Artificial Life IX: Proceedings of the Ninth International Conference on the Simulation and Synthesis of Living Systems, J. Pollack, M. Bedau, P. Husbands, T. Ikegami, and R. A. Watson, eds., 262-267, Boston, MA, 2004, MIT Press.

Complex Genetic Evolution of Self-Replicating Loops

Chris Salzberg[1,2], Antony Antony[3] and Hiroki Sayama[1]

[1]Department of Human Communication, University of Electro-Communications, Tokyo 182-8585, Japan
[2]Graduate School of Arts and Sciences, University of Tokyo, Tokyo 153-8904, Japan
[3]Section Computational Science, Universiteit van Amsterdam, 1098 SJ Amsterdam, the Netherlands
chris@cx.hc.uec.ac.jp antony@phenome.org sayama@hc.uec.ac.jp

Abstract

It is generally believed that self-replication models constructed on cellular automata have quite limited evolutionary dynamics in both diversity and adaptative behavior. Contrary to this view, we show that complex genetic diversification and adaptation processes may occur in self-replicating loop populations. Applying newly developed tools for detailed genetic identification and genealogy tracing to evoloop populations, we uncovered a genotypic permutation space that expands combinatorially with replicator size. Within this space populations demonstrate broad behavioral diversity and non-trivial genetic adaptation, maximizing colony density while enhancing sustainability against other species. We also found a set of non-mutable subsequences enabling genetic operations that alter fitness differentials and promote long-term evolutionary exploration. These results reveal the amazing potential of cellular automata to re-create complex genetic evolution of self-replicators in a simple, deterministic framework.

Introduction

Since von Neumann's seminal work on self-reproducing automata (von Neumann 1966), models of artificial self-replicators based on cellular automata (CA) have formed one of the mainstreams in Artificial Life (Langton 1984; Reggia et al. 1993; Sipper 1998). Recent developments indicate that simple CA systems can reproduce natural selection processes occuring on different self-replicating structures (Sayama 1999). Their evolutionary dynamics, however, are generally believed to be quite limited in both diversity and adaptative behavior (Sayama 1999; McMullin 2000; Suzuki et al. 2003). Previous results point to a seemingly well-defined fitness landscape in which optimization converges to a single global maximum: homogeneous populations dominated by a single species of the smallest size and shortest replication time.

Contrary to these earlier observations, here we show that complex genetic diversification and adaptation processes may occur in such simple CA. We investigate a system of evolving self-replicating loops (evoloops) (Sayama 1999) in which replication, variation and natural selection emerge solely from local rules. Applying newly developed tools capable of sophisticated genetic identification and genealogy tracing to evoloop populations (Salzberg 2003; Salzberg, in press), we uncovered a genotypic permutation space that expands combinatorially with replicator size. Within this space populations demonstrate broad behavioral diversity and non-trivial genetic adaptation, maximizing colony density while enhancing sustainability in the presence of other competing species. Such adaptation was observed even within species of the same size, thought to be of equal fitness in previous treatment. Intriguing genetic features were also found that may parallel issues in molecular genetics, including the discovery of non-mutable subsequences enabling genetic operations that alter relative fitness differentials. Simulations with such "genetically modified organisms" demonstrate continuously changing, long-lasting evolutionary behavior. These results reveal the amazing potential of CA to re-create complex genetic evolution of self-replicators in a simple, deterministic framework.

Model

The evoloop (Sayama 1999) we investigate is a deterministic nine-state 2D CA model with von Neumann neighborhoods, designed after Langton's self-replicating loop (Langton 1984). An evoloop individual contains an identifiable modular structure describing the shape of offspring (genotype) and an external structure of its own body (phenotype). The former is a sequence of moving signal states (genes) and the latter is a looped sheath of square or rectangular shape, with an arm thrust outward [Fig. 1(a)]. A viable gene sequence contains several '7' states for straight growth of the arm and a pair of consecutive '4' states to control left turning of the arm. In a process of self-replication, cyclic propagation of signal states coordinates the external arm to create a new structural entity. The growing arm is guided through three successive turns and eventually meets its own root, causing tip and root to bond together to form a new, separate loop [Fig. 1(b)]. The truncated arm then retracts, completing the self-replication process.

