BioSystems 101 (2010) 187–193

Contents lists available at ScienceDirect

BioSystems

journal homepage: www.elsevier.com/locate/biosystems

A hypothesis on the role of transposons

Alessandro Fontana

IEEE, Via Morante 24, Vanzago, Italy

ARTICLE INFO

Article history:
Received 31 May 2010
Received in revised form 6 July 2010
Accepted 8 July 2010

Keywords:
Artificial Embryology
Evo-devo
Transposable elements
Junk DNA

ABSTRACT

Genomic transposable elements, or transposons, are sequences of DNA that can move to different positions in the genome; in the process, they can cause chromosomal rearrengements and changes in gene expression. Despite their prevalence in the genomes of many species, their function is largely unknown: for this reason, they have been labelled "junk" DNA. "Epigenetic Tracking" is a model of development that, combined with a standard evolutionary algorithm, become an evo-devo method able to generate arbitrary shapes of any kind and complexity (in terms of number of cells, number of colours, etc.). The model of development has been also shown to be able to produce the artificial version of key biological phenomena such as the phenomenon of ageing, and the process of carcinogenesis. In this paper the evo-devo core of the method is explored and the result is a novel hypothesis on the biological role of transposons, according to which transposition in somatic cells during development drives cellular differentiation and transposition in germ cells is an indispensable tool to boost evolution. Thus, transposable elements, far from being "junk", have one of the most important roles in multicellular biology.

1. Introduction

Previous work in the field of Artificial Embryology (see Stanley and Miikkulainen, 2003 for a comprehensive review) can be divided into two broad categories: the grammatical approach and the cell chemistry approach. In the grammatical approach development can be guided by sets of grammatical rewrite rules, instruction trees, or directed graphs. L-systems were first introduced by Lindenmayer (1968) to describe the complex fractal patterns observed in the structure of trees. The cell chemistry approach draws inspiration from the early work of Turing (1952), who introduced reaction and diffusion equations to explain the striped patterns observed in nature (e.g. shells and fur of animals). This approach attempts to simulate cell biology at a deeper level, going inside cells and reconstructing the dynamics of chemical reactions and the networks of chemical signals exchanged between cells.

"Epigenetic Tracking" is the name of an Artificial Embryology model applied to morphogenesis, i.e. the task of generating arbitrary 2d or 3d shapes, described in Fontana (2008). From this initial work, two lines of research are possible. One tries to make use of the method as a general-purpose tool to solve real-world problems; another line of research tries to bridge the gap between the model and real biology. This second research direction has been pursued in Fontana (2009, 2010), where the model has been shown able to provide insights into key aspects of biology such as the phenomenon of ageing and the process of carcinogenesis and will be continued in this paper, which deals with another feature of biological systems: the ubiquitous presence in the genomes of most species of pieces of DNA called "transposable elements", capable of moving between different chromosomal loci. This paper is organised as follows: Section 2 describes concisely the model of development; Section 3 describes a procedure called "Germline Penetration", which lies at the heart of the evo-devo process; Section 4 analyses analogies and differences between driver cells and stem cells; Section 5 outlines the biological role of transposons; Section 6 draws the conclusions.

2. The Model of Development

This section provides a concise description of the model of development. Shapes are composed of cells deployed on a grid; development starts with a cell (zygote) placed in the middle of the grid and unfolds in N age steps, counted by the variable "Age Step" (AS), which is shared by all cells and can be considered the "global clock" of the organism. Cells belong to two distinct categories: "normal" cells, which make up the bulk of the shape and "driver" cells, which are much fewer in number (a typical value is one driver for each 100 normal cells) and are evenly distributed in the shape volume. A shape can be "viewed" in two ways: in "external view" cells are shown with their colours (or grey shades); in "internal view" colours represent cell properties: blue (dark grey) is used for normal cells alive, orange (light grey) for normal cells just (i.e. in the current age step) created, grey for cells that have just died, yellow (very light grey) for driver cells (regardless of when they have been created).

E-mail addresses: alessandro.fontana@ieee.org, cnd12001@yahoo.it.

0303-2647/$ – see front matter © 2010 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.biosystems.2010.07.002

Driver cells have a genome and a variable called cellular epigenetic type (CET, an array of integers). While the genome is identical for all driver cells, the CET value is different for each driver cell. In this way, it can be used by different driver cells as a "key" to activate different portions of the genome. The CET value represents the source of differentiation during development, allowing driver cells to behave differently despite sharing the same genome. The genome is an array of "instructions", which are essentially models of biological genes. Each instruction is composed of a left (or "if") part, encoding a condition, and a right (or "then") part, encoding an action: once the condition is verified, the action is executed; this is in analogy with biological genes, which are transcribed once the transcription factors attach to their binding sites.

An instruction's left part is composed of the following elements: an activation flag (AF), indicating whether the instruction is "structurally" active or not; a variable called XET, of the same type as CET; and a variable called XS, of the same type as AS. At each step, for each instruction and for each driver cell, the algorithm tests if the instruction's XET matches the driver's CET and if the instruction's XS matches AS. In practise, CET plays more or less the role of a transcription factor which, upon binding to its genomic counterpart (the XET), triggers the action encoded by the gene; XS behaves like a "timer", imposing a further condition on the instruction activation, namely that the clock reaches a certain value. If both conditions are verified, the instruction's right part is executed. The right part codes for three things: change event type (EVT), shape (SHP) and colour (CO). Instructions give rise to two "types" of "change events": "proliferation instructions" produce "proliferation events", which cause the matching driver cell (called "mother cell") to proliferate in the volume around it (called "change volume") and "apoptosis instructions" produce "apoptosis events", which cause cells in the change volume to be deleted from the grid. The parameter "shape" specifies the shape of the change volume, in which the proliferation/apoptosis events occur, choosing from a number of basic shapes called "shaping primitives". In case of proliferation, the parameter "colour" specifies the colour of the newly created cells.

