Honda Research Institute EU
Honda Research Institute Europe GmbH
https://www.honda-ri.de/

Evolution and Analysis of Genetic Networks for Stable Cellular Growth and Regeneration

Lisa Schramm, Yaochu Jin, Bernhard Sendhoff

2012

Preprint:

This is an accepted article published in Artificial Life. The final authenticated version is available online at: https://doi.org/[DOI not available]

Evolution and Analysis of Genetic Networks for Stable Cellular Growth and Regeneration

Lisa Schramm[*,**]
Technische Universität Darmstadt

Yaochu Jin[†]
University of Surrey

Bernhard Sendhoff[*,‡]
Honda Research Institute Europe

Abstract A computational model is presented that simulates stable growth of cellular structures that are in some cases capable of regeneration. In the model, cellular growth is governed by a gene regulatory network. By evolving the parameters and structure of the genetic network using a modified evolution strategy, a dynamically stable state can be achieved in the developmental process, where cell proliferation and cell apoptosis reach an equilibrium. The results of evolution with different setups in fitness evaluation during the development are compared with respect to their regeneration capability as well as their gene regulatory network structure. Network motifs responsible for stable growth and regeneration that emerged from the evolution are also analyzed. We expect that our findings can help to gain a better understanding of the process of growth and regeneration inspired by biological systems, in order to solve complex engineering problems, such as the design of self-healing materials.

Keywords
Cellular growth, development, regeneration, stability, gene regulatory network, motifs, evolutionary strategy

A version of this paper with color figures is available online at http://dx.doi.org/10.1162/artl_a_00075. Subscription required.

1 Introduction

As engineering processes and systems become more and more complex, it is increasingly difficult to guarantee stability and robustness while enhancing efficiency and flexibility. Therefore, new approaches that can augment conventional engineering methods must be sought. To a large degree the organization of natural systems is shaped by a combination of evolutionary and developmental processes whose potential for designing complex self-organizing engineering systems is just starting to be explored in greater detail. One important source of inspiration is the growth and operation of biological cellular systems.

Cellular growth processes could be applied as novel methods—for example, for three-dimensional topology optimization in a multidisciplinary and multiphysics framework. The growth of irregular inner structures could lead to novel designs that allow the concurrent optimization of (for example) fluid dynamics, chemical reaction rates, weight, and material costs. Adaptive manufacturing and rapid prototyping techniques would allow us to physically realize such complex topologies. Self-organization

* Contact author.
** Technische Universität Darmstadt, Karolinenplatz 5, 64289 Darmstadt, Germany. E-mail: lschramm@rtr.tu-darmstadt.de
† Department of Computing, University of Surrey, Guildford, Surrey, GU2 7XH, United Kingdom. The work was conducted while Yaochu Jin was with the Honda Research Institute Europe. E-mail: yaochu.jin@surrey.ac.uk
‡ Honda Research Institute Europe, Carl-Legien-Str. 30, 63073 Offenbach, Germany. E-mail: bs@honda-ri.de

© 2012 Massachusetts Institute of Technology

could help to enable the system to cope with changes imposed during the design phase as well as during the operational phase.

Whereas self-repair during the operational phase requires advances in material science (e.g., material self-repair [8]), the realization of self-repair capabilities during the design and development phase will be more straightforward, because they can be embedded into a purely digital process. It is common that constraints, system properties, and requirements change during the development phase. Systems that can autonomously react to such changes will be very beneficial and cost-effective. They would also foster an overall more robust engineering process. A simple example would be the self-repair of mesh geometries for fluid dynamics or structural dynamics simulations after different parts have joined together or after the structure has been changed drastically during an optimization process. Such methods have not yet been realized based on models of biological growth processes, but they are conceivable and they would have a large impact on engineering systems and processes.

It is the purpose of this article to outline a model of biological cellular growth and regeneration that allows an analysis of the underlying evolved functionality, which in turn will help us to identify the optimal abstraction level and the best strategy to transfer principles from computer simulations of biological systems into technical solutions.

Artificial development, often referred to as artificial embryogeny [6, 36] or artificial ontogeny, simulates the biological developmental process using different levels of abstraction. Our main purpose in studying computational models of development is to advance understanding of stable growth and regeneration in technical systems. Nevertheless, such models have often been used to enhance our understanding of biological systems. For example, computational models have become widely accepted in systems biology [23], and they are frequently referred to as in silico experiments (in contrast with in vitro and in vivo). Due to the enormous complexity of biological systems and the coupling of many different spatiotemporal scales, the level of detail that is required in the simulation to make verifiable predictions and to provide meaningful analysis is often difficult both to specify and to realize.

In this article, we outline a simulation environment that evolves a model of a gene regulatory network (GRN) combined with a cellular model with physical interactions. Using this model, we investigate aspects of dynamically stable growth and the ability of biological organisms to regenerate.

The basic setup of the model builds upon the work by Eggenberger Hotz et al. [11], with two main extensions. First, the cellular interactions occur in a continuous *physical* space (which is particularly important if we want to avoid an a priori grid, e.g., for topology optimization). Secondly, we vary the fitness calculations distributed over the developmental process. The abstraction level of our model allows us to look in more detail into the evolved genetic networks that enable dynamically stable growth and into the evolution of such networks, in particular with respect to their modularization into subnetworks or structural network motifs [1].

A concise review of artificial development has recently been compiled by Harding and Banzhaf [16]. In particular, there are several approaches that simulate biological development based on gene regulatory networks and cell growth; see [10, 11, 21, 22, 37]. Zhan et al. use gene regulatory networks to design electronic circuits [41]. Federici and Downing have also evolved fault-tolerant developmental systems using a recursive neural network with different developmental stages [12]. They increase the number of developmental stages during evolution and use different chromosomes at the end of the evolution for each developmental step. Cell lineages using a gene network model are developed in [15]. In Section 1.1 we review several analyses of gene regulatory networks.

In most reported developmental models, simulations are stopped after a fixed number of steps. A stable development, where the development stops by itself or a balanced cell growth and apoptosis is achieved, remains a challenging task. Models proposed for examining stability and self-repair are reviewed in Section 1.2.

In biology, some position-dependent signals are available during the regeneration process that can guide the regenerating tissue, which is also important for simulations. In a first step, these patterning signals have to be locally self-generated as a response to lesions, and in a second step they have to be used by the regulatory system to trigger cell proliferation and differentiation. In our simulation, we restrict our analysis to the second aspect, namely, how the regulatory system responds to cell loss,

assuming that an appropriate signal is locally available. Therefore, in our simulation, we use (and artificially maintain) prediffused maternal gradients that guide the embryogenesis of the organism and their regeneration capacity. Maternal gradients in the early development of organisms have been observed consistently [35]. Kaneko's group present different developmental models of multicellular organisms. They discuss, for example, the pattern formation in one-dimensional chains of cells [40] and analyze the regulation that leads to ordered spatial patterns of two-dimensional organisms [14].

In our framework, we are able to evolve individuals that reach a dynamically stable state during development without using additional functions like contact inhibition or additional information like the number of surrounding cells, as most other models do. In our experiments, cell turnover is in an equilibrium, that is, cell proliferation is of the same order as cell apoptosis. When some cells of an individual are removed, regeneration is possible in some individuals. We will outline our experimental setup in Sections 2 and 3 and present and discuss the results in Sections 4 and 5, where we also analyze the resulting GRN structure with respect to recurring patterns (motifs) during the evolutionary process.

1.1 The Analysis and Evolution of Gene Regulatory Networks

According to the review by Babu et al. [4], analysis of gene regulatory networks can be divided into three different approaches. First, analysis of connections between transcription factors and target genes, and thus the activations of the different genes. Second, analysis of network motifs, which are patterns in the GRN that are overrepresented; the review by Alon [1] identifies network motifs that are found in bacteria (e.g., *E. coli*), yeast, and some other plants and animals. Third, analysis of large modules and connectivity hubs.

Ashe and Briscoe [3] review the role of morphogen gradients in the early development. They outline different signaling pathways and some strategies for the regulation of genes. The number of responses a morphogen gradient can generate, which is equal to the number of different concentrations that cause different behavior of cells, was analyzed empirically. Between three and seven distinct thresholds were identified. Cooper et al. [9] predict the types of different feedforward loops in *E. coli*. François and Hakim [13] evolve bistable switches and oscillators and compare their results with real biological networks. They find small networks with those behaviors that display both known and original designs. Jin and Meng [18] showed the way in which genetic loops are coupled plays a key role in robustness and that the most robust motifs emerge from in silico evolution.

