

# Emergent Patterning Phenomena in 2D Cellular Automata

---

**Eleonora Bilotta**

Department of Linguistics  
Cube 17/B  
University of Calabria  
87036 Arcavacata di Rende  
Calabria  
Italy  
bilotta@unical.it

**Pietro Pantano**

Department of Mathematics  
Cube 30/B  
University of Calabria  
87036 Arcavacata di Rende  
Calabria  
Italy  
piepa@unical.it

**Abstract** This article discusses mechanisms of pattern formation in 2D, self-replicating cellular automata (CAs). In particular, we present mechanisms for structure replication that provide insight into analogous processes in the biological world. After examining self-replicating structures and the way they reproduce, we consider their fractal properties and scale invariance. We explore the space of all possible mutations, showing that despite their apparent differences, many patterns produced by CAs are based on universal models of development and that mutations may lead either to stable or to unstable development dynamics. An example of this process for all possible one-step mutations of one specific CA is given. We have demonstrated that a self-replicating system can carry out many slightly different but related entities, realizing new different growth models. We infer that self-replicating systems exist in an intermediate regime between order and chaos, showing that these models degrade into chaotic configurations, passing through a series of transition stages. This process is quantified by measuring the Hamming distances between the pattern produced by the original self-replicator and those produced by mutated systems. The analysis shows that many different mechanisms may be involved in patterning phenomena. These include changes in the external or internal layers of the structure, substitution of elements, differential rates of growth in different parts of the structure, structural modifications, changes in the original model, the emergence of different structures governed by different CA rules, and changes in the self-replication process.

---

## Keywords

Cellular automata, self-replicating structures, patterning phenomena, fractals

---

## 1 Introduction

An organism's morphogenesis—its development—as it adapts to its environment consists of an ongoing sequence of changes involving gene regulation, changes in cell shape, intercellular cooperation, and other elementary processes [8]. Of key importance is cellular self-replication. But what are the mechanisms underlying self-replication? Might it be possible to use studies in the artificial domain to gain insight into the working of biological systems? Possible answers to these questions can be found in the works of D'Arcy Thompson and Alan Turing, two of the first scientists to concern themselves with morphogenesis. Thompson [7] pointed out that the form of an organism is determined by the way it grows. He went on to argue that just as the shape of nonliving structures depends on the properties of inorganic matter, so the form of new biological structures is a consequence of the mathematical and physical properties of living matter. Form is a mathematical problem, and growth is a physical problem. Turing [23] proposed a hypothetical molecular mechanism for pattern formation. Turing's reaction-diffusion system begins in a homogeneous

state. As chemical substances diffuse through the system, they react with each other, producing patterns. Extending this model to the biological world, Turing hypothesized that when those morphogens are present in sufficient concentration, they give rise to organs.

Today, Turing's reaction-diffusion models are well known to mathematical biologists [14, 17]. In real chemical systems, patterns emerge from the interaction between the length scale implicit in the reaction-diffusion dynamics and the geometry of the space in which they are embedded. Turing's model is a paradigm for modern studies of morphogenesis. Much theoretical work has used Turing's reaction-diffusion model to explain patterning phenomena in morphogenesis. Natural [11, 15] and artificial experiments [9], however, did not produce clear evidence that the Turing model is at work in the biological world. Although specific patterns, such as skin coloring in animals, have been modeled using Turing's approach, researchers have yet to produce a model that captures the essence of morphogenesis.

Morphogenesis is closely related to proliferation via self-replication. One of the main goals and challenges of contemporary research is to reproduce this process, generating lifelike forms in digital media. Much of this work is based on von Neumann's [26] cellular automata (CAs): discrete dynamical systems that display complex behavior and the ability to self-replicate [12]. They are able to represent the "logic of life," as well as being capable of universal computation [27], in close relationship with the idea that life is based on informational processes.

Research on self-reproducing structures can be divided into broad categories. Studies in the 1950s and 1960s [26, 6, 25] drew on von Neumann's concept of a self-reproducing automaton. Later research, initiated by Langton [12], attempted to identify the minimum system capable of nontrivial reproduction [3, 22, 21, 16, 20]. During the 1990s, many studies investigated the computational capabilities of self-replicators [18, 5]; very recent research has concentrated on self-reproducers' emergence and evolution mechanisms [13, 4]. Other work has looked into ways of overcoming the limitations imposed by the use of a two-dimensional lattice space [24]. One way of achieving this goal is to use so called *graph automata*. These automata are not restricted to a particular class of lattice space, but can be used to describe processes with variable numbers of elements and variable topologies.

In this article, we investigate emergent patterning phenomena in self-replicating 2D CAs generated by genetic algorithms [1]. In the systems we describe, self-replicators emerge spontaneously from the primeval soup and are very stable, reemerging after collisions and remaining dominant over long periods of time. During the self-replication process, these systems display changes in scale while maintaining invariance in form. They reveal an algorithmic logic in reproduction, which can proceed at different rates, adopting different patterns of spatial organization. The duplicated information, increasing the total quantity of the system at a microscopic level, creates global changes at a macroscopic level, which modify the quality of the system. These changes may affect the form and function of self-replicators and the time scales on which they operate. In the same way as the lengthening of DNA may facilitate the evolution of new functionality in biological organisms, the increase in the quantity of information in the system facilitates qualitative and functional change. The increase does not in itself produce any significant new information. Nor does it directly cause an increase in complexity. But it nonetheless provides raw material for the production of new information in the future.

The replication of structures can be seen as the creation of a sort of redundancy in the system. Self-replication allows the reproduction of the characteristics of a structure within a system that is more complex than the system in which it originally evolved, thereby providing a basis for the emergence of novel functionality. Replication, we believe, involves a combination of multiple phenomena. Given the dynamic context in which self-replication takes place, information tends to multiply until a structure acquires dynamics, allowing it to achieve novel functionality. When a structure reproduces, it may produce a very similar structure, one that is slightly different, or one that is very different. The differences may be quantitative (e.g., the number of cells in the structure) or qualitative (e.g., different states or different combinations of these states, different behavior). But in some way, on a larger scale, they are synchronized. The mathematical language used to describe these

phenomena reflects the characteristics of the replication algorithm: a product of evolution that we have rediscovered in artificial systems.

In what follows we will argue that self-replication is a basic morphogenetic process, underlying growth in both living and artificial systems. If this is so, self-replicators can be seen as a model for the basic informational and computational capabilities underlying life, even if this is not the only key aspect. Section 2 discusses formal aspects of two-dimensional CAs using  $k$ -totalistic rules. We describe how it is possible, starting from a single system, to obtain a family of derived structures. Section 3 introduces the concept of self-replicating structures and the mechanisms through which they reproduce themselves. The discussion focuses on the fractal nature and scale invariance of these structures. It is shown that, despite their apparent differences, many patterns are organized around universal models of growth, which can be stable or can be rendered unstable by perturbations that lead the systems towards chaotic configurations. In Section 4 we explore the space of all possible single-step mutations experimented on, and in Section 5 we describe the patterning phenomena occurring when the system undergoes mutations.

## 2 Formal Aspects of $k$ -Totalistic Cellular Automata

From a formal point of view, a CA can be thought as a tuple

$$A = (d, S, N, f) \tag{1}$$

where  $d$  is a positive integer that indicates the dimension of the CA (1, 2, 3, or more),  $S = \{0, 1, \dots, k - 1\}$  is a finite set of states,  $N = (x_1, \dots, x_n)$  is a neighborhood vector containing the  $n$  elements of  $Z^d$ , and  $f$  is a local rule defined as follows:

$$f : S^n \rightarrow S. \tag{2}$$

In our case  $d = 2$ , and the neighborhood consists of all the cells with a local interaction ray  $r$ , so Equation 2 associates each of the  $(2r + 1)^2$  elements in  $S$  with another element of  $S$ , that is,

$$\begin{pmatrix} \dots & \dots & \dots \\ \dots & s_{ij}(t) & \dots \\ \dots & \dots & \dots \end{pmatrix} \mapsto s_{ij}(t + 1) \tag{3}$$

with  $s_{ij} \in S$ .

A rule that discriminates all possible cases of the neighborhood is expressed in exhaustive form, considering all  $k^{\binom{2r+1}{2}}$  possible cases. In this article we consider two-dimensional CAs, using a particular form of rule, called hereafter the  $k$ -totalistic rule. Such rules do not distinguish the positions of neighbors in the surrounding area, but consider only the number of such cells that are in a given state.

