

Evolution of form and function in a model of differentiated multicellular organisms with gene regulatory networks

Daniel Lobo*, Francisco J. Vico

*Departamento de Lenguajes y Ciencias de la Computación, Universidad de Málaga
Severo Ochoa 4, 29590 Málaga, Spain
Phone: +34 952137037*

Abstract

The emergence of novelties, as a generator of diversity, in the form and function of the organisms have long puzzled biologists. The study of the developmental process and the anatomical properties of an organism provides scarce information into the means by which its morphology evolved. Some have argued that the very nature of novelty is believed to be linked to the evolution of gene regulation, rather than to the emergence of new structural genes. In order to gain further insight into the evolution of novelty and diversity, we describe a simple computational model of gene regulation that controls the development of locomotive multicellular organisms through a fixed set of simple structural genes. Organisms, modeled as two-dimensional spring networks, are simulated in a virtual environment to evaluate their steering skills for path-following. Proposed as a behavior-finding problem, this fitness function guides an evolutionary algorithm that produces structures whose function is well-adapted to the environment (i.e., good path-followers). We show that, despite the fixed simple set of structural genes, the evolution of gene regulation yields a rich variety of body plans, including symmetries, body segments, and modularity, resulting in a diversity of original behaviors to follow a simple path. These results suggest that the sole variation in the regulation of gene expression is a sufficient condition for the emergence of novelty and diversity.

Keywords: Evolutionary novelty, Artificial regulatory network, Artificial development, Boolean networks, Evolutionary computation

Accepted Manuscript

Lobo D. and Vico F.J.

“Evolution of form and function in a model of differentiated multicellular organisms with gene regulatory networks”

BioSystems 102(2-3), pp. 112-123, 2010.

DOI: 10.1016/j.biosystems.2010.08.003

*Corresponding author

Email addresses: dlobo@geb.uma.es (Daniel Lobo), fjvico@geb.uma.es (Francisco J. Vico)

1. Introduction

Nowadays, our planet is populated by some 1 to 20 million animal species. Quite remarkably, they represent less than 1% of the total number of animal species that have ever existed (Carroll et al., 2004). This astonishing diversity of forms and behaviors has emerged by the evolution of novel features among animal species, a process not fully understood yet, which remains as a fascinating and challenging topic of research (Carroll et al., 2004; Moczek, 2008). Biological evidence suggests that the sources of novelty might have to do with a complexification in the regulation of gene expression (Levine and Tjian, 2003). In this sense, it has been pointed out that evolutionary change in body plans devolves from change in the architecture of developmental regulatory programs (Davidson, 2006), suggesting that diversity can be better explained by variation in the regulation of gene expression than by variation in the structural genes (Davidson and Erwin, 2006). Moreover, the developmental process seems to be a key component in the evolution of diversity (Borenstein and Krakauer, 2008). However, due to the limitations to perform experiments in biological evolutionary processes, it has not been demonstrated yet that the reason for evolutionary emergence of developmental novel features and diversity is in fact the variation in the regulation of gene expression, rather than the variation in the structural genes.

On the other hand, theoretical models of biological phenomena are a valid alternative to experimentation, and have been extensively used to prompt new questions and research directions, especially in biological fields not suited to experimentation, such as evolutionary development. The work presented here subscribes to this approach. We show that a computer model, including genetic regulation of developmental processes, placed in a scenario of artificial evolution provides information about the evolutionary emergence of novelty and diversity.

Several theoretical models and formalisms have been proposed to describe genetic regulatory systems (see (de Jong, 2002) for a review). Among them, the Boolean networks proposed by Kauffman (1969) have been extensively used, and allow the simulation of large regulatory networks (de Jong, 2002). Furthermore, a recent study (Davidich and Bornholdt, 2008) has demonstrated a good correspondence between Boolean networks and more realistic models based on differential equations of chemical kinetics. Similarly to Boolean networks, other network-level models focus on a statistical analysis of network properties and patterns. When these models are embedded in an evolutionary context, mutation is typically implemented as changes in the connectivity and in the nodal output functions. These transformations have little to do with the effects derived from biological mutations and impose limitations to the way networks, and hence phenotypes, do evolve (Watson et al., 2004). Thus, in order to apply realistic mutation operators in network-level models, an encoding of the network in a sequence-based genome is needed. Among such models, the Artificial Genome proposed in (Reil, 1999) has attracted much attention. An Artificial Genome encodes a regulatory network in a sequence of digits, being the dynamics of this regulatory network equivalent to a Boolean network that limits the possible Boolean functions in its nodes (Willadsen and Wiles, 2003).

Similarly, theoretical models have also been proposed to model biological development, experiencing a considerable growth as a subfield of evolutionary computation. The main reasons of such advances are the benefits brought about by these models in scalability, adaptability, and evolvability (Hornby and Pollack, 2002) in a wide range of problems (see (Stanley and Miikkulainen, 2003) for a review). Within this emerging discipline, some models have been proposed at the network-level for developmental regulation. Fleisher and Barr (1993) presented a developmental model based on genetic encoding (hand-

coded), chemical diffusion, and mechanical interactions, formalized by ordinary differential equations, which were coupled with if-clauses for cell differentiation. Unfortunately, evolutionary developmental properties could not be studied, since this model was not embedded in an evolutionary process. Dellaert and Beer (1994) proposed a model where organisms are made up of two-dimensional squares, which develop by square division and differentiation through regulation by a Boolean network. Although the model included complex regulation, the phenotypes based on square divisions were inadequate for the emergence of novelty. Sims (1994b) presented a system for the evolution of physically-simulated virtual creatures made of articulated rigid parts, effectors, and sensors, and controlled by an extended neural network. Several tasks were optimized, resulting in a considerable variety of morphologies and behaviors. However, the morphology and the controller were encoded separately in two recurrent directed graphs, what does not really model biological development. Eggenberger (1997) described a growing phenotype made up of spherical modules, connected by articulated joints. A parametric regulatory network model was used, including diffusion concentration and diffusion sites of genes. The evolved forms presented limited variability, emerging only bilaterality. Bongard and Pfeifer (2003) extended that model by adding a neural controller, that was intended to evolve agents that developed directed locomotion and block pushing. The evolved agents managed to perform the assigned tasks, although with a limited variability in their characteristics. Hogeweg (2000) proposed a morphogenetic model of 2D multicellular organisms where cells behaved according to a multiscale cellular automaton. Although the phenotypes presented interesting developmental dynamics, the simplicity of the organisms made the results hard to use in studies of novelty. Kumar and Bentley (2003) proposed a computational model of development where a regulatory network controlled the synthesis of proteins, and embryos

with spherical forms were evolved. Here, again, the simplicity of the evolved phenotypes is not enough for studies of novelty emergence. Roth et al. (2007) presented a model of developmental multicellular organisms based on an artificial genetic regulatory network and chemical diffusion of morphogens. In this work, squares in a lattice represent cells that can differentiate into motors and sensors, connected by a simple wiring strategy. However, the model lacks an evolutionary component. Watson et al. (2008) proposed a model of artificial development and evolution of early land plants in 3D. This model employs an artificial genome to regulate the timing of bifurcation events and its rotation angles, yet the evolved phenotypes are too simple for the emergence of appreciable novelties. Doursat (2008) proposed a model of growing multicellular development, where a 2D lattice of cells proliferates and self-patterns into differential domains orchestrated by a gene regulatory network. Although the model produced substantial results, the process was not studied in an evolutionary perspective. Chavoya and Duthen (2008) proposed a model for 2D cell pattern generation based on a gene regulatory network, which controls a cellular automaton. The phenotypes generated by the model represented simple flag-like patterns, which are not adequate for novelty studies. Andersen et al. (2009) proposed a model of developmental cellular systems in 3D based on signaling and gene regulatory networks. Evolved embryos showed particular stable shapes and high capacity for self-repairing; however, the shapes presented by the phenotypes were too simple, rectangular or spherical, for the emergence of novelty. Finally, Zhan et al. (2009) presented an evolutionary developmental system based on cell signalling and artificial genetic regulatory networks focused on engineering design: electronic circuits design. In summary, the theoretical developmental models based on genetic regulation presented in the literature are not completely adequate for the study of the emergence of evolutionary novelty and diversity.

In this paper, we propose and analyze the results obtained by a theoretical model intended to gain further insight into the evolution of novel features and diversity. More precisely, these results suggest that the evolution of genetic regulation could be a sufficient condition for the emergence of novelty and diversity. The model is based on an Artificial Genome that encodes a Boolean network. Regulating the expression of a fixed elementary set of structural genes, the network controls the development of locomotive multicellular organisms. Organisms develop form and function simultaneously during the developmental process, resulting in a phenotype that integrates seamless morphology and control. An evolutionary algorithm is implemented to evolve organisms that succeed in following a path. We show that, despite the simplicity and invariability of the structural genes, the evolution of gene regulation yields a rich variety of novel body plans, including symmetries, body segments, and modularity. Moreover, the morphological diversity obtained yields a diversity of path-following behaviors.

Section 2 describes in detail the proposed artificial development model, from the description of the genome to the evolutionary algorithm. The results of the evolution are presented in Section 3. Finally, in Section 4 the conclusions derived from the results are discussed.

2. Description of the model

The model is described in several levels: (1) genome (encoding of the Boolean network), (2) development (genetic expression and cellular differentiation), (3) physical simulation of the organism and its interaction with the environment, and (4) the evolutionary algorithm.