Salazar-Ciudad et al. [30] created large random networks using (1) diffusion and (2) direct-contact induction processes and analyzed their network patterns and their ability to create spatial patterns in 1D and 2D. They found many different patterns, which are a combination of a few simple patterns.

Quayle and Bullock [29] model the evolution of gene regulatory networks and analyze the occurrence of different cyclic patterns. The distribution of their artificial genetic networks lies between those of random networks and scale-free networks. They evolve cyclic patterns of different lengths and show that the system can evolve short patterns but the performance decreases rapidly for longer patterns. Not only does the length of the pattern constrain the performance of the evolution, but also some pattern lengths seem to be more difficult. This suggests that evolution generates complex expression using building blocks of simple patterns. Crombach and Hogeweg use a Boolean threshold network, with varying evolutionary goals, to gain evolvability [10]. The entanglement of processes at different time and space scales and its influence on the development of organisms is analyzed in [17]. It is shown in [20] that evolvability is selectable in evolving genetic oscillators using a computational model of development.

Simulations of the generation of regulatory networks using duplication and mutation also show scale-free and small-world topologies [25]. Their distributions are comparable to those of *E. coli* and *S. cerevisiae* and differ from random networks.

The analysis of the regulatory networks evolved in some of the individuals that is presented in Sections 4.1 and 4.2 builds upon the work briefly reviewed in this section. In particular, the way in which genes are used, how they are activated, and how they can be grouped in motifs that emerge and govern the regeneration capability of our artificial organisms will be of interest for our analysis.

1.2 Evolution of Stable Growth and Regeneration

Miller [27] was one of the first to discuss the issue of stability in the context of artificial development. Instead of manually stopping the developmental process after a predefined number of steps, the target is to reach a steady state where the artificial system maintains its morphology and/or function. This stable state can be static or dynamic. In the static case, cells do not divide anymore, either because of an evolved mechanism in the regulatory system or because of additional cellular properties like contact inhibition. If the steady state is dynamically stable, cell proliferation and differentiation still occur; however, they are compensated for by cell apoptosis. This leads to a state of constant cell turnover as in the cnidarian *Hydra*. Of course, a combination of both mechanisms is conceivable, although supporting biological evidence remains to be revealed.

A frequently used simple benchmark problem is the evolution of a French flag [39] represented by cells that differentiate into stripes of color blue, white, and red. Miller showed that some flags reach a stable state after the developmental process. In his model, the cells are located on a fixed grid and cannot move. Some resulting designs are also capable of regeneration, although the regrown pattern shows scarring [26].

In the gene regulatory network model of Knabe et al. [24], cells consist of several pixels and are also used to represent the French flag.

Andersen et al. [2] outline a GRN model based on the approach by Eggenberger Hotz, where development converges to a statically stable state due to contact inhibition. In their framework, the cells are fixed on a grid and cells that are surrounded by other cells cannot divide anymore. Therefore, this work represents a typical example where stability is reached due to some additional (external) constraints.

Basanta et al. [5] evolve static and dynamic stability using a model based on cellular automata. The authors define a set of 100 rules, which directly depend on the development time (or steps), the number of neighboring cells and their positions, and the number of divisions the cell has already performed. Therefore, the basic ingredients, such as the cell positions or the number of cell divisions, are prespecified in order to enable convergence to a stable state. Nevertheless, the results are very interesting, and the authors evolved the first systems that exhibited regeneration.

In our previous work, we evolved behaving cellular animats [31] whose cell functions were represented using the same gene regulatory network model as outlined in the next section. Based on the same underlying model, the target of this article is to achieve dynamic stability and regeneration without additional constraints, using physical cells that interact in a continuous space.

2 Computational Model for Morphological Development

The growth of the animat morphology is under the control of a gene regulatory network (GRN) and cellular physical interactions; see Schramm et al. [31]. The morphological development starts with a single cell put at the center of a two-dimensional computational area (100×80). A cell is circular and defined by its center and radius ($r = 1$). Cells are not fixed to a grid and can divide or die, depending on their genes. Nonlinear adhesion and repulsion forces control the interaction between the cells. For distances that are smaller than the sum of the two radii (i.e., the cells overlap), they repel each other. Cells that have a small overlap adhere to each other. Cells that do not overlap attract each other with decreasing forces as the distance increases. For distances longer than five times the radii, there are no forces between the two cells. We add a small jitter, a random value, to the positions of all cells.

Foremost, the level of abstraction of a model is governed by the scientific questions one wants to answer by employing the model. At the same time, without knowing the answer, it is often not possible to unambiguously determine the optimal level. In our case, we are interested in a better understanding of growth and regeneration processes in order to be able to determine the appropriate abstraction level for transferring biological principles into technical solutions. Although some model

properties can be determined (e.g., the need for a continuous "cellular world"), others (e.g., the detail of the gene regulatory process) are not known beforehand. In the latter case, we have chosen a rather pragmatic approach by using an abstraction level whose computational complexity still allows us to couple it to an evolutionary optimization process.

Therefore, we based our model on the work of Eggenberger Hotz et al. [11], which starts from a single cell and develops to a morphology. In contrast with Miller's work, our cells are not fixed to a grid, and our genome structures are less abstract [26].

The cell growth model is slightly modified from the one in [19, 38]. The model uses a GRN to regulate the developmental process. The GRN is defined by a set of genes consisting of regulatory units (RUs) and structural units (SUs); refer to Figure 1. SUs define cellular behaviors, such as cell division and cell death, or the production of transcription factors (TFs) for intracellular and intercellular interactions. The activations of the SUs are determined by the associated RUs. Note that single or multiple RUs may regulate the expression of a single or multiple SUs and that RUs can be activating (RU^+) or repressive (RU^-).

Each RU and TF has a certain affinity value that determines whether the TF can influence the RU. If the difference between the affinity values of a TF and an RU is smaller than a predefined threshold ϵ (in this work ϵ is set to 0.2), the TF can bind to the RU. The affinity overlap ($\gamma_{i,j}$) between the ith TF and jth RU is defined by

$$\gamma_{i,j} = \max\left(\epsilon - \left|\mathrm{aff}_i^{\mathrm{TF}} - \mathrm{aff}_j^{\mathrm{RU}}\right|, 0\right). \tag{1}$$

If $\gamma_{i,j}$ is greater than zero, then the concentration c_i of the ith TF is checked to see if it is above a threshold ϑ_j defined in the jth RU:

$$b_{i,j} = \begin{cases} \max\ (c_i - \vartheta_j, 0) & \text{if } \gamma_{i,j} > 0, \\ 0 & \text{else.} \end{cases} \tag{2}$$

Thus, the activation level contributed by the jth RU (denoted by $a_j, j = 1,\ldots, N$) can be calculated as follows:

$$a_j = \sum_{i=1}^{M} b_{i,j}, \tag{3}$$

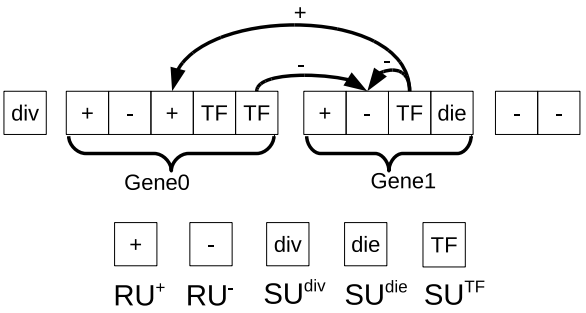


Figure 1. An example chromosome for the development. The first gene (gene 0) starts at the first RU (regulatory unit) of the genome. Each transition from an SU (structural unit) to an RU defines a boundary between two genes. The connections between SU^{TF}s and RUs are determined by their affinity overlap $\gamma_{i,j}$ as defined in Equation 1. Whether the interaction is activating or inhibiting is determined by the sign of the RU. All genes with all their connections result in a GRN, which can be visualized (see, e.g., Figure 7 in Section 4.1.1).

where M is the number of TFs that regulate the jth RU. Assuming the kth structural gene is regulated by N RUs, the expression level of the gene can be defined by

$$\boldsymbol{\alpha} = \boldsymbol{g}(\boldsymbol{c}), \tag{4}$$

$$g_k(\boldsymbol{c}) = 100 \sum_{j=1}^{N} l_j a_j (2s_j - 1), \qquad s_j \in (0, 1), \tag{5}$$

where $s_j \in [0, 1]$. Here $2s_j - 1$ determines the sign (positive for activating and negative for repressive) of the jth RU, and l_j is a parameter representing the strength of the jth RU. If $\alpha_k > 0$, then the kth gene is activated ($\delta_k = 1$) and its corresponding behaviors coded in the SUs are performed.