Let  $b_s(t)$  be the number of cells in the neighborhood, in state  $s$  at time  $t$ . Let  $V$  be the set of all possible configurations of the neighborhood, with each configuration represented as a string of integers  $(b_0 b_1 \dots b_{k-1})$ , satisfying the constraint

$$b_0 + b_1 + \dots + b_{k-1} = (2r + 1)^2, \quad b_i \geq 0, \quad i = 0, 1, \dots, k - 1. \tag{4}$$

By definition, a  $k$ -totalistic rule  $T$  is a function that associates each configuration  $\nu \in \mathcal{V}$  with an element of  $\mathcal{S}$ :

$$T : \mathcal{V} \rightarrow \mathcal{S}. \tag{5}$$

The function is equivalent to a CA rule table, where

$$(b_0 b_1 \dots b_{k-1}) \in \mathcal{V} \mapsto T(b_0 b_1 \dots b_{k-1}) \in \mathcal{S}.$$

The table associates each of the first  $k - 1$  totals with a number between 0 and  $k - 1$ ; the sum of these numbers satisfies the constraint in Equation 4. For example, for a  $k$ -totalistic CA with  $k = 4$  and  $r = 1$ , the string (240) indicates that two elements of the nine-cell neighborhood are in state 0, four are in state 1, none are in state 2, and three are in state 3. The rule associates the string (240) with a single element  $s \in \mathcal{S}$ . In this case, there will be 220 rules in the rule table. The constraint in Equation 4 restricts the analysis to a much smaller number of configurations. The number of possible configurations is given by the number of ways in which it is possible to place  $(2r + 1)^2$  indistinguishable objects in  $k$  containers. The rule table in Equation 2 can be represented as a sequence of  $N_T$  numbers:

$$l = (s_1, s_2, \dots, s_{N_T}), \tag{6}$$

with  $s_i \in \mathcal{S}$  and  $i = 1, 2, \dots, N_T$ . Taking Equation 6 as the CA rule,  $N_T$  is given by

$$N_T = \binom{k + (2r + 1)^2 - 1}{(2r + 1)^2} = \frac{(k + (2r + 1)^2 - 1)!}{(2r + 1)^2!(k - 1)!}. \tag{7}$$

To associate a given configuration  $(b_0 b_1 \dots b_{k-1}) \in \mathcal{V}$  with a specific  $s_i \in \mathcal{S}$ , we establish the convention that  $s_i$  corresponds to the  $i$ th configuration in lexicographical order. To process the rule algorithmically, we represent it as a decision tree [10] in which the values of  $b_j$  are intermediate nodes, the alternatives are paths between nodes, and the  $s_i$  are the leaves, or terminal nodes. In this case, a configuration  $(b_0 b_1 \dots b_{k-1}) \in \mathcal{V}$  represents a directed path leading to the leaf  $s_i$ . For  $k = 2$ , the  $k$ -totalistic rules are equivalent to the well-known totalistic rules for Boolean CAs. In what follows, we will limit our considerations to *legal rules* where the last character in the string is 0 (0 is the quiescent state). If all elements of the CA have the value 0, they will take the value 0 in the next time step.

### 3 The Evolution of a Self-Replicating System and the Process of Mutation

A CA may be considered as a machine that, given an input, generates an output. The generation of outputs from inputs can be interpreted as a form of information processing. Given an initial configuration  $\mathbf{x}(0)$ , a special device uses the rule table to generate a sequence of configurations in the form  $(\mathbf{x}(0), \mathbf{x}(1), \dots, \mathbf{x}(t))$ . This sequence of CA states is called a *trajectory* and can be represented as a spatial-temporal pattern, providing in turn different configurations in the time steps of the simulation.

Starting from the string  $l$  (Equation 6), we can create a new string  $l'$  of length  $N_T$  (which we call the *string of consultation*). This string collects the characters that are not consulted by a CA in the rule table, using a neutral symbol  $y$ , and puts in evidence the characters that are consulted in the evolution of the systems. The string appears in the following form:

$$l' = (y, y, s_i, y, \dots, s_j, y, \dots, s_k, y, \dots). \tag{8}$$

In this way, it is possible to know exactly which characters are involved in the realization of the system's evolution.

Let us define  $L$  as the set of all possible strings  $l$  of length  $N_T$ , and  $L'$  as the set of all possible consultation strings. The sets  $L$  and  $L'$  constitute languages, defined respectively over the alphabets  $S = \{0, 1, 2, \dots, k - 1\}$  and  $\{S' = 0, 1, 2, \dots, k - 1, y\}$ . Here  $l'$  is a function of time. Consider an initial state  $\mathbf{x}(0)$ , in which the majority of elements have the value 0. Over time, the number of characters in  $l'$  that differ from  $y$  will tend to grow. In some cases, the string will achieve a steady state. That is, from time  $T$  onwards,

$$l'(t) = l'(T) \quad \forall t \geq T. \tag{9}$$

In conditions where Equation 9 is true,  $l'$  can be used to construct a *consultation network*, whose vertices are constituted by the set of characters with values other than  $y$ . We label the vertices of the graph using the symbol  $R$  followed by a number that indicates the position of the character in  $l'$ . The value of each vertex is the value of the corresponding character. We define the *instantaneous consultation string*  $l''(t)$  as the consultation string obtained by using  $\mathbf{x}(t)$  as the initial data and computing a single time step of the CA system's evolution. If we define  $L''$  as the set of all instantaneous consultation strings,  $L''$  will be a language over  $S'$ . In the case shown in (Equation 9), the instantaneous consultation strings represent a limit cycle:

$$\exists \bar{T} \in N : l''(t) = l''(t + \bar{T}), \tag{10}$$

where  $\bar{T}$  is the period of the consultation network.

We define a *sum of strings* as

$$l(t) = l''(0) \oplus l''(1) \oplus \dots \oplus l''(t), \tag{11}$$

where the operator  $\oplus$  is defined as

$$\begin{aligned} \oplus : L'' \times L'' &\rightarrow L', \\ (l''_1, l''_2) &\mapsto l' = l''_1 \oplus l''_2, \end{aligned}$$

and where the elements  $c_i \in l'$  are as follows:  $c_i = y$  if  $s_{2i} = s_{1i} = y$ , and  $c_i = s_{1i}$  or  $c_i = s_{2i}$  if at least one of the corresponding elements of the strings  $l''_1, l''_2$  has a value other than  $y$ . In instantaneous consultation networks derived from the same string  $l$ , if both elements have values other than  $y$ , then  $s_{2i} = s_{1i}$ .

A mutation is defined by the function

$$\begin{aligned}
 m_s^a &: L \rightarrow L, \\
 l \in L &\mapsto Ra_{-s} \in L
 \end{aligned}
 \tag{12}$$

that associates an  $l \in L$  with a new string  $Ra_{-s} \in L$  possessing the same characters as  $l$ , except for the character in position  $a$ , which is assigned the new value  $s$ . If  $Ra_{-s} = l$ , the mutation is the *identity* element. A string  $l'$  with  $|l'| = \varkappa$  (where  $\varkappa$  is the number of characters of  $l'$  different from  $y$ ) that displays the features described in Equation 9 will contain  $\kappa\varkappa$  mutations, of which  $\varkappa$  are *identities*. The strings  $Ra_{-s} \in L$  generated by substituting characters  $0, 1, \dots, \kappa - 1$  in the string  $l$ , in position  $a$ , whenever the characters of  $l'$  have values other than  $y$ .

The set  $M_l^j$  of strings  $Ra_{-s} \in L$  is the space of all possible mutations:

$$M_l^j = \{m_s^a(l) = Ra_{-s} \in L\}.
 \tag{13}$$

### 4 Self-Replicating Systems

In a previous article [1], we used the string in Equation 6 as the genotype for a genetic algorithm, and defined various indices of complexity as a fitness function, based on the input entropy [29]. For step  $t$  in a simulation, the input entropy  $S^t$  is defined as

$$S^t = - \sum_{i=1}^{N_i} \frac{Q_i^t}{n^2} \times \log \left( \frac{Q_i^t}{n^2} \right),
 \tag{14}$$

where  $n^2$  is the number of cells in the CA grid, and  $Q_i$  is the frequency of consultation of the  $i$ th character of string  $l$ .

We performed many experiments, and we found different kinds of self-reproducers [1]. In almost all evolved rules, self-replicating systems are always present. It thus appears that artificial matter has strong potential for self-replication and evolution. Many self-replicating structures display stable consultation networks [2]. In order to examine the patterning phenomena, let us consider a  $\kappa$ -totalistic CA where  $\kappa = 5$  and  $r = 1$ , with

$$\begin{aligned}
 l = &000300000000000000120000220013000000000404440002004000000102 \\
 &00000000000400014213000000023000042130000300000000004000420 \\
 &00140000000120010300000000000020010010000010000010401430000 \\
 &00000000400300001000401004003020040000003000000000000400000 \\
 &0000000210000040000000410001031030000000000000431000210000 \\
 &2002300400420010000000000000000000010001000002442010000000 \\
 &4220100000000001000000000013000000100000000000000001000000 \\
 &00200000000320000410030300000310334330002400414000042141001 \\
 &30030044110124002000000000002404401020104002004300020020000 \\
 &43100000000000010030034000001303413000000000400011300000043 \\
 &00000301300000100300002000134002002000100000100004000100000 \\
 &01000304040100000000040140004230040003000004400000000020100 \\
 &0130000.
 \end{aligned}
 \tag{15}$$

Then  $\mathbf{x}(t)$ , the state of the CA at time  $t$ , can be represented by an  $n \times n$  matrix whose elements have values lying between 0 and 4.

Consider now the following initial states:

$$\mathbf{x}(0) = \begin{pmatrix} 00\dots\dots\dots00 \\ \dots\dots\dots\dots\dots\dots \\ 0..01110..0 \\ 0..01110..0 \\ \dots\dots\dots\dots\dots\dots \\ 00\dots\dots\dots00 \end{pmatrix}. \tag{16}$$

The matrix  $\mathbf{x}(0)$  consists exclusively of 0's except for two sequences of three 1's, in two different rows. This means that only six cells of the system are active (these are the only cells whose values are different from the quiescent state). Furthermore,

$$\mathbf{x}' = \begin{pmatrix} 00000 \\ 01110 \\ 01110 \\ 00000 \end{pmatrix}$$

is a  $4 \times 5$  submatrix of  $\mathbf{x}(0)$ .