Loops are destroyed by the appearance and propagation of the dissolver state '8' through contiguous loop structures. Triggered by local configurations non-integral to the normal self-replication cycle, this process of structural dissolution typically arises from shortage of space due to overcrowding and exhibits highly complex dynamics. Its spread is af-

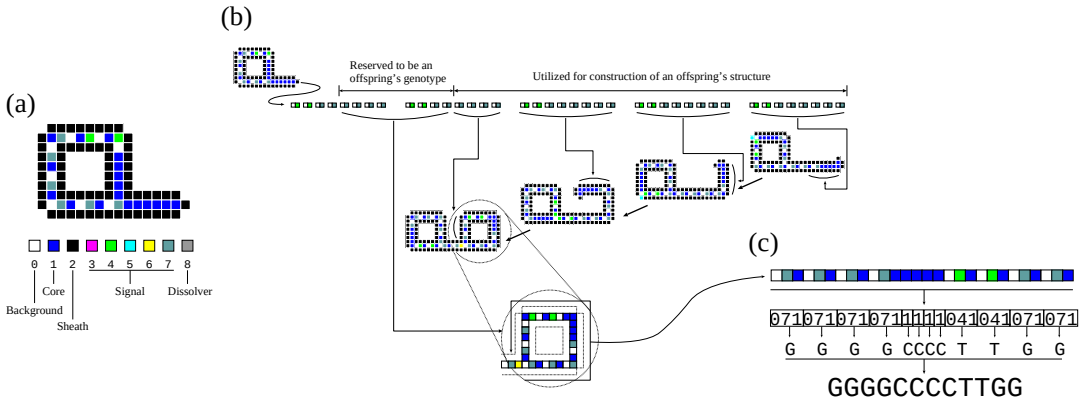

Figure 1: (a) An evoloop individual. (b) Self-replication of an evoloop. Gene sequence is utilized five times during replication, first to construct the umbilical cord (omitted in the figure) and four times to construct the offspring loop. Following loop closure, the truncated arm retracts towards the parent loop and both loops commence the next replication cycle. (c) Labeling scheme of gene sequence of an evoloop. Starting from the bonding location, the mapping transforms '071' triplets to G's, '041' triplets to T's, and '0' states to C's.

fected by minor variations in gene sequence permutation and spacing, often producing leftover sheath cells that form a static, reactive environment. Collisions of loop sheath structures during replication often lead to a change in the gene sequence of offspring. Resting solely on the local interactions of states on a CA grid, such "mutation" events result in an emergent process of variation and natural selection that collectively shapes the path of evolution.

This distinctly bottom-up feature of the evoloop system distinguishes it from other well-studied artificial evolutionary systems of computer programs (Yedid and Bell 2002; Lenski et al. 2003); in these systems a central system manager tracks living organisms and applies probabilistic mutations to their genomes so that explicit control is partially possible. In contrast, the evoloop system supplies no universal structural cues; everything down to the separation between replicator and its environment has to be specified by the system observer. The challenge of analyzing such a potentially "messy" system — coupled with the belief that its evolution always converges homogeneous populations of smallest-sized, fastest-replicating species — has left the detail of its dynamics practically untouched to date.

Methods

We attempt a complete analysis by structurally identifying every birth and death event at the highest level of detail (Salzberg 2003; Salzberg, in press). Unique local configurations are used as markers for detecting such events: the appearance of an umbilical cord dissolver (state '6') for birth detection and the disappearance of an inner sheath (state '2') for death detection. The detection mechanism was embedded in simulator software as an event-driven function requiring almost no additional computational overhead.

