Genetic Diversification and Complex Genealogy of Self-Replicators
Discovered in Simple Cellular Automata: A Preliminary Report

Chris Salzberg†
chris@cx.hc.uec.ac.jp

Antony Antony⋆
antony@phenome.org

Hiroki Sayama†
sayama@hc.uec.ac.jp

† Department of Human Communication
University of Electro-Communications
1-5-1 Chofugaoka, Chofu
Tokyo 182-8585, Japan

⋆ Section Computational Science
Universiteit van Amsterdam
Kruislaan 403, 1098 SJ Amsterdam
The Netherlands

Abstract

Cellular automata based models of self-replication have been considered capable of producing rather limited evolutionary dynamics only. Counter to this view, we have discovered genetic diversification and complex genealogy emerging in a simple deterministic cellular automata model, the evoloop. For this purpose we use a newly developed set of analyzing tools that are capable of far more sophisticated genetic identification and genealogy tracing. This paper gives a brief summary of our recent findings that demonstrate the richness of evolutionary dynamics of the model, implying the importance of detailed analysis in developing a deeper understanding of emergent evolution of abstract models in Artificial Life.

Keywords: self-replication, cellular automata, evoloop, genetic diversity, complex genealogy

1 Introduction

We have recently been witnessing a moderate resurgence of theoretical studies on self-replicating and evolving structures in Artificial Life [1, 2, 3, 4, 5]. Among them are the efforts on models of template-based self-replication [4, 5]. They aim at simulating the emergence of living (i.e., self-replicating and possibly evolving) structures from a soup of non-living components under virtual physics laws, which are of more physical realism than those used in traditional pure abstract models of self-replication such as cellular automata and rewriting systems. Template-based replication imitates several important aspects of RNA/DNA molecules, which is directly relevant to the understanding of the origin of life and necessary or sufficient conditions for it to occur.

Besides the physical plausibility of transition rules, one may notice that there is another key difference between those template-based models and other traditional models of self-replication delegated by von Neumann's universal constructor [6, 7]: the dependence of self-replication mechanisms upon the structure of self-replicators. In the template-based self-replication models, as well as some of cellular automata based ones like [2], almost any kind of structure can replicate itself, since virtual physics laws are solely responsible for replication processes in those models. On the other hand, in self-replication models based on universal construction capabilities, the actual shape of self-replicating structures is crucial in the execution of self-replication tasks. This difference can be manifested when one tries to see if randomly created structures can self-replicate in those models; they can indeed in the former, but not in the latter, because the latter class of models generally requires a carefully designed complex structures that could virtually never arise in a probabilistic arrangement. This problem roots on the same issue as discussed in the context of the origin of cells in real biology.

To bridge this gap between these two extremes in complexity has been a big challenge in self-
replication studies. Langton’s well-known self-replicating loop [8] and its descendant models, reviewed in [9], were mostly intended to explore possibilities in this gap, yet remaining unsuccessful in showing it as a continuum seamlessly connecting the two ends in the complexity spectrum. Evolutionary growth of complexity has been considered the most promising resolution for bridging this gap. To date, however, theoretical models proposed so far seem to gravitate rather quickly to simple predictable attractors in their phase space with no continuous evolutionary change [12, 13].

In this paper, we try to bring up a positive reinterpretation regarding those earlier efforts, implying that such earlier models with seemingly simple dynamics may have produced far more complex evolutionary behaviors than we thought. We have reached this implication by implementing more detailed and sophisticated methods to identify species and trace genealogical links between them in the evolutionary dynamics of systems, and by applying it to a specific self-replication model, the evoloop [12].

In the original analysis focusing on size difference described in [12], the evolutionary behavior of the evoloop was believed to be so simple that the smallest species always dominate. Using the new methods of observation, however, we have discovered, in the original evoloop model with no modification added to its transition rules, the emergence of huge genetic and behavioral diversity, and possibly, the fitness landscape in a much more complex shape than we originally assumed. The evolutionary dynamics of evoloops is therefore not solely determined by size difference, but rather strongly depends on several other subtle, nontrivial issues (just as it is in real biology) including the topology of the possibility space that accounts for the evolutionary accessibility from one type to another [14].