At the next time step of the simulation, the CA takes on states

$$\mathbf{x}(1) = \begin{pmatrix} 0 \dots \dots \dots 0 \\ \dots \dots \dots \dots \dots \dots \\ \dots \dots \dots \dots \dots \dots \\ 0 \dots 0010100 \dots 0 \\ 0 \dots 0134310 \dots 0 \\ 0 \dots 0134310 \dots 0 \\ 0 \dots 0010100 \dots 0 \\ \dots \dots \dots \dots \dots \dots \\ \dots \dots \dots \dots \dots \dots \\ 0 \dots \dots \dots \dots 0 \end{pmatrix}, \quad \mathbf{x}(2) = \begin{pmatrix} 0 \dots \dots \dots 0 \\ \dots \dots \dots \dots \dots \dots \\ 0 \dots 000010000 \dots 0 \\ 0 \dots 000000000 \dots 0 \\ 0 \dots 101343101 \dots 0 \\ 0 \dots 101343101 \dots 0 \\ 0 \dots 101343101 \dots 0 \\ 0 \dots 000000000 \dots 0 \\ 0 \dots 000010000 \dots 0 \\ \dots \dots \dots \dots \dots \dots \\ \dots \dots \dots \dots \dots \dots \\ 0 \dots \dots \dots \dots 0 \end{pmatrix}$$

$$\mathbf{x}(3) = \begin{pmatrix} 0 \dots \dots \dots 0 \\ \dots \dots \dots \dots \dots \dots \\ \dots \dots \dots \dots \dots \dots \\ 0 \dots 111000111 \dots 0 \\ 0 \dots 111000111 \dots 0 \\ \dots \dots \dots \dots \dots \dots \\ \dots \dots \dots \dots \dots \dots \\ 0 \dots \dots \dots \dots 0 \end{pmatrix}$$

It may be observed that after three time steps, the system generates two new structures, which show the same features of the initial  $\mathbf{x}'$ . This is self-replication.

Figure 1 illustrates the evolution of this structure over time. At steps  $T = 3, T = 9,$  and  $T = 15$  the initial configuration repeats itself, generating respectively first two, then four, then four more copies of the initial data. This self-replicator appears rapidly and spontaneously in the primeval soup, and is very stable, regardless of the initial  $\mathbf{x}'$ . In this kind of self-replicator, slight modifications in the structure of the self-replicator have no effect on its overall behavior. Although the dynamics can generate configurations of great complexity, the basic development models are very simple.

If we examine the 2D view in Figure 2, we observe that the final configuration is obtained from basic pieces that, following simple rules, realize a fractal pattern: each piece is repeated every time, in different but related steps, during the evolution of the self-replicating system. The way in which this process proceeds suggests that the patterns produced by the self-replicator are constrained by hidden rules specifying the way in which they can be put together. The final pattern is repeated many times, at different levels, with substructures repeating larger patterns on a smaller scale. In short, the system displays self-similarity, as indeed many self-replicating systems do. Figure 3a, for example, is similar to the Sierpiński triangle.

The logic underlying the self-replicating behavior can be described in precise algorithmic terms. Many subalgorithms can be nested together, giving life to complex growth dynamics.

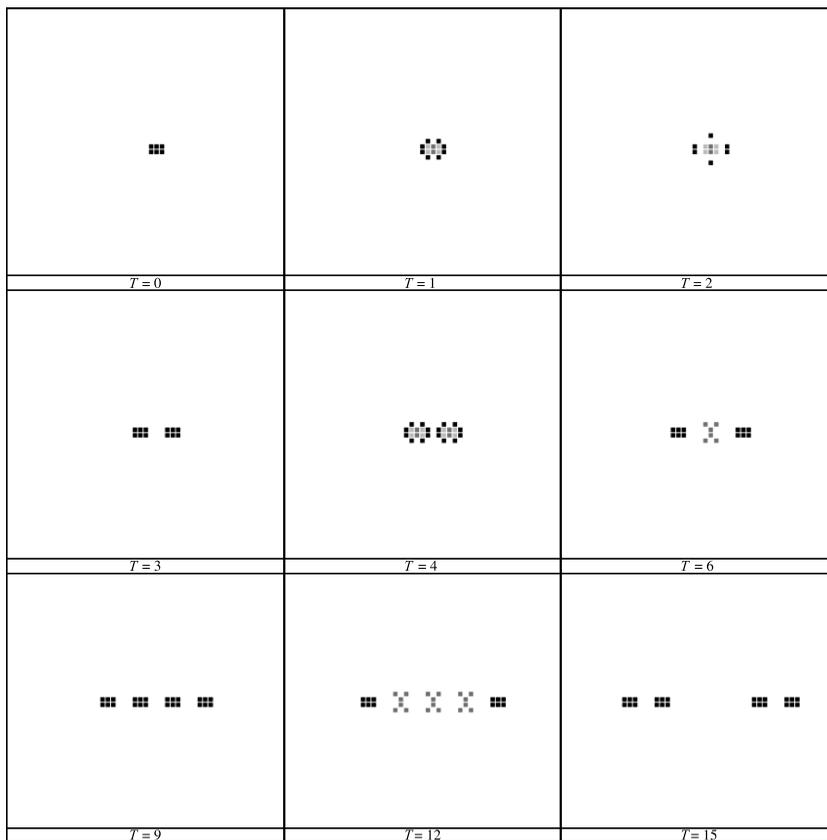


Figure 1. Development of a self-replicator. The figure shows the state of the self-reproduction process at times  $T = 3, T = 9,$  and  $T = 15$ .

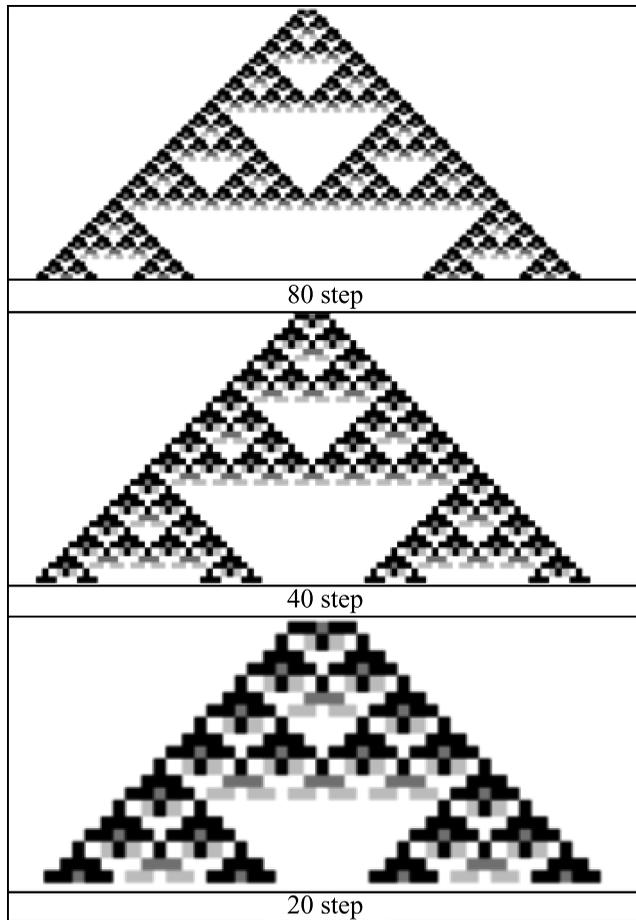


Figure 2. Fractal growth occurs at specific stages in the simulation. The 2D spatial-temporal structure, generated after 20 steps of simulation, is multiplied in the following 20 steps and multiplied again after 80 steps.

This kind of fractal patterning is one of the universal models nature uses in regulating the growth of natural organisms and is frequently observed in the development of leaves and trees [19].

In order to explain this phenomenon, let us consider the image in Figure 3a. Figure 3b shows the results when we apply the following rule: if all Moore neighbors are filled, generate a filled cell; if at least one Moore neighbor is empty, generate an empty cell. This procedure allows the removal of the external configuration and the production of a skeleton image, in this case a Sierpiński triangle. By applying the inverse rule (generate a filled cell if the Moore neighbors contain at least one filled cell), we obtain the image in Figure 3c, showing a larger-scale Sierpiński-like pattern.

This phenomenon is well known in the elementary CA (ECA) literature. According to Wolfram [28], some CAs have the ability to emulate the behavior of others. For example, ECA rule 90 emulates the behavior of rule 22 and (with specific initial data) rule 126. This process is made possible by the dynamics of specific spatial-temporal structures, or blocks.

But, if we examine Figure 4, we can find patterning phenomena that are more complex than this: the Sierpiński-like pattern (which we refer to here as the *external shell*) contains another structure

(the *internal core*), which develops as a  $k=2, r=2$  1D CA. The rule table for this structure is the following:

```

11111111111111110000000000000000
11111111000000001111111100000000
11110000111100001111000011110000
11001100110011001100110011001100
10101010101010101010101010101010

00000001000101100001011101111100
    
```

The rule table can be interpreted as follows: if the cell's neighbors include more than two filled cells, output 0; if the neighbors include exactly two filled cells, output 1; if the neighbors include one filled

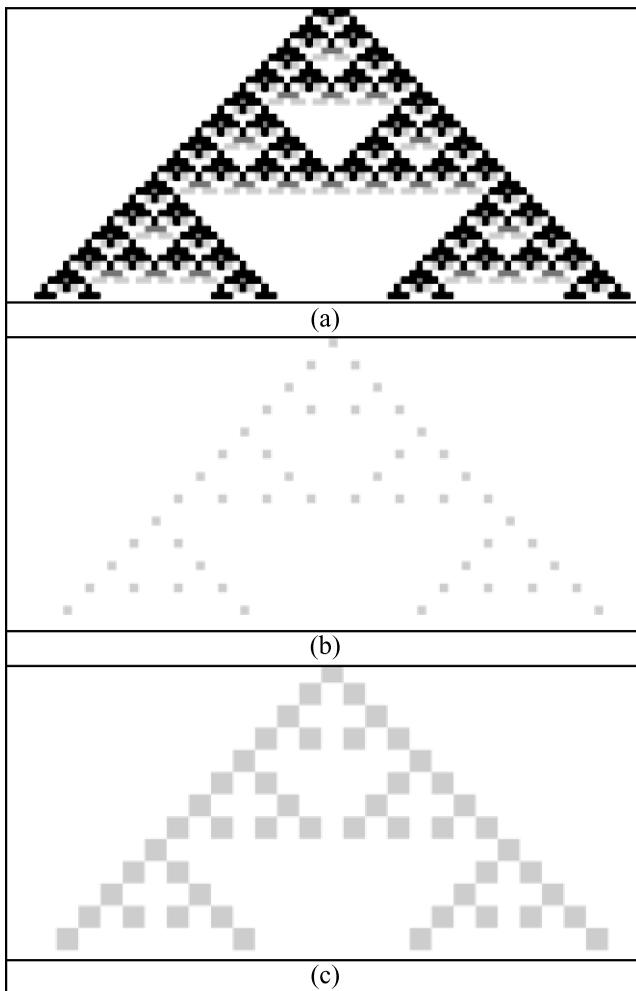


Figure 3. Three stages in the identification of the internal organization of a complex pattern. (a) The initial configuration of the self-replicator. (b) A skeleton image, organized as a Sierpiński triangle. The image was produced by using a rule to remove the external configuration of the first pattern. (c) The image produced when the inverse rule is applied.