At birth, the detection mechanism extracts information about evolutionary identity of the newborn loop, i.e. a genotype corresponding to the configuration of genes in its gene sequence traced counter-clockwise starting at the location of the umbilical cord dissolver [Fig. 1(b)] and a phenotype describing the size (length and width) of its sheath structure. A pair of genotype and phenotype describes a species. To write a gene sequence, we represent a triplet '071' that describes a gene for straight growth by G, a triplet '041' that describes a gene for left turning by T, and a single core state '1' that fills in the sheath by C. For example, the gene sequence of the newborn in Fig. 1(b) is written as GGGGCCCCTTGG [Fig. 1(c)].

Each different species observed during a run is assigned a unique integer label. As a run progresses, a database is compiled containing the mapping between species labels and their evolutionary identities (gene sequence and loop size). Each newborn loop appearing in the CA space is first checked with all the species registered in the database; if its identity is not matched, it is assigned a new label and added to the database. Then we record each such birth event with the labels of both parent and offspring with a time stamp indicating the moment of loop closure. From this level of detail, entire genealogical histories may be reconstructed and every evolutionary transition precisely pinpointed. This new analysis scheme has enabled us to discover a richness of evolutionary phenomena in the evoloop system that were largely overlooked in earlier studies.

Results

Genetic and behavioral diversity

In the birth event records compiled during our simulation runs, self-replicating species have the same labels for both parent and offspring and are hence easily identified. We collected gene sequences of all self-replicating species and discovered, to our surprise, a far larger and more diverse set than expected beforehand. From this set we extracted the following constraints imposed to the sequences for successful replication: (1) the sequence must include the same number of G's as the size of its phenotype, (2) the sequence must

Size	Number of species	Size	Number of species
4	15	12	646,646
5	56	13	2,496,144
6	210	14	9,657,700
7	792	15	37,442,160
8	3,003	16	145,422,675
9	11,440	17	565,722,720
10	43,758	18	2,203,961,430
11	167,960	19	8,597,496,600

Table 1: Number of different self-replicating loop species for a given loop size. We estimate this number by calculating the number of possible gene arrangements in a fixed-length sequence within the constraints for self-replication described in text. For a loop of size n, this estimate amounts to ${}_{2n-2}C_{n-2}$ different species (Salzberg 2003). Loops of size 3 or less cannot self-replicate as there is insufficient space in their genome to fit the required G and T genes.

include a pair of T's, (3) the two T's must have no intervening G between them, and (4) the trailing T must be immediately followed by G. Within these constraints, permutation of G's, T's and C's amounts to a set of viable genotypes whose number grows combinatorially with loop size. We analytically derived an estimate of this number to be ${}_{2n-2}C_{n-2}$ where n is the loop size (Salzberg 2003), listed in Table 1 for sizes from 4 to 19. By size 18, this figure already amounts to over two billion different viable self-replicators. Such huge genetic diversity has been totally ignored in the earlier classification based on loop size only.

Each genotype in this large possibility space may have quite different behavioral patterns. We carried out exhaustive simulation runs up to size-9 loops to make sure that all the permutated genotypes counted in the above estimate are actually self-replicating. Figures 2 and 3 show the results for size-4 and size-6 loops, demonstrating a striking behavioral diversity within the same-sized loops that were considered as a single species in earlier treatment. Note that these patterns — though seen at larger scales than individual loop bodies — are solely dictated by their gene sequences through their non-trivial interactions via transition rules, and thus should be considered as an "extended phenotype" (Dawkins 1990) of each species that can also be subject to natural selection.

Genetic adaptation

The rapid convergence toward smallest self-replicators, quite commonly found in artificial evolutionary models including evoloops (Sayama 1999; Yedid and Bell 2002), tells that the replication time is clearly one of the key quantities being optimized through evolution. However, the huge genotypic permutation space presented in the previous section implies that there may be more room for fine tuning in genetic adaptation, even among the same-sized loops that basically share the same replication time. Whether such microevolution occurs in the evoloop system has remained unresolved to date. To answer this question, we focus on two characteristic quantities for each species and evaluate how they evolve in actual simulation runs.