An SU that produces a TF (SU^{TF}) encodes all parameters related to the TF, such as the affinity value, the decay rate D_i^c, and the diffusion rate D_i^f, as well as the amount of the TF to be produced:

$$\boldsymbol{A} = \boldsymbol{h}(\boldsymbol{\alpha}),$$

$$h_i(\alpha_k) = \begin{cases} \beta\left(\frac{2}{1+e^{-20f\alpha_k}} - 1\right) & \text{if } \alpha_k > 0, \\ 0 & \text{otherwise,} \end{cases} \tag{6}$$

where f and β are both encoded in the SU^{TF}.

A TF produced by an SU can be partly internal and partly external. To determine how much of a produced TF is external, a percentage ($p^{ex} \in [0, 1]$) is also encoded in the corresponding gene. Thus, $p^{ex}A$ is the amount of external TF, and $(1 - p^{ex})A$ is that of the internal TF, to be produced.

External TFs are put on four grid points around the center of the cell, which undergo first a diffusion process and then a decay process:

$$\text{Diffusion:} \qquad \boldsymbol{u}_i^*(t) = \boldsymbol{u}_i(t-1) + 0.1 D_i^f (\boldsymbol{G}\boldsymbol{u}_i(t-1)), \tag{7}$$

$$\text{Decay:} \qquad \boldsymbol{u}_i(t) = (1 - 0.1 D_i^c)\ \boldsymbol{u}_i^*(t), \tag{8}$$

where $\boldsymbol{u}_i$ is a vector of the concentrations of the ith TF at all grid points, and the matrix $\boldsymbol{G}$ describes which grid points are adjoining. Internal TFs bear only a decay process:

$$c_i^{int}(t) = \left(1 - 0.1 D_i^c\right) c_i^{int}(t-1). \tag{9}$$

All internal and external concentrations of TFs are limited to the interval [0, 1].

In our experiments we defined two prediffused, external TFs without decay or diffusion in the computation area. The predefined TFs simulate the maternal gradients available in the early development of many animals. The first TF has a constant gradient in the x direction, and the second in the y direction, so positional information is available for all cells.

The SUs encode cellular behaviors and the related parameters. The SU for cell division encodes the angle of division, indicating where the daughter cell is placed. A cell with an activated SU for cell apoptosis dies at the developmental time step at which it is activated. When both cell apoptosis and cell division are active at the same developmental step, only cell apoptosis is performed. There are two additional SUs for other possible actions, which are not used in this work; therefore, it can happen that genes are activated but no action is performed.

For the complete development of an individual, several developmental steps are computed. The number of developmental steps is determined in the following section. Each developmental step consists of different parts. Firstly, the concentrations of all internal and external TFs simulating diffusion and decay are computed. After that, the activations of all cells are computed, depending

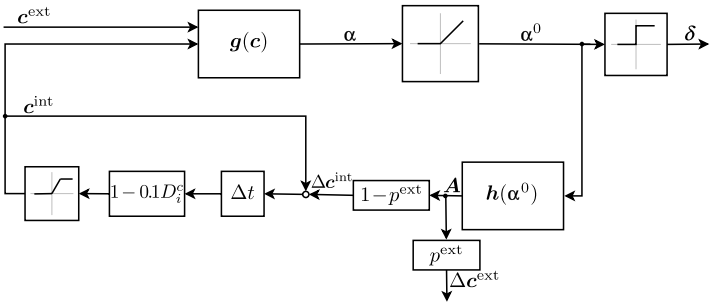


Figure 2. Block diagram of the model of a single cell. The concentrations **c** of the TFs inside and outside the cell determine an expression level α of each gene. If α > 0, the activation of the associated gene, δ, is set to one and the gene becomes active. The expression level α is also used to define the strength of the activation of a gene and therefore the amount of a TF to be produced, ***A***. All internal TFs underlie a decay process and are used in the next developmental step to compute the activations of the genes again. The external TFs underlie a diffusion and a decay process, which are not depicted in the figure.

on their genes and the concentrations of the TFs. Lastly, the actions are executed and the mechanical interactions between the different cells are computed.

Figure 2 shows a block diagram of the main components of a GRN in one cell, describing the cell dynamics. The cell dynamics can become coupled through external transcription factors, which undertake a diffusion process (Equation 7) and a decay process (Equation 8) and are position-dependent. The number of TFs involved in gene regulation of the cellular behaviors is determined by the genome and the parameters in the resulting GRN as well. The number of cells also changes during development; we start with a single cell and two external TFs. The maximum number of cells is limited to 700 for reducing the computational cost.

The most relevant parameters encoded in the RUs and SUs are summarized in Tables 1 and 2.

3 Evolution of the Developmental Model for Stable Growth

We use an extended evolution strategy (ES) with $\mu = 30$ and $\lambda = 200$ and with three elitists (see [7] for an introduction to ESs). Three elitists means that the best three individuals of the current generation can be passed to the next generation. The strategy parameter σ is fixed to 10^{-4}. Let k_i be a real-valued parameter encoded in the chromosome. A mutation operation changes its value as follows:

$$k_i' = k_i + z_i, \tag{10}$$

Table I. Five real-valued parameters encoded in the regulatory units (RU_j).

g_1	Affinity value for the RU, aff_j^{RU}
g_2	Threshold for the concentration ϑ_j of the TF that reacts on the RU
g_3	Sign: [0, 0.5] → inhibitory (–) [0.5, 1] → activating (+)
g_4	Strength l_j of the RU
g_5	—

Table 2. Nine real-valued parameters encoded in the structural units (SU_i).

Paramater	TF	Divide	Die
s_1	[0, 0.2]	[0.2, 0.4]	[0.4, 0.6]
s_2	—	—	—
s_3	—	—	—
s_4	TF affinity aff_i^{TF}	—	—
s_5	Decay rate D_i^c	—	—
s_6	Diffusion rate D_i^f	—	—
s_7	Expression rate β	—	—
s_8	f (see Equation 6)	Division angle	—
s_9	Externality p_{ext}	—	—

where

$$z_i \sim \mathcal{N}(0, \sigma_i). \tag{11}$$

In addition to the standard operators, we use gene duplication, gene transposition, and gene deletion in order to be able to structurally change the genome. Gene duplication randomly copies a sequence of RUs and SUs in the chromosome and then inserts it into the chromosome at a different randomly chosen position. In the case of gene transposition or deletion, a random sequence is picked out of the RUs and SUs and is moved to another randomly chosen site on the chromosome, or is simply removed. The corresponding probabilities are $p_{\text{dup}} = 0.05$, $p_{\text{trans}} = 0.02$, and $p_{\text{del}} = 0.03$.

Due to the physical interactions between the cells, circular shapes are easy to develop using the model described above. Hence, in order to solve a more demanding task and to be conceptually closer to the *Hydra* morphology, we decided to evolve an elongated shape; see Figure 3.

The fitness is defined as follows:

$$f = p_1 - p_2 - \min\left\{\min_i\{y^i\}, -5\right\} + \max\left\{\max_i\{y^i\}, 5\right\}, \tag{12}$$

where (x^i, y^i) represents the position of the ith cell,

$$p_1 = \begin{cases} 70 - \min_i\{x^i\} & \text{if } \min_i\{x^i\} < -40, \\ -30 & \text{if } -40 < \min_i\{x^i\} < -30, \\ \min_i\{x^i\} & \text{otherwise,} \end{cases} \tag{13}$$

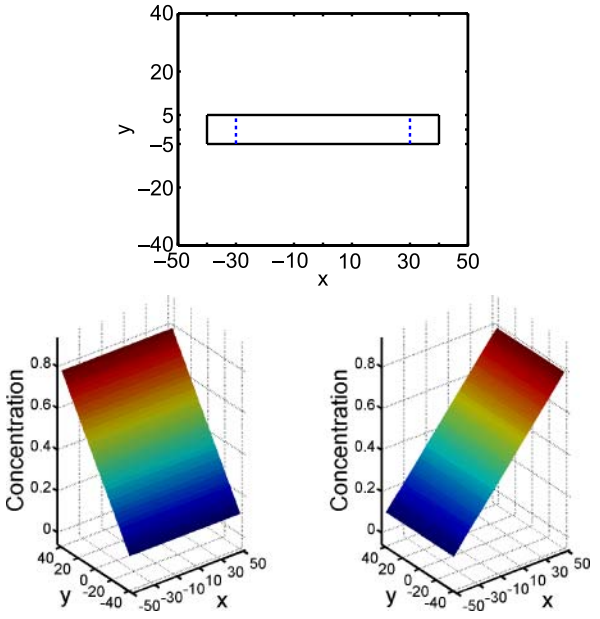


Figure 3. In the electronic version, top: Optimal shape of the individuals. All cells should be connected, and the individual should be longer than the blue, dashed lines. Cells lying outside the black, solid box are penalized. Bottom: The concentrations of the prediffused transcription factors on the continuous space (x,y) are shown on the z axis and by the color code (red = 0.8, blue = 0.1).

and

$$p_2 = \begin{cases} -70 - \max_i\{x^i\} & \text{if } \max_i\{x^i\} > 40, \\ 30 & \text{if } 40 > \max_i\{x^i\} > 30, \\ \max_i\{x^i\} & \text{otherwise.} \end{cases} \tag{14}$$

We add a penalty term if the number of cells n_c is below 10 ($f = 600 - n_c$) or above 500 ($f = n_c$). If the individual is composed of unconnected parts (i.e., one or several cells have a large distance to all other cells), the fitness is set to 50.