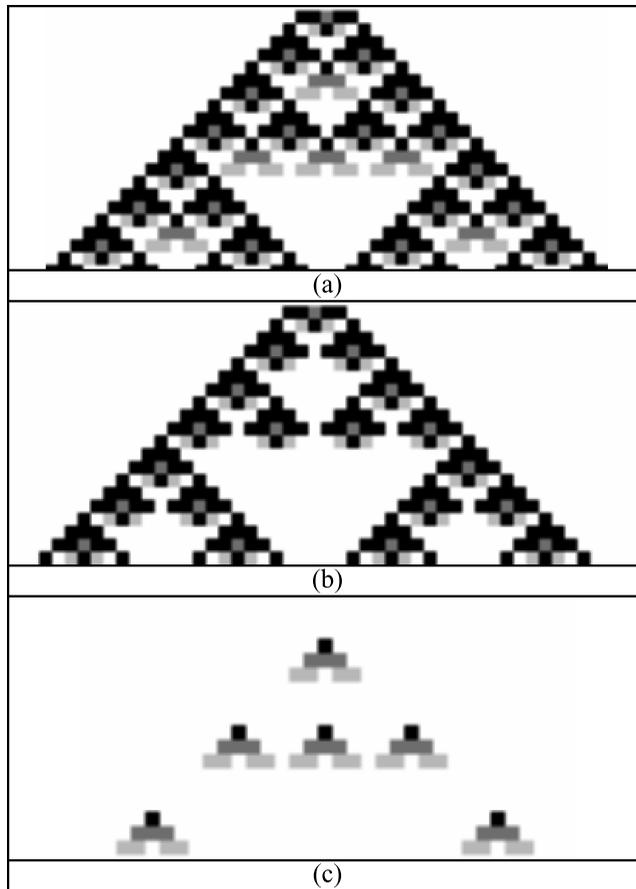


Figure 4. The process used to extract the external shell and the internal core from the pattern generated by the self-replicator. (a) The pattern generated by a self-replicator  $\mathbf{x}'$ . (b) The external core, as simulated by ECA rule 90. (c) The self-replicator's internal core as simulated by a  $k = 2, r = 2$  totalistic rule.

cell on either end of the row, output 0; if the single filled cell is elsewhere, output 1; if the neighbors include no filled cells, output 0. The two nested growth processes, driving the self-replicating system  $\mathbf{x}'$ , generate a multifractal structure.

The first structure has a time dimension of 3, while the second has a time dimension of 6. In reality, many self-replicating systems have the ability to bring together two different growth processes, generating nested structures. When the structures generated in this way are in tune, development of the CA systems is synchronized. A lack of synchronization leads to chaotic patterns. In the self-replicating system  $\mathbf{x}'$ , we have  $|I'| = 31$ ; the network is stable during its development. Figure 5 shows the consultation network  $\rho$  for the self-replicating structure. The vertices of the network and their values are given in Table 1.

## 5 Genetic Mutations

The similarity between certain features of artificial, 2D self-replicators and specific characteristics of biological organisms suggests an analogy between the dynamics of these systems and those of DNA in the biological world. We argue that artificial self-replicators can be considered as proto-organisms, whose genomes (networks of interconnected genes) exactly code their form, function,

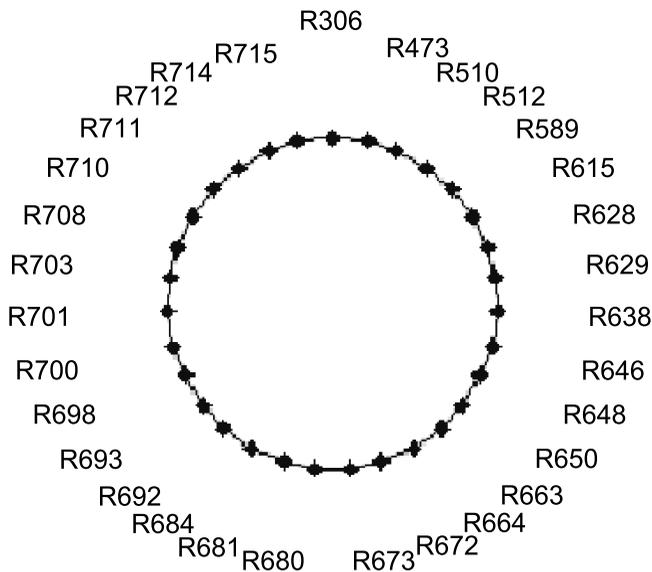


Figure 5. Graph of the consultation network for the self-replicator. In this network  $|I'| = 31$ .

and behavior. Mutations, by changing the genome, can produce significant changes in these traits, giving rise to complex growth dynamics and the emergence of internal clocks, activating or inhibiting specific forms of genetic expression at specific stages during the proto-organism’s development.

Starting from the string  $l$  and using the network of consultation  $\rho$  of the system we have described above (which we shall now call the *genetic network*), we can build up the space of all the mutations  $M_\rho = M'_\rho$  the self-reproducer supports. This space contains 155 elements, of which 31 are identical strings obtained by application of the identity operator. By allowing the self-replicating structure to develop for a certain number of steps (say, 32),  $\forall l \in M_\rho$  we construct an  $l' \in I'$ . Let  $M'_\rho$  be the set of all  $l'$  constructed in this way. We can now represent the space  $M'_\rho$  using a grayscale code, in which each character of a string is identified by a particular gray tone, as in Figure 6. We call this image the *spectrum* of the self-replicating

Table I. Vertices and values of the consultation network  $\rho$ .

R306 → 4	R629 → 1	R664 → 0	R692 → 0	R708 → 0
R473 → 3	R638 → 0	R672 → 0	R693 → 4	R710 → 1
R510 → 0	R646 → 0	R673 → 1	R698 → 0	R711 → 3
R512 → 0	R648 → 0	R680 → 3	R700 → 0	R712 → 0
R589 → 4	R650 → 0	R681 → 0	R701 → 0	R714 → 0
R615 → 0	R663 → 0	R684 → 0	R703 → 0	R715 → 0
R628 → 0				

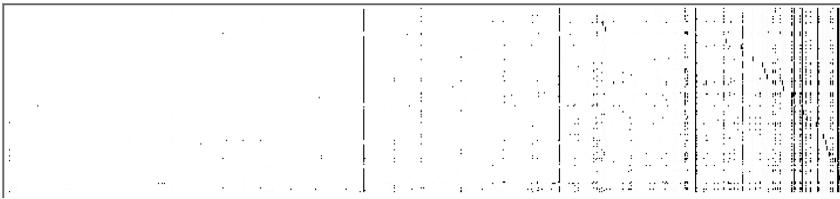


Figure 6. The self-replicator spectrum, showing the space of all possible mutations  $M_\rho$ . The space contains 155 strings; 31 are generated by identity mutations. Each string is represented by a row, and the points of which each row is composed represent the values of the consulted characters.

system  $\mathbf{x}'$ . Whenever a character of the string  $l$  is consulted, a dot is displayed in the white space. Different gray-shaded dots map to different CA values. It is interesting to observe that many characters represented on the left side of the space are never consulted during the development of the system. The right side, on the other hand, is impressively active.

The way in which patterns develop is closely correlated with the number of characters consulted during the development process. Changes to the consultation strings lead to variations in behavior; the trajectories produced by novel strings diverge from those in  $\mathbf{x}'$ . In what follows we describe several different regimes of system behavior (we recall that  $|l'|$  is the number of characters of  $l'$  different from  $y$ ):

$|l'| < 19$ . The only three strings to satisfy this condition are R710\_0, R710\_3, R710\_4. After a very few steps, all patterns are extinguished and the system reaches a fixed point  $\mathbf{0}$ , with 0 active elements.

$19 \leq |l'| \leq 30$ . The set of strings of length 19–30 contains 14 patterns (see Figure 7). Almost all these systems maintain the self-replicating capabilities of  $\mathbf{x}'$ , displaying varied but substantially Sierpiński-like patterns. The only exceptions are R638\_4, which develops into two related gliders, and R714\_3, which disappears after a few steps.

$31 \leq |l'| \leq 40$ . For strings in this range, the structure and function of the basic self-replicator are robust to mutation. During our exploration we found 18 modified self-replicators, which consult up to 30 characters, and a single self-replicator that develops into two related gliders (see Figure 8). The phenotypes of R512\_2, R680\_1, R680\_4, R701\_2, R712\_3, and R712\_4 display modified functionality with respect to the original self-replicator. In these phenotypes the number of characters consulted varies over time. The patterns produced by these systems appear to be very complex, and may at times be completely chaotic.

