A Modular Gene Regulatory Network Model of Artificial Ontogenesis

Amer Ghanem and Ali A. Minai
University of Cincinnati
ghanemar@email.uc.edu
Ali.Minai@uc.edu

All but the simplest multicellular organisms are highly heterogeneous, with specific arrangements of organs and structures central to their functionality. At the level of individual organisms, this organization must be extremely robust against variations in the developmental process and in the environment. At the same time, evolution requires sufficient variation *across* organisms in order for selective pressures to work. Thus, evolving systems face a critical tension: How to maintain a stable heterogeneous organization while allowing sufficient phenotypic variation.

Recent work in evolutionary developmental biology has demonstrated that modularity is the key enabler for balancing variation and stability in animals. The deployment of the same developmental modules in different spatial and temporal combinations (called heterotopy and heterochrony, respectively) can generate a wide range of forms without the need for much change within the modules. Thus, useful – and sometimes critical – features evolved with great difficulty can be conserved as modules, while large changes in the regulatory mechanisms underlying their use can provide a large part of the variation needed for evolution.

In this paper, we combine the idea of regulatory variation in a modular developmental system with Kauffman's hypothesis of cell types as attractors. We present a simple model for the ontogenesis of two-dimensional shapes where an attractor network of regulatory genes robustly controls the deployment of genetic network modules that determine phenotypic features. We use this model to show that a wide variety of forms can be generated purely through changes in the connectivity and weights of the regulatory network while conserving the rest of the system. We also investigate the effects of varying switching conditions and initial asymmetries in maternal proteins.

1 Introduction

The balance between variation and stability is a critical component of *evolvability*, i.e., the viability of the evolutionary process in living systems [10, 33, 25]. Change is necessary for the evolution of fitter forms, but it cannot come at the expense of disrupting structures or processes that are critical to survival. This is especially true in multicellular organisms which depend on specific heterogeneous organizations (e.g., body forms) for their functionality. In this paper, we present a simple developmental model to explore how modularity and attractor dynamics enable robust phenotypic variation in heterogeneous multicellular organisms.

Gene regulatory networks (GRNs) have been used extensively to model phenotypic development. Such models have ranged from detailed ones [15, 11, 18, 28], which are used to make quantitative predictions and come up with empirical hypotheses, to abstract models focused on high-level properties of gene regulation [14, 16, 13]. The GRN model we propose falls in the latter class, and attempts to address the important issues of robustness and variation at an abstract level.

2 Background and Motivation

There is significant evidence from animal development that nature conserves good ideas (e.g., appendage generation) as *modules*, using them in different variations, combinations, sequences and multiplicities to produce the variety needed for evolution [9, 8, 26, 5]. By varying the *relative positions* of the same structures (e.g., limbs) – termed *heterotopy* – or the *relative timing* of the same processes or events (e.g., growth of the body vs. inception of limbs) – called *heterochrony* – a developmental program can be made to produce very different phenotypes, much as a change in rhythm can make a musical composition sound different [35, 29, 12]. The modularity of this paradigm enhances developmental robustness and evolvability [27, 21, 32] by offering three distinct advantages: 1) It is more likely to produce viable organisms because it conserves critical components; 2) Its phenotypic variation is more productive because it is using "tried-and-tested" components (albeit in novel ways); and 3) It can generate and sustain faster evolution because much of the variation is arising from relatively "simple" or "superficial" changes (i.e., re-arrangement of modules). In a sense, the system is able to partially overcome the classic exploration-exploitation trade-off by decoupling the processes – letting the modules "exploit" while using regulatory variation to "explore". Furthermore, as illustrated beautifully by Simon in his parable of Tempus and Hora [30], modularity is a key enabler of open-ended complexity because it allows productive exploitation of combinatorial growth – a fact well understood by engineers, planners and organizational experts.

Another major issue for living organisms is the necessity for heterogeneity within the phenotype. All but the simplest multicellular organisms have highly differentiated bodies with remarkably stable architectures. Again, experimental

data indicates that modularity is a key enabler of this heterogeneity [8], with variant deployments of the same modules used to produce diverse structures such as limbs, antennae, etc. Each structure and organ has characteristic cell populations expressing specific proteins, and it has been proposed that these patterns of expression represent attractors in the phase space of gene expression dynamics [23, 24, 20]. Our view is that such attractors can help explain the relative stability and robustness of heterogeneous structures seen in metazoan phenotypes.