We designed three experiments in order to test the system's ability for stable growth and regeneration. Equation 12 expresses the fitness at one developmental step. We define three different evaluation procedures that compute the fitness at different developmental steps. The first setup evaluates individuals once a certain number of developmental steps is completed. In the second setup, more developmental steps are simulated, and the fitness is the mean of several evaluations after a certain number of developmental steps. In the third setup, the regeneration capacity of our artificial organism is tested by removing cells after a certain number of developmental steps:

- *Setup 1* (S1): Simulation of 16 developmental steps and evaluation after step 16.
- *Setup 2* (S2): Simulation of 30 developmental steps and evaluations after steps 16, 17, 18, 19, 20 and 26, 27, 28, 29, 30. Thus, the fitness is the mean of the 10 evaluations.
- *Setup 3* (S3): Same as setup 2, except that the individual is injured at its center once after step 20 by removing cells with $-5 < x < 5$.

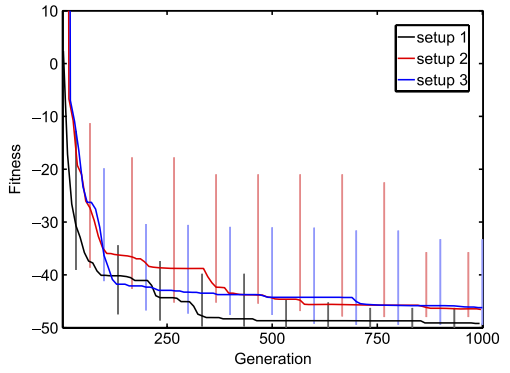
According to the criteria in Equations 12–14 the best achievable fitness is −50; for setups 2 and 3 the same fitness is computed 10 times and averaged. Ten trials with different random seeds for each

setup represents the upper limit to what is computationally feasible. We stopped the evolution if the global optimum of −50 was reached or after 1000 generations.

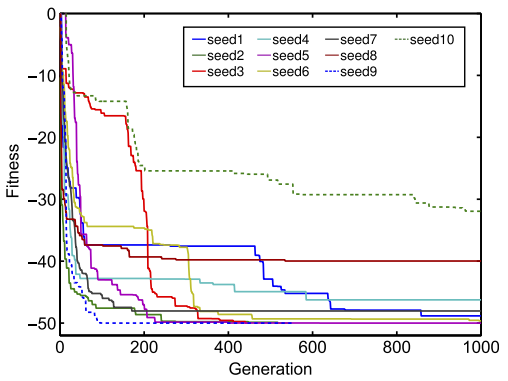
Experimental Framework The program was written in C++ and parallelized using MPI. The hardware environment consisted of 30–60 compute nodes (dependent on availability), each with two Opteron 250 single-core CPUs at 2.4 GHz. Depending on the number of compute cores, the number of developmental steps, and the number of cells, one complete evolutionary run took between 2–3 hours and 2–3 days.

4 Results

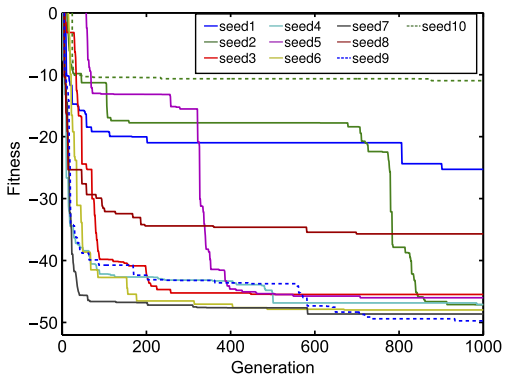
The resulting fitness curves are shown in Figure 4. Figure 4a shows a boxplot (median and 25th and 75th percentiles) of all 10 evolutionary runs for the three setups. The mean fitness of S1 is better than the ones of S2 and S3, which confirms the assumption that the first setup is easier to evolve. The experiment using only 16 developmental steps (S1) shows that 4 out of 10 experiments reach the global optimum (−50); only one experiment fails completely, with a fitness of −31.94. During the experiments without removing cells (S2), only one experiment reaches the global optimum, but



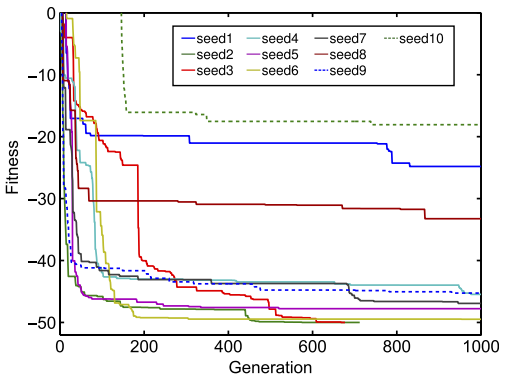
(a) Boxplot (median and the 25th and 75th percentiles) of the fitness curves of the three setups.



(b) Fitness curves of setup 1

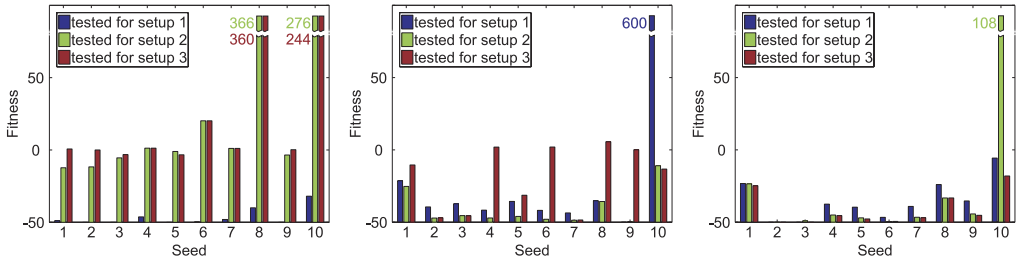


(c) Fitness curves of setup 2



(d) Fitness curves of setup 3

Figure 4. (a) Mean and variance of the fitness for the three experiments described in the text (setups 1–3). Fitness curves for 10 runs with different random seeds with (b) setup 1, (c) setup 2, and (d) setup 3.



(a) evolved for setup 1 (b) evolved for setup 2 (c) evolved for setup 3

Figure 5. The fitness of the best 10 individuals of the 10 evolutionary runs for each setup, computed for all three setups. For example, in the electronic version, the green bars in (a) show the fitness results for the best 10 individuals resulting from the evolution with setup 1, tested for setup 2. It is clear that the best results are always achieved by those individuals that have been evolved for the setup they are tested with, that is, blue bars (a), green bars (b), and red bars (c).

still only two setups fail, with a fitness worse than −40. The experiments with removing cells in developmental step 20 (S3) show that only 2 of the 10 experiments reach the global minimum. Three runs result in low performance.

To analyze the developmental stability and the regeneration capability of the different evolved artificial organisms, we compare the resulting individuals of all runs on all evaluation setups. The results are shown in Figure 5. Figure 5a shows the fitness values of the 10 individuals evolved using the fitness defined in S1 reevaluated using the fitness function in S1, S2, and S3, respectively. Similarly, individuals evolved using the fitness function in S2 and S3 are reevaluated using the fitness functions in the three setups, and their fitness values are presented in Figure 5b and Figure 5c, respectively. Recall that the three setups are of increasing complexity and that S2 includes aspects of S1, and S3 includes aspects of S1 and S2. We expect that individuals evolved in S3 can still perform well in S1 and S2 and individuals from S2 and S3 can perform well in S1. Indeed, the results in Figure 5 confirm this expectation, although averaging over 10 different developmental steps (in S2 and S3) slightly reduces the selection pressure to have a perfect individual after step 16, which is reflected by the slightly worse performance of individuals evolved for S2 and S3 on S1 than that of individuals evolved for S1 on S1. As is evident from Figure 5a, individuals evolved for S1 do not reach a dynamically stable state in any of the 10 runs. Thus, from our experiments, we can conclude that dynamically stable growth is evolvable only under explicit selection pressure toward it.

