow \begin{cases} (a, b) & \text{with } p_1(a, b) \\ (A, b) & \text{with } p_2(a, b) \\ (a, B) & \text{with } p_3(a, b) \\ (A, B) & \text{with } p_4(a, b) \end{cases} & (A, b) &\longrightarrow \begin{cases} (A, b) & \text{with } p_1(A, b) \\ (a, b) & \text{with } p_2(A, b) \\ (A, B) & \text{with } p_3(A, b) \\ (a, B) & \text{with } p_4(A, b) \end{cases} \\
 (a, B) &\longrightarrow \begin{cases} (a, B) & \text{with } p_1(a, B) \\ (A, B) & \text{with } p_2(a, B) \\ (a, b) & \text{with } p_3(a, B) \\ (A, b) & \text{with } p_4(a, B) \end{cases} & (A, B) &\longrightarrow \begin{cases} (A, B) & \text{with } p_1(A, B) \\ (a, B) & \text{with } p_2(A, B) \\ (A, b) & \text{with } p_3(A, B) \\ (a, b) & \text{with } p_4(A, B) \end{cases}
 \end{aligned} \tag{5}$$

Dispersal. To describe offsprings dispersal, we will be interested in empty cells after the survival process. We construct an offsprings matrix produced after the mating step and denoted by $(r(i, j))_{i,j}$. The considered rule consists for each cell (i, j) whose state is given by $\tilde{s}_t(i, j) = *$ in selecting randomly one of its neighbouring mated individuals and taking its offspring. Let $N : \mathcal{L} \longrightarrow \mathcal{L}^n$ define the neighbourhood type in the considered CA model and $n = |N(i, j)|$ its size defined by its cardinality. Consider for each cell (i, j) the number $m_t(i, j) \leq (n - 1)$, of neighbouring cells occupied by a mated organism at time t . Each offspring produced in the neighbourhood has a probability $\frac{1}{m_t(i, j)}$ to colonize the empty cell (i, j) . It will die if it is surrounded only by occupied cells. The rule summarizing this step is expressed by :

$$s_{t+1}(i, j) = \begin{cases} \tilde{s}_t(i, j) & \text{if } \tilde{s}_t(i, j) \neq * \\ r(i', j') \text{ with probability } \frac{1}{m_t(i, j)} & \text{if } m_t(i, j) \neq 0 \\ * & \text{otherwise} \end{cases} \tag{6}$$

where s_{t+1} denotes the final state after a complete transition, $(i', j') \in N(i, j)$ and $r(i', j') \neq *$.

3 Simulation Results

The landscape is represented by a square grid of 50×50 cells. Accordingly, the maximal population size is 2 500 individuals. Three matrixes are then constructed : habitat matrix which is given or randomly generated and is unchanged during the simulation. Individuals occupation matrix representing the ecological and mating characters which are initially randomly generated. The individuals ab, aB, Ab and AB are represented by 0, 1, 2 and 3 respectively. An empty cell is represented by 4. An offsprings matrix which is initialized at each iteration after the mating process.

We used two types of landscape either a highly or a non-spatially structured habitat as shown in Fig. 1. In both cases, we varied the strength of disruptive selection s and the strength of assortative mating p between 0.5 and 1. We followed the evolution of the genetic structure in space and time.

To make the understanding of our theoretical results easier, we first observe the different types of outcome while varying the model parameters : *Extinction*,

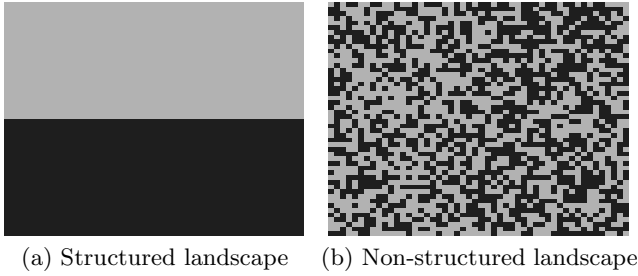


Fig. 1. Highly and non- spatially structured considered landscapes

Fixation, Genetic Polymorphism and Speciation. We then present the set of parameter values allowing for these possible outcomes.

