

Facilitating Evolutionary Innovation by Developmental Modularity and Variability

René Doursat

Institut des Systèmes Complexes, CNRS / CREA, Ecole Polytechnique
57-59, rue Lhomond
75005 Paris, France
+33 1 42 17 09 99

rene.doursat@polytechnique.edu

ABSTRACT

Natural complex adaptive systems show many examples of self-organization and decentralization, such as pattern formation or swarm intelligence. Yet, only multicellular organisms possess the genuine *architectural* capabilities needed in many engineering application domains, from nanotechnologies to reconfigurable and swarm robotics. Biological development thus offers an important paradigm for a new breed of “evo-devo” computational systems. This work explores the evolutionary potential of an original multi-agent model of artificial embryogeny through differently parametrized simulations. It represents a rare attempt to integrate both self-organization *and* regulated architectures. Its aim is to illustrate how a developmental system, based on a truly *indirect* mapping from a modular genotype to a modular phenotype, can facilitate the generation of variations, thus structural innovation.

Categories and Subject Descriptors

I.2.11 [Artificial Intelligence]: Distributed Artificial Intelligence – coherence & coordination, intelligent agents, multiagent systems

General Terms

Algorithms, Design, Experimentation, Reliability, Theory.

Keywords

Artificial Embryogeny, Evolutionary Development, Bio-Inspired Engineering, Spatial Computing, Complex Systems, Systems Design, Self-Organization, Modularity, Robotics, Architectures.

1. TOWARD EVO-DEVO ENGINEERING

Engineering is torn between two seemingly contradictory objectives. On the one hand, it strives to deliberately design devices and systems with full control of their structure, on the other hand, it wishes them to be functionally and structurally autonomous, self-repairing and adaptive—or “intelligent”. Today, even our most sophisticated contraptions (computers, AI, robots) must be literally “spoon-fed” at every stage of their existence, i.e., entirely designed,

built, and programmed, constantly monitored, repaired, and upgraded. We are still far from the heralded future of engineering in which artificial systems would exhibit all the highly desirable “self-*x*” properties that are found in many natural systems: *decentralization* (emergence, redundancy), *autonomy* (self-organization, homeostasis), *adaptation* (learning, evolution).

1.1.1 Inspiration from Natural Complex Systems

While we are wondering how to emancipate human-made systems, our natural environment already hosts a profusion of autonomous systems, whether physical patterns, biological cells, organisms, animal societies, ecosystems, or—in a broader sense of “natural”—spontaneously emerging super-structures indirectly caused by humans (societies, the economy, Internet, etc.). These decentralized, unplanned systems are probably the most pervasive, efficient, and robust type of structures. It is centrally planned systems that are unique, costly to build, and fragile, because they fundamentally require another “intelligent” (but sometimes incompetent) system to exist and operate them.

In this context, natural “complex adaptive systems”, in particular biological (developmental, neural, evolutionary) and emergent-social (complex networks), have an important role to play in providing a powerful source of inspiration for emerging technologies. Understanding these systems could help create a new generation of artificial systems with the above-mentioned properties still largely absent from traditional engineering. Now, taking a closer look at natural complex systems, it is a striking fact that they are all made of a myriad of elements that interact locally in large networks and produce an emergent collective behavior at a macroscopic scale. These elements follow individual rules or laws of dynamics that can be more or less sophisticated.

Thus decentralization over a myriad of agents seems to be an essential condition for autonomy in natural systems. To some extent, this is already mirrored by current trends in information and communication technologies. Segmentation and distribution of large computing systems over a multitude of smaller and relatively simpler components is both a growing need and an inevitable fact in many domains of computer science & engineering, AI, and robotics. Faced with an explosion in size and complexity of computing systems at all levels of organization, including hardware (integrated parts), software (program modules), and networks (applications and users), engineers are gradually led, more or less willingly, to rethink these systems in terms of complex systems [27], [20]. It means that, instead of rigidly designing things in every detail, engineers would need to step back and only “meta-design” them, i.e., focus on

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

GECCO'09, July 8–12, 2009, Montréal, Québec, Canada.
Copyright 2009 ACM 978-1-60558-325-9/09/07...\$5.00.

generic conditions allowing endogenous growth, function, and evolution.

1.1.2 The Need for Precise Morphogenetic Abilities

Yet, as the prolific creativity and spontaneity of natural complex systems is examined, one realizes that in most cases they do not exhibit the truly *morphogenetic* and *architectural* capabilities that are needed in machine design. Relatively complex collective behavior can result from simple agent-based rules, a fact often touted as the hallmark of complex systems, however, it is “complex” only to a certain point. In many well-known families of systems (e.g., pattern formation, swarm intelligence, complex networks [1]), the emerging patterns are “statistical” in nature, i.e., they are either mostly stochastic or mostly shaped by boundary conditions, or a mix of both (e.g., spots and stripes, ant trails, meandering bird flocks, hub-forming social encounters).

Complex systems that are “morphological” in nature, i.e., exhibit an *intrinsic architecture* that is neither repetitive nor imposed by the environment, are much less frequent or studied. One monumental exception is the development and evolution of biological forms, especially multicellular organisms. Embryogeny represents a unique case of combined self-organization *and* elaborate structures. Organisms are made of segments and parts arranged in specific ways that resemble engineered devices. Yet, they also self-assemble in a decentralized fashion, under the precise control of genetic and epigenetic information stored in the zygote. This endows cells with a repertoire of non-trivial behaviors and allows them to differentiate depending on a notion of relative positional information. In other words, biological development is a prime example of *programmable* complexity. It demonstrates that complex systems can be much more than “homogeneous”, “monolithic” or “random”: they can contain a wide diversity of agents creating *heterogeneous* patterning and shaping; they can be *modular*, hierarchical, and architecturally detailed at multiple scales; they can be *reproducible* because they rely on programmable agents.