In this work, we present a simple model for the ontogenesis of two-dimensional shapes, combining the idea of regulatory variation in modular developmental systems with Kauffman's hypothesis of cell types as attractors [23, 24, 20]. In our model, an attractor network of regulatory genes controls the deployment of genetic network modules that explicitly determine growth and phenotypic attributes. We study the effect of various systematic variations on the model's propensity to generate diverse, internally heterogeneous phenotypes.

3 Mathematical Description of the Model

In our model, an *organism* is defined as a group of cells organized in a two-dimensional square lattice with a specified *anterior-posterior (A-P) axis* and *left (L)* and *right (R)* lateral directions. Each simulation of the ontogenesis model begins with a *zygote*, which is a small (usually rectangular) group of cells at the center of the lattice, and grows under the control of the *gene regulation network* (GRN) described below.

Each cell in the organism has the same set of N *genes*, comprising the organism's *genome*, and the same GRN. As in real organisms, heterogeneity in the organism arises purely through self-organization in response to small asymmetries in the initial distribution of maternally supplied proteins and environmental influences [34]. This self-organization is driven by the differential expression of genes in different parts of the organism. Thus, the organism's GRN is a primary determinant of its final form. It is described in the next subsection.

3.1 GRN Model

The structure of the GRN model is shown in Figure 1. It comprises five interacting layers: 1) the maternal layer, 2) the signaling layer, 3) the control layer, 4) the labeling layer, and 5) the phenotypic layer. Each layer includes specific genes, and the connectivity between layers determines the regulatory relationships. Each arrow in Figure 1 corresponds to a specific *weight matrix*. The control layer is internally configured into several modules and sub-layers whose architecture is discussed later.

The system has a total of N genes and a set of μ maternal proteins that represent the inputs to the system. We assume that all genes in the model genome are "coding genes", i.e., each gene's expression produces a specific protein. In particular, the expression of gene i at time t results in the production of protein

i. The level of protein i in cell j at time t is denoted by $P_i^j(t)$, and the state of the cell is characterized by the level of all N proteins, $\{P_i^j(t)\}$.

The expression of gene i in cell j at time t is controlled by the levels of certain proteins in that cell at the immediately preceding time step. The *stimulus* to i is given by:

$$Y_i^j(t) = \sum_{k=1}^{N} w_{ik} P_k^j(t-1) \tag{1}$$

where w_{ik} is the *weight* or strength of the regulatory effect of protein k on gene i. The *activation state* of gene i is given by:

$$G_i^j(t) = \begin{cases} 1 & Y_i^j(t) > \theta_i \\ 0 & else \end{cases} \tag{2}$$

Thus, i is only activated in cell j if it is sufficiently stimulated. The *expression level* of gene i in cell j at time t is then given by a sigmoid function:

$$X_i^j(t) = \frac{G_i^j(t)}{1 + e^{-(Y_i^j(t) - \gamma_i)}} \tag{3}$$

where γ_i is an offset for i, determining the amount of stimulus necessary for significant expression.

The level of protein i in cell j at time t is given by the update equation:

$$P_i^j(t) = \alpha_i X_i^j(t) + (1 - \alpha_i) P_i^j(t-1) \tag{4}$$

where $0 \geq \alpha_i \leq 1$, the rate parameter.

Typically, the α_i, θ_i and γ_i parameters are fixed for all simulations, and only the w_{ik} parameters are varied to generate different organisms. Thus, these weights comprise the essential parameters of the GRN, and their values determine which genes regulate the expression of each gene to what degree. Signaling proteins follow a somewhat modified forms of the above model, and maternal proteins only undergo passive decay.

Next, we describe each layer of the GRN and its role in ontogenesis.

3.2 Maternal Layer

Each node in this layer represents a maternal protein [34] with a specific distribution along the A-P and L-R axes in the zygote. Together, the distribution pattern of maternal proteins defines locations on the zygote, and provides the initial condition for development. We assume that no new maternal proteins are produced during development, and they decay gradually towards 0 with time.