In the following sections, we give a brief summary of these new findings to promote discussions on the potential of abstract models as a medium for reconstructing evolution, which may possibly fill in the above gap in the complexity spectrum.

2 Methods

The evoloop [12] we choose as our subject is a simple, scalable model of self-replication and evolution, taking the form of nine-state cellular automata with a von Neumann neighbourhood. An evoloop individual is composed of two basic structures: an inner and outer sheath of square or rectangular shape and a gene sequence of moving signal states. The gene sequence contains several state ‘7’ genes for straight growth of a construction arm of the loop and a pair of state ‘4’ genes for left turning of the arm. After three times of such left turning, the arm collides into itself; this makes the tip and the root of the arm bond together to complete self-replication (Fig. 1). Loops are destroyed by the appearance and propagation of the dissolver state ‘8’ through contiguous loop structures, which is triggered by local configurations non-integral to the normal self-replication cycle. This typically arises from shortage of space due to overcrowding. Evoloop populations mutate through direct interaction (collision) of their sheath structures, leading to a change in the gene sequence of offspring loops. This results in a uniquely emergent process of evolution, one which has been considered to generally favor smaller-sized loops due to their robustness and high replication rate [12].

We analyze the evolutionary dynamics of evoloops in full detail by attempting a complete trace of birth and death of every individual loop that ever appeared throughout the run, where each event is associated with a complete description of genotypic and phenotypic identities of the participating loops (both parent and offspring). To detect birth and death events, we notice the fact that there are particular local configurations that arise only when such events occur; we use the appearance of an umbilical cord dissolver state ‘6’ for birth detection, and the disappearance of inner sheath state ‘2’ for death detection, each of which can be implemented in an event-driven manner, with almost no computational overhead needed.

At birth, loops are assigned a genotype corresponding to the configuration of genes in their gene sequence traced counterclockwise starting at the location of the umbilical cord dissolver, and a phenotype describing the size (length and
width) of their sheath structures. The description about their parent loops is also attached to the newborn loops when ancestral relationships are traced. While the gene sequences are usually expressed in a compressed hexadecimal format for the purpose of extensive analysis [15, 16], here we adopt a plainer expression using alphabetical symbols for intelligibility to a broader range of readers. We represent a triplet ‘071’ that describes a gene for straight growth of an arm by G, a triplet ‘041’ that describes a gene for left turning of an arm by T, and a single core state ‘1’ that fills in the sheath structure by C, respectively. For example, for the birth event in the middle frame of Fig. 1, identification reveals that the genotype is GGGGCCGTTGCCCCG and the phenotype is 8×8, hence the newborn loop is a member of species GGGGCCGTTGCCCCG/8×8. Such a pair of genotype and phenotype contains the sufficient information to reconstruct a loop in the exact configuration as it was when it was born. For more details of the methods we refer to [15, 16]. The number of G’s in its genotype must be equal to the length (and the width) of its phenotype for self-replicating species, so the phenotype information will mostly be omitted for simplicity in what follows.

3 Observations

The new tools introduced above have enabled us to discover the richness of genetic diversification and complex genealogy emerging in the evoloop world, which have been largely overlooked in the earlier studies on self-replication models. One of the fundamental findings is that there are significant genetic and behavioral diversity within loops of the same size (and hence they were considered as a single species in the original study) because of differences in their gene sequences. These species with different genotypes can show quite different evolutionary outcomes. Figure 2 compares the growth of colonies of three different loops of the same size 6×6 with different gene sequences. Their reproductive behavior differs surprisingly, as do their evolutionary characteristics such as fitness, long-term stability, and mutatability. Such behavioral diversity in growth patterns is caused by the nontrivial differences in the spread of dissolving state ‘8’ affected by minor variations in gene sequence permutation and spacing, which was already present in the original evoloop model.