There is a great demand for such precise self-formation capabilities in a variety of distributed engineering systems, and morphogenetic approaches have also been proposed in some applications: arrays of micro-processors [8], mobile sensors [2] internet security hosts [15] reconfigurable modules [14], [18], [16], or swarm robots [6]. It is also an important challenge in complex techno-social networks made of myriads of mobile devices, software agents and human users, all relying on local rules and peer-to-peer communication. How to foster a programmable, yet adaptive design of self-reconfiguring manufacturing plants, self-regulating energy grids, or self-deploying emergency taskforces? [12]

1.1.3 Let Grow, then Evolve

Traditional engineering is currently following a “deliberate design” (DD) paradigm. It is expected to progressively shift toward a “biological” paradigm, in which systems (organisms) are the pure products of undesigned evolution (UE). This would proceed via intermediate stages of “developmental meta-design” (DMD) and “evolutionary meta-design” (EMD). In DMD, designers focus on creating generative mechanisms rather than the systems themselves. In EMD, even more disengaged meta-architects create laws of variation and selection, prepare primitive ancestor systems and step back to let evolution invent the rest.

In the variation/selection couple of Darwinian evolution, variation has become the poor child of biology’s Modern Synthesis, while most of the attention was focused on selection. It is only in recent years that the understanding of variation as the *generation of phenotypic* innovation became the main concern of evolutionary development, or “evo-devo”, a rapidly expanding field of biology comparing the development of different species from the molecular and genetic level to the anatomy [1], [17]. The link from genotype to phenotype cannot remain an abstraction if we want to unravel the generative laws of evolution [17]—and eventually transfer them to self-organized artificial systems. On the artificial side, there is a similar paradigm shift, as new trends in evolutionary engineering, such as artificial embryogeny (AE) [23], [3], [19] also emphasize developmental variations as a prerequisite for the evolution of complex structures.

In the framework of genetic algorithms and evolutionary computation, this means an *indirect* or *implicit* mapping (as opposed to direct or explicit) from genotype to phenotype. The advantages of a developmental representation that are generally put forward [23], [3], [13] are compactness and reduction of search space (how to describe complex structures using the fewest parameters possible), reusability and modularity (how to describe structures as a hierarchy of parts and subparts), self-organization and adaptability (how the phenotype can be influenced by the environment in which it grows). Most importantly, there is a consensus that developmental systems are especially well positioned to exhibit a high degree of *evolvability*. It is felt and, in some examples demonstrated [16], that they inherently have greater variational and adaptive power, as they allow combinatorial tinkering on highly redundant parts. However, there is no unified or standard evo-devo methodology. Rather, there exist many different evolutionary computation works of developmental inspiration, from abstract grammars to biologically detailed models (see a taxonomy in [23]).

The present work wishes to contribute to the field of artificial embryogeny by exploring the evolutionary potential of an original model of architected morphogenesis. It represents a rare attempt to integrate both self-organization *and* sophisticated, regulated architectures. In particular, its aim is to illustrate how a self-organized developmental system, based on a modular genotype giving rise to a *modular phenotype*, can facilitate the generation of variations, thus evolutionary structural innovation. Part 2 summarizes and elaborates upon the developmental mechanisms taken from the “embryomorphic” systems of Doursat [9], [10], [11], while part 3 showcases an array of evolutionary experiments based on hand-made mutations of the genotype. Part 4 outlines the potential of this approach and its future directions.

2. A MODEL OF SELF-ORGANIZED MODULAR MORPHOGENESIS

Embryomorphic systems rely on a computational, multi-agent model of artificial embryogeny that integrates *self-assembly* (SA) and *pattern formation* (PF)—metaphorically, “sculpting” and “painting” [7]—in a decentralized swarm of interacting cells. The embryo’s shape or *phenotype* of the collective behavior is guided by a *genotype G* stored inside each cell. *G* provides a compact developmental representation of the system, i.e., a set of genetic parameters that rely on generative developmental rules for their interpretation and transformation into an organism (see 2.1 for a list of these mechanisms). *G* is composed of pairs of specialized

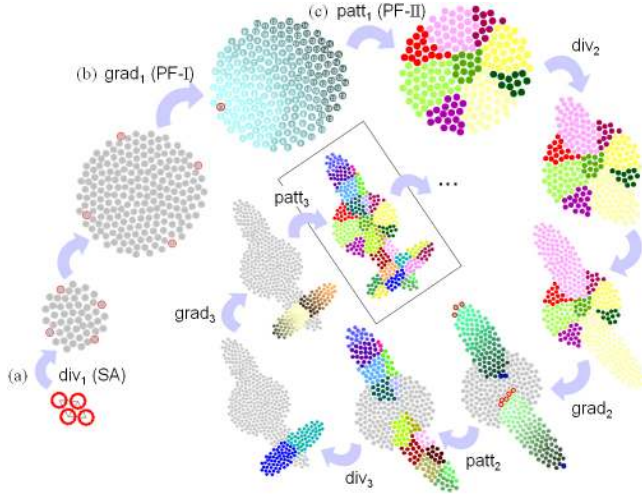


Figure 1: A three-stage embryogeny simulation (see text).

parameter subsets, (G_{SA}, G_{PF}) , that respectively direct SA and PF behavior and also influence each other. On the one hand, the mechanical properties of a cell (SA: division rate and adhesion forces) are determined by its local surrounding region of genetic identity, or “cell type”, within the global pattern of gene expression. On the other hand, these patterns of identity regions (PF: gene regulation network) further segment themselves into subregions, subsubregions, etc., under the expansion and deformation caused by mechanical constraints. The complete genotype G of a cell is made of multiple *modules* $\{(G_{SA}^{(k)}, G_{PF}^{(k)})\}$ that are organized in a tree graph and create complex morphologies by recursive refinement of details. At any time, each cell c of the embryo is characterized by a set of state variables $Q^c(t)$, also composed of SA-specific and PF-specific subsets: $Q^c(t) = (Q_{SA}^c(t), Q_{PF}^c(t))$.