The first pattern we obtain is called *Extinction*. The numbers of individuals of the different genotypes go down to 0.

The second obtained pattern is what we call *Fixation*. In this case, after a few generations, all the individuals tend to have the same genotype.

The more interesting cases are the ones where individuals of different genotypes persist. In these cases, simulations end up with either genetic polymorphism or speciation. To make the difference between these two potential outcomes, we used a common quantity called linkage disequilibrium (LD), and defined as :

$$LD(t) = \frac{N_0(t) * N_3(t) - N_1(t) * N_2(t)}{(|\mathcal{L}| - N_4(t))^2} \tag{7}$$

where $N_i(t)$ is the number of cells in state i at time t and $|\mathcal{L}|$ designates the total number of lattice cells.

Since genetic polymorphism corresponds to one of the two following structures $\{AB, aB\}$ or $\{Ab, ab\}$, we expect LD to tend towards 0. On the contrary, when speciation proceeds, LD is expected to converge to 0.25. More specifically, the simulation stop criterion for speciation corresponds to the following fulfilled condition :

$$|LD(t) - 0.25| < \varepsilon, \text{ for } t > T \tag{8}$$

when ε is a given tolerance and T is chosen to be big enough.

The first two patterns observed in Fig. 2 are called *Genetic Polymorphism* and were obtained with the two types of landscape of Fig. 1. In this case, two pools of individuals persist, each limited to the part of the landscape it is adapted to. As exemplified in Fig. 2a,b, individuals bearing allele a and A leave in habitat H_1 and H_2 , respectively. But, these two pools of individuals do not correspond to two species since they are still reproducing one with another. This can be seen looking at the locus encoding for reproductive isolation. At this locus either allele b or allele B is fixed, so that all the individuals (whatever their ecological phenotype) enter the same mating pools. Which of allele b or allele B get

fixed depend on initial conditions. Both cases correspond to genetic polymorphism as indicated by the null value of LD .

The two following pattern we observed in Fig. 2c,d, correspond to *speciation*. In this case the two pools of individuals are fixed for alternative alleles at both the ecological and the mating loci. For instance, the upper part of the landscape is occupied by ab individuals while the lower part is occupied by AB individuals. We also observed a spatial splitting between aB and Ab individuals, which also is identified as an instance of speciation.

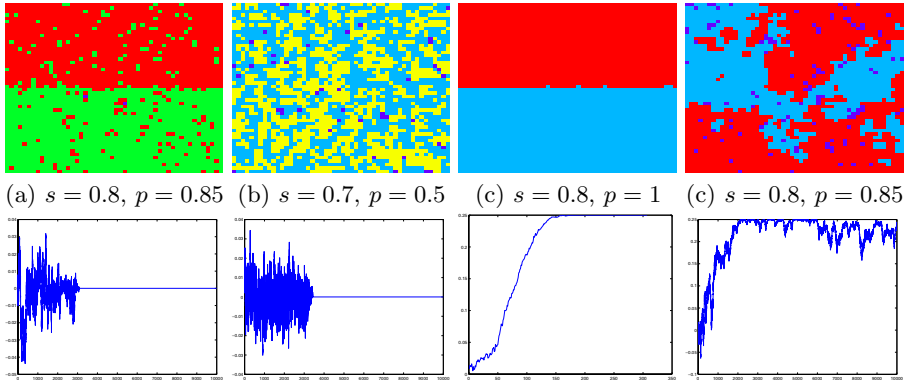


Fig. 2. Final patterns for different probabilities s and p and the corresponding linkage disequilibrium graphs

Interestingly, speciation is much faster in a highly structured landscape. Furthermore, in a random landscape the LD goes on varying in a large range of values while it does not when the habitat is structured. This is because in the former case the border between the areas occupied by the two species are much bigger. This gives opportunities for the frequency of the two genotypes to vary and, consequently, for the LD to fluctuate.