2.1. Model of genome

The genome is represented by a vector of digits, similarly to the Artificial Genome proposed in (Reil, 1999), where template matching determines the encoded regulatory network (Fig. 1a). Digits in an artificial genome

correspond to bases in a real genome, and four bases are used (represented by decimal digits from 0 to 3). A gene is a sequence of digits of a specified length N (equal to four, in our case) preceded by a promoter (the word '0101'). A word placed between the promoter of a gene and the previous gene plays the role of regulatory region of the former gene. The product of a gene is a sequence of N digits, and it is obtained by increasing every digit in a gene by one and then taking the modulo over the number of bases. Gene products regulate only those genes whose regulatory region matches. A protein can act as an enhancer, activating the gene, or as an inhibitor, blocking its activation. Similarly to previous works, the rule implemented in this model is that proteins ending with the base 0 are inhibitory, otherwise they behave as enhancers. While the presence of a single enhancer will be enough to activate a gene, inhibition blocks enhancement.

The function implemented with this genome is equivalent to a Boolean network. Genes in the sequence correspond to nodes in the network, what limits the set of activation functions (Fig. 1b). Both, random Artificial Genomes and random Boolean networks have a number of out-connections distributed in a Poissonian way, but they differ in the distribution of in-connections: uniform in random Boolean networks, and Poissonian in random Artificial Genomes (Willadsen and Wiles, 2003).

Within this approach, the genetic information of an organism is encoded in an Artificial Genome. Not using a Boolean network directly has the advantage that a genomic representation allows the evolution of organisms with bioinspired mutation operators, instead of network-level mutations. In this way, mutations raise a wide variety of network-level changes, that finally project onto the morphology (Watson et al., 2004).

2.2. Model of development

In our model, an organism is represented by a connected and directed two-dimensional geometrical graph.

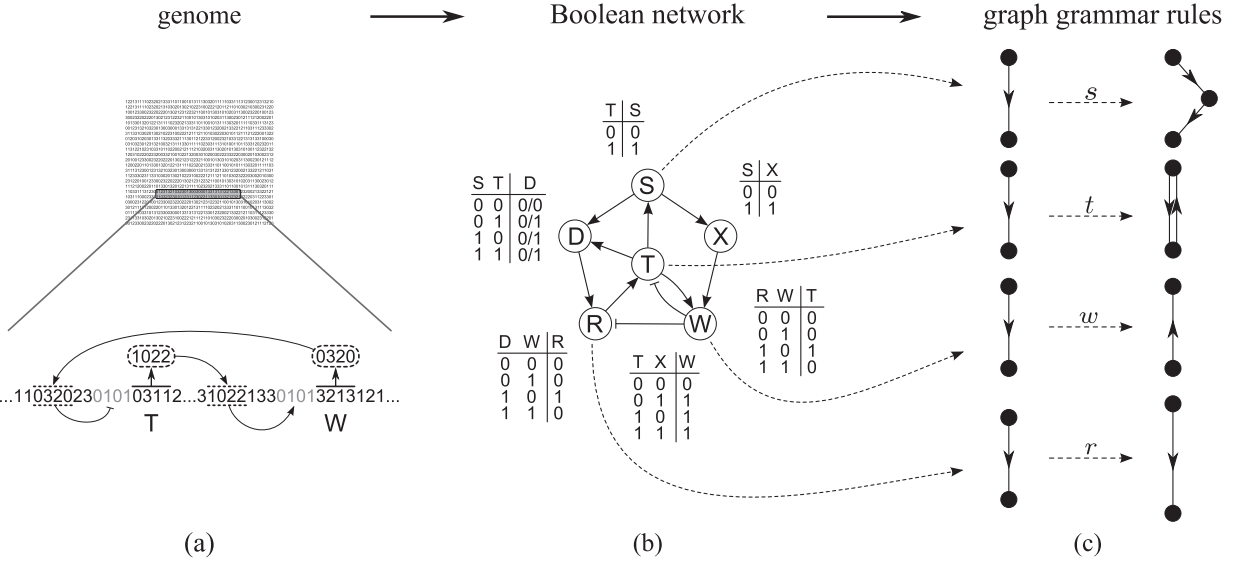


Figure 1: Morphogenetic model consisting on a derivation of a graph grammar regulated by a Boolean network encoded in a sequential genome. (a) The genome is represented by a sequence of digits. Below it is shown a detail of the sequence, with the regulation between genes T and W. Gene T is an enhancer of gene W while gene W is an inhibitor of gene T. (b) Boolean network encoded in the genome. Node D is the differentiation gene. Nodes S, T, W, and R are genes mapped to split, duplication, swap, and resize rules, respectively. Finally, node X is a regular gene (type node omitted for clarity). (c) Graph grammar rules set, being s the split rule, t the duplication rule, w the swap rule, and r the resize rule.

In this graph, edges can be interpreted as cells (black lines in Fig. 2, and green, red and blue-colored ellipsoids in Figures 4 to 7), and the nodes are junctions where cells get attached to each other (black dots in Fig. 2), in order to form a compact multicellular organism.

Edges (like cells do) perform developmental actions during embryogenesis, when structural genes get active. These actions are formally specified by a graph grammar. Each rule in the production system of this grammar corresponds to an action, and the language specified by the grammar determines the search space of possible morphologies. The proposed rules (Fig. 1c) have been chosen as to realistically match cellular transformations:

- Split rule (s): analogously to the mitosis process (responsible of the eukaryotic cell division), this rule rewrites one edge with two edges of half-length and perpendicular direction.

- Duplication rule (t): similar to the split rule, the duplication rule produces two edges, that have the same length and are arranged in parallel.
- Swap rule (w): this rule changes the direction of an edge, i.e. after this transformation the nodes of an edge remain connected, but in the opposite direction.
- Resize rule (r): similarly to the way in which real cells change their volume, this rule changes the length of an edge in a discrete way: a 25% increase or decrease in the original length.

Every edge in an organism embeds a Boolean network that regulates the application of the grammatical rules to transform the edge. The genome (and so, the connectivity of the Boolean network) is the same for all the edges in an organism, similarly to living organisms, where all the cells contain roughly the same genetic material.

Although all edges are governed by the same Boolean network, each one has its own expression state during development, thus allowing cell differentiation.

Each rule is controlled by a specific node (a structural gene) in the Boolean network. The genome specifies an order for the nodes, that are then mapped to the rules. During development, one edge is transformed according to a rule when the corresponding node of its Boolean network is activated. In this way, a step of development in an organism comprises the following actions: (1) updating the state of the Boolean network in each edge, and (2) transforming the edges according to their active nodes. The updating of the edges is done sequentially, from the oldest to the newest. Also, if several nodes are active in a Boolean network, the application of the rules is done in a sequential manner, following the ordering of the nodes.

The first node of the Boolean network is assigned a role of differentiation. When split or duplication rules are applied (division rules), one of the descendant edges will set this node to an active state, while the other resets it. The fixed mapping between nodes and rules is made starting from the second node. This asymmetry introduces a slight difference in the future expression patterns of both cells, allowing cell differentiation. In practice, the option of following or not differentiating ways is implemented by including two different versions (mapped to different nodes in the Boolean network) of the split and the duplication rules; one version sets different values for this node, and the other deactivates the node in both edges.

Similarly to differentiation in biological multicellular organisms, the model includes a cellular type that determines how the edge will behave in the physical simulation: motor edges, sensor edges, and structural edges. This is implemented again with a special node in the Boolean network, the type node, which determines the type of the edge. Each edge embeds a counter that accounts for the number of times this node has been active during

development. This counter acts as a signaler that induces the differentiation of the edge. One edge becomes a sensor if it has accumulated more than three quarters of the maximum activations of an edge in the organism. It becomes a motor edge if it accumulates less than a quarter of that amount (and it has been active at least once). And it becomes a structural edge otherwise. In this way, the function and the form of an organism are implemented by edges; consequently, the model makes no distinction between the control (i.e., how the function is commanded) and the morphology of the organism.

Development starts with a graph of a single edge (resembling the zygote in living beings). In grammatical terms, this graph is the axiom of the Boolean-network-regulated graph grammar, from where the resulting graph derives after a number of productions. The nodes of its Boolean network are initially inactive, except for the first node, which initiates the dynamics of the network. The developmental process ends when one of these conditions verifies: (1) all the edges have ended their expression (i.e., all nodes in the Boolean network of every edge remain inactive), (2) the expression of an edge enters in a loop without division rules, or (3) the organism has exceeded a given number of edges (in our simulations it is limited to 20 edges). Finally, mimicking biological competition and cellular death processes at the cellular level, the resulting graph is simplified by pruning duplicated edges (those that connect the same pair of nodes). If edges of different types connect the same pair of nodes, they are deleted in this order: structural, sensor and motor edges.

The described genetic expression model defines a family of infinite 2D connected graphs. Fig. 2 resembles the morphogenetic process of an organism regulated by the genome in Fig. 1. Each edge is labeled with its network state (the nodes are ordered 'DRSTWX'). The derivation starts with the graph to the left (the zygote). The first step is performed updating the state of this edge according to the Boolean network. This does not alter the

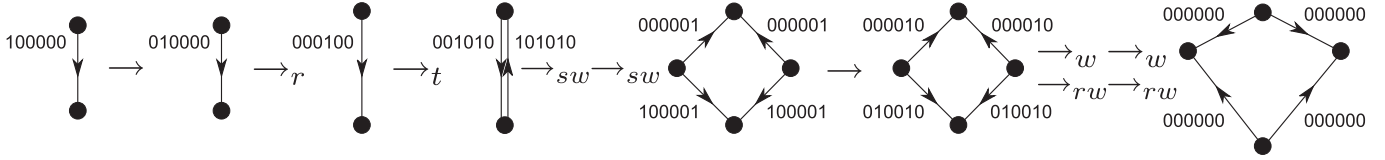


Figure 2: Example of morphogenesis of an organism regulated by the genome in Fig. 1. The first graph to the left is the zygote. Each edge has been labeled with its expression state.

graph but changes the state of its network. In this state, node R is active, so its corresponding rule (r , the resize rule) is applied to the edge in the next step, changing its length as showed in the third graph. Its state is updated again, and morphogenesis proceeds until a halting condition verifies.