A global view of the embryogeny cycle is given in Figure 1. Starting from a small clump of cells, three main operations are applied iteratively and recursively: (a) “div”, representing the SA behavior of the swarm: the division, proliferation and spatial rearrangement of cells; (b) “grad”, representing the first half of the PF behavior: the diffusion of morphogen proteins across the swarm (coded with integer counters; see below); and (c) “patt”, representing the second half of the PF behavior: the readout and transformation of the gradient signals into various regions of gene expression levels (via the response of a gene regulatory network; see below). In the numerical simulation, these three operations run in parallel: cells continuously divide (div) while at the same time they exchange gradient signals (grad) and immediately interpret these signals to calculate their new type (patt). A fundamental aspect of the embryomorphic model is its hierarchical organization in developmental *stages*, where each stage repeats the execution of the (div, grad, patt) trilogy at a lower spatial scale. The example of Figure 1 shows three such stages: $(div_1, grad_1, patt_1)$, $(div_2, grad_2, patt_2)$ and $(div_3, grad_3, patt_3)$. Stage 2 is applied to two of the nine regions that appeared during stage 1, forming “limbs” that extend up and down. Stage 3 is executed inside two of the eight regions of the lower limb, creating two “digits”. This recursive process may continue to later stages, in a multiscale, fractal fashion resembling generative

grammatical systems, such as L-systems [22]—although in a self-dissimilar way (and with other differences described in part 4).

2.1 Developmental Mechanisms and Genetic Parameters

2.1.1 Self-Assembly by Division and Adhesion (SA)

An embryomorphic swarm in 2-D is composed of cells with neighbor interactions calculated from a Delaunay-Voronoi tessellation. They follow two major laws of cellular biomechanics in a simplified format: (i) cell *division*, coded by a uniform probability p for any cell to split into two, and (ii) cell *adhesion*, represented by elastic forces derived from a quadratic potential V with resting length r_e , hard-core radius r_c , and scope of visibility r_0 [24] (Figure 2a). Under these laws, cells tend to form a quasi-regular hexagonal mesh. At any time, each cell c is characterized by its 2-D coordinates (x, y) and its interaction links with neighboring cells $\{l^{cd}\}$, where $l = 0$ or 1. An extra parameter g regulates the proliferation time: cells stop dividing when one of their gradient values reaches g (see next section). In summary, the SA genetic parameters (common to the whole region) and the SA state variables (in each particular cell c) are:

- $G_{SA} = (p, r_c, r_e, r_0, g)$
- $Q_{SA}^c(t) = (x, y, \{l^{cd}\})(t)$

2.1.2 Positional Information by Gradients (PF-I)

The model distinguishes between two types of PF signals that cells continuously exchange and process: *gradient* variables (PF-I) and *pattern* variables (PF-II). Gradient values propagate from neighbor to neighbor and establish *positional information* [26] across the swarm (Figure 2b). Each cell receives a counter g from neighboring cells and increments it by 1, resulting in a wave pattern of counter values [8], [2], [21]. Two waves W and E running in opposite directions create a midline WE where cells have approximately equal g_W and g_E values. The model uses four gradients forming a coordinate system based on two midlines WE and NS (the four sources W, E, N, S are not placed by hand but self-position by migrating away from each other inside the swarm). These midlines in turn become sources and generate two roughly perpendicular planar gradients, alternatively denoted by X and Y . In summary, there are no PF-I genetic parameters, while the PF-I state variables in each particular cell c are:

- $G_{PF-I} = \emptyset$
- $Q_{PF-I}^c(t) = (g_W, g_E, g_{WE=X}, g_N, g_S, g_{NS=Y})(t)$

2.1.3 Patterning by Gene Expression Levels (PF-II)

Pattern values correspond to gene expression levels calculated on top of the X and Y gradient values to create different *cell types* (Figure 2c). This calculation relies on a gene regulatory network (GRN), whose weights constitute the genetic parameters of the PF process. Patterning represents the emergence of *heterogeneity*, i.e., the segmentation of the swarm into “identity regions” corresponding to high levels of expression of particular genes I_k . A well-known example is the early striping of *Drosophila*, controlled by a five-layer hierarchy of segmentation genes along the anteroposterior axis (maternal, gap, primary and secondary pair-rule, segment polarity) [1].

The present model relies on a three-layer caricature of the same principle in 2-D, along the two intersecting X and Y axis. The

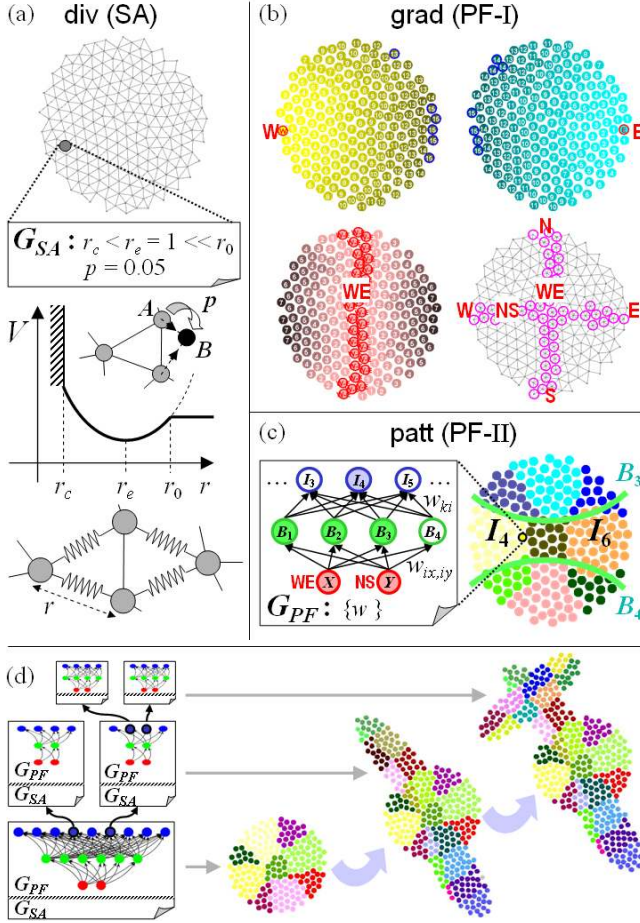


Figure 2: Overview of the developmental mechanisms of the multiagent model (see text, part 2). (a) Self-assembly by division and elastic interactions (top: mesh of edges). (b) Gradient diffusion. (c) Patterning by gene regulation. (d) A 3-stage genotype gives rise to a 3-stage phenotype.