Maternal proteins are divided into two types: Anterior-Posterior(AP) proteins whose distribution varies only along the A-P axis, and L-R proteins, which have symmetrically varying lateral distribution around the central axis.

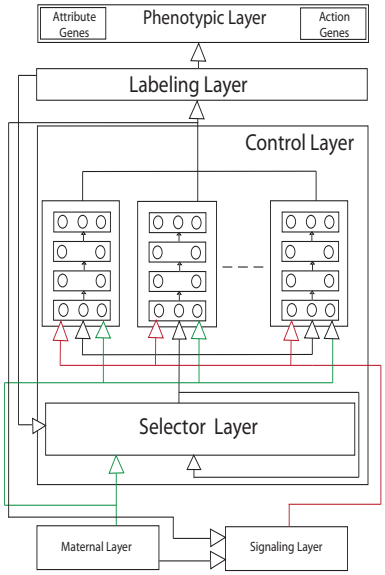

Figure 1: Detailed network

3.3 Signaling Layer

Proteins produced by genes in this layer are responsible for signaling between cells. These genes are stimulated by proteins from the maternal and labeling layers, and provide output to the Control Layer. Unlike other proteins, signaling proteins diffuse among neighboring cells, which models the signaling process. Diffusion of the signaling proteins is the only way in which cells in our model communicate with each other.

3.4 Control Layer

The Control Layer comprises the genes directly responsible for regulating phenotypic attributes and growth, and is the primary locus of interest in the model. It is further divided into two stages: 1) The Selector Network; and 2) The Control Module Level (CML). Genes in the CML are organized into M distinct modules, each modeled as a feed-forward network with four sub-layers – an input sub-layer, two hidden sub-layers, and an output sub-layer. The input sub-layer integrates information from the Maternal and Signaling layers. This is processed through the hidden sub-layers to produce the output of the module at the output sub-layer. These Control Layer outputs drive the Signaling Layer and the Labeling Layer (see below), which can be seen as integrating them into a control signal for the phenotypic genes.

The Selector Network is a single layer of $m > M$ genes embedding several patterns of activity as fixed point attractors [19, 24]. All genes in a given module,

q, are gated by a small unique set, S_q, of genes in the Selector Level (here, we assume this set to be of size 1 for convenience). Thus, the set of CML modules active at a given time depends on the current activation pattern of the Selector Network. The set of active modules corresponding to each Selector Network attractor is called its *functional configuration (FC)*. Some genes in the Selector Network also encode changes in the axis of symmetry.

The Selector Network takes input from the Maternal and Signaling Layers, as well as recurrent input from the Labeling Layer. The Selector Networks remains "locked" in an attractor until it is destabilized by a specific "key" pattern of activity in the Labeling Layer [22], resulting in the stabilization of a new attractor and a new regime of active modules. This "lock-and-key" arrangement provides a systematic but exceptionally rich regulatory mechanism for phenotypic variation.

3.5 Labeling Layer

The function of the Labeling Layer is to represent the integrated regulatory state of the cell. It gets input from the Control Layer and projects to the Phenotypic Layer, determining the latter's activity. It also provides feedback to the Selector Network, and thus mediates the switching between functional configurations. The update rules for proteins in this layer follow the standard equations 2, 3, and 4.

3.6 Phenotypic Layer

Proteins in this layer are responsible for the phenotypic structure and attributes of the cell. They get input only from the Labeling Layer and their output does not go to any layer; it is observed through changes in the functions or the attributes of the cell. The update rules for proteins within this layer are described in equations 2, 3, and 4. Proteins in this layer are divided into two types:

Growth Proteins These proteins are responsible for growth actions. In the default case, four proteins are used and the decision of the growth action is based on the corresponding gene activation. Four genes with binary expression value give 16 different permutations that encode five different growth actions: grow forward, grow backward, grow away from the axis of symmetry, grow toward the axis of symmetry, and death (*apoptosis*)

Attribute Proteins Proteins of this type are responsible for the cell's attributes. Attributes can include color, texture, shape, size, etc. However, in the current implementation, we have considered color as the only attribute. In the default case, three genes are used to determine the color of the cell. Each gene produces a protein that has a value between 0 and 1. The triplet of the three protein levels represents a Red, Green, Blue (RGB) value that is used to determine the color of the cell.