2.3. Physics of the model

After development has completed, an organism is physically simulated in a flat world where they have to follow a path and go as far as possible in a constant time. An organism interacts with the environment by sensing and acting: it is propelled by its motor edges, and senses the path borders with its sensor edges, in a chemotactic way. Colliding has not been implemented in this virtual world.

The physics assigned to the graph are: edges have been modeled as damped springs, and nodes are free movable joints that have friction with the medium. All edge springs have the same physical parameters (spring and damping constants). The connectivity and rest lengths of springs are the connectivity and geometrical length of the corresponding edges in the developed graph. Spring dynamics are simulated according to the Hook's law, the damping force, and the friction with the medium. For the class of structures used in this work, a 4th order Runge-Kutta integrator is suitable for the former equations of motion in the physical world.

The three types of edges that can make an organism have different properties in the physical model. Apart from the forces of a damped spring, motor edges implement an additional force \vec{F}_{t+1} that pushes the edge in

the direction defined by the edge. The magnitude of this force is proportional to the actual length of the edge, accordingly to the following equation:

$$\vec{F}_{t+1} = \alpha L_t \vec{u}$$

being $\alpha > 0$ the motor strength parameter, L_t the length of the motor edge in the current time-step, and \vec{u} is the unitary vector of the direction. Consequently, the whole organism moves as a result of the motor edges pushing forward in a continuous way. Edges differentiated as sensors transduce the physical world information to the organism. They have the regular forces of a damped spring, but their spring rest length l_t is dynamically upscaled accordingly to the following equation:

$$l_{t+1} = (r_t (\beta - 1) + 1) l_o$$

being $\beta > 1$ a continuous gain parameter that regulates the upscaling factor, $r_t \in [0, 1]$ is the proportion of the edge that falls outside of the path in the current time step, and l_o is the original rest length of that edge in the graph that results from morphogenesis. In this way, if a sensor edge is completely inside of the path, its rest length equals its original rest length. On the contrary, when falling completely outside the path the rest length equals the original rest length scaled by the gain parameter β (so it gets bigger). Intermediate situations are linearly scaled by the amount of edge falling outside the path. Notice that a sensor edge transduces sensory information (how much it falls outside the path) to mechanical information (its rest length). This mechanical information is propagated to its adjacent edges, in the same way

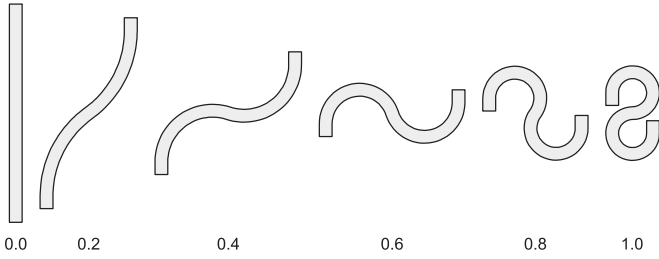


Figure 3: Difficulty of the path as given by the value of parameter γ (below).

as muscular cells propagates a change in length to adjacent cells. Finally, structural edges are normal springs without any particular effect.

The paths used in the simulations are made of two equal curves but in opposite directions discretized by a closed polygon. Each path curve is formed by two circular arcs that form the left and right path borders respectively. A path is defined by three parameters: $\gamma \in [0, 1]$ is its difficulty, ω is its width, and λ is its length. The difficulty determines the sharpness of the bend, being a path with $\gamma = 0$ a straight line, and a path with $\gamma = 1$ is the sharpest one (Fig. 3). The actual curves have an angle $a = \frac{3}{2}\pi\gamma$. The segments needed to build the whole path will have positive angle if the curve is to the left, and negative if it is to the right. A radius of $\frac{\lambda}{2a} + \frac{\omega}{2}$ units will apply for the external border, and $\frac{\lambda}{2a} - \frac{\omega}{2}$ for the internal border. Finally, the extremes of the path are extended with a beginning and an end (straight segments of length 2ω).

Evolution is intended to obtain efficient path-followers, i.e., organisms that, when placed at the beginning of the path, can follow it until the end. The fitness of an individual is determined by the length of path traveled in a constant simulation time. The path is divided into consecutive sections (similar to tiles) in order to quantify how well and far an organism moves along it. A simulation starts by developing the individual from its genomic information, and placing the resulting organism at the beginning of the path. The physics are then run for a

fixed number of steps, and it stops if the organism arrives in the end. During the simulation, a new section of the path is labeled as visited if the centroid of the organism (computed as the average position of its nodes) steps on it. In order to prevent high scores in organisms that do not interact with the environment (e.g. by starting with a trajectory that simply fits with the path), the fitness is the minimum between two runs: in the second simulation the path is flipped along the horizontal axis.

2.4. Evolutionary algorithm

A genetic algorithm has been implemented to evolve the structure and function of organisms. The initial population is made of 200 random organisms with short genomes (256 bases). On average, 256 bases contain just a single gene. In each generation only 25% of the population is mutated. Biologically inspired sequence-level mutation operators are used:

- Single-point: a single nucleotide is replaced by another nucleotide.
- Duplication: a segment of the genome is randomly chosen and copied immediately after the original (tandem duplication).
- Transposition: a segment of the genome is deleted and copied in a random location.
- Deletion: a segment of the genome is randomly chosen and removed.
- Inversion: a segment of the genome is randomly chosen and re-written in reverse order.

In all cases, the size of the segment to be mutated was fixed to 256 bases; however, genomes can vary their lengths during evolution, as a direct consequence of mutations. The mutant individuals obtained are added to the population, and the next generation is obtained by deterministic tournament selection with size 2. This scheme

induces a low selection pressure, what has been compensated by elitism of one individual. The result is a good balance between exploitation and exploration that favors the evolution of different strategies of locomotion.

3. Experiments and results

In order to evolve a variety of path-followers, the genetic algorithm has been run 21 times, comprising 7 evolutionary runs for each different path ($\gamma = 0.2$, $\gamma = 0.4$, and $\gamma = 0.6$). The total computing time was 175 hours in a cluster of 48 CPUs at 2 GHz. On average, 1 minute was the computing time for the creation and evaluation of one generation, being the physical simulation of the organisms the most time-consuming part.

In spite of the simple building blocks available for the organisms, 4 clearly different steering behaviors have evolved (classified by hand). Below, we present a representative organism for each resulting behavior, including their epigenic and ontogenic history, and their characteristic behavior. They have been labeled from A to D, with a short description of the particular technique that they use. We note that these techniques ‘emerge’, to clearly distinguish the evolutionary methodology used here from other approaches where a designer builds the parts in, as it is the case with classical Braitenberg’s Vehicles (Braitenberg, 1984). The relative frequencies of emergence in each behavior, among all evolutionary runs, were about 50, 10, 30, and 10 percent respectively. The best resultant organism in 76% of the evolutionary runs completed the entire path (fitness 1).

In these experiments we have observed a common pattern in the evolution of the population: initial growth of the length of the genomes, followed by a refinement of the structures. The initial population is made of very simple genomes (256 bases), that develop into simple organisms, limiting their structure and function. Hence, these organisms move very little or not at all, yielding low fitness values. Further duplications during the evolution allow

genomes to become larger, and the organisms get complex enough to start moving around. From this point, the increase in length slows down, and the solutions in the population starts getting refined. The evolutionary process stops when an organism has traveled the entire path or after reaching a maximum number of generations (1,000).

3.1. Behavior A: emergence of bilateral sensors

The simplest path-follower we can think of would include sensors in both sides to correct the direction, and a motor in between. This type of behavior has evolved on about half of the evolutionary runs. Fig. 4 describes the results of an evolution where a behavior inspired in this principle has been obtained. Fig. 4a shows the curves for the best and mean fitness in each generation. The trajectory of the organism while traveling the paths is shown in Fig. 4b, where the gray line represents the sequence of points drawn by the centroid. In both runs, after been positioned at the beginning of the path, the organism travels the paths and reaches the end. Note that, in the second case, the organism makes a large loop outside the path before reentering it with a different initial direction, what is not penalized, since the organism completes the entire path before the time expires.

The lineage of the best evolved organism in the experiment is shown in Fig. 4c. Lineages in these experiments are linear (instead of tree-like) since sexual reproduction has not been considered. Two aspects are worth mentioning in this figure. Firstly, the segmentation in the morphology of the ancestor in generation 40, which is made up of four equal segments, each of them consisting in two motor edges connected to two sensor edges. Secondly, as in the ancestor in generation 122, two main modules can be distinguished in the phenotypes: an upper module made of one structural edge connecting two motor edges which propel the organism forward, in addition to a second lower module made of two sensor edges that control

the direction of motion by varying their lengths. These two modules are connected through two structural edges in the ancestor in generation 122, in the same way that in the final organism; however, the ancestor in generation 152 displays the two modules directly connected. This is an example of the versatility of an indirect encoding based on regulatory networks, which allows the omission, modification through neutral mutations, and later reappearance of part of the organism’s morphology during the evolution.