bottom layer of the GRN contains the two positional variables g_x and g_y . The middle layer contains “boundary” genes B_i that segment the embryo into horizontal and vertical half-planes of strong and weak expression levels via 2-D step functions: $B_i = \sigma(w_{ix}g_x + w_{iy}g_y - \theta_i)$, where $\{w_{ix}, w_{iy}\}_{i=1..n}$ are the regulatory weights from g_x and g_y to B_i , parameter θ_i is B_i ’s threshold and σ is a sigmoid function $\sigma(u) = 1/(1 + e^{-\lambda u})$. The top layer contains the identity nodes I_k derived from positive and negative products of B_i ’s, i.e., various intersections of the B_i half-planes: $I_k = \prod_i w_{ki} B_i + (1 - w_{ki})/2$, where $w_{ki} \in \{-1, 0, +1\}$ are ternary weights from B_i to I_k , meaning that B_i ’s factor inside I_k can be either $(1 - B_i)$ or B_i . Two of these regions, I_4 and I_6 , are the location of further growth and patterning in stage 2 in most of the examples presented here. In summary, the PF-II genetic parameters (common to the whole region) and the PF-II state variables (in each particular cell c) are:

- $G_{PF-II} = (\{w_{ix}, w_{iy}, w_{ki}, \theta_i\})$
- $\mathcal{Q}_{PF-II}(t) = (\{B_i, I_k\})(t)$, where $k' = \operatorname{argmax} \{I_k\}$ is the cell’s type (and codes for its color in the figures)

2.1.4 Simultaneous Growth & Patterning (SA & PF)

SA and PF behaviors are combined to create growing patterns at every stage. Cells continually adjust their positions according to the elastic SA constraints, *while* continually exchanging gradient values and PF signals over the same dynamic links. This dual dynamics is guided by the combined genotype $G = (G_{SA}, G_{PF})$, where $G_{PF} = (G_{PF-I}, G_{PF-II})$. Daughter cells inherit all the attributes of mother cells, including G and the internal \mathcal{Q}_{PF} variables (current gradient counters and gene levels). The \mathcal{Q}_{SA} variables (coordinates and edges) are recalculated from a position close to the original cell. Both sets of variables are immediately updated, as the newly born cell starts contributing to SA forces and the traffic of PF gradients that maintain the pattern’s consistency at all times in the swarm.

2.1.5 Modular, Recursive Development (SA & PF)- k

Embryological patterns do not develop in one shot but in numerous incremental stages. An adult organism is produced by modular, recursive growth and patterning [7], [1]. In *Drosophila*, regions of the embryo that acquire leg, wing or antenna identity (“imaginal disc”) develop local coordinate systems of morphogen gradients to support the pre-patterning and construction of the planned organ. The present model includes a pyramidal *hierarchy* of network modules able to generate patterns in a recursive fashion (Figure 2d). First, the base network $G_{PF}^{(0)}$ establishes main identity regions, then subnetworks $G_{PF}^{(k)}$ triggered by the identity nodes I_k further partition these regions into smaller, specialized compartments at a finer scale. Other shapes than blobs can also be created by inhomogeneous and anisotropic SA dynamics, i.e., by varying and diversifying the cells’ G_{SA} parameters (cell division p and elastic adhesion V) as a function of their PF type and spatial position. For example, limb-like structures can grow by imitation of meristematic plant offshoots. In this process, only the tip or “apical meristem” of the organ is actively dividing at any time. (Cells forming the tip can self-identify as the local maxima of a gradient generated by the base of the limb.)

3. THE GENERATION OF VARIATION BY MODULES

The embryomorphic model presented in Part 2 demonstrates that entire *complex, programmable* and *reproducible* morphologies can develop and self-organize from a handful of cells following local rules of reproduction and differentiation. These morphologies are architecturally complex because they can be made of any variety of modules and parts that are not necessarily repeated in any periodic or self-similar way. They represent programmable phenotypes because they emerge from the same given genotype carried by every cell of the swarm. They are reproducible, as their structure and shape are not left to chance but tightly controlled by the genotype. Naturally, the exact positions of cells at the microscopic level are still random, but not the mesoscopic and macroscopic regions that they form.

Moreover, the modularity of the phenotype is a direct reflection of the modularity of the genotype. The hierarchical SA and PF dynamics recursively unfolds inside the different regions and subregions that it creates. Each module $G^{(k)} = (G_{SA}^{(k)}, G_{PF}^{(k)})$ can have different internal genetic SA and PF parameters, potentially giving each region a different morphodynamic behavior and

different gene activity landscape. The integration between SA and PF is controlled by the identity nodes I_k : these nodes switch on the execution of subordinate modules $G^{(k)}$, i.e., their gene expression activity (parametrized by $G_{PF}^{(k)}$) to create new local segmentation patterns, and their mechanical behavior (parametrized by $G_{SA}^{(k)}$) to create new morphodynamical behaviors.

This part presents a few experiments involving hand-made *mutations* of the genotypes of embryomorph systems and their corresponding phenotypes. For now, these systems are purely developmental and do not serve a specific function. No organism “fitness” is defined (neither structural, nor functional) and no selection is performed during a systematic search. The goal here is rather to illustrate the link between genotype modularity and phenotype modularity, and the programmable and predictable effect that mutating the former can have on evolving the latter.