At the same time, we observe from Figure 5b that in three cases individuals evolved for S2 also perform well for S3. Thus, once the genetic mechanisms for reaching a dynamically stable state have been evolved, the regeneration capability follows almost without additional effort. In our experiments, regeneration is also observed in 3 out of 10 cases in which stable growth is evolved.

4.1 Analysis of the Development of Three Individuals

In this section, we analyze the dynamics of the GRN and the structure behind the process of dynamically stable growth and regeneration for three individuals. We picked one of the best individuals that were evolved for each of the three setups. We are particularly interested in what happens during the developmental process—for example, which genes are used and how the genes are activated.

4.1.1 First Individual: Setup 1, Run 9

Figure 6 shows the cellular development of the morphology of the best individual evolved from S1, run 9 (S1R9). At the beginning of the development, the cells in the middle (y direction) divide, and cells in the upper part die. The cells in the lower part perform no action. After developmental step 12, cell apoptosis becomes active in the lower part. At developmental step 15, cellular function

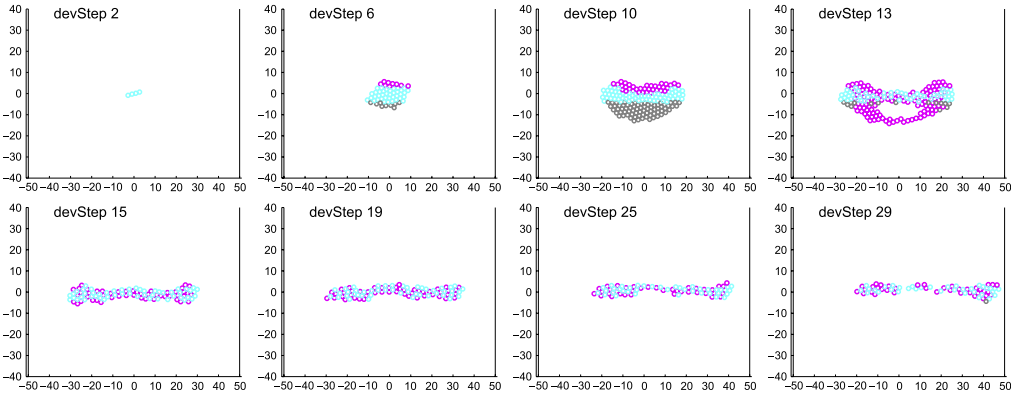


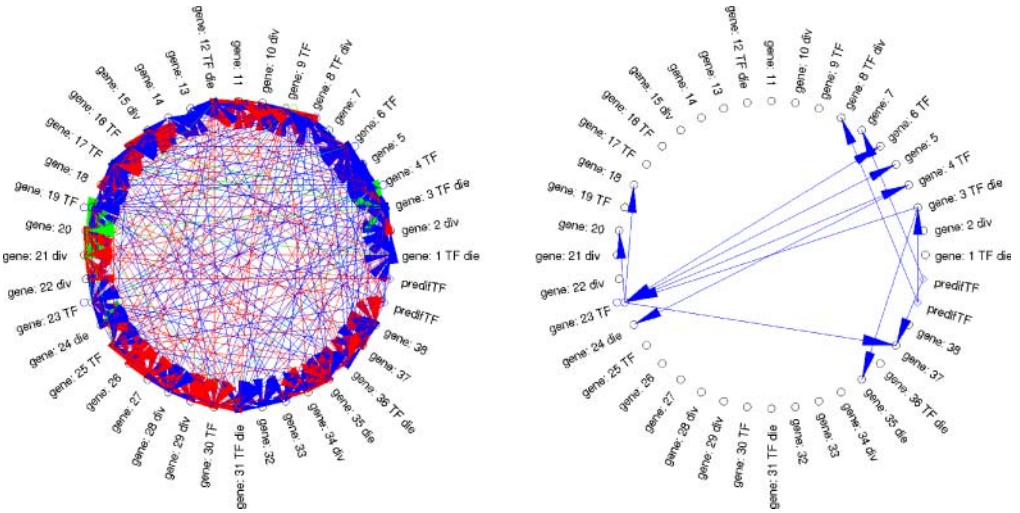
Figure 6. The development of the best individual of setup 1, run 9. In the electronic version, cells that will divide in the next developmental step are marked in blue, dying cells in magenta, and idle cells in gray.

changes and every second cell dies. (Note that this individual is only evaluated at developmental step 16.) Then each cell first divides and dies in the next developmental step.

The complete gene regulatory network, shown in Figure 7a, is highly connected. However, if we remove all connections that are never activated during development, only a few remain, as can be seen from Figure 7b.

Remarkably, there are no inhibitory connections in the pruned GRN (the part of the GRN that is actually activated). We can see that only one gene for cell division is used (gene 8) and three genes for cell death are used (genes 3, 24, and 35).

Comparing the individual gene activations (data not shown) with the growing phenotype reveals that the gene for cell division is active in cells above a fixed x position; see also Figure 6.



(a) Complete GRN (b) Pruned GRN

Figure 7. GRN of the best individual of setup 1, run 9. Dots represent the different genes, diamonds the two prediffused TFs. Arrows symbolize the connections between the different TFs and genes. In (a) we show all possible interactions coded in the genome; in (b) we show only the interactions that are used during development. In the electronic version, red arrows represent an inhibiting connection, blue an activating one, and green arrows represent both an activating and an inhibiting influence.

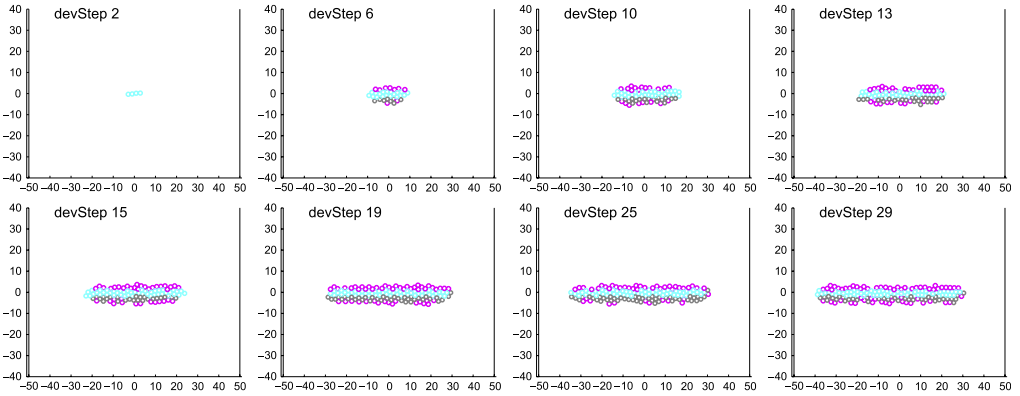


Figure 8. Development of the best individual of setup 2, run 2. Refer to Figure 6 for explanation of the color coding in the electronic version.

As apoptosis is performed prior to cell division (if the GRN activates both functions in a cell), gene 3 (cell death) is also activated by the first predefined TF and active above a horizontal line. The activation of gene 35 is coupled to the activation of gene 3, which has no further influence on the development. The activation of gene 24 is more complex; it depends on the activation of gene 4, which itself depends on gene 23. The activation of gene 4 starts at developmental step 10; the concentration of the TF produced by gene 4 is high in the middle of all cells producing it. Every cell needs two developmental steps to produce enough of the TF to activate gene 24.

4.1.2 Second Individual: Setup 2, Run 2

The second example is the best individual from the second evolutionary run on S2, whose development is shown in Figure 8. Cellular function is position-dependent. All cells above and below a defined y value die, and cells in between divide, which is a simple but effective way of achieving a dynamically stable state in the y direction.

Again after removing all unused connections, the pruned GRN shown in Figure 9 can be analyzed.