3.7 The Ontogenesis Process

The process of development begins with a rectangular zygote that has an A-P axis of symmetry. Initially, all cells of the zygote are in the same functional configuration (i.e., have the same Selector Network attractor activated). This is termed the *root functional configuration* (RFC), and it drives a process of growth and patterning through the activation of labeling and phenotypic genes. This goes on until cells with certain labels – analogous to imaginal disks in real organisms [34, 31] – have their functional configuration switched via the GRN. This switch, which is typically accompanied by a reorientation of the axis of symmetry for the switched cells, results in new patterns of development in these specific cell groups, e.g., producing new organs, or outgrowths such as limbs or wings. Essentially, each organ grows like a new "body" within a limited part of the overall organism. Over time, different groups of cells come to have specific regulatory states, attributes, growing patterns, etc., creating the heterogeneous body of the organism. Eventually, each cell reaches a *terminal functional configuration (TFC)*, which is defined as an FC incapable of producing a labeling output that can destabilize its attractor.

4 Results

We generated different organisms from various genetic configurations. For each configuration, we systematically introduced gradual variation in various data and/or model parameters, such as the initial conditions or the connections between the layers in the GRN, and generated other organisms from the modified configurations. The resulting organisms were compared to show that : a) A little variation can generate a wide variety of forms; and b) the difference in form depends systematically but nonlinearly on the variation. Our results also provide information on the effectiveness of different mechanisms for generating phenotypic variation.

4.1 Comparison Method

We compared the phenotypes obtained from the baseline and modified configurations by determining the similarity in their structure and color (attributes). The result is a heuristically defined similarity metric between 0 (maximally dissimilar) and 1 (identical), where 90% of the metric's value depends on the structural comparison and 10% on the attribute comparison (see [17] for details).

Armed with the similarity metric defined above, we studied how (and how much) phenotypes change as various aspects of the system are varied. Note that in all manipulations of one baseline phenotype, the actual control modules were *not* varied. Indeed, our goal was to study how much variation is possible without varying the modules.

To set a baseline for variation, we first considered how much variation is possible between two configurations with random, independently chosen weights,

but using the same control modules. The systematic variation was applied as follows:

4.2 Variation Due to Changes in the Weight Matrix

Changing Connections to Control Layer

In this variation, the change affected only in the connections to the Input Layer of the modules; the amount of change was determined by the number of connections that were changed from those layers to the Control Layer. Three different levels of variation – 10%, 50%, and 100% – were applied to different organisms to capture the effect of such variation.

Changing Connections to the Selector Network

This type of variation involved changing the connections to the selector genes, which work as gates for the modules. The variation was applied in three different amounts; 10%, 50%, and 100%.

Changing The Attractors

In this type of variation, the set of attractors embedded in the Selector Network were changed. The set of attractors determines which modules are active at a specific time, and their switching patterns. Three different amounts of variation, 10%, 50% and 90%, were applied to capture the effect of changing the attractors.

After applying each amount of variation 10 times on the same organism, the relationship between the comparison result and the amount of variation is shown in figure 2. The average over three different phenotypes after applying the same amount of variation is shown in figure 3

4.3 Variation Due to Changes in the Maternal Proteins Distribution

The distribution of maternal proteins determines the initial relative location of the cells in the zygote. In this case, five maternal proteins are responsible for this task. While the Left-Right proteins – those whose gradient changes only in the left-right direction – have to be symmetric about the central axis to ensure symmetry, maternal proteins can have almost unlimited variation, resulting is heterotopic and heterochronic variation. A simple scenario for variation is to change one maternal protein distribution at a time while keeping the rest of the distributions unchanged. For simplicity, the distributions were changed by taking the 1's complement of the functions, i.e. if the value was 0 at the anterior end and 1 at the posterior, the 1's complement function would be 1 at the anterior end and 0 at the posterior.