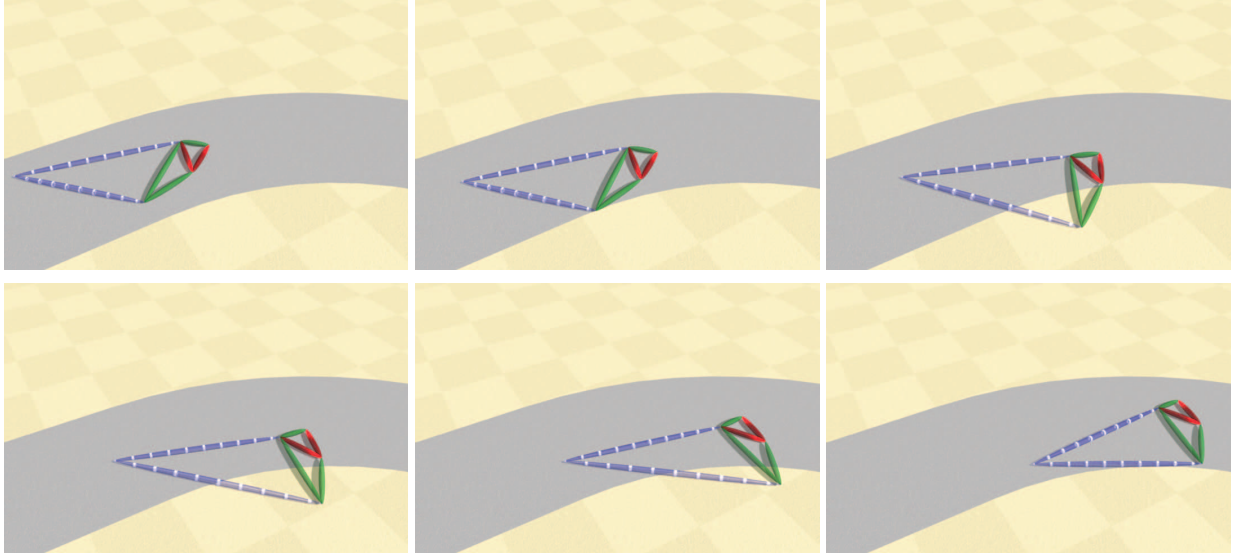
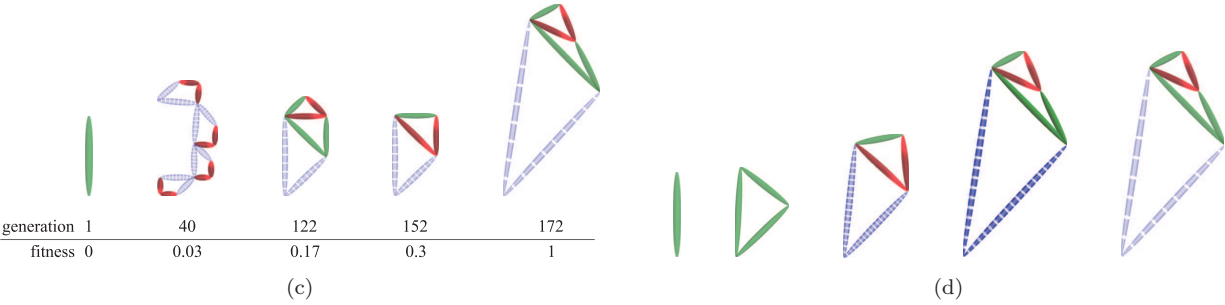
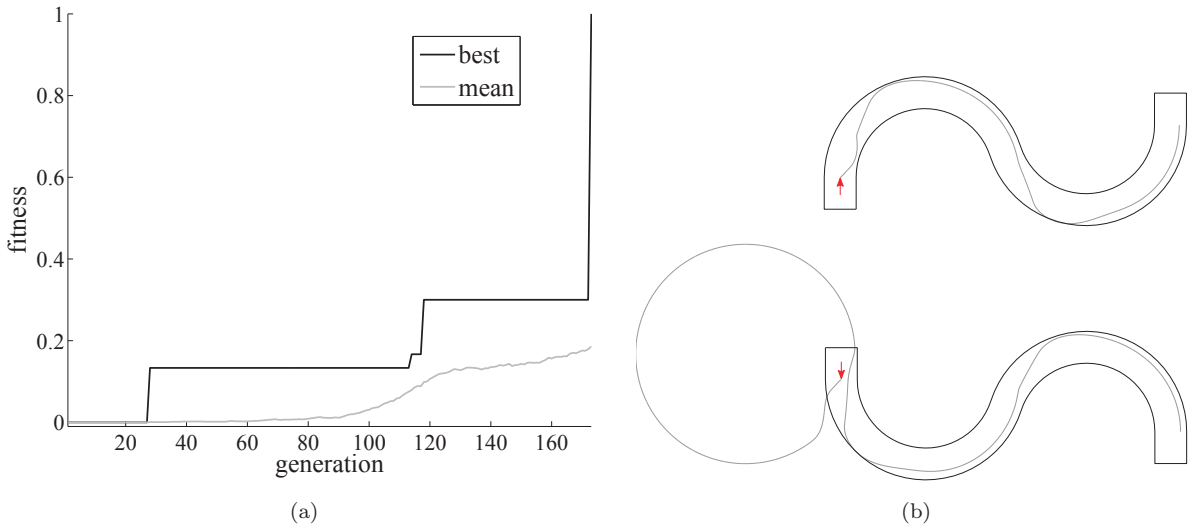
Fig. 4d shows the morphogenetic process (the development) of the best organism. Morphogenesis starts with the graph to the left, and edges are rewritten as the graph rules are applied according to the Boolean network that is encoded in the genome. This produces the graph in the right-side, its definite phenotype. More precisely, the morphology is developed by means of an edge’s activation of a duplication rule followed by a split rule during three consecutive steps. Each application of this pair of rules forms a new triangle-shaped module in the developing organism. Besides, in the third step, a resize rule is activated in both lower edges, increasing their lengths equally. During morphogenesis, four edges differentiate due to the differential activation of their type nodes, originating two sensor and two motor edges.

Finally, Fig. 4e illustrates the steering behavior with a sequence of snapshots (see also Movie S1, available as supplementary material). When the organism is on the path (in gray), the forces of its two motor edges are compensated, resulting in a straight movement. When one of the sides exits the path, the sensor becomes longer, transmitting a positional change to the motor edges. This corrects the direction of movement, pointing now to the interior of the path. This process repeats every time the organism transgresses a path border, allowing it to stay inside the path.

3.2. Behavior B: emergence of turning by friction

This is an interesting behavior that exploits a completely different aspect of the physics. The morphology of the organism integrates a more sophisticated sensory system (8 sensor edges), only one motor edge, and it shows symmetry with respect to the motion direction axis. It moves straightforward while inside of the path. When the organism starts exiting the path, the external skeleton of structural edges forces the sensors to reconfigure internally, and the symmetry breaks down due to the elongation of some sensors. In this asymmetrical configuration, more nodes concentrates in the side opposite to the exiting border, producing a higher overall friction on that side that generates a bent movement towards the path. When the organism gets back in the path it recovers the symmetry.

Fig. 5 shows an organism exhibiting behavior *B*. This and further figures are organized similarly to Fig. 4. Note that the ancestor in generation 94 almost managed to travel the full path, but it took 234 generations more to reach an organism with fitness 1. During this period mutations had a neutral effect, and the fitness increased when a mutation produced a bigger organism. Morphogenesis (Fig. 5d) reveals why becoming bigger was not easy. A straightforward way is to upsize the zygote, so that the resulting organism is proportionally bigger, but any change in an early stage will propagate through the morphogenetic process, amplifying its effect and disrupting the phenotype. Hence, it takes more time for evolution to modify these early stages of organism development. The final phenotype is composed of a central motor edge surrounded by four equal segments arranged in radial symmetry. Every segment grows from a single sensor edge, which in turn develops into a triangle formed by an external structural edge and two internal sensor edges (each sensor edge superimposes with a sensor edge of an adjacent module). It is also worth mentioning that the sensors develop separately, giving rise to four independent



(e)

Figure 4: Behavior A: bilateral sensors. (a) Best and mean fitness of the population in each generation during the evolution. (b) Paths and trajectories described by the best evolved organism. (c) Lineage of the best evolved organism. (d) Morphogenesis of the best evolved organism. (e) Illustration of the steering behavior as a sequence of snapshots (from left to right, and from top to bottom). Motor edges are represented in red, cyan edges are sensors (with white bands to better compare relative lengths), structural edges are the green ones, and the path is the area in gray. It can be seen how the elongation of the sensor that exits the path steers the pair of motor edges towards the path, correcting the direction and bringing the organism back to the path.

nodes in the center of the organism, what is important for the behavior in order to have enough difference of friction during its asymmetrical configuration to provoke the turn (only nodes are responsible for friction).

The trajectory followed by the organism and some snapshots are shown in Fig. 5b and Fig. 5e, respectively (see also Movie S2). The images illustrate how the initial straight movement changes when the sensors of the right segments elongate as they move away from the path. This makes the organism to turn to the left due to a higher friction on that side. When the organism is completely back on the path, it adopts again its symmetrical form and moves straightforward along the path.