Only one gene (15) represents cellular division, which is activated by the first predefined TF and deactivated by the second. Therefore, gene 15 is active above a horizontal line, but deactivated at the

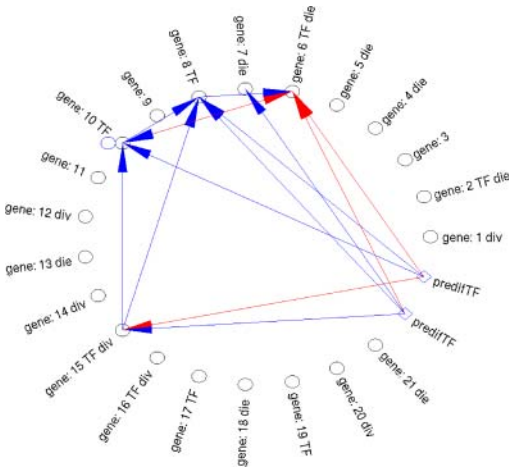


Figure 9. Pruned GRN of the best individual of setup 2, run 2. See Figure 7 for a detailed description of the nodes.

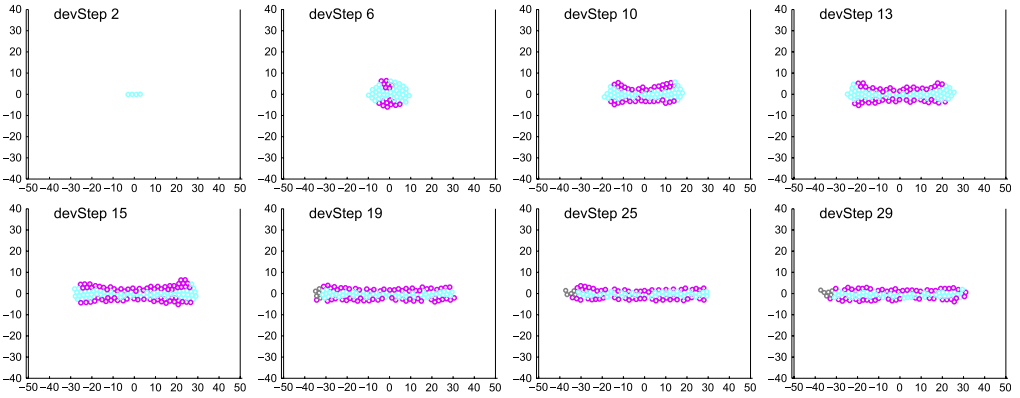
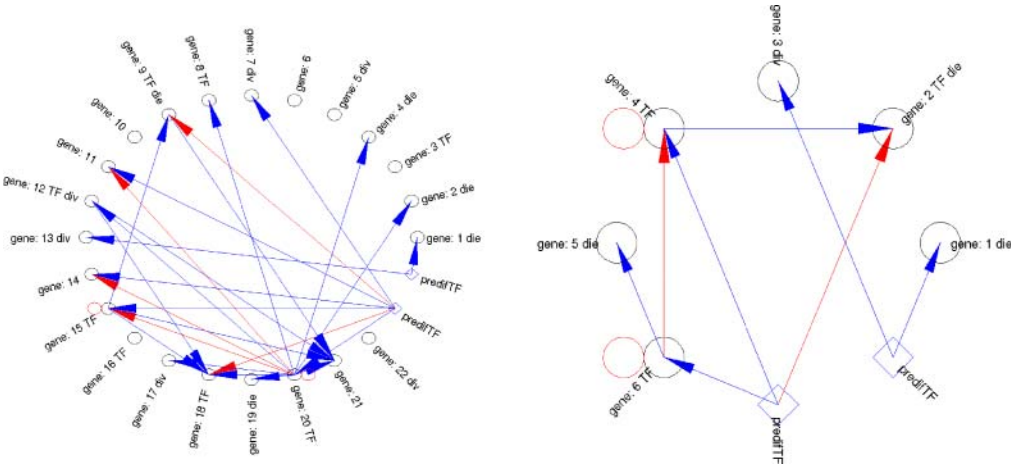


Figure 10. Development of the best individual of setup 3, run 2. Refer to Figure 6 for explanation of the color coding in the electronic version.

right side of the individual. Genes 7 and 6 (cell apoptosis) are activated above and below a horizontal line, respectively.

4.1.3 Third Individual: Setup 3, Run 2

The development of the third example individual from S3 is shown in Figure 10. The development is comparable to that of the second individual, but the pruned GRN is different; see Figure 11. Since the pruned GRN is still rather complex, we further simplify it by removing all connections that do not influence the fitness of the individual even though it is activated during development. The pruned and simplified GRN is shown in Figure 11b. Gene 13 and genes 1, 9, and 20 represent cellular division and apoptosis. Gene 13 (division) is activated by the second predefined TF and is active in all cells except the far left ones. Gene 1 (apoptosis) is also activated by the second predefined TF and is only active in cells at the right end of the individual. Gene 9 (apoptosis) is active at the bottom of the individual. It is activated by the TF produced by gene 15 and deactivated by the first predefined TF. Gene 15 is active in all cells. Above a horizontal line, the TF of gene 20 activates it. Gene 20 (apoptosis) itself is also only active at the top and activated by the first predefined TF.



(a) Pruned GRN (b) Prunded and further simplified GRN

Figure 11. GRN of the best individual of setup 3, run 2. For a detailed explanation see Figure 7.

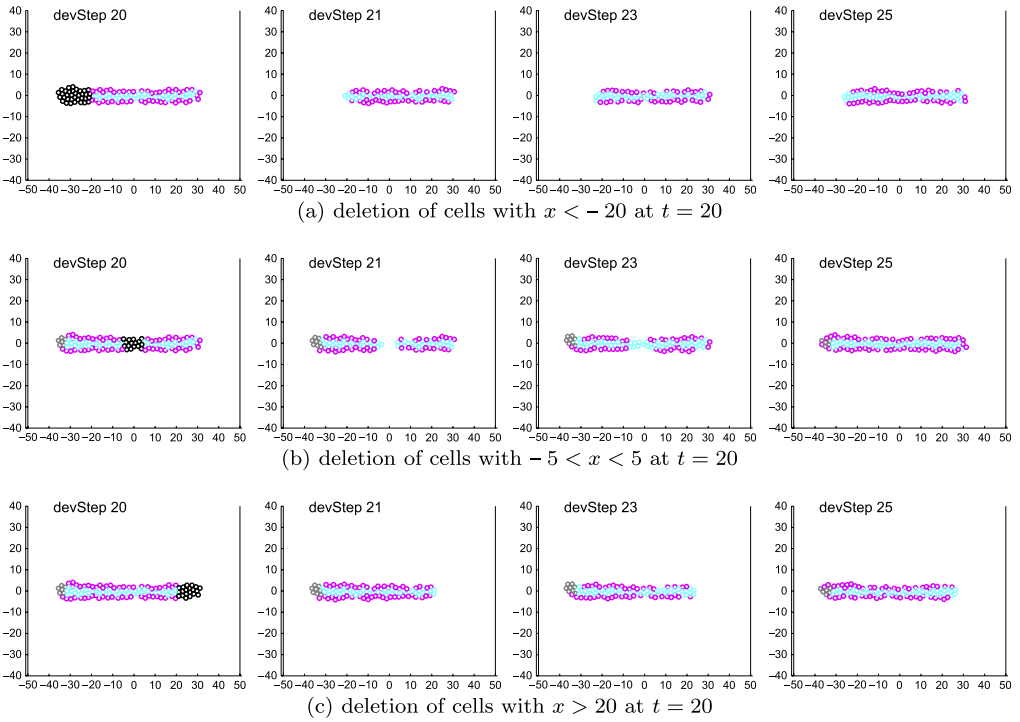
Note, however, that the pruned and simplified GRN is less stable against mutations than the pruned GRN. Therefore, the redundancy improves the evolvability of the genetic representation; see [32] for a more detailed discussion.

Next, we deleted some cells at developmental step 20 in order to observe the regeneration process. In Figure 12a, 12b, 12c we removed cells on the left, the center, and the right of the individual.

Figure 12 shows that in all three cases the regeneration process takes about five developmental steps. Although only deletion at the center of the artificial organism was included in the evaluation in S3, the resulting GRN function works well in all three cases. The high morphological plasticity is not locally confined, and therefore it does not matter where cells (or how many cells) are removed. We use the term *morphological plasticity* to denote the ability of an artificial organism to change its morphology (i.e., in our case the cellular composition of the phenotype) after the development process.