Figures 3-6 show several examples of modular embryogeny and how certain mutations in the genotype correlate with quantitative or qualitative changes in the phenotype. The organism of Figure 3a is taken as the reference or “wild type”. Its genotype is composed of two different modules: a base module establishing the body plan (lower module) and a specialized module in charge of growing a limb-like appendage (upper module). As described previously, each module consists of two types of “genes” or genetic parameters: self-assembly genes G_{SA} , coding how cells divide and spread spatially, and pattern formation genes G_{PF} , coding how cells acquire their types. In the simplified representation of Figures 3-6, the gene regulatory network of G_{PF} is not shown. Instead, only the type of checkered pattern it produces (explained below) and the switch identity genes are displayed.

In the example of Figure 3a, the G_{SA} part of the body plan imposes a uniform division rate $p = 0.05$ and a stop signal at gradient value $g = 15$. This means that each cell divides on average every twentieth time step, and stops dividing when one of their gradient counters g_W, g_E, g_N or g_S reaches 15 (via a stop

signal that the border cells propagate across the swarm). This results in a regular disc-shaped swarm about 15 cells wide. Meanwhile, the G_{PF} part of the same body-plan module produces a 3×3 checkered pattern, i.e., nine different types of cells $I_1 \dots I_9$. (Incidentally, this relies on two horizontal and two vertical boundary genes, B_1, B_2 and B_3, B_4 , not shown here; see Figure 2c). Then, during stage 2, the execution of the limb module is triggered twice, in two different regions of the body plan pattern, by identity genes I_4 and I_6 (resp. the middle left and middle right regions). In the limb module, the G_{SA} part also imposes a 0.05 Hz frequency of division, however it is restricted to the *tip* of the growing region, i.e., cells of maximal g_W or g_E values (see 2.1.5). All other cells in the stem of the limb have a zero rate. As for the G_{PF} part of the limb, it does not involve any further internal patterning, thus is equivalent to a 1×1 checkered pattern with a unique identity gene.

3.1 Quantitative Variations

3.1.1 Varying Limb Thickness by GRN Weights (PF)

In Figure 3b, the same organism has been affected by a “thin-limb” mutation of the base body plan. Although not shown, the weights of the base module’s gene regulatory network G_{PF} have been modified in such a way that they now create a checkered pattern with a narrower central row allowing less space for the limbs to grow, hence making them actually thinner. This is accomplished by modifying the thresholds of the B_3 and B_4 genes (see Figure 2c) so that they are respectively less and more sensitive, i.e., shrink and increase their regions of expression, causing both boundary lines to shift closer to the center. The 0.5 coefficient encodes this drift (compared to 1 for the wild type). The reverse, “thick-limb” mutation is shown in Figure 3c, with coefficient 2 (B_3 and B_4 boundaries shifted *away* from the center). This is a good example of the compactness of the genotype and its large-scale effect on the phenotype: just varying the sensitivity of a couple of genes can result in dramatic morphogenetic changes.

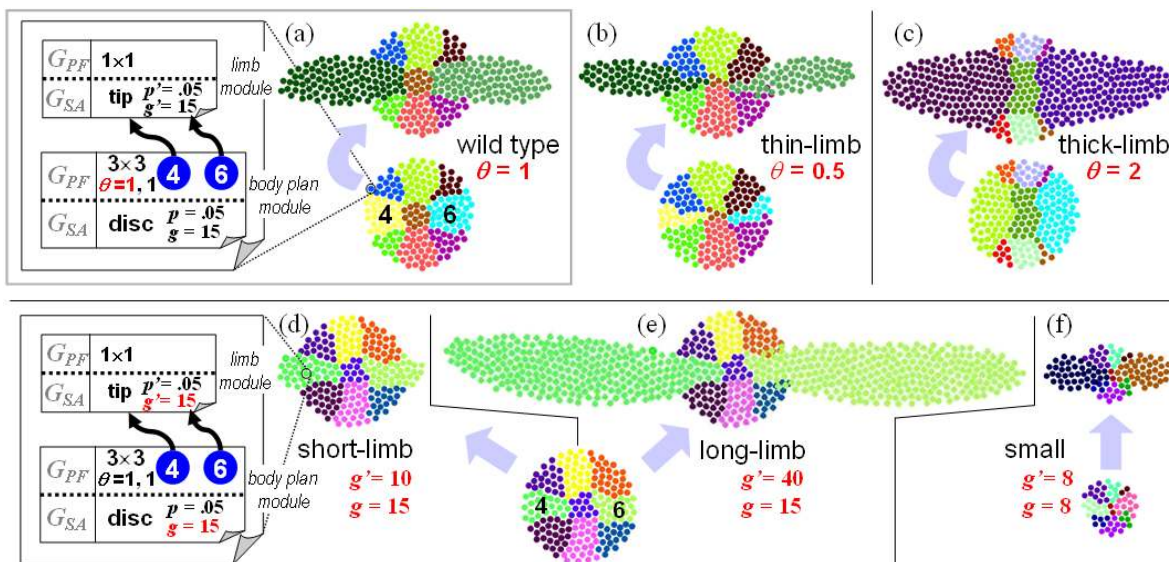


Figure 3: Simulations from the multiagent model showing quantitative variations. (a)-(c) Varying limb thickness by modifying GRN weights (see text, 3.1.1). (d)-(f) Varying length and size by stopping division earlier/later (see text, 3.1.2).

3.1.2 Varying Limb Length by Division Signals (SA)

By modifying the division rate and/or the stop conditions of proliferation, the size of various parts of the embryo can also be varied. For example, in Figure 3f, both body plan and limbs stop growing beyond gradient values $g^* = 8$, producing a phenotypic shape that is proportionally smaller to the wild type. In Figures 3d and 3e, cell proliferation is regulated only in the limbs, respectively by stopping it sooner ($g^* = 10$) and later ($g^* = 40$). Note that similar effects can also be achieved by decreasing or increasing the probability of division p , while keeping the maximum gradient values constant (see Figure 4c).