3.3. Behavior C: emergent spinning

Contrary to what could be expected, the second behavior preferred by evolution had to do with spinning organisms. A combination of sensor and motor edges arranged in a sort of quadrilateral pattern favors a rotational movement. Typically, a small asymmetry is important for this type of organisms, since it allows it to actually start moving and reach a path border, preventing from an endless rotation around the starting point. In this way, when a path border is transgressed for the first time, the organism follows it due to an iterative elongation of its sensors during the rotation, while keeping its centroid inside the path most of the time.

Fig. 6 shows an organism that has implemented this behavior. Graphical results are arranged as in behavior A. Its lineage (Fig. 6c) features a structure simplification in generation 55, and almost the final structure reappears in generation 80. From there, small refinements are fixated during more than 300 generations to reach an organism with fitness 1. This way of evolving the general structure fast, and refine it slowly has been shown also in the former behavior (Fig. 5). In this case, the refinements also include a size increment implemented by a mutation that affected early stages of development.

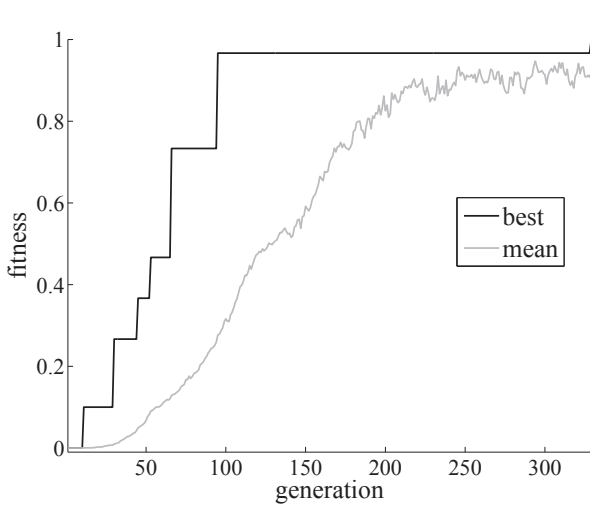
The morphogenesis of the final organism is divided in two phases (Fig. 6d). First, the zygote elongates and divides during two steps (new edges do not show in the figure because they are superimposed). Secondly, several edges split and turn, yielding the final structure: four segments distributed in two quasisymmetric pairs on both sides of an additional central motor edge. Each segment is made up of two connected edges: a motor edge whose direction is controlled by the elongation of a sensor edge.

Organisms showing spinning behavior describe a typical cyclic trajectory along one of the path borders during the simulation (Fig. 6b). Snapshots of the organism performing a complete spin are shown in Fig. 6e (see also Movie S3). In this case, the organism spins counter-clockwise, but clockwise spinning is also common in other experiments. Notice how in snapshot 1 the exterior motor edges are aligned at roughly 45° with respect to the interior motor edges. In snapshot 3 this angle has increased to some 90° due to the elongation of the sensor edges that fall outside the path. Repetitive transition between these two configurations allows the organism to steer following the path's border.

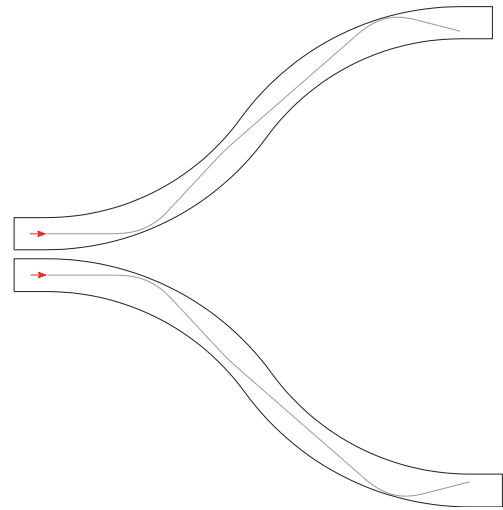
3.4. Behavior D: emergent rectification

Finally, some organisms revealed a much more elaborate behavior. Remarkably, this behavior emerged with the simplest possible sensory system: one sensor edge. While the organism is inside of the path, and as a result of balanced motor actions, all the edges get arranged in a single line, and the organism follows a straightforward movement. When the sensor exits the path, its elongation breaks the previous configuration, initiating a long sequence of actions (the rectification) which force the organism to go backwards, return to the path, and start another trajectory, shifted some degrees (around 40°) with respect to the original one.

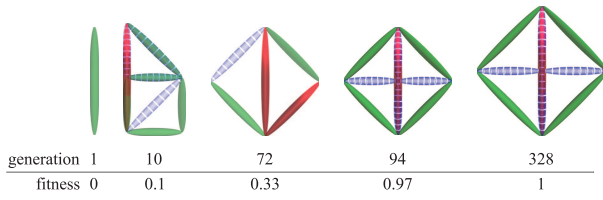
Fig. 7 shows an organism with such a behavior. The evolution stopped after reaching the maximum number



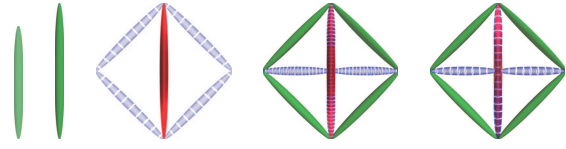
(a)



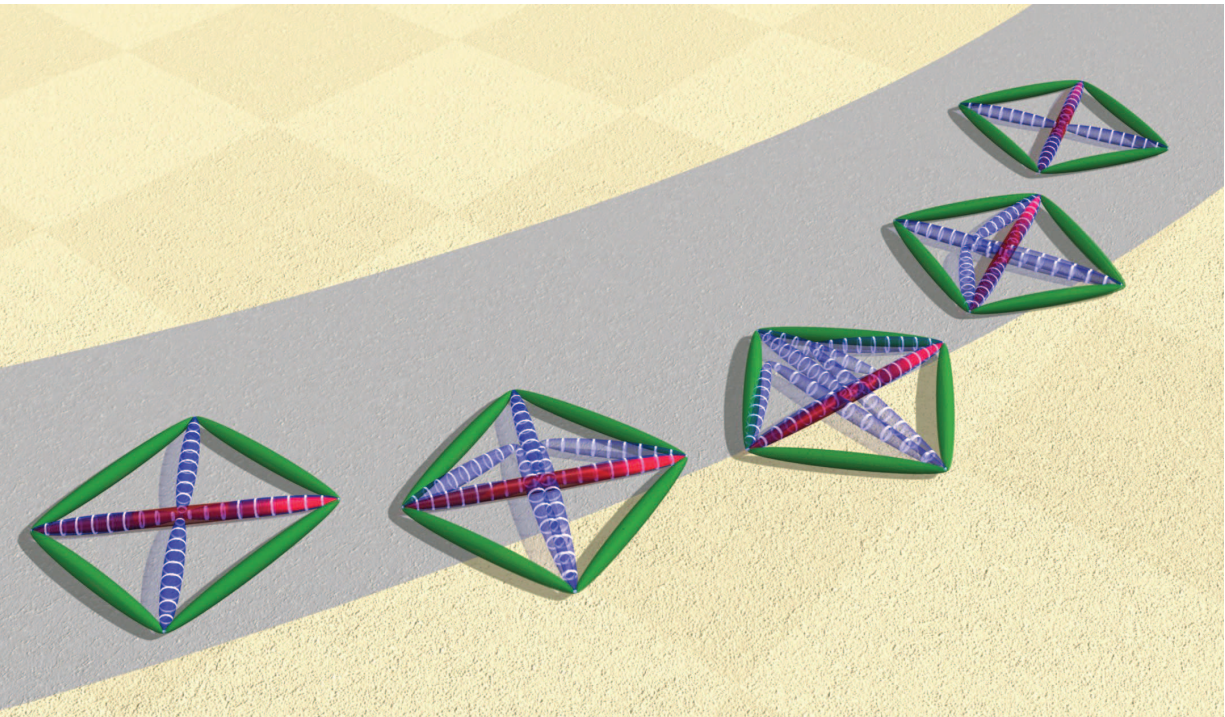
(b)



(c)



(d)



(e)

Figure 5: Behavior *B*: turn by friction. (a) to (d) as in Fig. 4. (e) Illustration of the steering behavior as a superposition of snapshots. As some sensors exit the path, their change in length pushes some nodes towards the path. This shifts the forces of friction in a way that corrects the direction of the organism, recovering its original configuration when it travels again over the path.