4.2 The Role of Morphogen Gradients and the Evolution of Motifs

In most of the individuals, cellular growth is structured by controlling cell apoptosis through predefined (morphogen gradients) or produced transcription factors. Therefore, we first count the activations of an apoptosis gene above the individual with $25 < y < 45$, which we term motif 1 (M1). The activation can be directly regulated through the predefined TF or indirectly through a TF of a gene that is activated by the predefined TF. Of course, deeper cascades are possible; however, mutational stability is then reduced. The second motif (M2) performs the same function, namely, controlled cell apoptosis on the right end of the individual with $0 < x < 10$. These motifs define the dependence of activations of specific genes on the predefined TFs and are not directly related to the structural network motifs such as feedforward loops. Different pathways of activation can lead to the same motif. We analyze the structural network motifs in [34].



(a) deletion of cells with $x < -20$ at $t = 20$

(b) deletion of cells with $-5 < x < 5$ at $t = 20$

(c) deletion of cells with $x > 20$ at $t = 20$

Figure 12. Regeneration: Development of the best individual of setup 3, run 2. After developmental step 20 cells have been removed at different positions: (a) left, (b) middle, (c) right side of the individual.

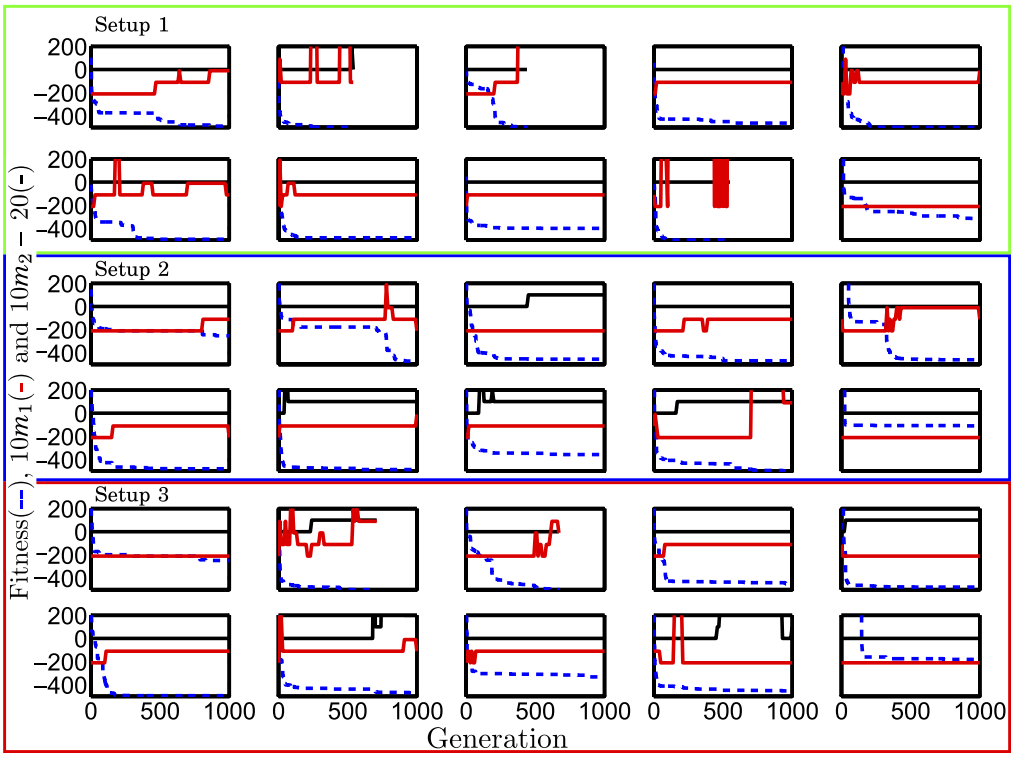


Figure 13. Activations of the morphogen gradients are displayed for setup 1 in the first two rows, for setup 2 in rows 3 and 4, and for setup 3 in the last two rows. The *x* axis shows the number of generations. In the electronic version, the blue, dashed line shows the fitness, the red solid line shows the number of motifs M1, and the black solid line shows the number of motifs M2 in the genome. Refer to the text for a definition of motifs M1 and M2. The motifs are counted every tenth generation.

The plots in Figure 13 show the fitnesses of the different individuals and the occurrences of M1 and M2, where m_1 is the number of used paths activating an apoptosis gene at $25 < y < 45$, and m_2 is the number of used paths activating an apoptosis gene at $0 < x < 10$. The numbers of motifs are scaled in Figure 13, so $10m_1$ and $10m_2 - 20$ are plotted. M1 occurs in most runs early on during evolution, and its appearance is often accompanied by a fitness increase. If M1 is not present, either the fitness is low as in the individuals in setup 1, run 10 (S1R10), S2R10, S3R1, and S3R10, or an alternative mechanisms has evolved as in S2R3, S3R5, and S3R9:

- S2R3: Apoptosis is activated by an internal TF that is always produced. The cells that divide never reach the threshold of the TF to activate apoptosis, because during division the concentration of the internal TF is halved.
- S3R5: Cells divide only at the left and right ends of the individual, so no apoptosis is required in the *y* direction. The artificial organism does not reach a dynamically stable state, because growth on the left side is uncontrolled. M2 controls cellular growth on the right side of the individual.
- S3R9: At the beginning of the development a blob forms, followed by an elongated shape. Most cells first divide and then die in the next developmental step, comparable to the first analyzed individual of the previous section.

The second motif does not occur in S1, because after 16 developmental steps the individuals have not reached the left or the right border of the morphological fitness constraint. In S2 and

S3, 4 out of 10 runs evolved the second motif, whose occurrence is often accompanied by a fitness increase, as in S1R3, S1R4, S2R5, S2R7, S3R4, S3R5, and S3R6.

Therefore, controlled cellular growth leading to a dynamic equilibrium of high morphological plasticity and regeneration capability is governed mostly by the interaction of predefined TFs (morphogen gradients) with cellular apoptosis and, in fewer cases, with cellular division. In all cases, cellular apoptosis is position-dependent.

5 Discussion

In this article, we have proposed a framework for the evolution of a cellular morphology of an artificial organism whose growth process is governed by a gene regulatory network. Without having to rely on additional functions like contact inhibition or on additional information like the number of surrounding cells, we have been able to evolve individuals that reach a dynamically stable state of high morphological plasticity during development. In this state, cell turnover is in equilibrium, that is, cell proliferation is on the same order as cell apoptosis. Dynamic stability could only be reached after distributing the evaluation of each individual over a number of developmental steps (in our experiments, 10).

The regeneration ability of the artificial organisms is often a result (3 out of 10 cases) of high morphological plasticity. We speculate that once morphological plasticity has been evolved but regeneration is not yet realized, additional evolution of the functional regulatory networks to enable regeneration is straightforward, or at least easier than it was to discover dynamically stable growth in the first place.

The detailed analysis of the genetic regulatory networks revealed that most individuals achieve dynamically stable growth and regeneration through the position-dependent control of cellular apoptosis using prediffused transcription factors (morphogen gradients). Stable growth and controlled cell removal leads to high morphological plasticity and to the desired regeneration capability. These are achieved, however, at the expense of higher energy costs, which we did not take into account in our experiments. Two motifs that represent such a control can be identified that occur in most successful artificial organisms.

The individuals evolved for the evolutionary framework S1 do not reach a dynamically stable state in any of the 10 runs. Therefore, it seems to be more difficult to evolve dynamical stability than to evolve a specific morphology. Of course, alternative evaluation schemes are conceivable, such as evaluating after different steps and recording the best values. This way, one could decouple the evaluation from one specific step but still only evaluate one step.

Furthermore, in our framework we circumvent the important question of signal generation by relying on the prediffused morphogen gradients to guide both the developmental and the regeneration processes. By introducing (and artificially maintaining) the maternal gradients, we provide the scaffolding for regeneration and we provide a global coordinate system to the artificial organism. A model without such a priori information would both be more realistic, and it would also have a higher explanatory value. However, the first experiments that we conducted suggest that it is very difficult to evolve gene regulatory networks that can lead to dynamical stability and regeneration without the constraints.

The framework and the experiments presented in this article have been inspired by the amazing regeneration capability of biological organisms. At the same time, as we elaborated in the introduction, we aim at novel approaches for the design of complex engineering systems. The target of this phase in our research was not to devise a new method, but to improve our understanding and to identify the most promising directions for the next phase. We can conclude:

- It is possible to evolve cellular growth of a simple morphology without introducing hard constraints on the physical world (fixing cells to a grid, contact inhibition, etc.). Therefore, in principle the system can be applied to topology optimization. However, it remains to be seen whether it can scale up to more complex geometries.