The first quantity we choose is the sustainability of each species in the presence of other competing species. We characterize this by a relative population ratio of that species after a given period of time in competition with another species, each of which starts from one ancestor. If the given time period is not too long, this ratio captures a shapshot of the population composition under gradual dominance by one over the other, which quantitatively indicates the competitive strength and evolutionary stability of the species against the competitor. Computing an average of such ratios with all the possible competitors would give a mean *survival rate* of that species in the melee of various other species in the "wild". To actually compute this rate, however, one has to restrict the competitor candidates in a practical number. We thus limit ourselves to size-4 species only, assuming that their possible competitors are also of size 4 due to the natural selection favoring shortest replication time. We carried out a round robin among all the fifteen size-4 species and used the results to obtain the mean survival rate for each species, which is shown in Fig. 4. It is clearly seen that there are significant differences of sustainability within the same-sized species, even of the smallest size-4 ones. We note that two species (1 and 15) show particularly low sustainability due to their evolutionary instability; they quickly evolved into other species in most cases.

The second quantity being measured is the colony density of each species. We characterize this by a quadratic coefficient of a parabola[1] fitted by the least-squares method to the population growth curve of that species in an infinite domain. Specifically, we fit a parabola $p(t) = at^2 + bt + 1$ to the population curve and used a as a characteristic quantity of colony growth, which we call *colony density index*. This quantity can be easily measured and defined to each species for its own. It depends, however, on the choice of time range of data point sampling for fitting from the population growth curve. We have tested 0–1500, 0–2000, 0–3000 and 0–5000 updates for the sampling time range. The results with 0–2000 are shown in Fig. 5, reflecting a diversity of growth patterns illustrated in Fig. 2.

These two quantities are found to positively correlate with each other (Fig. 6). Their correlation coefficient varies with different time ranges used for the measurement of colony density index (0.420 with time range 0–1500, 0.674 with 0–2000, 0.423 with 0–3000, and 0.274 with 0–5000) and is highest when the range 0–2000 is chosen. This implies that the sustainability of a population is determined by natural selection acting at a time scale around 2000 updates in the evoloop system. This can be understood in that time scales shorter than this would produce no significant difference in colony structure and time scales longer than this would not be relevant for selection since such a large colony would rarely appear in actual evolutionary processes.

Interestingly, the above two quantities both increase during evolution of loops *in vivo*. Figure 7 shows an exam-

[1]Note that a population of evoloops grow parabolically, not exponentially, due to the geometric constraint of the 2D space.