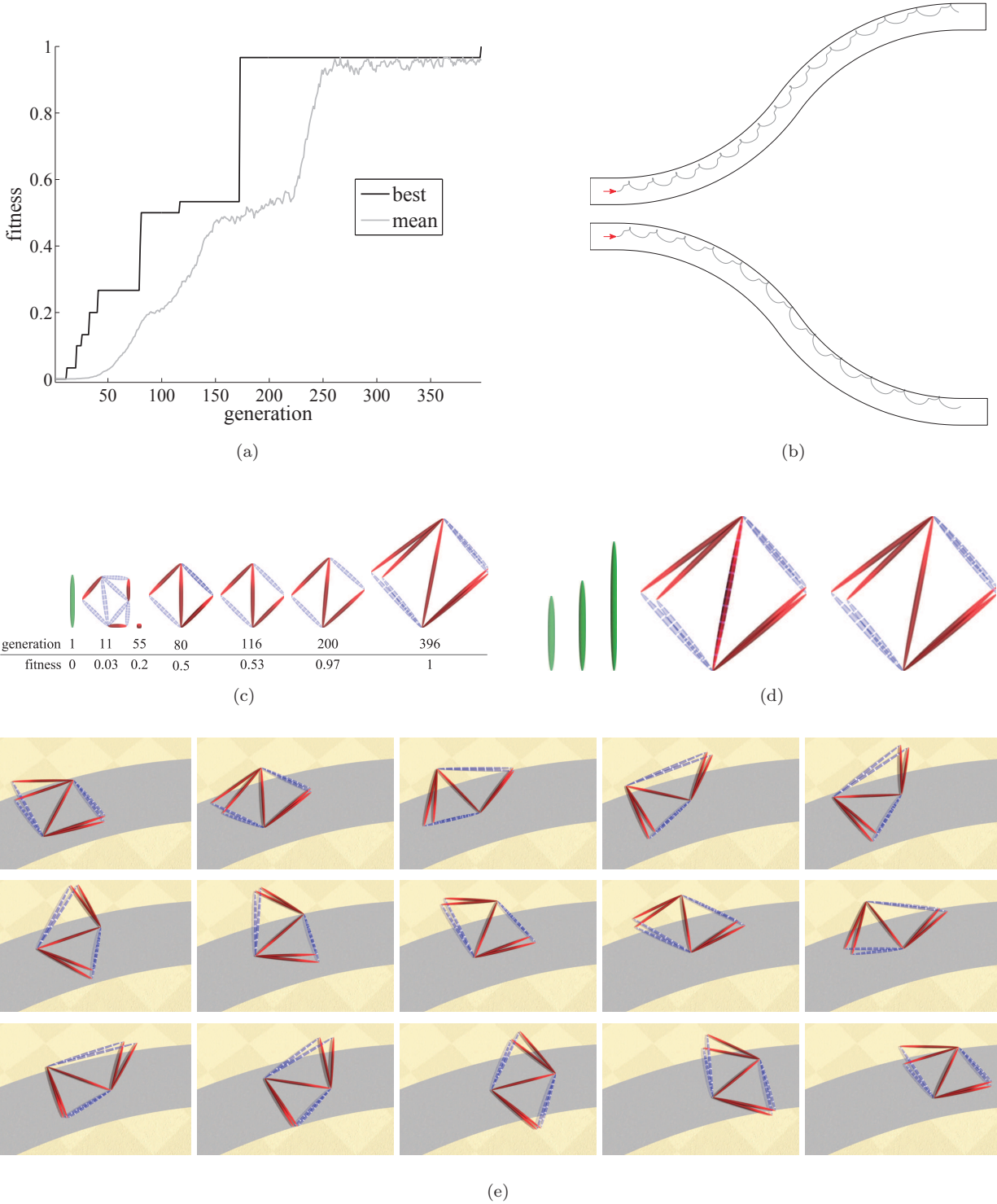


Figure 6: Behavior *C*: emergent spinning. (a) to (d) as in previous figures. (e) From left to right, and from top to bottom: one complete cycle in the counter-clockwise spinning behavior of the organism. Differently from other strategies, this one attaches the organism to one border of the path, changing the angles of the external motor edges with respect to the central motor edge. This angle varies from some 45° (snapshot 1) to about 90° (snapshot 4), providing the organism with a net movement that tracks the border of the path.

of generations. At that point, the best evolved organism had a fitness of 0.5 (Fig. 7a). Though, given enough simulation time, the organism managed to complete the path (Fig. 7b). This organism moves comparatively much slower than the others, and maneuvers in a complex way to correct the direction. The lineage shows a more diversified set of ancestors (Fig. 7c), what meant an intricate evolution, compared to previous behaviors. The morphogenesis (Fig. 7d) starts with a duplication phase during the first two steps, followed by a split-and-turn step, and finishes with some refinement. The final morphology shows two equal segments made up of two motor edges, and connected by a pair of structural edges, showing the ensemble bilateral symmetry. An extra homologous segment in the upper part of the organism breaks the balance of motor forces. This extra segment is attached to a sensor edge, whose elongation causes a rotation in the extra segment, leading to a steer which rectifies the trajectory.

Fig. 7e shows in detail how the organism performs the rectification (see also Movie S4). The structure of the organism includes two pairs of motor edges that push forward, and another pair that pushes backwards (hence the overall slowing down). The net effect results in an alignment of the edges, and a straight movement in the direction of the two pairs of motor edges. When the organism exits the path, the sensor elongates, forcing one of the leading pairs of motor edges to rotate and push backwards. In this configuration, the net movement is backwards, taking the organism back to the path. As the sensor enters the path again it shortens, provoking the pair of motor edges to return to its original aligned arrangement. While this happens, the organism tilts to one side, correcting the original direction. Finally, the organism keeps moving straightforward.

3.5. Generalization capability of the behaviors

Behaviors described above were obtained under particular settings (three path difficulties and constant fric-

tion). The resulting organisms have been simulated for a range of values of the difficulty and friction parameters in order to test the robustness of their behaviors. In each different setting, the organisms were simulated with a time limit of 5,000 steps. Fig. 8a shows the fitness (i.e., how much of the path was actually traveled) of the four organisms described in the previous section, along six different paths with a difficulty that increases from 0 to 1 (as showed in Fig. 3). This reveals that the more complex behavior is also the most robust to changes in the curvature of the path: behavior *D* performs well in any path, from the simplest to the most complex. Fig. 8b displays the average performance with different friction constant. In this case, the performance degrades in all cases as it gets more slippery, since sensor edges fail to steer when the motor edges propel the organism too fast. Behavior *A* is the only one that performs well for small friction.

Fig. 9a shows how behavior *D* deals with the two cases of maximum difficulty. Surprisingly, it can also generalize for very narrow paths, as shown in Fig. 9b. While the organism fails to track the path when it first exits it, it manages to wander around until it reenters the path, but this time it does it backwards (instead of the behavior showed in Fig. 7). Then it starts to travel the path going from one border to the other, until the end is reached (see Movie S5). Some behaviors demonstrate this ability to reenter the path after quitting it. What is remarkable because this capacity was not expected, since simulation time during the evolution is too short to develop reentering skills. This suggests that more complicated behaviors might be obtained if a longer time is allowed for fitness evaluation. Fig. 10 shows some snapshots illustrating how the organism with behavior *B* manages to reenter the path. The organism follows a straight trajectory when outside the path, with the sensory system arranged in a star-like configuration. As some sensors enter the path, their change in length pushes some nodes

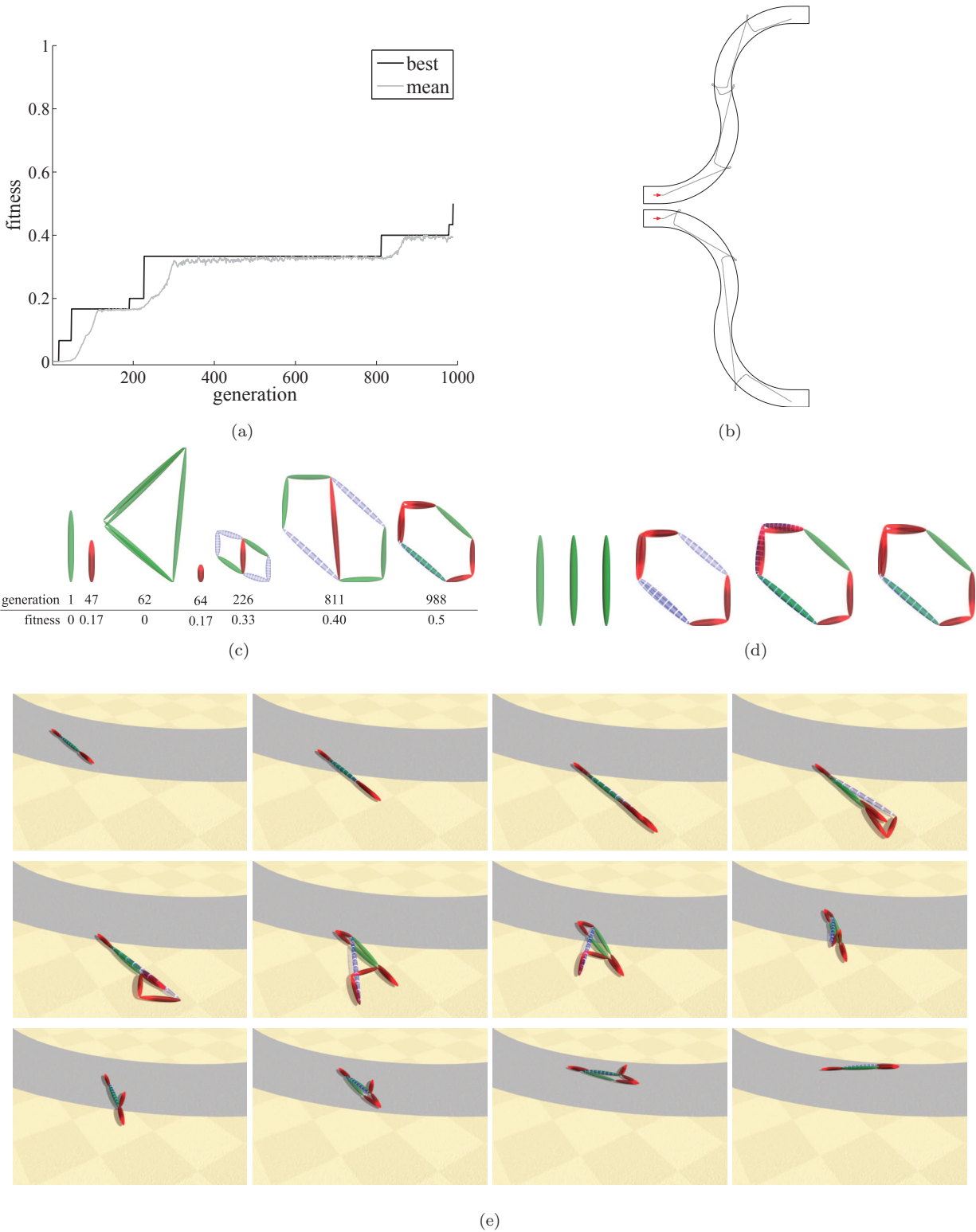
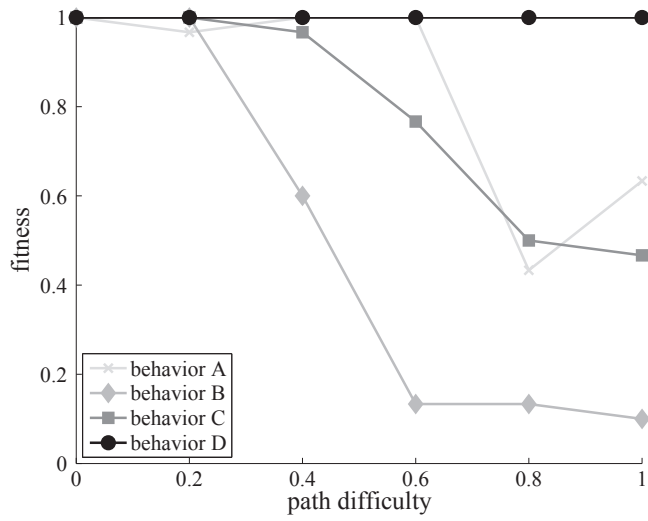
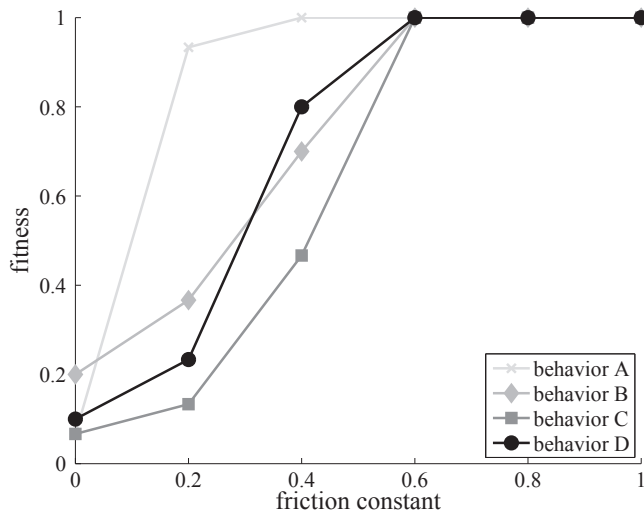