As seen in Fig. 2, the possible genetic diversity within evoloops of the same size is achieved by permutation of genes (G, T) and core states (C) in the sequence. We have conducted a detailed calculation for the estimation of the number of possible self-replicating species [17]. The result is listed in Table 1 for sizes from 4 to 18. The number of possible genotypes grows along combinatorial explosion of the number of possible permutation of genes. More importantly, each single genotype in such huge possibility space may have different behavioral patterns just as depicted in Fig. 2. Such striking diversity has been totally ignored in the earlier classification of loops.

With the attention to the existence of such diversity in the evoloop world, we have visualized their ancestral relationship and its time evolution in a series of directed graphs. Figure 3 shows an examplar case, where a species is mapped to a point in the two-dimensional visualization space whose x-axis is a hash value of its gene sequence and whose y-axis represents its size [16]. Dominant self-replicating species are marked by circles, and mutations from one species to another are represented by lines, with a thickness proportional to their accumulative frequencies during a time window. Even though the evolution eventually gravitates to the smallest loops of size 4, there are a number of different intermediate species emerging during this process, exploring the genotypic
Figure 2: Difference in patterns of colony growth between different species of the same size 6×6. Genotypes are: (a) GCCCCGGGTTGG, (b) GGCGTTGCGCC, and (c) GGGTTGCCCCG.

Table 1: Number of possible self-replicating species as a function of loop size.

<table>
<thead>
<tr>
<th>Size</th>
<th># of species</th>
<th>Size</th>
<th># of species</th>
<th>Size</th>
<th># of species</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>15</td>
<td>9</td>
<td>11,440</td>
<td>14</td>
<td>9,657,700</td>
</tr>
<tr>
<td>5</td>
<td>56</td>
<td>10</td>
<td>43,758</td>
<td>15</td>
<td>37,442,160</td>
</tr>
<tr>
<td>6</td>
<td>210</td>
<td>11</td>
<td>167,960</td>
<td>16</td>
<td>145,422,675</td>
</tr>
<tr>
<td>7</td>
<td>792</td>
<td>12</td>
<td>646,646</td>
<td>17</td>
<td>565,722,720</td>
</tr>
<tr>
<td>8</td>
<td>3,003</td>
<td>13</td>
<td>2,496,144</td>
<td>18</td>
<td>2,203,961,430</td>
</tr>
</tbody>
</table>

/phenotypic possibility space. The total number of different species that appeared through this run accumulated up to 668 (including non-self-replicative ones). At the end of this simulation, interspecific competition is still in action between four distinct size-4 species with roughly equal fitness. It would require a much longer time period to see one of these four finally dominate the whole system. One can notice from these plots, when comparing it with a traditional genealogy analysis based on the size of loops only [12], that most of the links plotted in Fig. 3 were just overlooked in such traditional treatments. These links are the very source of evolutionary transitions, events which happen locally yet decide the path of global trends in evolution collectively.

Moreover, we have experimentally found a couple of intriguing features of the evoloop’s gene sequences that may parallel issues in real molecular genetics. One is the discovery of non-mutable subsequences in the evoloop genomes. A typical example is that any spacing between two T’s filled with C’s in a sequence (e.g., TCCCCCCT) is always conserved intact throughout the run, surviving all mutations leading to other self-replicating species. We have identified an entire class of subsequences including the above that share the same mutation resistance [17]. At this point, however, we have not yet obtained a rigorous explanation why this conservation against mutation occurs. Another feature that we have found quite fascinating is the possibility of genetic operation to inject such non-mutable subsequences for producing ‘genetically modified organisms’ that have resistance
Table 2: Sequence of dominant species during the evolution of ‘genetically modified’ evoloops, in the order in which they first become dominant.