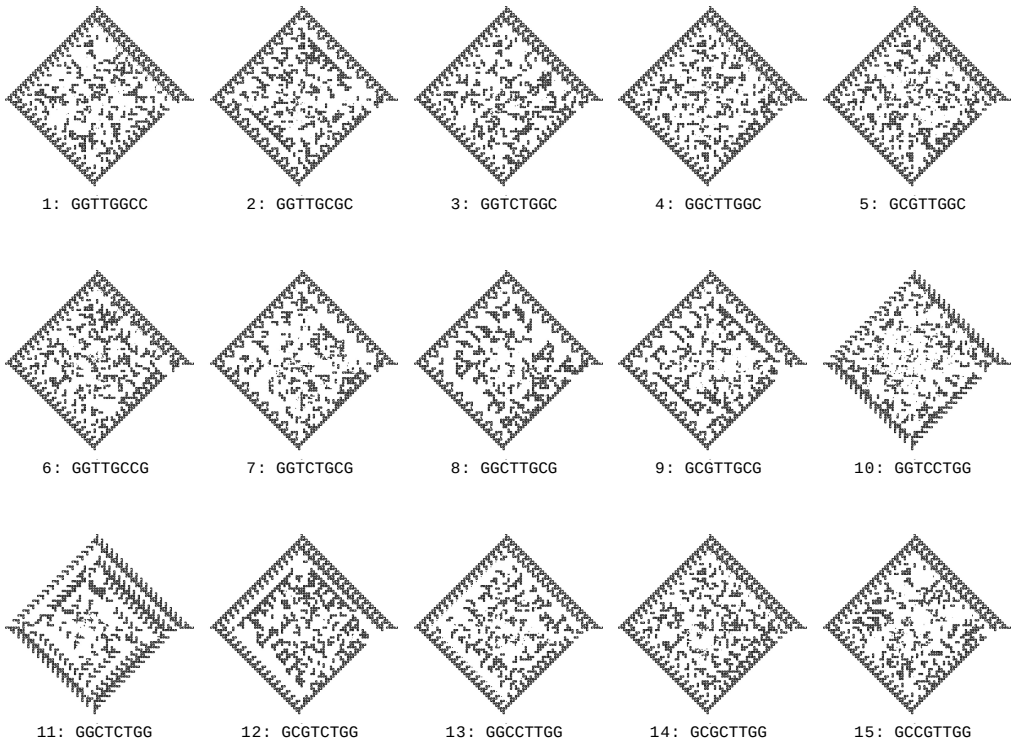

Figure 2: Growth patterns of all the size-4 evoloop species capable of self-replication. The total number of such species is ${}_{2\times4-2}C_{4-2} = 15$. Each snapshot is taken after 5000 updates starting from one ancestral loop. An integer label is attached to each species, which will be used in the following figures.

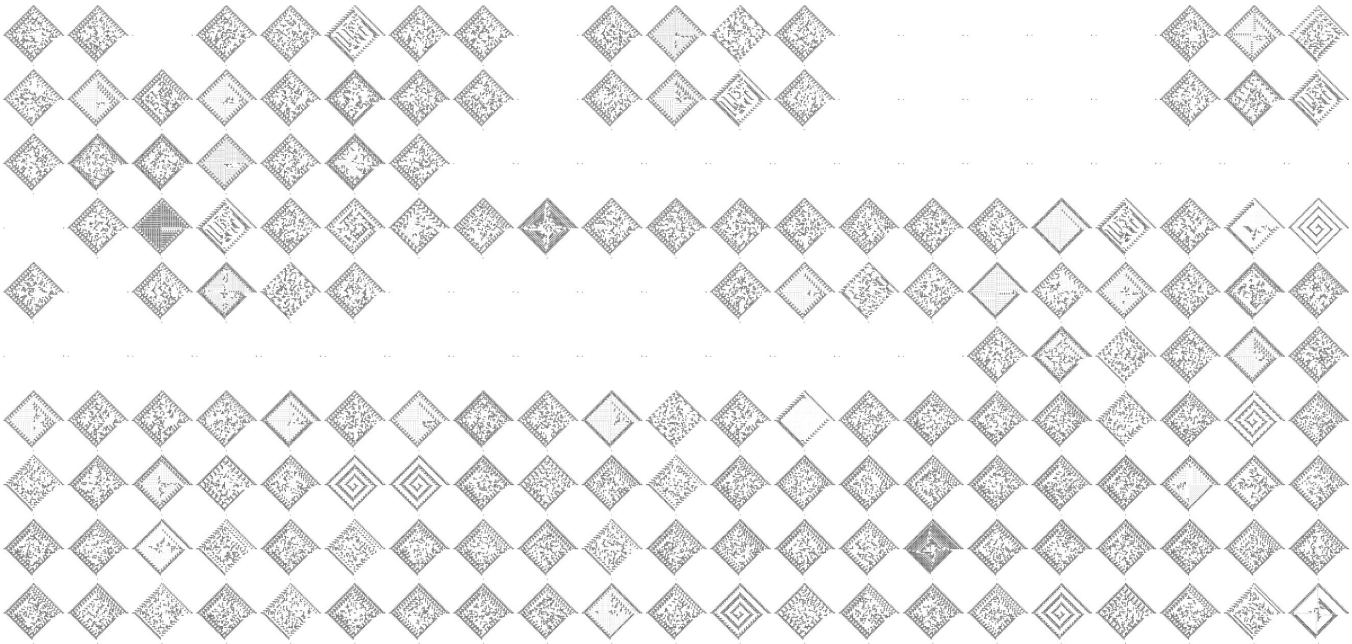

Figure 3: Growth patterns of all the size-6 evoloop species capable of self-replication. The total number of such species is ${}_{2\times6-2}C_{6-2} = 210$. Each snapshot is taken after 5000 updates starting from one ancestral loop. Empty areas indicate unsuccessful species that can self-replicate just once in their lifetime so that there is always only one individual alive in the space. More results for different sized loops can be found at http://complex.hc.uec.ac.jp/loops/.