During proliferation, both normal cells and driver cells are created: normal cells fill the change volume and driver cells are "sprinkled" uniformly in the change volume. To each new driver cell a new, previously unseen and unique CET value is assigned (consider for example the driver cell labelled with A triggered to proliferate in Fig. 1), obtained by starting from the mother's CET value (the array [0,0,0] in the figure, labelled with A) and adding 1 to the value held in the ith array position at each new assignment (i is the current value of the AS counter). With reference to the figure, the new driver cells are assigned the values [0,1,0], [0,2,0], [0,3,0], . . ., labelled with B, C, D, etc. (please note that labels are just used in the figures for visualisation purposes, but all operations are made on the underlying arrays). In practise a proliferation event does two things: first it creates new normal cells and sends them down a differentiation path (represented by the colour); then it creates other driver cells, one of which can become the centre of another event of proliferation or apoptosis, if in the genome an instruction appears, whose XET matches such value. Fig. 1 reports an example of hand-coded (i.e. not evolved with an evolutionary algorithm) development.

The described model of development, coupled with a standard evolutionary algorithm, becomes an evo-devo method to generate arbitrarily shaped 2d or 3d cellular sets; specifically, a standard genetic algorithm is used. The choice to use a genetic algorithm is arbitrary: any evolutionary algorithm would produce the same results. The method evolves a population of genomes that guide the development of the shape starting from a zygote initially present on the grid, for a number of generations. At each generation, all genomes in the population (one at a time) guide the development

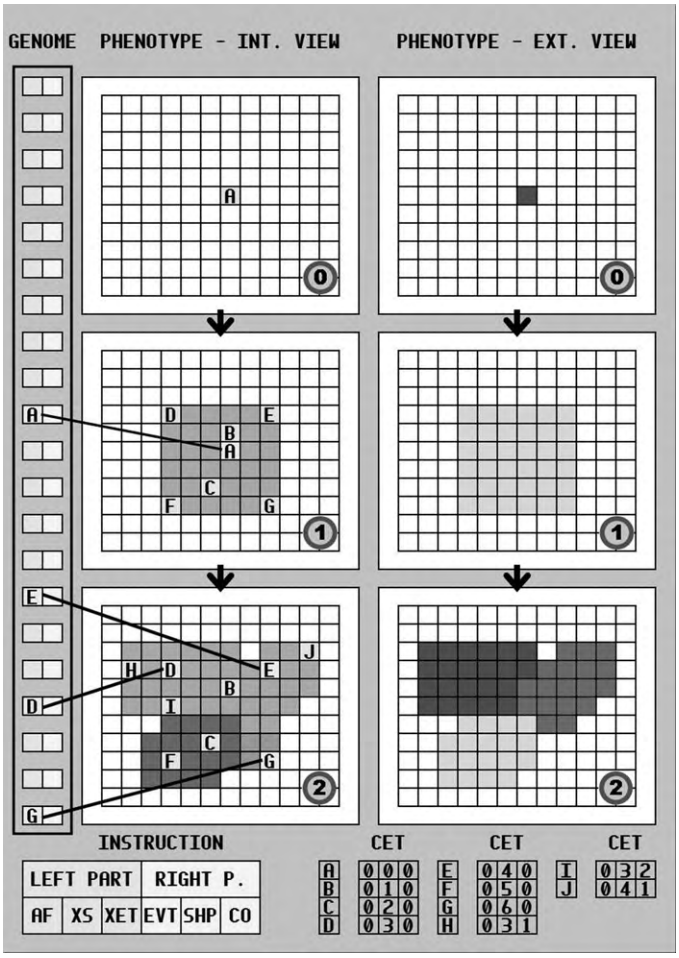

Fig. 1. Example of development in three steps (AS = 0,1,2): in step 1 the driver cell labelled with A is triggered to proliferate, in step 2 the driver cell labelled with D is triggered to undergo apoptosis and the driver cell labelled with E is triggered to proliferate. Internal view on the left, external view on the right.

of the shape from the zygote stage until, after a number of steps, a final shape is obtained. The adherence of such a shape to a predefined target shape is employed as fitness measure. This is repeated for all genomes, so that eventually each genome is assigned a fitness value. The genetic population is composed of 500 individuals (represented as strings of quaternary digits), undergoing elitism selection for up to 20–30,000 generations. The parameters of the genetic algorithm are 50% single point crossover and a mutation rate of 0.1% per digit. Different values for the parameters (e.g. 40% and 60% for the crossover probability, 0.05% and 0.2% for the mutation rate) have been tried and no significant change in performance has been observed. The fitness function formula is the same as that adopted by De Garis (1999):

$$F = \frac{\text{ins} - \text{outs}}{\text{des}} \tag{1}$$

where ins is the number of cells of the evolved shape falling inside (and matching the colour of) the target shape, outs is the number of cells of the evolved shape falling outside the target shape, and des is total number of cells of the target shape.

"In silico" experiments (see example in Fig. 2) have proved the effectiveness of the method in devo-evolving any kind of shape, of any complexity (in terms of e.g. number of cells, number of colours, etc.). Using shape complexity as a metaphor for organismal complexity, such simulations have established the potential of the method to generate the complexity typical for biological systems.