$41 \leq |l'| \leq 50$ . Strings R650\_3, R673\_3, R680\_0, R684\_3, R684\_4, R698\_3, and R711\_4 change the function of the original self-replicator; as in the cases cited earlier, the number of rules consulted varies as the patterns evolve. The patterns generated by these rules range from organized to chaotic (see Figure 9).

$51 \leq |l'| \leq 60$ . At first sight, the phenotypes produced by strings in this interval do not appear to be self-replicators. However, further examination showed that R673\_2 creates a self-replicating structure. Two additional systems generate highly complex, gliderlike patterns (see Figure 10).

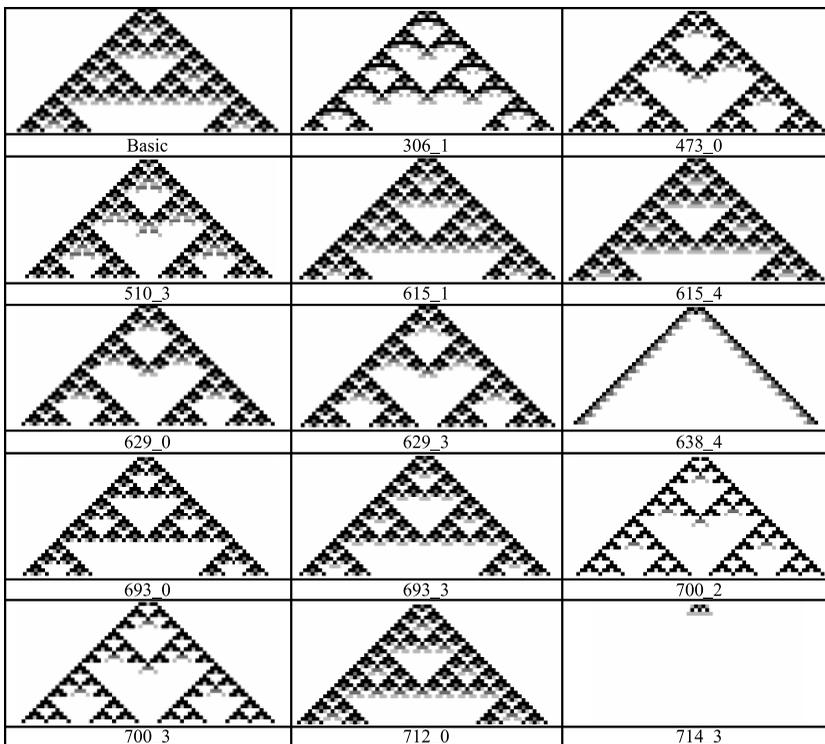


Figure 7. Patterns generated by strings of length 19–30. The first is the basic self-reproducing system.

$61 \leq |I'|$ . In this range, we found only one self-replicator, realized by the string R306\_0, which consults 80 characters, and an additional system that evolves into two related, very complex glider-like configurations.

The analysis of the set of all possible mutations shows that the patterns produced by mutated CA strings can be classified in terms of the presence or absence of self-replication and the number of elements in their respective genetic networks.

Self-replicators were identified using a hybrid automatic and manual procedure. In the first (automatic) phase of the analysis, we identified all self-replicators with a period of less than 16 ( $T < 16$ ). To achieve this goal, we ran each simulation 19 times from step 0 to step 16 and collected the consultation strings. The results of this initial round of simulation were used as input for a second 16-step round. At the end of this round, consultation strings were collected again. The first set of consultation strings was compared with the second. Rules whose consultation strings did not differ between the first and second rounds were identified as self-replicators. In a second phase of the analysis, self-replicators with periods greater than 16 were identified by manual inspection. The analysis produced the results described below.

- *Self-replicating systems.* Exploration of the space of all possible mutations discovered 33 strings with self-replication periods less than 16. The consultation networks for these strings contained between 20 (R700\_3) and 44 characters (R629\_2). The second manual phase of the analysis identified a number of self-replicators with  $T > 16$ : R306\_0 ( $T = 34, |I'| = 80$ ); R306\_3 ( $T = 18, |I'| = 51$ ); R510\_3 ( $T = 21, |I'| = 30$ ); R615\_3 ( $T = 21, |I'| = 33$ ); R673\_2 ( $T = 24, |I'| = 56$ ); R698\_3 ( $T = 24, |I'| = 49$ ).

- *Non-self-replicating systems.* Although some of these strings produced regular patterns; they do not display recognizable self-replication. Strings belonging to this category can be divided into four subgroups:

A. The subgroup of strings (R710\_0, R710\_3, R710\_4, R714\_3, R714\_4) generates ordered behavior. These strings have consultation networks of size 17, 14, 16, 27, and 44 respectively.

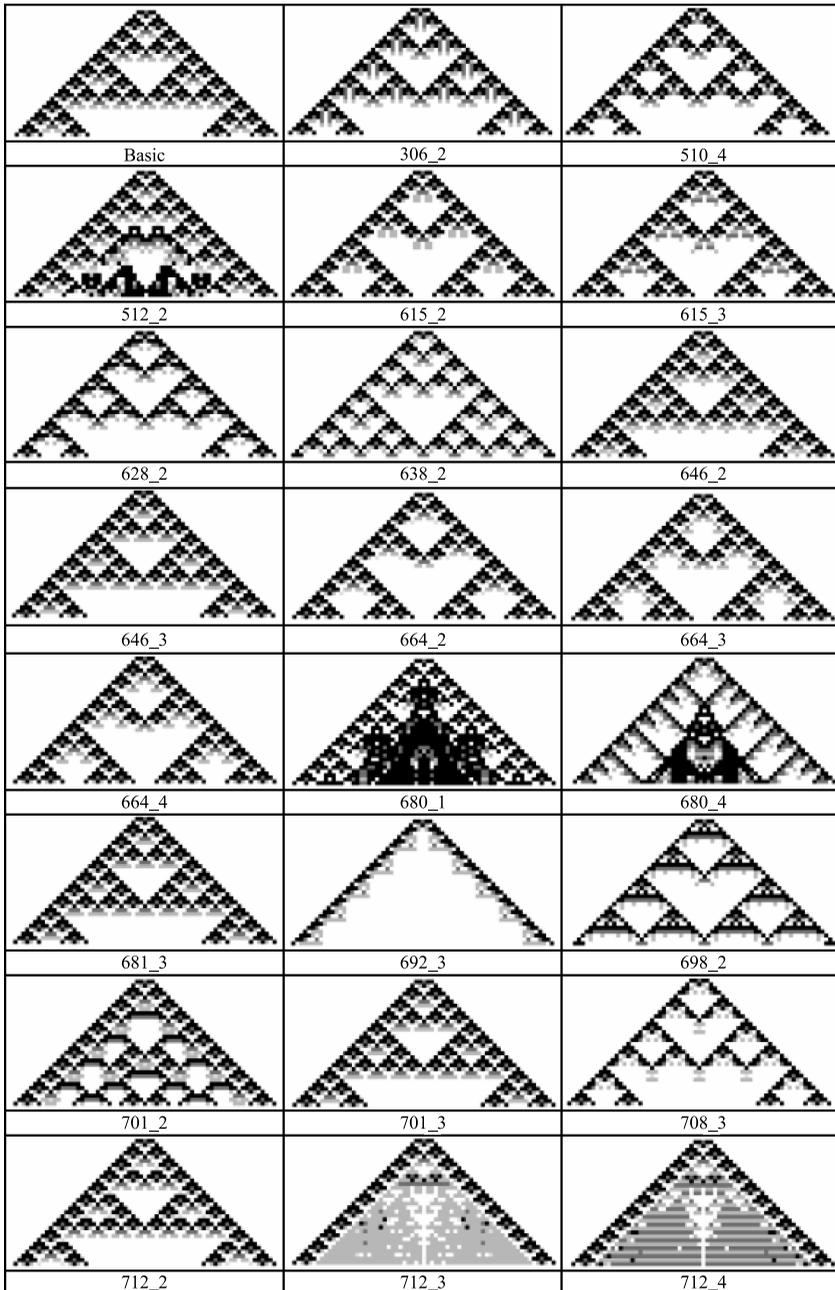


Figure 8. Patterns generated by strings of length 31–40.

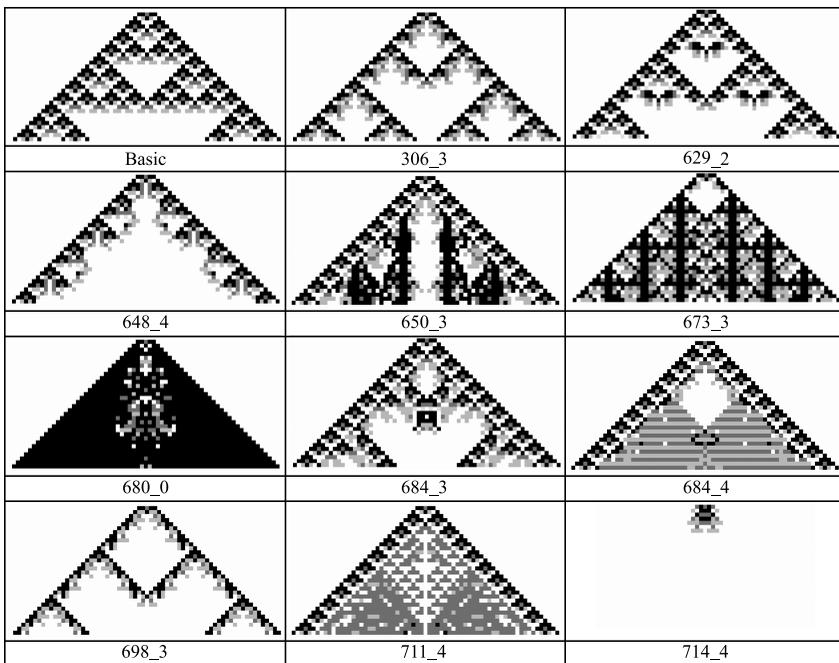


Figure 9. Patterns generated by strings of length 41–50.