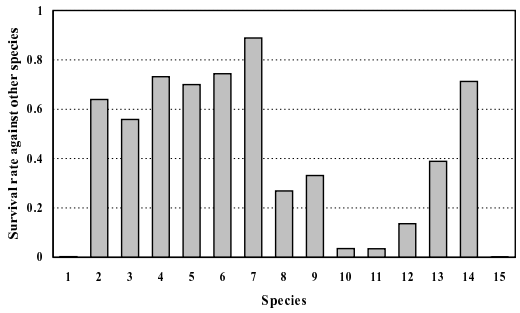

Figure 4: Mean survival rate of size-4 species obtained from the results of the round robin among all the fifteen size-4 species. Relative population ratios are measured after 100000 updates in each competition. The space used is of 1000×1000 grid with opponents placed at opposite ends of the periodic space. We also ran another set of experiments on a 500×500 grid, confirming these results.

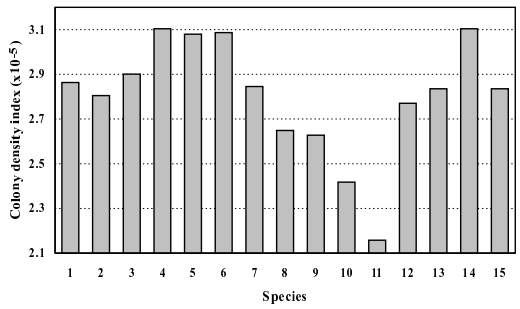

Figure 5: Colony density index of size-4 species. Sample data points for parabola fitting are taken between 0 and 2000 updates at intervals of 10 from the population growth curve of each species.

ple of such processes starting from a size-8 ancestral loop. The evolutionary transition of dominant species in this run is mapped onto Fig. 6; the population moves diagonally in the plot to optimize both quantities. This result gives a clear-cut answer to the question we posed above: there *is* microevolution taking place in the evoloop system, even among the same-sized loops with the same replication time. Natural selection not only favors short replication time but also increases colony density of loops and enhances sustainability against other species through non-trivial genetic adaptation.

Non-mutable subsequences

Moreover, from extensive simulation results, we recently discovered empirically that any subsequence of the form G{C}T{C}TG, where {C} represents any number of C's, will always survive mutations leading to other self-replicating species. Such non-mutable subsequences are a non-trivial outcome of dynamic properties of the evoloop's CA rules and have yet to be rigorously explained. Their existence implies that the genetic state space is partitioned into distinct groups of self-replicating species, each possessing the same conserved subsequence, between which no connecting evolutionary path exists. Each group enforces a minimum loop size for which exact self-replication is possible;

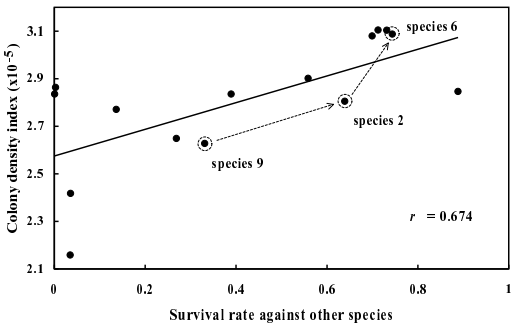

Figure 6: Correlation between survival rate and colony density index plotted in Figs. 4 and 5, respectively. The dashed arrows represent the actual evolutionary transition of dominant species seen in the experiment in Fig. 7.