<table>
<thead>
<tr>
<th>Size</th>
<th>Gene sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>GGGGCGGGGGGGGG TCCCCCCCCCCCCT GG</td>
</tr>
<tr>
<td>15</td>
<td>GGGGCGGGGGGGGG TCCCCCCCCCCCCT GG</td>
</tr>
<tr>
<td>16</td>
<td>GGGGCGGGGGGGGG TCCCCCCCCCCCCT GCCG</td>
</tr>
<tr>
<td>17</td>
<td>GGGGCGGGGGGGGG TCCCCCCCCCCCCT GCCG</td>
</tr>
<tr>
<td>15</td>
<td>GGGGCGGGGGGGGG TCCCCCCCCCCCCT GG</td>
</tr>
<tr>
<td>14</td>
<td>GGGGCGGGGGGG TCCCCCCCCCCCCT GG</td>
</tr>
<tr>
<td>15</td>
<td>GGGGCGGGGGGG TCCCCCCCCCCCCT GG</td>
</tr>
<tr>
<td>13</td>
<td>GGGGCGGGGGGG TCCCCCCCCCCCCT GG</td>
</tr>
</tbody>
</table>

against mutations toward species smaller than any arbitrarily chosen minimal size. More precisely, we can make all the loops of sizes smaller than this threshold non-self-replicative, because of the lack of space on their genome to contain those injected subsequences as well as an enough number of genes needed for self-replication.

Figure 4 shows an example of the evolutionary dynamics of such ‘GMO’ evoloops. A subsequence TCCCCCCCCCCCCT is injected to a single size-14 ancestor loop, making its entire sequence GGGGCGGGGGGGGTCCCCCCCCCCCCTGG. This subsequence sets the minimum size of viable loops to 13, and only one genotype GGGGCGGGGGGGGTCCCCCCCCCCCCTGG would be viable as a size-13 species with self-replication ability. Since the fitness differences among loops of such large sizes are relatively weak, the system shows a continuously changing, long-lasting evolutionary behavior; it takes more than 2.7M iterations until it finally finds the genotype of the optimal size-13 species.

The sequence of dominant species in this exploration process is shown in Table 2, in which one can confirm that the injected subsequence is preserved as is. Interestingly, it seems to show the presence of some global pattern in the genetic modification process. At first, the dominant genotype simply extends with a G added to its head and a C inserted between the last two G’s at its tail. Once a smaller (size-15) species happens to be reached, then the subsequence before the injected part is optimized in a nontrivial manner, with no change to the tail GG. Why genetic modification proceeds this way remains an open question.
Figure 4: Population dynamics of ‘genetically modified’ evoloops. Only species whose count exceeds 30 are plotted. This case is run on a $500 \times 500$ grid with periodic boundary conditions, beginning with a single size-14 loop with gene sequence $\text{GGGGCGGGGGGGGTCCCCCCCCCCCTGG}$.

4 Conclusion

The methods and observations presented in this article are strongly based on a specific model, and thus their implications are not easily generalizable to address real biological questions. Nevertheless, it is a surprise, especially for those who know well the capabilities and limitations of cellular automata, that even a simple model with nine-state five-neighbor transition rules can produce such complex genetic diversity and nontrivial evolutionary behavior as its emergent property. Note that this discovery required a significant effort paid to the development of the “high-resolution” methods to analyze details of the evolutionary process of model systems. This fact manifests the importance of sophisticated observation and interpretation of model behavior to capture the true richness of lifelike behavior and the hierarchical complexity increase emerging at multiple scales within the model. This has long been underestimated compared to model construction in self-replication studies.

Finally, we would like to emphasize that we have not yet reached any succinct explanation or principle that governs the entire dynamics of the evoloop world, even though it is just a well-defined discrete cellular automata system whose transition rules are completely known. Each time we tried to construct some neat mathematical formulations to describe its evolution, we soon found other minor but crucial factors coming in to bring the whole system toward unpredictable directions. Such a messy nature of the evoloop world may well capture what biologists are struggling with real life. In this sense, we believe that what we perceive is real in this abstract model qualitatively agrees with biologists’ reality they find in real organisms, and that this is the most profound proposition underlying the studies on abstract lifeforms.

Acknowledgments

This work is supported in part by grants from the Hayao Nakayama Foundation for Science, Technology and Culture, and the International Information Science Foundation.

References