One important biological implication of these results is that completion of speciation (when it happens) depends on the structure of the landscape.

These phenomena are illustrated by implementing the CA rule given by eq (6) starting from an arbitrary initial condition and considering a von Neumann neighbourhood of radius $r = 1$. The tolerance ε is equal to 0.05. The codes are written in Matlab using its graphical interface for visualization. In Table 1 and 2, we give the probabilities values for which the speciation hold in the two types of landscape.

Clearly, selectivity for the mating pool has to be very high for speciation to happen, while requirements on the strength of disruptive selection are much lower. We proceed to the same sensitivity analysis considering a non-structured landscape where habitat H_1 and habitat H_2 are randomly attributed to each cell of the grid.

Table 1. Survival and mating probabilities for speciation with highly structured landscape

Mating probability p	Survival probability s	Speciation time
0.95	0.65	2243
0.95	0.7	1043
0.95	0.75	624
0.95	0.8	535
0.95	0.85	466
0.95	0.9	578
0.95	0.95	995
1	0.55	299
1	0.6	246
1	0.65	263
1	0.7	274
1	0.75	277
1	0.8	307
1	0.9	549
1	0.95	676

Table 2. Survival and mating probabilities for speciation with non-spatially structured landscape

Mating probability p	Survival probability s	Speciation time
1	0.5	1345
1	0.55	1881
1	0.6	5175

When the landscape is non spatially structured, the set of parameters allowing for speciation is even lower. Especially, speciation requires a perfect choice for each of the mating pools.

It should be noted that the obtained results in Table 1 and 2 for speciation time which corresponds to the first time the condition 8 has met, are very sensitive to the initial condition which is random for soil occupation and population distribution. Nevertheless, the values of s and p allowing speciation did not change between several simulations.

4 Discussion

In the *Assortative mating genes* scenario we investigated, individuals tend to mate assortatively with respect to genes non-involved in ecological adaptation. Different models have been set up to investigate the joint evolution of ecological specialization and reproductive isolation, all including different refinements in the description of the survival and mating processes (see [3,6,11] for reviews).

Our cellular automata model confirms that speciation is hard under the *Assortative mating gene* scenario since it requires (when the landscape is highly structured) a very high level of mating pool preferences ($p > 0.9$). Nevertheless, in those conditions, the required level of disruptive selection (that is the level of adaptation of the two morphs to each type of habitat) is not as high as previously reported.

An important feature of our modelling is that the spatial distributions of two types of habitat are modelled explicitly and that individuals do not choose the habitat where their offspring will grow up. This differs from previous models where females are assumed to preferentially lay their eggs in the habitat where they were raised (e.g., Maynard-Smith [17]) or where other types of niche preferences are allowed to evolve [7,10,13,14]. Evolution of such preferences is shown to be an important requirement for speciation to proceed. In our model, offspring are more likely to live in the same habitat as their parents, although we did not model it explicitly. This can happen because we assumed offspring to occupy one of the neighbouring cells around the spatial location occupied by female before mating. A typical example of such process is reproduction of plants where seeds are passively dispersed around the trees where they come from. Whether an offspring is likely to live in the same habitat as its mother then depends of the structure of the landscape. A highly structured landscape is expected to produce an involuntary habitat preference while a random habitat is expected to result in no such by product of limited dispersal.

Indeed, speciation occurs more easily and faster when a strong spatial structure is included in the model, but it still occurs even in landscape when both types of habitat are randomly distributed. In this case, unexpectedly, there are strong spatial genetic structures which do not match with the random distribution of the two type of environment each species is adapted to. These results provide us with interesting perspectives to explain sympatric speciation in plants by adaptive processes as recently demonstrated for palm trees [21].