B. The seven strings in this subgroup (R638\_1, R638\_3, R638\_4, R648\_4, R672\_3, R692\_3, R698\_4) generate gliderlike patterns. The sizes of their respective networks lie in the range between 28 (R638\_4) and 66 (R638\_1 and R648\_4).

C. This group of strings (R673\_4,  $|I'| = 68$ ; R692\_4,  $|I'| = 56$ ) generates glider guns. R692\_4 is noteworthy for its very complex dynamics.

D. This fourth subgroup contains 67 chaotic CAs. These systems display highly unpredictable behavior and produce the broadest range of patterns observed in any of the systems examined. The sizes of their genetic networks range from 63 (R711\_4 and R589\_3) to 316 (R714\_2).

Figure 11 illustrates the space of all possible single-step mutations, showing the relationship between the number of nodes in the consultation network and CA behavior. Inspection shows that in one region of this space, the self-replicators develop continuously; in a second region they die (at the bottom of the figure); in a third region they generate chaotic patterns (at the top of the figure).

### 6 Patterning Phenomena

Let us consider

$$M''_{\rho} = \{\lambda' \in M'_{\rho} / |\lambda'| = \text{const}, \bar{T}16\}, \tag{17}$$

the set of all strings  $\sigma \in M''_{\rho} \subset M'_{\rho}$  for which it is possible to construct a consultation network  $\varphi \in \Phi$  (where  $\Phi$  is the set of all consultation networks), the self-replication period  $\bar{T} < 16$ , and  $\lambda$  are the strings defining the self-replicating CA.  $M''_{\rho}$  contains 33 elements. The values of  $|\mathcal{N}'|$  range from 20 (two strings generated by mutations  $m_3^{700}$  and  $m_0^{693}$ ) to 44 ( $m_2^{629}$ ) (see Figure 12).

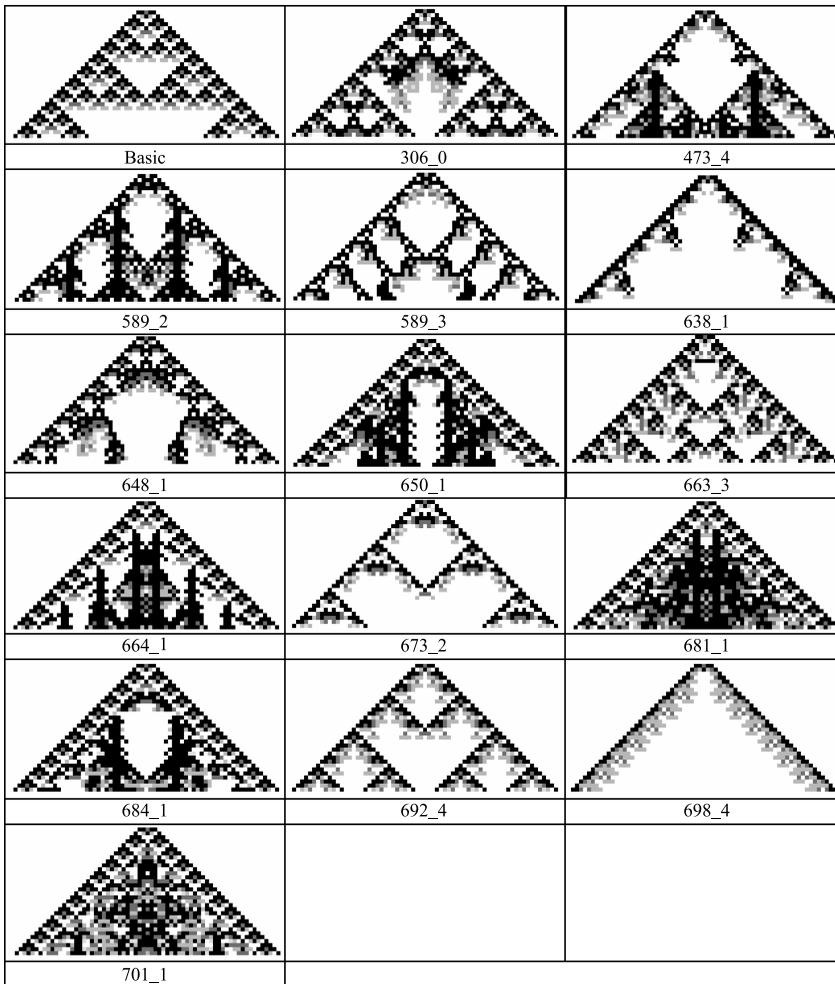


Figure 10. Patterns generated by strings of length 51–60.

Table 2 lists strings in  $M''_{\rho}$  and the sizes of their associated consultation networks. The mean value for all  $|\mathcal{N}'| \in M''_{\rho}$  is approximately 31. To identify the structural transformations underlying this classification, we compared the patterns generated by mutated self-replicating systems with those produced by  $\mathbf{x}'$ . As described earlier, the initial self-replicator produced a Sierpiński-like pattern, with an external shell and an internal core, governed respectively by ECA rule 90 and by a  $k = 2, r = 2$  CA rule. All the self-reproducing systems obtained by mutating this first self-reproducer are in turn Sierpiński-like configurations, which contain interesting processes related to patterning phenomena.

The differences between the patterns realized by mutated systems and those generated by the initial model were identified using a technique based on Hamming distances. For each system, we performed 32 simulation steps on a  $100 \times 100$  grid. The resulting 3D time-space patterns were then projected onto a 2D space. The Hamming distance was computed by comparing the pattern generated with that produced by the initial model and counting the number of discordant cells. Using Hamming distance as an evaluation function, we constructed a hierarchy of mutational processes.

The graph in Figure 13 shows the Hamming distance between each of the self-replicating patterns and the pattern generated by the initial self-replicator. As the graph shows, some strings (R712\_1;

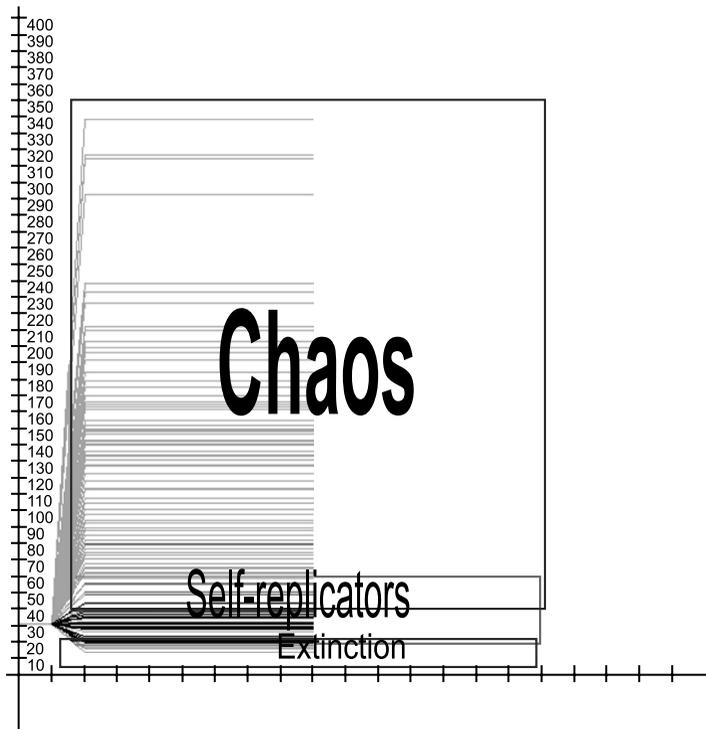


Figure 11. Representation of the space of all possible mutations. The  $y$  axis shows  $||'$ . The line on the far left corresponds to the consultation network for the original self-replicator. The tree structure represents the complete set of  $||'$  values. Clustered in the bottom segment of the figure, there are systems that do not self-replicate, above which self-replicating systems are found.

628\_3; 615\_1) produced patterns whose two-dimensional projections were identical to those for the initial model. It should be noted, however, that these systems are based on very different consultation networks:  $d = 1, d = 7, d = 13$ , respectively, where  $d$  is the difference between the number of vertices in the network of consultation of the basic self-reproducers and the corresponding number for the modified system.

For a second group of six patterns, the Hamming distances lie in the range 8–15, the differences between patterns are visible, and the consultation networks are close to the network for the original model (minimum:  $d = 1$ ; maximum:  $d = 2$ ).

Three patterns (R646\_2, R512\_4, and R711\_0) have Hamming distances in the range 30–60; the consultation networks are still fairly similar ( $d = 5, d = 1, \text{ and } d = 8$ ). Substitutions in the external shell and in the internal core are evident, but do not affect the general configuration of the pattern.

Three patterns (R693\_3, R693\_0, and R629\_2) have Hamming distances between 105 and 133 and display larger differences in their consultation networks than the patterns with smaller Hamming distances.

R629\_2 has an internal core that is completely different from the core in the initial model. The first row in Figure 14 displays the two patterns for comparison. The first panel in the second row shows the results when the the second pattern is subtracted from the first; the second panel shows the results of the inverse operation where the first pattern is subtracted from the second. Six patterns have Hamming distances between 194 and 264, with network differences  $d = 16, d = 18, \text{ and } d = 21$ . All these patterns present modifications to the internal core.

