On the Performance of some Bioinspired Genetic Operators in Complex Structures Evolution

María J. Casas
Grupo de Estudios en Biomimética
ETSI Informática
Universidad de Málaga (Spain)
mariajcasas@geb.uma.es

Francisco J. Vico
Grupo de Estudios en Biomimética
ETSI Informática
Universidad de Málaga (Spain)
fjv@geb.uma.es

ABSTRACT

Indirect encoding methods demand operators that minimize the disruption of conventional operators, widely studied in direct encoding approaches. While some efforts have already been done in this direction, the growing field of Genetics sheds new light on the dynamics of the nucleic acids, and their implications in the evolution of life on Earth. Here we model basic mechanisms of gene duplication and horizontal gene transfer, presenting preliminary results of its application to L-systems evolution. The first interesting finding is that, in the particular simplified framework proposed, most of these operations are only slightly disruptive allowing the structures to evolve without loosing what has been gained in the past. Large populations of L-systems have been evolved to meet simple restrictions on their phenotypic readout. A case-study is described: evolution of a form under changing conditions. Genotypic and phenotypic evolutions are discussed.

Categories and Subject Descriptors: I.2.11, I.6.5.

General Terms: Algorithms, Performance, Theory

1. A SIMPLE ENCODING FRAMEWORK

In this work we will focus on the evolution of deterministic 0L-systems. It will be shown how an adequate use of the duplication operator can introduce neutral non-disruptive genetic alterations. This type of change, combined with traditional point mutation, will provide the genetic substrate for evolving new structures, which will help in fitting the individual to the landscape.

In the spirit of Prusinkiewicz and Lindenmayer's seminal work [1], lets assume a single nonterminal (G), corresponding to a line of a given length in the graphical object, and lets define a rule to rewrite it. Since only one non-terminal symbol is allowed in the proposed D0L approach, the right-hand side string of the rule that rewrites this symbol will refer to the corresponding L-system, as well as to the structure that is derived after a pre-established number of productions.

The starting direction of the axiom clearly favors verticality, then flat, horizontally distributed arrangements would be harder to find. Also, the individuals in the initial population will have short genomes, giving rise to poorly branched structures (for a small constant derivation length). Combined, these two objectives shape a non-trivial fitness function to make L-systems evolve; the target in our testbed will be searching for flatty, branched L-systems.

The aspect ratio of an L-system will be defined as the height over the width, relative to the bounding box of the structure that results from a fixed-length derivation of such L-system. The branches of the L-systems are counted as the different segments that are read-out from the derived string, taking into account that long linear branches are made of several segments, and that overlapping segments are counted as one.

2. BIOINSPIRED GENETIC OPERATORS

Bracketing in L-systems allows our particular genetic approach the delimitation of areas of expression. Such areas can undergo duplications, resulting in genomes that develop indistinguishable phenotypes, while containing redundant genetic material. These additional copies constitute the genetic substrate for new structures.

Following this strategy, a duplication operator can be defined. Such operator rewrites a given bracketed expression into two concatenated copies of itself. These new genomes develop into the same functional structure as the original one, as represented in a 2D graph hence we recall their indistinguishable nature. However, the phenotype of the duplicated one is not identical, since parts of the structure are replicated (but do overlap with the original ones). Future mutations (either alterations or additions) operating on this redundant parts of the genome will provide new structures.

Three variants of the duplication mechanism have been modeled, where a bracketed expression (a segment, from now on) is copied:

- Random duplication: the copied segment is placed somewhere in the genome, without restrictions.
- Leveled duplication: the copied segment is placed somewhere in the genome at the same nesting level ($...[...[...[...\gamma...]...]...]... \rightarrow ...[...[...\gamma...]...]...[...[...\gamma...]...]...$, where γ is a segment).
- Contiguous duplication: the copied segment concatenates the original one (e.g. $...\gamma... \rightarrow ...\gamma\gamma...$, being γ a segment).

Horizontal gene transfer was modeled by allowing a genome to incorporate, in any location, a bracketed expression from another one. Apart from duplication and gene transfer, three classical operators have been added: insertion, deletion, and point-mutation.

3. RESULTS

Here we have studied the evolution of an unfitted complex structure (optimized for an aspect ratio close to 1) to a form that is optimized for flatty and branchy arrangements (generations 1 to 15), and its fast adaptation (generations 16 to 26) to an environment where high and narrow structures are favored. Since the initial population was homogeneous (no variability), it made no sense to expect speciation, then a high selective pressure was applied to speed up the simulation.

The first row in Table I shows the original genome that is spread out along the initial population. In the second generation the genome has expanded with a random duplication (new genes in bold), and this same gene has been selected to be copied randomly in the fourth and fifth generations. Again, once the genome has reached a size that allows variability, deletions and point-mutations do the rest of the work: from generation 6 to generation 15, only small mutations take place, tuning the structure in the favored direction. The last column measures the shape value (s), which integrates how flat and dendritic the structure is.

Table I. L-systems evolving in a changing environment. Bold: duplications/insertion, rectangle: deletion/mutation.