Variation in the distribution of the maternal proteins does not imply any variation in the genetic configuration of the organism, only in the initial conditions for development. A combination of changing more than one distribution

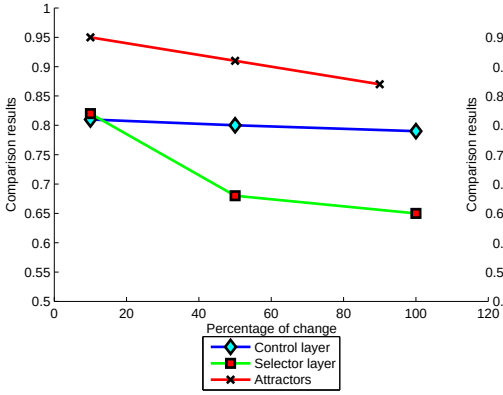

Figure 2: The results after applying changes to all the layers 10 times for the same phenotype

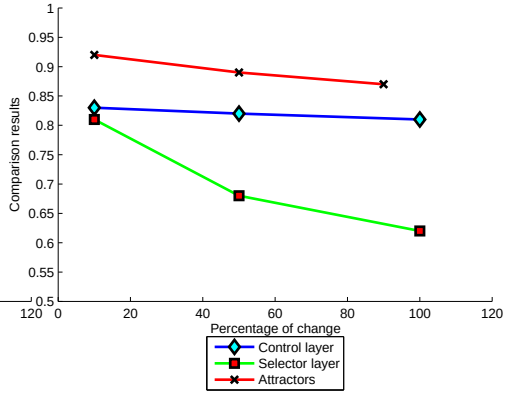

Figure 3: The results after combining the averages of the three phenotypes

at a time can generate a wider variety of organisms, though they all share the same genetic configuration. However, the remarkable aspect of the results in this section is the relative *stability* of the phenotypes generated even after very drastic changes to the input/initial conditions (maternal protein distribution). We believe that this is due, at least in part, to the stabilizing effect of attractor-based regulation. However, the issue of phenotypic robustness against genetic and extrinsic variation is a very complex one, and has been the subject of much investigation [1, 2, 3, 4, 6, 7].

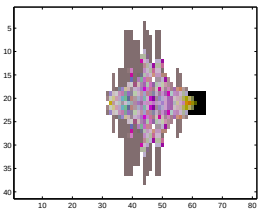

Figure 4: Original phenotype

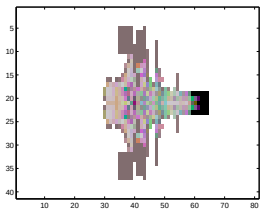

Figure 5: change in AP1

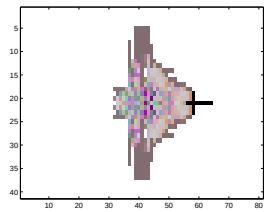

Figure 6: Change in AP2

4.4 Variation Due to Other Parameters

Changing the Activation Threshold θ

The activation threshold θ determines the amount of regulatory protein activation needed to stimulate the genes in the network. A systematic change in the

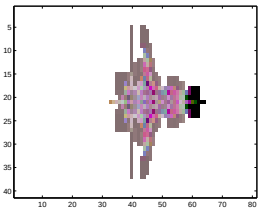

Figure 7: Change in AP3

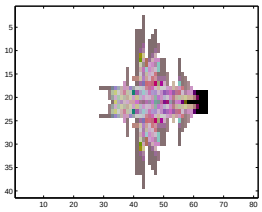

Figure 8: Change in LR1

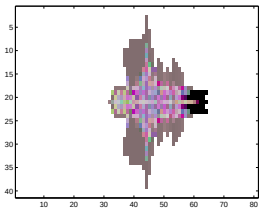

Figure 9: Change in LR2

value of this parameter leads to systematic variation in the resulting phenotype.

The relationship between the activation threshold θ and the comparison results after applying 10 different amounts of variation is shown in figure 10. The average result over three different phenotypes after applying the same amounts of variation is in figure 11.