References

1. Bagnoli F., Bezzi M.: Speciation as Pattern Formation by Competition in a Smooth Fitness Landscape. *Phys. Rev. Lett.* 79, (1997) 3302–3305
2. Chopard B., Droz M.: Cellular Automata Modeling of Physical Systems. Cambridge University Press, (1998)
3. Coyne, J.A. and Orr, H.A.: Speciation. Sinauer Associates Sunderland, MA (2004)
4. Culik K., Hurd L. P., S. Yu.: Computation theoretic aspects of cellular automata. *Physica D*, 45, (1990) 357-378
5. Darwin, C. On the Origin of Species by Means of Natural Selection, or The Preservation of Favoured Races in the Struggle for Life. John Murray, London, (1859)
6. Dieckmann U, Doebeli M, Metz JAJ& Tautz D (eds): Adaptive Speciation. Cambridge University Press (2004)
7. Diehl, S. R., Bush G. L.:The role of habitat preference in adaptation and speciation. In D. Otte and J. A. Endler (eds.), Speciation and its consequences. Sinauer Associates, Sunderland, Massachusetts (1989) 345-365

8. El Yacoubi S., El Jai A.: Cellular automata and spreadability. *Mathematical and Computational Modelling*, vol. 36, (2002) 1059–1074
9. Ermentrout G. B., Edelstein-Keshet L.: Cellular automata approaches to biological modeling. *Journal of Theoretical Biology*, 160, (1993) 97–133
10. Fry, J. D.: Multilocus models of sympatric speciation: Bush versus Rice versus Felsenstein. *Evolution* 57 (2003) 1735–1746
11. Gavrillets, S.: *Fitness Landscapes and the Origin of Species*. Princeton University Press, *Monographs in Population Biology*, vol 41, (2004)
12. Green D. G.: Simulated effects of fire, dispersal and spatial pattern on competition within forest mosaics, *Vegetation*, Vol.82 (1989) 139–153
13. Kawecki, T.J.: Sympatric Speciation Driven by Beneficial Mutations. *Proceedings Royal Society of London* Vol. 265 (1996) 1515–1520
14. Kawecki, T.J.: Sympatric Speciation via Habitat Specialization Driven by Deleterious Mutations. *Evolution* 51 (1997) 1751–1763
15. Mange D., Tomassini M. (Eds.): *Bio-Inspired Computing Machines*, Presses Polytechniques et Universitaires Romandes (1998)
16. Margolus N.: Physics-like models of computation. *Physica D*, 10 (1984) 81–95
17. Maynard Smith J.: Sympatric speciation. *Am. Nat.* 100 (1966) 637–650
18. Mayr, E.: Systems of ordering data. *Biol. Phil.* 10 (1995) 419–434
19. Murray J.D.: *Mathematical Biology*, *Biomathematics Texts*. Springer (1993)
20. Perrier J.Y., Sipper M., Zahnd J.: Toward a viable, self-reproducing universal computer. *Physica D*, Vol.97 (1996) 335–352.
21. Savolainen V., Anstett M.C., Lexer C., Hutton I., Clarkson J.J., Norup M.V., Powell M.P., Springate D., Salamin N. and Baker W.J.: Sympatric speciation in palms on an oceanic island. *Nature* 441 (2006) 210–213
22. Sipper M.: Non-Uniform Cellular Automata: Evolution in Rule Space and Formation of Complex Structures. In R. A. Brooks and P. Maes, editors, *Artificial Life IV*, The MIT Press (1994) 394–399
23. Toffoli T.: Cellular automata as an alternative to (rather than an approximation of) differential equations in modeling physics. *Physica D*, Vol. 10 (1984) 117–127
24. Vichniac G.: Simulating physics with cellular automata. *Physica D* 10 (1984) 96–115
25. von Neumann J.: *Theory of Self-Reproducing Automata*. University of Illinois Press, Illinois, Edited and completed by A.W. Burks (1966)
26. Wolfram S.: *Cellular automata and complexity: collected papers*. Addison-Wesley Publishing Company (1994)