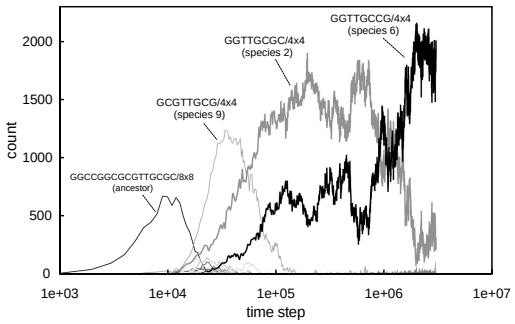

Figure 7: Example of evolution of loops *in vivo* starting from a size-8 ancestor with gene sequence GGCCGGCGCGTTGCGC. Time axis is log-scaled since the selection becomes slow as time proceeds. The transition of dominant species is mapped onto Fig. 6. The space used is of 1000×1000 grid. The same experiments were also performed on size 800×800 and 1200×1200 grids to verify the robustness of the observed dynamics to boundary conditions.

shorter gene sequences cannot contain both the conserved subsequence and the sufficient number of G genes required for exact self-replication.

We use this property to configure "genetically modified organisms", species which cannot evolve below a given minimum threshold size. Experimenting with this threshold enabled us to reduce the size-based fitness differential normally leading to strong competitive exclusion. Figure 8 shows evolutionary dynamics starting from a size-15 "GMO" evoloop injected with the subsequence GCCCCCTCCCCCCCCTG, enforcing a minimum size of 15 on all viable descendants. Although size-based fitness alone favors this minimal-sized species, the fitness differential in this case is relatively weak and gives way to other, emergent behavioral characteristics. As a result, the system fluctuates between dominant species, demonstrating continuous, long-lasting evolutionary behavior, covering over six million iterations and ending with incidental extinction. The progression of major species appearing in this exploration process is shown in Fig. 9. Interestingly, this progression seems to show the presence of some general pattern in the

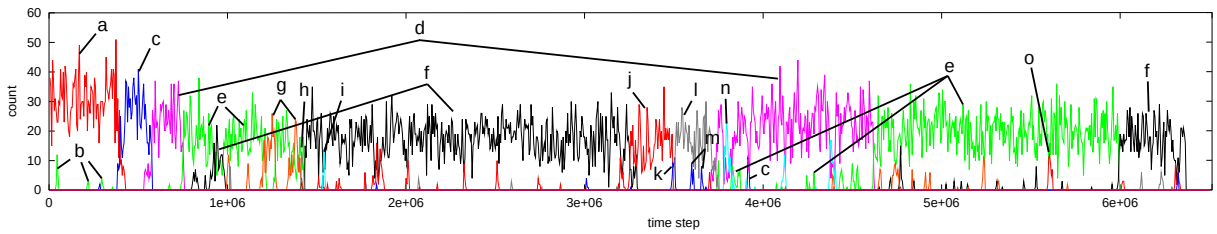

Figure 8: Evolutionary dynamics of "GMO" evoloops with subsequence GCCCCCTCCCCCCCCTG injected to set minimal viable loop size to 15. The space used is of 401x401 grid. Of the total 7106 species observed during this run (including 58 different self-replicating species), only exact self-replicators with populations exceeding 10 individuals are plotted. The same experiments were also performed on size 399x399 and 400x400 grids to verify the robustness of observed dynamics to boundary conditions.

genetic modification process. In this and other experiments, C states are preferentially inserted between G genes alongside the conserved subsequence, producing a general evolutionary tendency towards larger species. This trend is at least weakly reversible, as evidenced by the emergence of certain species with added C states in the middle of their sequence (species n and o) and by the re-appearance of certain smaller species (e.g. species d and e).

Conclusion

The complexity and diversity of CA dynamics has been well known to many for long. Still, it is quite surprising, especially to researchers well-acquainted with the capabilities and practical limitations of CA, that a system so simple can produce such genetic and behavioral diversity of self-replicators and their complex genetic evolution as an emergent property solely arising out of local transition rules. Our findings manifest the importance of developing sophisticated observation and interpretation techniques to capture the full richness of evolutionary phenomena emerging at multiple scales within the system, which has long been underestimated compared to model construction in self-replication studies.