3.2 Structural Variations

3.2.1 Changing Limb Position by Module Switching

In Figure 4, the modularity of the limb component is demonstrated through various mutations reminiscent of experiments on biological organisms such as *Drosophila*. The identity genes marking the regions (“imaginal discs”) responsible for the growth of a specific appendage can be literally turned on or off in new regions with respect to the wild type of Figure 3a. For example, in Figure 4a, a virtual case of “antennapedia” (the growth of a leg where there should be an antenna) is obtained by connecting a new identity region to the limb module, here region I_2 instead of region I_6 . This means *rewiring* the gene regulatory network G_{PF} to reflect the fact that the limb genes’ regulatory sites in the DNA have mutated and now accept gene I_2 ’s proteins as a promoters instead of gene I_6 ’s proteins. In the three-limb mutation of Figure 4b, these regulatory sites have duplicated themselves before mutating, accepting gene I_2 *in addition* to gene I_6 (not just in replacement), so that the limb module is now executed three times instead of twice.

3.2.2 Serial Homology by Duplication & Divergence

Later in the course of evolution, similar copies of the same organ can diverge and acquire specialized characteristics, as Figure 4c illustrates. In this scenario, three copies of the entire limb module were produced by duplication. In the DNA, this corresponds to actually copying the section of genome that contains all of the limb’s genes and their regulatory sites. Then, these copies can mutate independently from each other. In this artificial example, the upper module of the genotype appears now three times with three

different growth rates $p^* = 0.05, 0.03, \text{ and } 0.1$, respectively equal, lower and greater than the wild type’s rate, therefore creating a shorter and a longer limb. Serial homology is a major evolutionary process, resulting from *duplication* followed by *divergence*. Biological organisms often contain numerous repeated parts in their body plan. This is most striking in the segments of arthropods (several hundreds in millipedes) or the vertebrae, teeth and digits of vertebrates. After duplication, these parts tend to diversify and evolve more specialized structures (lumbar vs. cervical vertebrae, canines vs. molars, etc.). Homology exists not only within individuals but also between different species, as classically shown by comparing the forelimbs of tetrapods from the bat to the whale. Homology could also be explored as an important routine of artificial self-developing systems.

3.2.3 Adding Limbs by Body Plan Expansion

In the scenario of Figures 4d-e, new limbs are generated not by reusing the same body plan differently (Figure 4a-b) or by duplicating the limb module (Figure 4c), but rather by expanding the gene regulatory network G_{PF} of the body plan in order to create new regions of gene identity that can host limb growth. Practically, this means expanding the B_i and I_k layers of the GRN (see Figure 2c) by periodically copying and appending some of their nodes. This increases the embryo’s geography from a $3 \times 3 = 9$ -type checkered pattern to a $5 \times 3 = 15$ -type (Figure 4d) or $9 \times 3 = 27$ -type pattern (Figure 4e). In addition to identity regions I_4 and I_6 , the limb module is now executed in regions I_{10} and I_{12} in the first case, and I_{16}, I_{18}, I_{22} and I_{24} in the second case. The SA part of the body plan is also slightly modified to accommodate these new regions. It assumes an oval shape resulting from a *nonuniform* distribution of the division rate p that follows the NS midline gradient (see Figure 2b), i.e., greater toward the north and south poles ($p = 0.1$) and lower in the center ($p = 0.05$).

3.2.4 Adding Digits by Modular Hierarchy

Finally, along the same principles, Figure 5 shows a few cases of simulations of *three-tier* organisms. Figure 5a is taken as the new wild type. After the usual development of two limbs from the 3×3 body plan, extra “digits” grow from these limbs, guided by the top module of the hierarchical genotype. In these examples, the digits’ shape is similar to the limbs at a smaller scale (same SA behavior,

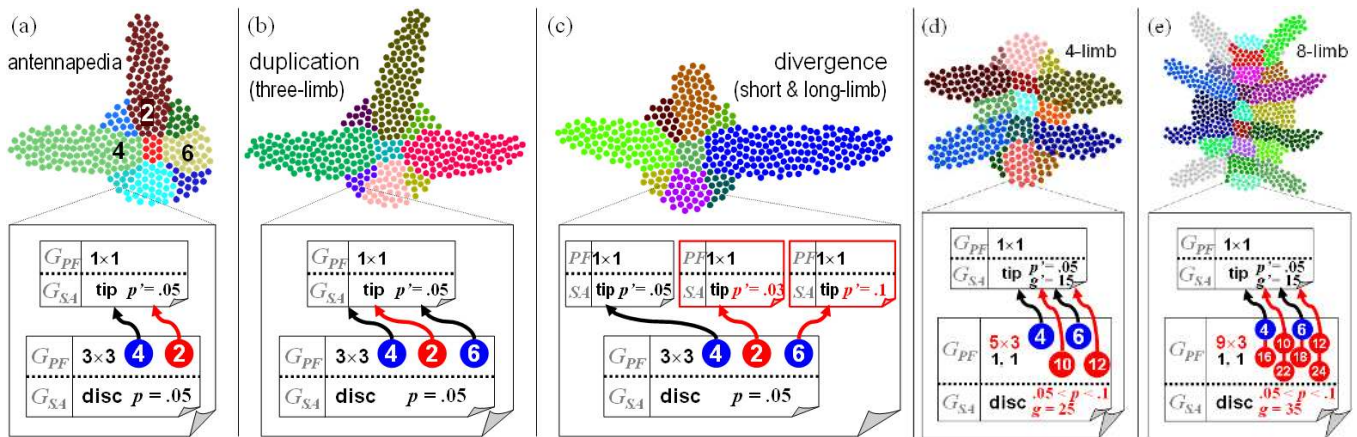


Figure 4: Simulations showing structural variations. (a)-(c) Changing limb configuration by switching the limb-triggering genes and/or duplicating the limb module (see text, 3.2.1, 3.2.2). (d)-(e) Adding limbs by body plan expansion (see text, 3.2.3).