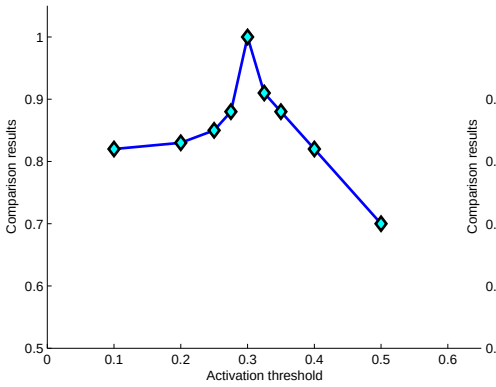

Figure 10: The relationship between θ and the comparison results for one phenotype

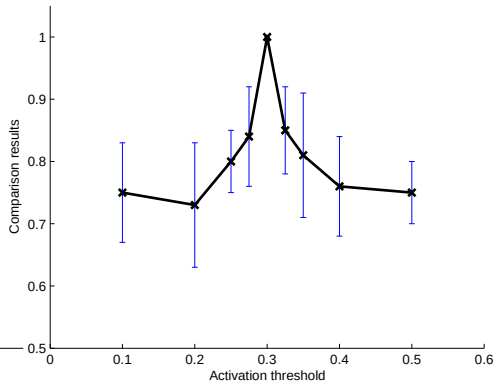

Figure 11: The results after combining the averages of the three phenotypes

5 Conclusion

In this paper, we have proposed a simple model for phenotypic development with the goal of exploring how switched modularity facilitates variation and robustness. Studying different scenarios of variation in the proposed model has led to the following conclusions:

- It is possible to generate wide variety of forms without a major change in the genetic configuration of the organism, simply by modifying regulatory mechanisms and parameters. The comparison results show that,

of the manipulations studied, changes in the connections to the Control Layer have the least effect on the resulting phenotypes, whereas changing connections to the selector genes has the largest effect. Changing the attractors underlying modules switching produces an intermediate degree of phenotypic variation.

- Systematic variation in the connections between the layers of the generic regulatory network generate corresponding variation in the phenotype. In most cases the smaller the variation, the smaller the average difference is between the resulting phenotypes. Although phenotypic effects are expected to be related monotonically to the degree of regulatory variation, the form and relative magnitude of these effects for each variation type are instructive, and can serve as the basis for future models of artificial development.

- Variation is also possible without any changes to the genetic configuration. For example, changing the distribution of the maternal proteins is able to produce a variety of forms with a "family resemblance". This represents a rather explicit form of heterotopy. Another example is the variation of the activation threshold θ. In most cases, the more the variation from the nominal value, the greater the variation seen in phenotypes.

Bibliography

[1] Albert, R., and G. Othmer, "The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in drosophila melanogaster", *Journal of Theoretical Biology* **223** (2003), 1–18.

[2] Austin, D., M. Allen, R. McCollum, D. Dar, J. Wilgus, N. Sayler, C. Samatova, and M. Simpson, "Gene network shaping of inherent noise spectra", *Nature* **439** (2006), 608–611.

[3] Azevedo, R., R. Lohaus, S. Srinivasan, Kristen. Dang, and C. Burch, "Sexual reproduction selects for robustness and negative epistasis in artificial gene networks", *Nature* **440** (2006), 87–90.

[4] Bar-Yam, Y., and R. Epstein, "Response of complex networks to stimuli", *PNAS* **101** (2004), 4341–4345.

[5] Benjamin, P., G. Nicolas, and S. Carroll, "Emerging principles of regulatory evolution", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, (May 2007), 8605–8612.

[6] Bergman, A., and M. Siegal, "Evolutionary capacitance as a general feature of complex gene networks", *Nature* **424** (2003), 549–552.

[7] Bergman, A., and M. Siegal, "Canalization of development by micro rnas", *Nature genetics* **38** (2006), 520–524.

[8] Carroll, S., "Evolution at two levels: On genes and form", *PLoS Biol* **3**, 7 (July 2005), e245.

[9] Davidson, E., J. Rast, P. Oliveri, A. Ransick, and C. Calestani, "A genomic regulatory network for development", *Science* **295**, 5560 (Mar. 2002), 1669–1678.

[10] Dawkins, R., "The evolution of evolvability", *Artificial Life: The Proceedings of an Interdisciplinary Workshop on the Synthesis and Simulation of Living Systems* (C. Langton ed.), (1989), 201–220.