Another set of 11 patterns has Hamming distances between 375 (R701\_2) and 442 (R700\_2). All these patterns display major modifications with respect to the initial model, showing changes both to

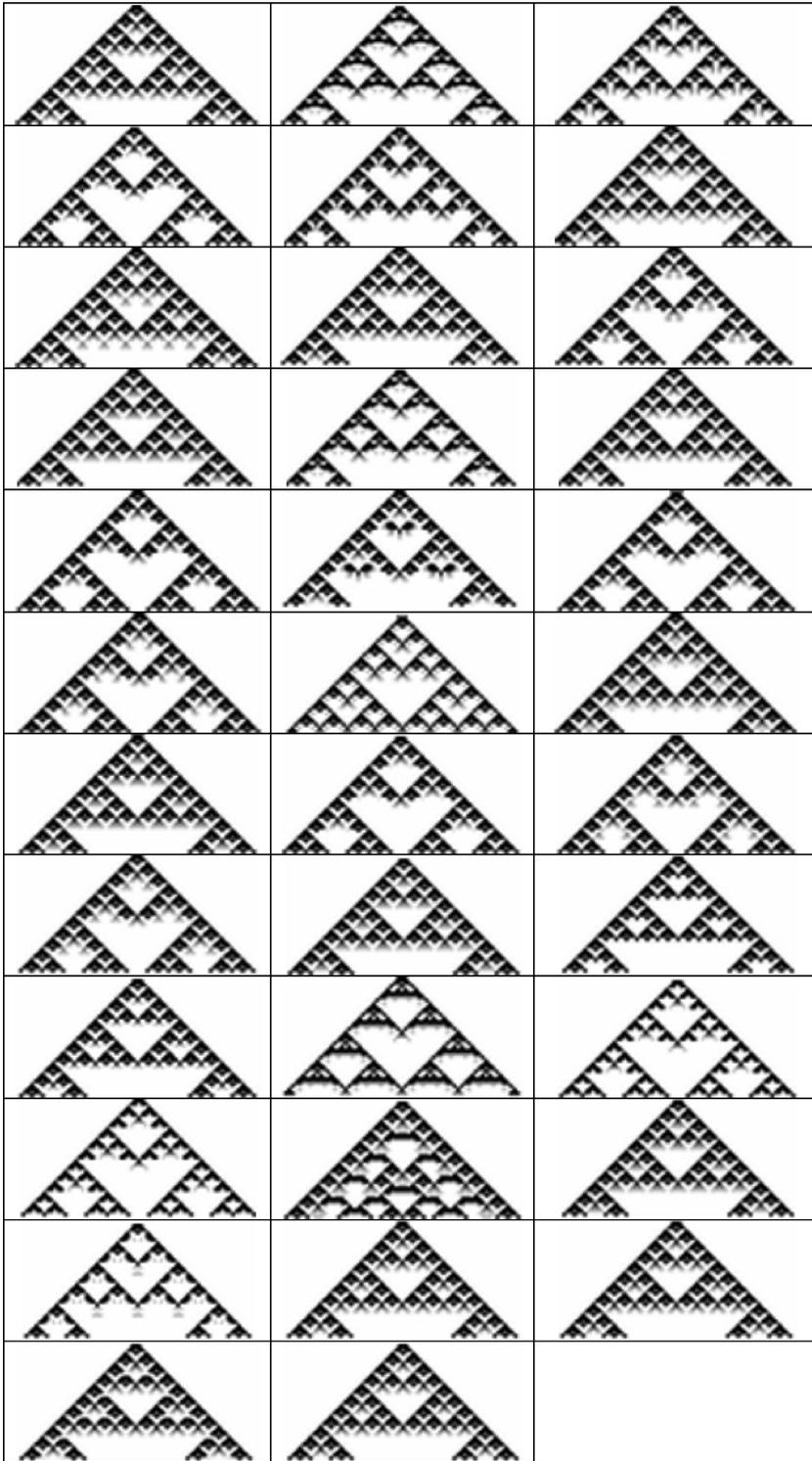


Figure 12. Collection of patterns generated by self-replicators and projected onto a 2D space. The first and last panels show the initial self-replicator model.

Table 2. Strings in  $M'_p$  and the size of associated consultation networks.

---

R306_1	→ 28	R615_2	→ 35	R629_4	→ 31	R681_3	→ 32	R701_3	→ 32
R306_2	→ 36	R615_4	→ 30	R638_2	→ 32	R693_0	→ 20	R708_3	→ 37
R473_0	→ 21	R628_2	→ 32	R646_2	→ 35	R693_3	→ 24	R711_0	→ 24
R510_4	→ 35	R628_3	→ 31	R646_3	→ 32	R698_2	→ 37	R712_1	→ 31
R512_3	→ 31	R629_0	→ 30	R664_2	→ 35	R700_2	→ 22	R712_2	→ 32
R512_4	→ 31	R629_2	→ 44	R664_3	→ 32	R700_3	→ 20		
R615_1	→ 29	R629_3	→ 30	R664_4	→ 32	R701_2	→ 39		

---

the external shell and to the internal cores. These are governed by a  $k = 3, r = 1$  1D CA rule with the following rule table:

```

222222222111111111000000000
222111000222111000222111000
210210210210210210210210
    
```

```

102102100021021021210210010
    
```

where 1 represents the external shell and 2 the internal core.

Figure 15 shows the pattern generated by R473\_0. Differences between these networks vary between a minimum of  $d = 2$  for R629\_0 and R629\_3) and a maximum of  $d = 24$ , for R700\_2. In the patterns generated by R700\_3 and R700\_2 the increased differences in the consultation networks (respectively  $d = 20$  and  $d = 24$ ) are due to substitutions in the external shell.

The patterns generated by R638\_2 and R698\_2 have Hamming distances of 634 and 645 respectively. The respective differences in the consultation networks are 12 and 25. The patterns

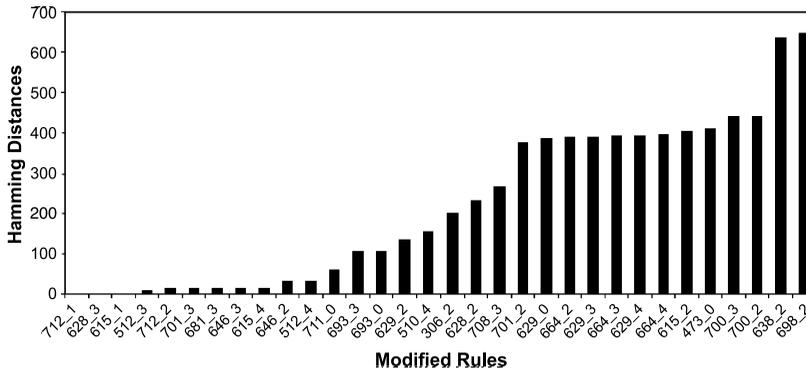


Figure 13. The Hamming distances between the 2D projection of the pattern generated by the initial self-replicator  $x'$ , chosen as model, and the patterns generated by mutated systems.

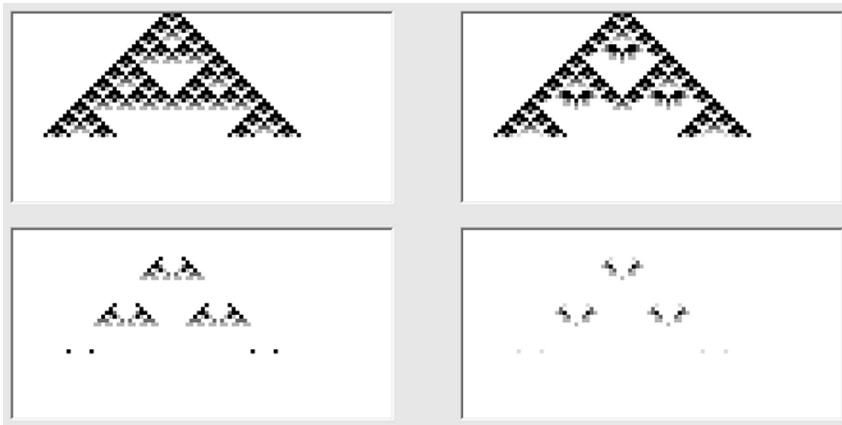


Figure 14. Comparison between the patterns produced by the prototype self-replicator (left) and *R629\_2* (right). The two systems differ significantly in the structure of the internal core.

generated by these rules are similar to those in the initial model, but the rates at which they appear are different (see Figure 16). These processes may be classified as follows:

1. Deletion or insertion of elements in the patterns, without modifying the original arrangement. These operations can be performed both on the external shell and on the internal core of the patterns as well.
2. Replacement of internal or external elements in patterns.
3. Small changes in the initial germs (or modules) that are amplified by changes in the temporal structure of the related patterns, where time affects the growth of the starting model.
4. Modifications of the original model and realization of different structures governed by different CA rules.
5. Dramatic changes in the self-replication process, determining a structural change in patterns.
6. Mixed changes of both temporal rhythms and structural modifications.

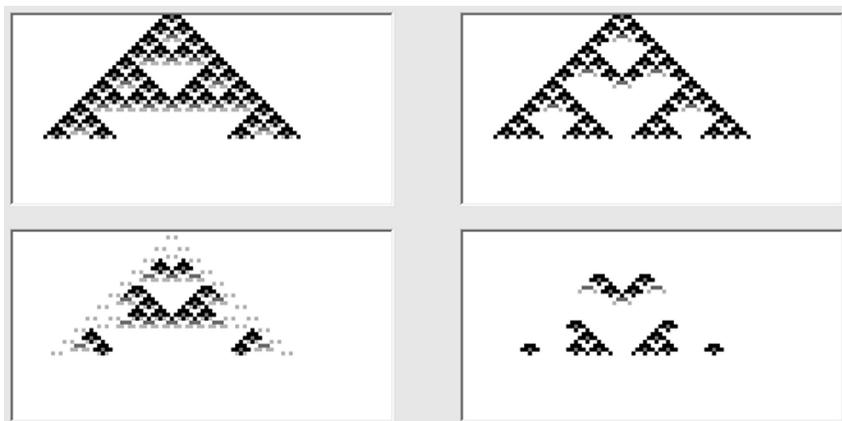


Figure 15. A comparison between the patterns generated by the initial self-replicator (left) and *R473\_0* (right). *R473\_0* presents modifications both to the external shell and to the internal core. The growth of the external shell and internal core is governed by a  $k = 3$ ,  $r = 1$  1D CA rule.