Acknowledgments

C.S. acknowledges financial support by grants from the International Information Science Foundation, the Netherlands Organization for International Cooperation in Higher Education (Nuffic), and the VSB Funds.

Label	Size	Gene Sequence
a	15	GGGGGGGGGGGG GCCCCCTCCCCCCCCTG G
b	16	GGGGGGGGGGGGG GCCCCCTCCCCCCCCTG CG
c	16	GGGGGGGGGGGG GC GCCCCCTCCCCCCCCTG G
d	17	GGGGGGGGGGGGG GC GCCCCCTCCCCCCCCTG CG
e	18	GGGGGGGGGGGGG GCGC GCCCCCTCCCCCCCCTG CG
f	19	GGGGGGGGGGGGG GCGCGC GCCCCCTCCCCCCCCTG CG
g	19	GGGGGGGGGGGGG GCGC GCCCCCTCCCCCCCCTG CGC
h	20	GGGGGGGGGGGGG GCGCGC GCCCCCTCCCCCCCCTG CGC
i	18	GGGGGGGGGGGGG GC GCCCCCTCCCCCCCCTG CGC
j	20	GGGGGGGGGGGG GCGCGCGC GCCCCCTCCCCCCCCTG CG
k	21	GGGGGGGGGGGG GCGCGCGCGC GCCCCCTCCCCCCCCTG CG
l	19	GGGGGGGGGGG GCGCGCGC GCCCCCTCCCCCCCCTG G
m	20	GGGGGGGGGGG GCGCGCGCGC GCCCCCTCCCCCCCCTG G
n	19	GGGGGGGG GCGC GGGGG GCCCCCTCCCCCCCCTG CCG
o	19	GGGG GC GGGGGGGG GCGC GCCCCCTCCCCCCCCTG CG

non-mutable subsequence injected into ancestor loop

Figure 9: List of gene sequences of self-replicating species that appeared in Fig. 8, shown in the order of their first appearance in large numbers. Sequences a – m all have C genes added between G genes along either side of the subsequence. Species n and o, which appear only briefly in large numbers, have C genes injected at positions further away from the subsequence.

References

Dawkins, R. 1990. *The Extended Phenotype.* Oxford, UK: Oxford University Press.

Langton, C. G. 1984. Self-reproduction in cellular automata. *Physica D* 10:135–144.

Lenski, R. E., Ofria, C., Pennock, R. T. and Adami, C. 2003. The evolutionary origin of complex features. *Nature* 423, 139–144.

McMullin, B. 2000. John von Neumann and the evolutionary growth of complexity: Looking backward, looking forward... *Artificial Life* 6:347–361.

Reggia, J. A., Armentrout, S. L., Chou, H. H. and Peng, Y. 1993. Simple systems that exhibit self-directed replication. *Science* 259:1282–1287.

Salzberg, C. 2003. *Emergent Evolutionary Dynamics of Self-Reproducing Cellular Automata.* M.Sc. Thesis. Section Computational Science, Universiteit van Amsterdam.

Salzberg, C., Antony, A. and Sayama, H. Visualizing evolutionary dynamics of self-replicators: A graph-based approach. *Artificial Life*, in press.

Sayama, H. 1999. A new structurally dissolvable self-reproducing loop evolving in a simple cellular automata space. *Artificial Life* 5:343–365.

Sipper, M. 1998. Fifty Years of Research on Self-Replication: An Overview. *Artificial Life* 4:237–257.

Suzuki, H., Ono, N. and Yuta, K. 2003. Several necessary conditions for the evolution of complex forms of life in an artificial environment. *Artificial Life* 9:153–174.

Yedid, G. and Bell, G. 2002. Macroevolution simulated with autonomously replicating computer programs. *Nature* 420:810–812.

von Neumann, J. 1966. *Theory of Self-Reproducing Automata.* Urbana, IL: University of Illinois Press.