- The evolved GRNs are overly complex. At the same time, the more detailed analysis identified a small number of active motifs that govern the process. In [33] we show that a certain level of redundancy is required for the evolutionary process. However, from a more practical viewpoint, it is recommended to exploit the motifs more strategically (e.g., by supplying a priori dynamics modules to the optimization algorithm). This is likely to improve the scaling behavior.
- Regeneration is a fascinating topic, and we are just at the beginning of understanding the processes involved. Such knowledge will allow us to use such models in an engineering context. The high morphological plasticity that we observe in our experiments could lead to highly dynamic simulation meshes in engineering simulations, which continuously improve and repair themselves when changes are made to the underlying geometries either by a shape optimization process or by joining different parts together. Methods for adapting meshes are already in use (see, e.g., [28]), and techniques for endowing individual cells with globally controlled autonomy are conceivable in the near future.

Acknowledgments

We thank Till Steiner for inspiring discussions. This work was supported by Honda Research Institute Europe.

References

1. Alon, U. (2007). Network motifs: Theory and experimental approaches. *Nature Reviews Genetics*, *8*, 450–461.
2. Andersen, T., Newman, R., & Otter, T. (2009). Shape homeostasis in virtual embryos. *Artificial Life*, *15*(2), 161–183.
3. Ashe, H. L., & Briscoe, J. (2006). The interpretation of morphogen gradients. *Development*, *133*(3), 385–394.
4. Babu, M. M., Luscombe, N. M., Aravind, L., Gerstein, M., & Teichmann, S. A. (2004). Structure and evolution of transcriptional regulatory networks. *Current Opinion in Structural Biology*, *14*(3), 283–291.
5. Basanta, D., Miodownik, M., & Baum, B. (2008). The evolution of robust development and homeostasis in artificial organisms. *PLoS Computational Biology*, *4*(3), e1000030.
6. Bentley, P. J., & Kumar, S. (1999). Three ways to grow designs: A comparison of embryogenies for an evolutionary design problem. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO '99)* (pp. 35–43).
7. Beyer, H.-G., & Schwefel, H.-P. (2002). Evolution strategies—A comprehensive introduction. *Natural Computing: An International Journal*, *1*(1), 3–52.
8. Blaiszik, B., Kramer, S., Olugebefola, S., Moore, J., Sottos, N., & White, S. (2010). Self-healing polymers and composites. *Annual Review of Materials Research*, *40*, 179–211.
9. Cooper, M. B., Loose, M., & Brookfield, J. F. (2008). Evolutionary modelling of feed forward loops in gene regulatory networks. *Biosystems*, *91*(1), 231–244.
10. Crombach, A., & Hogeweg, P. (2008). Evolution of evolvability in gene regulatory networks. *PLoS Computational Biology*, *4*(7), e1000112.
11. Eggenberger Hotz, P., Gómez, G., & Pfeifer, R. (2003). Evolving the morphology of a neural network for controlling a foveating retina—And its test on a real robot. In *Proceedings of the 8th International Conference on Artificial Life (ALife 03)* (pp. 243–251).
12. Federici, D., & Downing, K. (2006). Evolution and development of a multicellular organism: Scalability, resilience, and neutral complexification. *Artificial Life*, *12*(3), 381–409.
13. François, P., & Hakim, V. (2004). Design of genetic networks with specified functions by evolution in silico. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(2), 580–585.
14. Furusawa, C., & Kaneko, K. (2006). Morphogenesis, plasticity and irreversibility. *International Journal of Developmental Biology*, *50*, 223–232.

15. Geard, N., & Wiles, J. (2005). A gene network model for developing cell lineages. *Artificial Life*, *11*(3), 249–267.

16. Harding, S., & Banzhaf, W. (2008). Artificial development. In R. W. Würz (Ed.), *Organic Computing* (pp. 201–219). Berlin: Springer.

17. Hogeweg, P. (2002). Computing an organism: On the interface between informatic and dynamic processes. *BioSystems*, *64*(13), 97–109.

18. Jin, Y., & Meng, Y. (2011). Emergence of robust regulatory motifs from in silico evolution of sustained oscillation. *BioSystems*, *103*(1), 38–44.

19. Jin, Y., Schramm, L., & Sendhoff, B. (2008). A gene regulatory model for the development of primitive nervous systems. In *Proceedings of the 15th International Conference on Neural Information Processing (ICONIP)*, Vol. 5506/2009 (pp. 48–55).

20. Jin, Y., & Trommler, J. (2010). A fitness-independent evolvability measure for evolutionary developmental systems. In *IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology* (pp. 69–76).

21. Joachimczak, M., & Wròbel, B. (2009). Complexity of the search space in a model of artificial evolution of gene regulatory networks controlling 3D multicellular morphogenesis. *Advances in Complex Systems*, *12*, 347–369.

22. Joachimczak, M., & Wròbel, B. (2009). Evolution of the morphology and patterning of artificial embryos: Scaling the tricolour problem to the third dimension. In *Proceedings of the European Conference on Artificial Life (ECAL)* (pp. 33–40).

23. Kitano, H. (2002). Computational systems biology. *Nature*, *420*, 206–210.

24. Knabe, J. F., Schilstra, M. J., & Nehaniv, C. (2008). Evolution and morphogenesis of differentiated multicellular organisms: Autonomously generated diffusion gradients for positional information. In *Proceedings of the 11th International Conference on the Simulation and Synthesis of Living Systems (ALife)*.

25. Kuo, P. D., Banzhaf, W., & Leier, A. (2006). Network topology and the evolution of dynamics in an artificial genetic regulatory network model created by whole genome duplication and divergence. *BioSystems*, *85*(3), 177–200.

26. Miller, J. F. (2003). Evolving developmental programs for adaptation, morphogenesis, and self-repair. In *Advances in Artificial Life: 7th European Conference on Artificial Life (ECAL 03)* (pp. 256–265).

27. Miller, J. F. (2004). Evolving a self-repairing, self-regulating, French flag organism. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)* (pp. 129–139).

28. Quan, S. (2011). Simulations of multiphase flows with multiple length scales using moving mesh interface tracking with adaptive meshing. *Journal of Computational Physics*, *230*(13), 5430–5448.

29. Quayle, A. P., & Bullock, S. (2006). Modelling the evolution of genetic regulatory networks. *Journal of Theoretical Biology*, *238*(4), 737–753.

30. Salazar-Ciudad, I., Garcia-Fernández, J., & Solé, R. V. (2000). Gene networks capable of pattern formation: From induction to reaction-diffusion. *Journal of Theoretical Biology*, *205*, 587–603.

31. Schramm, L., Jin, Y., & Sendhoff, B. (2009). Emerged coupling of motor control and morphological development in evolution of multi-cellular animats. In *Proceedings of the European Conference on Artificial Life (ECAL)* (pp. 25–32).

32. Schramm, L., Jin, Y., & Sendhoff, B. (2011). Redundancy creates opportunity in developmental representations. In *Proceedings of the IEEE Symposium Series on Computational Intelligence 2011 (IEEE ALIFE 2011)* (pp. 203–210).

33. Schramm, L., Jin, Y., & Sendhoff, B. (2012). Quantitative analysis of redundancy in evolution of developmental systems. In *IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology*.

34. Schramm, L., Martins, V. V., Jin, Y., & Sendhoff, B. (2010). Analysis of gene regulatory network motifs in evolutionary development of multicellular organisms. In *Proceedings of the 12th International Conference on the Synthesis and Simulation of Living Systems (ALife XII)* (pp. 133–140).

35. Shvartsman, S. Y., Coppey, M., & Berezhkovskii, A. M. (2008). Dynamics of maternal morphogen gradients in the *Drosophila* embryo. *Current Opinion in Genetics and Development*, *18*(4), 342–347.

36. Stanley, K. O., & Miikkulainen, R. (2003). A taxonomy for artificial embryogeny. *Artificial Life*, *9*(2), 93–130.

37. Steiner, T., Olhofer, M., & Sendhoff, B. (2006). Towards shape and structure optimization with evolutionary development. In *Proceedings of the 10th International Conference on the Simulation and Synthesis of Living Systems (ALife)* (pp. 70–76).

38. Steiner, T., Schramm, L., Jin, Y., & Sendhoff, B. (2007). Emergence of feedback in artificial gene regulatory networks. In *Proceedings of the IEEE Conference on Evolutionary Computation (CEC)* (pp. 867–874).

39. Wolpert, L. (2004). *Principles of development*. Oxford, UK: Oxford University Press.

40. Yoshida, H., Furusawa, C., & Kaneko, K. (2005). Selection of initial conditions for recursive production of multicellular organisms. *Journal of Theoretical Biology*, *233*(4), 501–514.

41. 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.