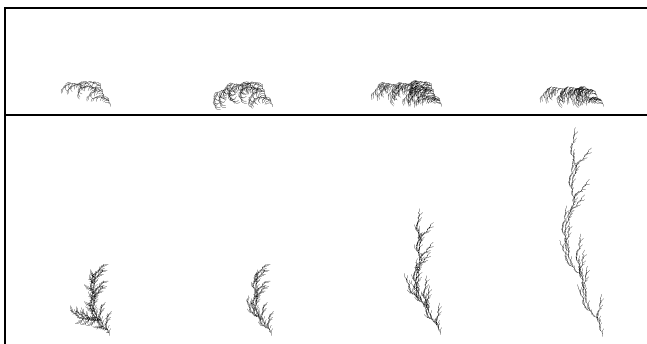| #	genotype	s
1	G+G-G[++G-GG][G]	.27
2	G+G-G[++G-GG**[++G-GG]**][G]	.13
4	G+G-G[++G-GG[++G-GG**[++G-GG]**]][G]	.10
5	G+G-G[**[++G-GG]**++G-GG[++G-GG[++G-GG]]][G]	.10
7	G+G-G[[++GGG]++G-GG[++G-GG[+ +G-GG]]][G]	.10
9	G+G-G[[++GGG]++G-GG[++G-GG[+G-GG]]][G]	.08
13	G+G-G[[++GGG]++G-GG[++G-GG[+G-GG]]][-]	.08
15	G+G-G[[++GGG]++G-G[+ +G-GG[+G-GG]]][-]	.07
16	G+G-G[[++GGG]+ +G-G[-+G-GG[+G-GG]]][-]	.31
17	G+G-G[[++GG G]+G-G[-+G-GG[+G-GG]]][-]	.36
19	G+G-G[[+ +GG-]+G-G[-+G-GG[+G-GG]]][-]	.44
21	G+G-G[[+GG-]+G-G[-+G-GG[+G-GG]]][-]	.62
23	G+G-G[[+GG-]+G-G[-+G-GG[+G-G**+**G]]][-]	.65
24	G+G-G[[+G G-]+G-G[-+G-GG[+G-G+G]]]**G**[-]	.72
25	G+G-G[[+G-]+G-G[-+G-GG[+G-G+G]]]G[-]	.93
26	G+G-G**G**[[+G-]+G-G[-+G-GG[+G-G+G]]]G[-]	1.0

Figure 1. Phenotypes of the evolved genomes in Table I: regular to flatty and branchy (first row), and towards high structures (last row).

Having reached this point, the surprising fact is that an abrupt change in the fitness landscape was quickly solved by small modifications of the genome: deletions (mainly) and insertions. This rapid adaptation can be appreciated in Figure 1 (second row) and Figure 2 (curve after the dotted line).

4. CONCLUSIONS

The main consequence derived from the simulation results is that the proposed operators result in a low disruption of the phenotype and allow a gradual evolution of forms towards the target shaping.
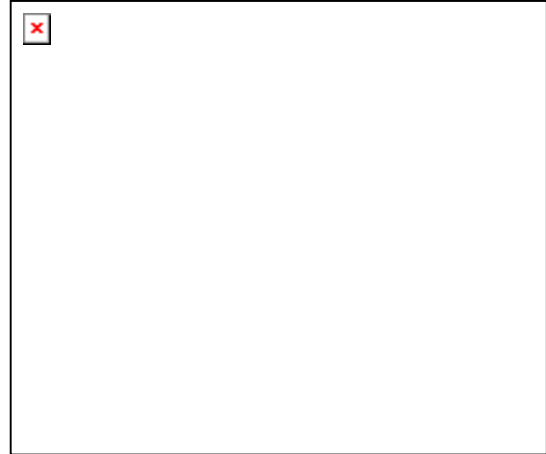

Figure 2. Evolution of the shape value for the two environments (delimited by a vertical dotted line).

Duplication, mutation and gene transfer play distinctive roles in the evolution, and then have different success depending on the precise time they are applied. Another consequence found in this study, directly derived from the type of operators used, is the generation of junk DNA. Only the portion of the genome that has been tuned to the fitness landscape is relevant to the structure, while a significant part of it (sometimes most of it) is of no use at all. Substrings like +-, [] or consecutive segments have no effect on the read out of the L-system, and could be deleted. Its importance, instead, is evolutive, since these symbols can play an important role in the future.

Gene duplication plays an important role in evolution [2]. Duplication of genes, and the consequent generation of paralogous genes, is a sort of silent mutation, in the sense that it is evolutionary neutral, and therefore free from selective pressure. These, and other mechanisms of DNA dynamics (like transposition), demand an extensive study in different frameworks of indirect encoding. Probably the knowledge obtained will be to a great extent linked to the definition of the particular framework, but it is possible that general principles can also be derived.

5. ACKNOWLEDGMENTS

6. REFERENCES

[1] Prusinkiewicz, P. and Lindenmayer, A. (1990). *The Algorithmic Beauty of Plants*. Springer-Verlag.

[2] Zhang, J. (2003). Evolution by gene duplication: an update. *Trends in Ecology & Evolution*, 18(6):292-298.