Figure 7: Behavior *D*: rectification. (a) to (d) as in previous figures. (e) From left to right, and from top to bottom: the organism moves straightforward while inside the path. When it exits the path the sensor elongates and provokes an unstable equilibrium, since a pair of motors moves forward. This equilibrium breaks at some point and forces the organism to return, since the pair of motors now points backwards. Back in the path, the sensor restores its rest length, and the pair of motor edges returns to the original position, correcting the previous direction, and keeping the organism traveling the path again.

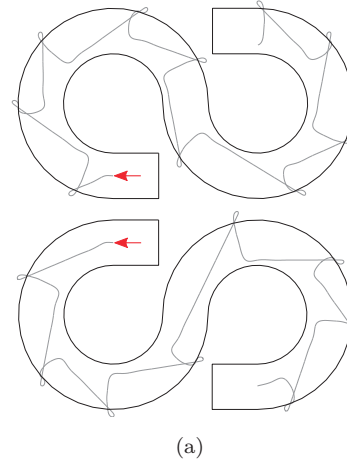


(a)

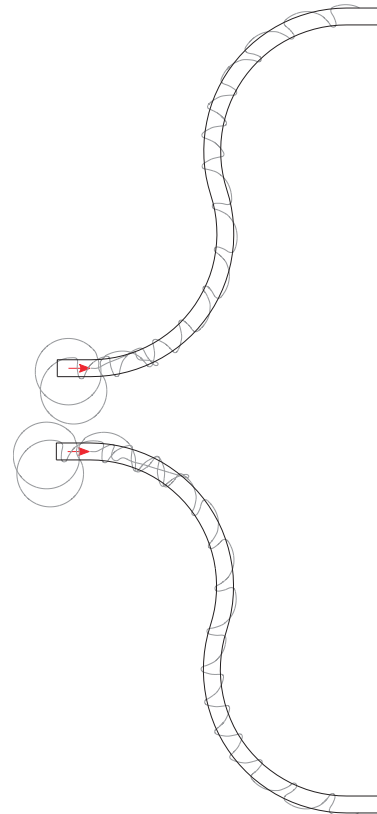


(b)

Figure 8: Study of generalization of the evolved behaviors in different environments. All the parameters are kept as they were during the evolution of each behavior, while path difficulty and friction of the medium are varied separately to find out the robustness of each organism to changes in the environment where it has evolved. (a) Performance of the behaviors for variable path difficulty. (b) Performance of the behaviors for variable friction constant.



(a)



(b)

Figure 9: Trajectories drawn by the organism that rectifies the trajectory (behavior D) in the cases of maximum difficulty of the path (a), and narrower path (b). Narrowing the path forces the organism to adopt a completely different strategy.

towards the path, changing friction forces and correcting the direction of the organism.

4. Conclusions and discussion

In order to provide arguments in support of the hypothesis that the evolution of developmental genetic regulation is indeed a sufficient condition for the emergence of novelty and diversity, we have defined and studied a model that integrates a considerable amount of biological features: (1) an encoding method based on sequence genotypes; (2) gene regulation by Boolean networks; (3) multicellular development through a fixed set of simple structural genes; (4) cell differentiation (three cellular types); and (5) evolution of morphologies and locomotive behaviors in a particular environment.

Despite the simple and fixed set of structural genes implemented, a rich variety of body plans have evolved, providing the organisms with appropriate steering strategies. In a recent work the authors have shown that symmetries, segments, and modules do not emerge when spring networks are encoded with direct methods (Lobo et al., 2010). This suggests that these properties, found in the described phenotypes, could emerge as a consequence of the genetic regulation of development that has been modeled.

Furthermore, apart from the rich variety of body plans, the evolved organisms presented a rich variety of behaviors. Indeed, the problem of path-following employed to test the model is unconnected to the traditional class of form-finding problems, in which the fitness of a solution is directly evaluated from its form. Instead, it is related to a new class of problems where the behavior is to be optimized, which we have named *behavior-finding* (Lobo, 2010). Behavior-finding problems search for solutions whose behavior (as the result from the interaction of its form with an environment) verifies a set of restrictions. Consequently, the fitness of a solution is evaluated as how it behaves in a concrete environment. Following a

path is but an example of this class of problems. The developmental model presented in this work has been shown to be adequate for the class of behavior-finding problems, as the resulting behaviors have demonstrated.

The behaviors obtained are diverse and complex, and successfully exploit very different aspects of the model: sensorial systems adapted to the geometry of the problem (behaviors A and B), physical aspects of the environment (behavior B), symmetry (behaviors B and C), or complex arrangements of edges (behaviors C and D). This diversity is remarkable if we attend to the simplicity of the resulting graphs (7 edges for behavior A , 17 for B , and 9 for C and D). Considering the very limited functionality of the cellular types that have been modeled, the performance and generalization capacity of the evolved organisms result from the expressive power of the genetic model and the high degree of adaptation to the environment reached by the organisms. The fact that behavior D can be obtained with only 9 edges and a single sensor edge is amazing considering the efficiency and generalization capacity demonstrated by this organism. Fig. 9b shows that this structure manages to travel backwards when the path is too narrow to be traveled forward. This suggests that the evolved structures implement not only the steering behavior, but also the ability to wander around the path, and reenter the path. In short, a remarkable result is how such a biological model obtains very simple structures that show a very complex behavior.

In the field of autonomous agents, an embodiment is employed to allow agents to interact with the environment. Such embodiments have been traditionally split in morphology and controller (Gruau, 1994; Sims, 1994a; Koza, 1995; Dellaert and Beer, 1996; Bongard and Pfeifer, 2001; Komosinski and Rotaru-Varga, 2002; Hornby et al., 2003), being the controller typically implemented by a neural network, and adjusted separately from the morphology. In contrast, the proposed model does not include an explicit controller, i.e., there is no clear sepa-