## 7 Conclusions

Maynard Smith and Szathmàry [14] have identified a series of major transitions in biological evolution, which they believe are responsible for the Earth's current biological diversity. The first transition involved the emergence of self-replicating molecules. Cooperation within populations of self-replicating molecules created an environment favoring the replication of other molecules.

In the biological domain, simple or very complicated organisms adapt themselves to different environmental situations throughout life. Here “adaptation” can denote three different things. The first definition refers to any alteration in the structure or function of an organism or any of its parts resulting from natural selection. The organism thus becomes better fitted to survive and multiply in its environment. The second refers to a form or structure modified to fit a changed environment. Finally, there is the ability of a species to survive in a particular ecological niche, because of alterations of form or behavior occurring through natural selection.

The two-dimensional self-replicators described in this article emulate many of the characteristics of biological self-replicators: their individual behavior is determined by their genomes; genomes take the form of a network of connections which algorithmically determine the shape into which the self-replicator develops. Mutations in the genome change this pattern of development. More specifically these changes affect the timing of development, activating or inhibiting the expression of specific portions of the genome. Many mutated self-replicators display changes in the pattern of self-replication, which may be organized in different geometrical and spatial forms.

In the experiments reported in this article nearly all consultation strings of size 19–39 maintain the external configuration and the functions of the original self-replicator. Consultation strings of size 40–150 produce highly complex configurations, though they fail to produce the kind of universal development model seen in ECA rules 90 and 150. Networks of size >150 produce chaotic patterns, including very interesting phenomena that resemble phase transitions.

Many configurations do not change their general shape (a Sierpiński-like triangle) but generate substructures within this basic configuration. Other patterns present very complex arrangements. Often different patterns manifest different rates of growth. The biological and adaptive significance of these variations is that a single basic model can give rise to many different patterns, rather in the same way as occurs in natural language when different speakers use the same basic elements to convey different meanings. Similarly, patterning phenomena constitute a semiotic process related to

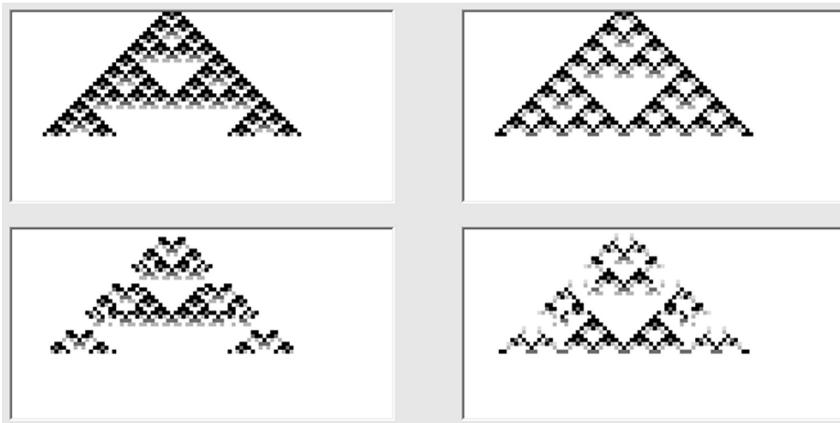


Figure 16. Row 1 shows the patterns generated by the initial self-replicator (left) and R638\_2 (right). Row 2 shows the differences between the two. It can be seen that many groups of cells are organized differently and grow at different rates.

the language of forms. The goal of morphogenesis is to generate meaningful information for biological communication.

All these situations are characterized by changes in phenotypes. The presence of some biological features in artificial 2D self-replicators would lead us to suppose that these systems are closely related to the genetic dynamics that DNA realizes in the biological world. These features are: the presence of genetic sequences, which individuate the self-replicators; the interplay among these genetic sequences, the shape of the self-replicators, and their behavior; the logic of the self-replicator realized by precise algorithms (these algorithms affect the time of the self-reproducing process, and many algorithms can be nested together to give rise to complex dynamics of growth); the fractal organization or geometry of the self-reproducing patterns; the genetic sequence, which is organized as a network of connections. The self-replicators utilize one of these networks. When they undergo mutation, they utilize other networks, thus following different paths in genetic space. Genetic mutations in the genetic CA rule space cause the system to obey new rules, changing in this way the logic of self-reproducer behavior and giving rise to completely different organisms according to the proportion of rules shared. The genetic sequence can be considered to be governed by an internal clock, which in the very precise steps of the organism's evolution can realize mechanisms of variation, activating, or inhibition of specific genetic groups. Furthermore, many variations yield different patterns of self-reproducing behavior.

## References

1. Bilotta, E., Lafusa, A., & Pantano, P. (2002). Is self-replication an embedded characteristic of artificial/living matter? In R. K. Standish, M. A. Bedau, & H. A. Abbass (Eds.), *Artificial life VIII* (pp. 38–48). Cambridge, MA: MIT Press.
2. Bilotta, E., Lafusa, A., & Pantano, P. (2004). Life-like self reproducers. *Complexity*, 9(1), 38–55.
3. Byl, J. (1989). Self-reproduction in small cellular automata. *Physica D*, 34, 295–299.
4. Chou, H. H., & Reggia, J. A. (1997). Emergence of self-replicating structures in a cellular automata space. *Physica D*, 110, 252–276.
5. Chou, H. H., & Reggia, J. A. (1998). Problem solving during artificial selection of self-replicating loops. *Physica D*, 115, 293–312.
6. Codd, E. F. (1968). *Cellular Automata*. New York: Academic Press.
7. Thompson, D. W. D. (1917). *On growth and form*. Cambridge, UK: Cambridge University Press.
8. Gierer, A., & Meinhardt, H. (1972). A theory of biological pattern formation. *Kybernetik*, 12, 30–39.
9. Kauffman, S. A. (1984). Pattern generation and regeneration. In G. M. Malacinski & S. V. Bryant (Eds.), *Pattern formation: A primer in developmental biology* (pp. 73–102). New York: Macmillan.
10. Knuth, D. E. (1997). *The art of computer programming*, Vol. I. New York: Addison-Wesley.
11. Kondo, S., & Asai, R. (1995). A reaction-diffusion wave on the skin of the marine angelfish *Pomacanthus*. *Nature*, 376, 765–768.
12. Langton, C. G. (1984). Self-reproduction in cellular automata. *Physica D*, 10, 135–144.
13. Lohn, J. D., & Reggia, J. A. (1995). Discovery of self-replicating structures using a genetic algorithm. In *Proceedings of IEEE International Conference on Evolutionary Computation* (pp. 678–683). Piscataway, NJ: IEEE.
14. Maynard Smith, J., & Szathmáry, E. (1995). *The major transitions in evolution*. Oxford, UK: W. H. Freeman.
15. Meinhardt, H. (1982). *Models of biological pattern formation*. London: Academic Press.
16. Morita, K., & Imai, K. (1996). Self-reproduction in a reversible cellular space. *Theoretical Computer Science*, 168, 337–366.
17. Murray, J. D. (1993). *Mathematical biology*. Berlin: Springer-Verlag.
18. Perrier, J. V., Sipper, M., & Zahnd, J. (1996). Toward a viable, self-reproducing universal computer. *Physica D*, 97, 335–352.

19. Prusinkiewicz, P. (1994). Visual models of morphogenesis. *Artificial Life*, 1, 67–74.
20. Sayama, H. (1998). Introduction of structural dissolution into Langton's self-reproducing loop. In C. Adami, R. K. Belew, H. Kitano, & C. E. Taylor (Eds.), *Artificial life VI: Proceedings of the Sixth International Conference on Artificial Life* (pp. 114–122). Cambridge, MA: MIT Press.
21. Sipper, M. (1998). Fifty years of research on self-replication: An overview. *Artificial Life*, 4, 237–257.
22. Sipper, M. (1994). Non-uniform cellular automata: Evolution in rule space and formation of complex structures. In R. A. Brooks & P. Maes (Eds.), *Artificial life IV: Proceedings of the 4th International Workshop on the Synthesis and Simulation of Living Systems* (pp. 394–399). Cambridge, MA: MIT Press.
23. Turing, A. (1952). The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society of London, Series B*, 237, 37–72.
24. Tomita, H., Kurokawa, H., & Murata, S. (2002). Graph automata: Natural expression of self-reproduction. *Physica D*, 171, 197–210.
25. Vitányi, P. M. B. (1973). Sexually reproducing cellular automata. *Mathematical Biosciences*, 18, 23–54.
26. von Neumann, J. (1966). *Theory of self-reproducing automata*. Edited and completed by A. W. Burks. Urbana, IL: University of Illinois Press.
27. Wolfram, S. (1984). Universality and complexity in cellular automata. *Physica D*, 10, 1–35.
28. Wolfram, S. (2002). *A new kind of science* (pp. 269–271). Champaign, IL: Wolfram Media.
29. Wuensche, A. (1999). Classifying cellular automata automatically: Finding gliders, filtering and relating space-time patterns, attractors, basins, and the Z parameter. *Complexity*, 4, 47–66.