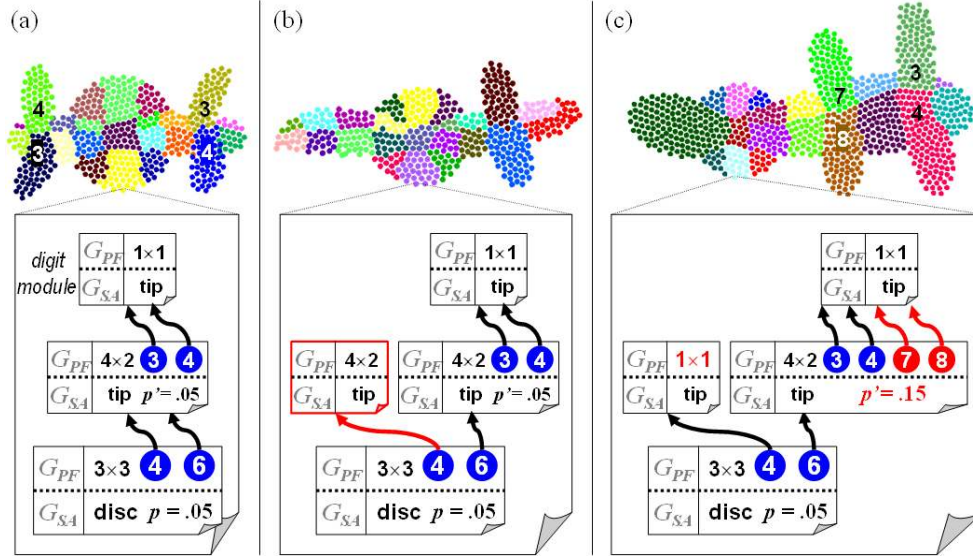


Figure 5: Adding digits via a third tier in the modular hierarchy of the developmental genotype (see text, section 3.2.4).

concentrating the actively dividing cells at the tip), although this is not a requirement. To make room and support the growth of these new digits, limbs have expanded their internal pattern from 1×1 to 2×4 (see previous section). Figure 5a presents a double bilateral symmetry, with respect to both horizontal and vertical axes. The duplication and divergence of the middle-tier limb module in Figure 5b prevents region I_4 's limb (left) from triggering the digit module and deprives it from its digits, compared to region I_6 's limb (right). This reduces the bilateral symmetry to the horizontal axis only. Figure 5c is a further mutation of Figure 5b, in which region I_6 's limb has accelerated its growth ($p' = 0.15$) and expanded into a 2×6 checkered pattern able to support the development of two new digits, whereas, on the contrary, region I_4 's limb has continued to regress back to an undifferentiated stump (division slowing down to $p' = 0.03$).

4. DISCUSSION

Figure 6 gives an overview of a possible phylogenetic tree based on the different forms detailed in part 3. Dashed branches suggest “convergent” speciation pathways. Embryomorph systems of the kind presented here fall into the “implicit embryogeny” category of [3], and are hybrid between “cell chemistry” and “grammatical” models according to [23], by their use of gradients and relative closeness to biology, but also by their rewriting scheme based on cellular division. However, the main differences with L-systems are the inherent self-dissimilarity (later stages are not necessarily copies of earlier stages), the fine-grain resolution of development, incorporating microscopic randomness (cells proliferate and spread irregularly inside regions) and, above all, the fact that development is perpetually regulated by context-dependent adjustment and spatial differentiation of the genetic parameters G . Thus, unlike context-free and open-ended generative systems, embryomorph systems actually constitute a programmable and regulated (self-limiting) type of development.

Another crucial aspect of embryomorph systems is *modularity*. Building a complicated image $I(g_x, g_y)$ directly on top of the initial global X and Y gradients would require maintaining a large number

of pattern variables in each agent to implement every detail, and thus would be difficult to evolve. Modularity, by contrast, is an essential condition of evolvability [25]. Most previous examples show that the body plan can be modified independently from the organs by mutating the appropriate module inside the genotype. Thus different genes can have very different qualitative effects on the phenotype at different architectural scales. Moreover, without modules it would not be possible to have *differential* SA behavior, necessary for the growth of new structures and shapes (limbs, digits) other than blobs. Finally, modules can be *reused*, as different identity genes I_k can trigger the same (G_{SA}, G_{PF}) block. In summary, modularity is a desirable feature in natural genotypes and phenotypes just as in any artificial architecture or system. It seems that biological evolution “discovered” this principle naturally [4]. An important future challenge of the present work will be to show how the modularity of the genotype, hence the phenotype, can in fact spontaneously evolve by duplication rules on GRN parts.

This study is inherently interdisciplinary, as it closely follows biological principles at an abstract level, yet does not attempt to model detailed data from real genomes or organisms. Thus it lies at crossroads between different families of works, from developmental and systems biology to artificial life, spatial computing, and evolutionary computation. It is an original attempt to integrate self-assembly and pattern formation under genetic control. Naturally, beyond the proof-of-concept simulations presented here, a more systematic evolutionary exploration by automated mass-production and analysis of virtual organisms is needed. It should involve the meta-design of a *functional* “fitness”, i.e., define what cell-agents and organ-regions represent in a practical applications, and to what degree their spontaneous collective arrangement is beneficial: processor-carrying micro-units, sensors and actuators, software agents, robot parts, mini-robots, and so on. Different selection strategies would be possible, either focusing on pre-specified forms (optimizing shape), pre-specified functions (optimizing performance), or allowing unspecified outcomes (open-ended evolution).