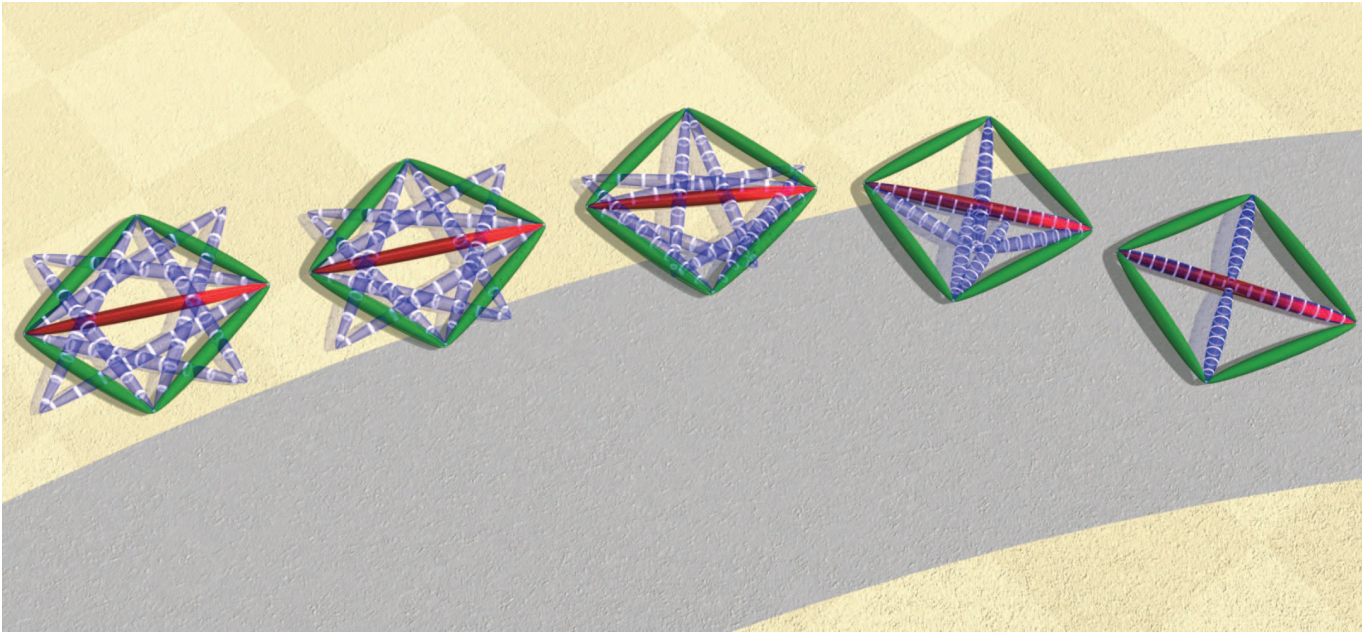


Figure 10: Capacity of behavior B to reenter the path after quitting it. The organism changes from a straight trajectory to a curve towards the path as some sensors reenter the path. Their change in length displaces some nodes towards the path, and the imbalance of friction forces tilts the organism towards the path. Once it travels again over the path, the organism recovers its symmetry and continues straightforward.

ration between the body and the brain that controls its behavior. A sensor is implemented in our model as an excitable element that alters its rest length depending on its position relative to the path. In this way, sensors transduce information of the environment by introducing a change in the geometrical state of the organism. This change propagates along the organism to adapt the response in a proper way, so as to accomplish what is favored by evolution: steering to keep following the path. Furthermore, since the controller and morphology are merged in the model, both of them develop seamless in the same process, simplifying the model as a reliable abstraction of biological development.

In conclusion, the theoretical results presented in this work support the latest biological hypothesis, suggesting that the sole variation in the regulation of gene expression is indeed a sufficient condition for the emergence of novelty and diversity of body plans. While the proposed model used a fixed and simple set of structural

genes that fired just a few basic developmental actions, the evolution of the regulation of gene expression originated a diversity of body plans, which managed to solve a path following problem with a rich variety of behaviors. This quite striking diversity of forms and behavioral strategies is connected to the diversity of forms and features found on earth, suggesting that such morphological richness should not be considered a surprising fact, but rather an inevitable consequence of the variability of gene regulation.

Acknowledgments

This work has been partially funded by the Sixth European Union Framework Program for Research and Technological Development, contract #028892. The authors are grateful to V. Canteli for graphical support and the anonymous reviewers for helpful comments and suggestions.

References

- Andersen, T., Newman, R., Otter, T., 2009. Shape homeostasis in virtual embryos. *Artificial Life* 15 (2), 161–183.
- Bongard, J. C., Pfeifer, R., 2001. Repeated structure and dissociation of genotypic and phenotypic complexity in artificial ontogeny. In: *The Genetic and Evolutionary Computation Conference, GECCO-2001*. Morgan Kaufmann publishers, pp. 829–836.
- Bongard, J. C., Pfeifer, R., 2003. *Evolving Complete Agents Using Artificial Ontogeny*. Springer-Verlag, pp. 237–258.
- Borenstein, E., Krakauer, D. C., 2008. An end to endless forms: Epistasis, phenotype distribution bias, and nonuniform evolution. *PLoS Computational Biology* 4 (10), e1000202+.
- Braitenberg, V., 1984. *Vehicles: Experiments in Synthetic Psychology*. The MIT Press.
- Carroll, S., Grenier, J., Weatherbee, S., October 2004. *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design*, 2nd Edition. Wiley-Blackwell.
- Chavoya, A., Duthen, Y., 2008. A cell pattern generation model based on an extended artificial regulatory network. *Biosystems* 94 (1-2), 95–101.
- Davidich, M., Bornholdt, S., 2008. The transition from differential equations to boolean networks: A case study in simplifying a regulatory network model. *Journal of Theoretical Biology* 255 (3), 269–277.
- Davidson, E. H., 2006. *The Regulatory Genome: Gene Regulatory Networks In Development And Evolution*. Academic Press.
- Davidson, E. H., Erwin, D. H., 2006. Gene regulatory networks and the evolution of animal body plans. *Science* 311 (5762), 796–800.
- de Jong, H., 2002. Modeling and simulation of genetic regulatory systems: a literature review. *Journal of Computational Biology* 9 (1), 67–103.
- Dellaert, F., Beer, R. D., 1994. Toward an evolvable model of development for autonomous agent synthesis. In: Brooks, R., Maes, P. (Eds.), *Artificial Life IV*. MIT Press Cambridge.
- Dellaert, F., Beer, R. D., 1996. A developmental model for the evolution of complete autonomous agents. In: *From animals to animats 4: Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior*. MIT Press, pp. 393–401.
- Doursat, R., 2008. Organically grown architectures: Creating decentralized, autonomous systems by embryomorphic engineering. In: Würtz, R. P. (Ed.), *Organic Computing*. Springer-Verlag, pp. 167–199.
- Eggenberger, P., 1997. Evolving morphologies of simulated 3d organisms based on differential gene expression. In: *the Fourth European Conference on Artificial Life*. MIT Press, pp. 205–213.
- Fleischer, K., Barr, A. H., 1993. A simulation testbed for the study of multicellular development: The multiple mechanisms of morphogenesis. Addison-Wesley, pp. 389–416.
- Gruau, F., 1994. *Genetic micro programming of neural networks*. MIT Press, Cambridge, MA, USA, pp. 495–518.
- Hogeweg, P., 2000. Evolving mechanisms of morphogenesis: on the interplay between differential adhesion and cell differentiation. *Journal of Theoretical Biology* 203 (4), 317–333.

- Hornby, G. S., Lipson, H., Pollack, J. B., 2003. Generative representations for the automated design of modular physical robots. *IEEE Transactions on Robotics and Automation* 19 (4), 703–719.
- Hornby, G. S., Pollack, J. B., 2002. Creating high-level components with a generative representation for body-brain evolution. *Artificial Life* 8 (3), 223–246.
- Kauffman, S., 1969. Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of Theoretical Biology* 22 (3), 437–467.
- Komosinski, M., Rotaru-Varga, A., 2002. Comparison of different genotype encodings for simulated three-dimensional agents. *Artificial Life* 7 (4), 395–418.
- Koza, J. R., August 1995. Gene duplication to enable genetic programming to concurrently evolve both the architecture and work-performing steps of a computer program. In: *IJCAI-95 Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*. Vol. 1. Morgan Kaufmann, Montreal, Quebec, Canada, pp. 734–740.
- Kumar, S., Bentley, P. J., 2003. Biologically inspired evolutionary development. In: *Evolvable Systems: From Biology to Hardware*. pp. 99–106.
- Levine, M., Tjian, R., 2003. Transcription regulation and animal diversity. *Nature* 424 (6945), 147–151.
- Lobo, D., 2010. Evolutionary development based on genetic regulatory models for behavior-finding. Ph.D. thesis, Universidad de Málaga.
- Lobo, D., Vico, F. J., Dassow, J., 2010. Graph grammars with string-regulated rewriting. Submitted to *Theoretical Computer Science*.
- Moczek, A. P., 2008. On the origins of novelty in development and evolution. *BioEssays* 30 (5), 432–447.
- Reil, T., 1999. Dynamics of gene expression in an artificial genome - implications for biological and artificial ontogeny. In: *European Conference on Artificial Life*. pp. 457–466.
- Roth, F., Siegelmann, H., Douglas, R. J., 2007. The self-construction and -repair of a foraging organism by explicitly specified development from a single cell. *Artificial Life* 13 (4), 347–368.
- Sims, K., 1994a. Evolving 3d morphology and behavior by competition. *Artificial Life* 1 (4), 353–372.
- Sims, K., 1994b. Evolving virtual creatures. In: *SIGGRAPH '94: Proceedings of the 21st annual conference on Computer graphics and interactive techniques*. ACM, New York, NY, USA, pp. 15–22.
- Stanley, K., Miikkulainen, R., 2003. A taxonomy for artificial embryogeny. *Artificial Life* 9 (2), 93–130.
- Watson, J., Geard, N., Wiles, J., 2004. Towards more biological mutation operators in gene regulation studies. *Biosystems* 76 (1-3), 239–248.
- Watson, J., Hanan, J., Wiles, J., 2008. Modeling the fitness of plant morphologies across three levels of complexity. *Biosystems* 94 (1-2), 182–190.
- Willadsen, K., Wiles, J., 2003. Dynamics of gene expression in an artificial genome. In: *2003 Congress on Evolutionary Computation*. IEEE Press, pp. 185–190.
- Zhan, S., Miller, J. F., Tyrrell, A. M., 2009. An evolutionary system using development and artificial genetic regulatory networks for electronic circuit design. *Biosystems* 98 (3), 176–192.