[11] De Jong, H., "Modeling and simulation of genetic regulatory systems: A literature review", *Journal of Computational Biology* **9** (2002), 67–103.

[12] Dequant, M., E. Glynn, K. Gaudenz, M. Wahl, J. Chen, A. Mushegian, and O. Pourqui, "A complex oscillating network of signaling genes underlies the mouse segmentation clock", *Science* **314** (2006), 1595–1598.

[13] Doursat, R., "The growing canvas of biological development: Multiscale pattern generationon an expanding lattice of gene regulatory networks", *Proceedings of the 6th International Conference on Complex Systems*, (2006).

[14] Eggenberger, P., "Genome-physics interaction as a new concept to reduce the number of genetic parameters in artificial evolution", *Proceedings of The 2003 Congress on Evolutionary Computation* (R. Sarker, R. Reynold, H. Abbass, K. Tan, R. McKay, and T. Essam, D. andGedeon eds.), vol. 1, (2003), 191–198.

[15] Furusawa, C., and K. Kaneki, "Complex organization n multicellularity as necessity in evolution", *Artificial Life* **6** (2000), 265–281.

[16] Geard, N., and J. Wiles, "A gene network model for developing cell lineages", *Artificial Life* **11**, 3 (2005), 249–267.

[17] Ghanem, A., "A modular gene regulation network model of artificial ontogenesis" (2007).

[18] Hasty, J., D. McMillen, F. Issac, and J.J. Collins, "Computational studies of gene regulatory networks: In numero molecular biology.", *Nature Reviews Genetics* **2** (2001), 268–279.

[19] Hopfield, J.J., "Neural networks and physical systems with emergent collective computational abilities", *Proceedings of the National Academy of Sciences USA* **79** (1982), 2554–2558.

[20] Huang, S., G. Eichler, Y. Bar-Yam, and D. Ingber, "Cell fates as high-dimensional attractor states of a complex gene regulatory network", *Physical Review Letters* **94** (2005).

[21] Ingolia, N., "Topology and robustness in the drosophila segment polarity network", *PLOS Biology* **2** (2004), 805–815.

[22] Istrail, S., and H. Davidson, "Logic functions of the genomic cis-regulatory code", *PNAS* **102** (2005), 4954–4959.

[23] Kauffman, S., *The origins of order: Self-organization and selection in evolution*, Oxford University Press Oxford, UK (1993).

[24] Kauffman, S., "The ensemble approach to umderstanding genetic regulatory networks", *Physica A* **340** (Jan. 2004), 733–740.

[25] Kirschner, W., and J. Gerhart, "Evolvability", *PNAS* **95** (1998), 8420–8427.

[26] Kirschner, W., and J. Gerhart, *The Plausibility of Life*, Yale University Press New Haven and London (2005).

[27] Kitano, H., "Biological robustness", *Nature* **5** (2004), 826–837.

[28] Kitano, H., S. Hamahashi, and S. Luke, "The perfect c. elegans project: An initial report", *Artificial Life* **4** (1998), 141–156.

[29] Rougvie, A., "Control of development timing in animals", *Nature Reviews - Genetics* **2** (2001), 690–701.

[30] Simon, H., "The architecture of complexity", *Proceedings of the American Philosophical Society*, vol. 106, (Dec. 1962), 467–482.

[31] Simpson, P., "Evolution of development in closely related species of flies and worms", *Nature Reviews - Genetics* **3** (2002), 1–11.

[32] Van Dassow, G., and E. Munro, "Modularity in animal development and evolution: Elements of a conceptual framework for evodevo", *Journal of Experimental Zoology* **285** (1999), 307–325.

[33] Wagner, G., and L. Altenberg, "Complex adaptations and the evolution of evolvability", *Evolution* **50**, 2 (1996), 967–976.

[34] Wolpert, L., *The Principles of Development*, Oxford University Press Oxford, UK (1998).

[35] Zelditch, L., and L. Fink, "Heterochrony and heterotopy: Stability and innovation in the evolution of form", *Paleobiology* **22**, 2 (1996), 241–254.