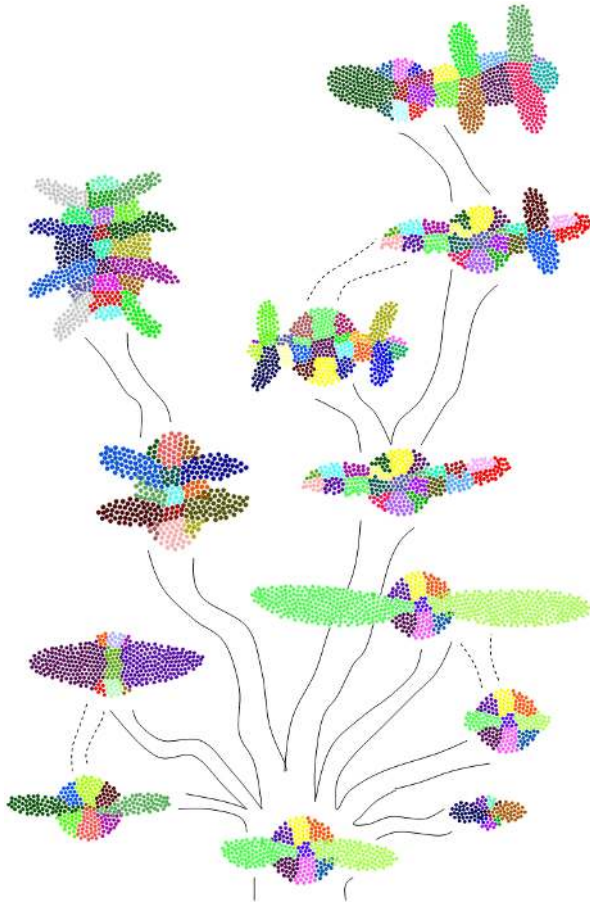


Figure 6: A possible phylogenetic tree (see text, part 4).

5. REFERENCES

- [1] Ball, P. *The Self-Made Tapestry*. Oxford Univ. Press, 1999.
- [2] Beal, J. and Bachrach, J. Infrastructure for engineered emergence on sensor/actuator networks. *IEEE Intell. Sys.*, 21, 2, 10-19, 2006.
- [3] Bentley, P. and Kumar, S. Three ways to grow designs: A comparison of embryogenies for an evolutionary design problem. In *Proceedings of the Genetic and Evolutionary Computation Conference* (Orlando, Florida), W. Banzhaf et al., Eds. Morgan Kaufmann, vol. 1, 35-43, 1999.
- [4] Callebaut, W. and Rasskin-Gutman, D., Eds. *Modularity*. MIT Press, Cambridge, MA, 2005.
- [5] Carroll, S. B., Grenier, J. K. and Weatherbee, S. D. *From DNA to Diversity*. Blackwell Scientific, Malden, MA, 2001.
- [6] Christensen, A., O'Grady, R. and Dorigo, M. Morphology control in a self-assembling multi-robot system. *IEEE Robotics & Automation Magazine*, 14, 4, 18-25, 2007.
- [7] Coen, E. *The Art of Genes*. Oxford University Press, 2000.
- [8] Coore, D. *Botanical Computing: A Developmental Approach to Generating Interconnect Topologies on an Amorphous Computer*, Ph.D. thesis, Dept. of Elec. Eng. & Computer Science, MIT, 1999.
- [9] Doursat, R. The growing canvas of biological development: Multiscale pattern generation on an expanding lattice of gene regulatory networks. *InterJournal: Complex Syst* 1809, 2006.
- [10] Doursat, R. Organically grown architectures: Creating decentralized, autonomous systems by embryomorphic engineering. In *Organic Computing*, R. P. Würtz, Ed. Springer-Verlag, Berlin, 2008, 167-200 (ch. 8), 2008a.
- [11] Doursat, R. Programmable architectures that are complex and self-organized: From morphogenesis to engineering. *11th Int'l Conference on the Simulation and Synthesis of Living Systems*, Winchester, UK, August 5-8, 2008b.
- [12] Doursat, R. and Ulieru, M. Emergent engineering for the management of complex situations. *2nd International Conference on Autonomic Computing and Communication Systems*, Turin, Italy, September 23-25, 2008.
- [13] Floreano, D. and Mattiussi, C. *Bio-Inspired A.I.: Theories, Methods, and Technologies*. The MIT Press, 2008.
- [14] Goldstein, S. C., Campbell, J. D. and Mowry, T. C. Programmable matter. *IEEE Computer*, 38, 6, 99-101, 2005.
- [15] Hofmeyr, S. A. and Forrest, S. Architecture for an artificial immune system. *Evolutionary Comput*, 8, 4, 443-473, 2000.
- [16] Hornby, G. S. and Pollack, J. B. Creating high-level components with a generative representation for body-brain evolution. *Artificial Life*, 8, 3, 223-246, 2002.
- [17] Kirschner, M. W. and Gerhart, J. C. *The Plausibility of Life: Resolving Darwin's Dilemma*. Yale University Press, 2005.
- [18] Lipson, H. and Pollack, J. B. Automatic design and manufacture of robotic lifeforms. *Nature* 406, 974-978, 2000.
- [19] Miller, J. F. and Banzhaf, W. Evolving the Program for a Cell: From French Flags to Boolean Circuits. In *On Growth, Form and Computers*, S. Kumar and P. Bentley, Eds., Elsevier Academic Press, 2003.
- [20] Minai, A. A., Braha, D. and Bar-Yam, Y. Complex engineered systems. In D. Braha, Y. Bar-Yam and A. A. Minai, eds., *Complex Engineered Systems: Science Meets Technology*. Springer Verlag, 2006.
- [21] Nagpal, R. Programmable self-assembly using biologically-inspired multi-agent control. *1st International Conference on Autonomous Agents*, Bologna, Italy, July 15-19, 2002.
- [22] Siero, P., Rozenberg, G. and Lindenmayer, A. Cell division patterns: syntactical description and implementation. *Comput Graphics and Image Processing* 18, 329-346, 1982.
- [23] Stanley, K. O. and Miikkulainen, R. A Taxonomy for artificial embryogeny, *Artificial Life*, 9, 2, 93-130, 2003.
- [24] Vicsek, T., Czirók, A., Ben-Jacob, E., Cohen, I. and Shochet, O. Novel type of phase transition in a system of self-driven particles. *Physical Review Letters*, 75, 1226-1229, 1995.
- [25] Watson, R. A. and Pollack, J. B. Modular interdependency in complex dynamical systems. *Artif Life*, 11, 4, 445-458, 2005.
- [26] Wolpert, L. (1969). Positional information and the spatial pattern of cellular differentiation development. *J. Theoret. Biology* 25, 1-47, 1969.
- [27] Würtz, R. P., Ed., *Organic Computing*. Springer, 2008