Shapes in the Shadow: Evolutionary Dynamics of Morphogenesis

P. Hogeweg
Theoretical Biology and Bioinformatics Group
Utrecht University
Padualaan 8, 3584 CH Utrecht, The Netherlands
ph@binf.bio.uu.nl

Abstract This article investigates the evolutionary dynamics of morphogenesis. In this study, morphogenesis arises as a side-effect of maximization of number of cell types. Thus, it investigates the evolutionary dynamics of side-effects. Morphogenesis is governed by the interplay between differential cell adhesion, gene-regulation, and intercellular signaling. Thus, it investigates the potential to generate complex behavior by entanglement of relatively "boring" processes, and the (automatic) coordination between these processes.

The evolutionary dynamics shows all the hallmarks of evolutionary dynamics governed by nonlinear genotype phenotype mapping: for example, punctuated equilibria and diffusion on neutral paths. More striking is the result that interesting, complex morphogenesis occurs mainly in the "shadow" of neutral paths which preserve cell differentiation, that is, the interesting morphologies arise as mutants of the fittest individuals.

Characteristics of the evolution of such side-effects in the shadow appear to be the following: (1) The specific complex morphologies are unique (or at least very rare) among the set of de novo initiated evolutionary histories. (2) Similar morphologies are reinvented at large temporal distances during one evolutionary history and also when evolution is restarted after the main cell differentiation pattern has been established. (3) A mosaic-like evolution at the morphological level, where different morphological features occur in many combinations, while at the genotypic level recombination is not implemented and genotypes diverge linearly and at a constant rate.

Keywor ds
evolution, morphogenesis, development, EvoDevo, fitness landscapes, cell adhesion, gene networks

1 Introduction

Morphogenesis is an inherently multilevel process involving processes at different time and space scales. It is conceptually, mathematically, and computationally convenient to treat such multilevel processes as hierarchical processes, which by separating time scales and/or space scales can be studied one hierarchical level at the time. This is how traditional-model formalisms used in studying morphogenesis operate. In contrast, we try to focus on the entanglements between levels. By this we mean the interplay between levels such that behavior on one level (co)determines and/or constrains the behavior of another level, and vice versa. In other words, we consider not only how microlevel "rules" give rise (via a self-structuring process) to (relatively) macrolevel be-

havior, but also how the macrolevel behavior determines the microlevel behavior (i.e., determines which "rules" are activated, and, through evolution, which rules exist). We think that such a reciprocal influence between levels is an essential characteristic for living, adaptable systems [7,8]. In such systems the levels cannot be separated, and thus these systems cannot be simplified in traditional ways. Nevertheless, we aim to study simple models with few parameters, as this is the only way to avoid undue *ad-hocness* and intractability. We can do this by exploiting the entanglement of levels as a research tool. Thus, instead of considering the entanglement as a complicating matter we will make it into an asset for modeling. To this end we proposed to use evolutionary processes and the type of "solution" they choose when confronted with a very general "problem" for which many solutions exist to map the interplay between levels, their mutual constraints, and their potential to generate complex dynamics [9]. Thus, the behavior studied is a side-effect of the maximization of a fitness criterion which is a prerequisite for it, but is itself not a sufficient condition for this behavior.

Using this methodology we have previously demonstrated the morphogenetic potential of the interplay between cell adhesion, cell differentiation and cell signaling, and the resulting cell migration, cell death, and cell growth/division [10]. We found gastrulation-like engulfing, meristematic growth, intercalation and stretching, and very complicated orchestration between cell growth, cell death, and cell differentiation, which dynamically cause "pseudo-isomorphic" outgrowth. All of these have been recognized in biological development as well.

In this article we focus on evolutionary dynamics at the level of the morphologies and their development. Because the morphogenesis arises as a side-effect of cell differentiation, as mentioned above, we in fact focus on the evolutionary dynamics of side-effects.

Clearly the mapping from genotype (a gene-regulation network) to phenotype (cell differentiation and morphology) is highly nonlinear. Evolutionary dynamics subject to a nonlinear genotype-phenotype mapping has been extensively studied in the context of RNA evolution. The genotype-phenotype mapping is, in this case, from RNA sequence to RNA secondary structure. (e.g., [3, 11, 12, 13, 25]). The most salient features of RNA evolution are the percolating neutral paths of the frequent secondary structures, and therewith the vicinity of such secondary structures to any arbitrary RNA sequence. Thus the path towards such a structure is typically short, and the subsequent neutral evolution can travel a long way. Evolution along a neutral path is "non-neutral:" the population moves to a relatively flat part of the landscape, that is, towards regions with many neutral neighbors. This was first shown in simulation experiments [11] and has recently been proven analytically [20]. Other interesting features of the neutral paths are, on the one hand, the existence of a "shadow," that is, a set of phenotypes which occurs frequently in the vicinity of the neutral path, wherever it is in genotype space, and, on the other hand, the ongoing potential for innovation along the neutral path: the number of new not previously encountered structures close to the neutral path does not level off [3,13].

We study how these concepts apply to the experiments reported here on the evolution of morphogenesis. We will show that many interesting shapes appear in the "shadow" of the neutral path on which the cell differentiation pattern is maintained. In fact we find that a large combinatorial set of morphologies occurs close to the neutral paths. The result is a mosaic-like evolution, in the absence of recombinations, reinvention of similar morphologies far apart in evolutionary time, and in parallel with different branches of the phylogeny.

This reoccurrence of shapes within one evolutionary history, or in independent branches after the main cell differentiation pattern has been established, is in sharp

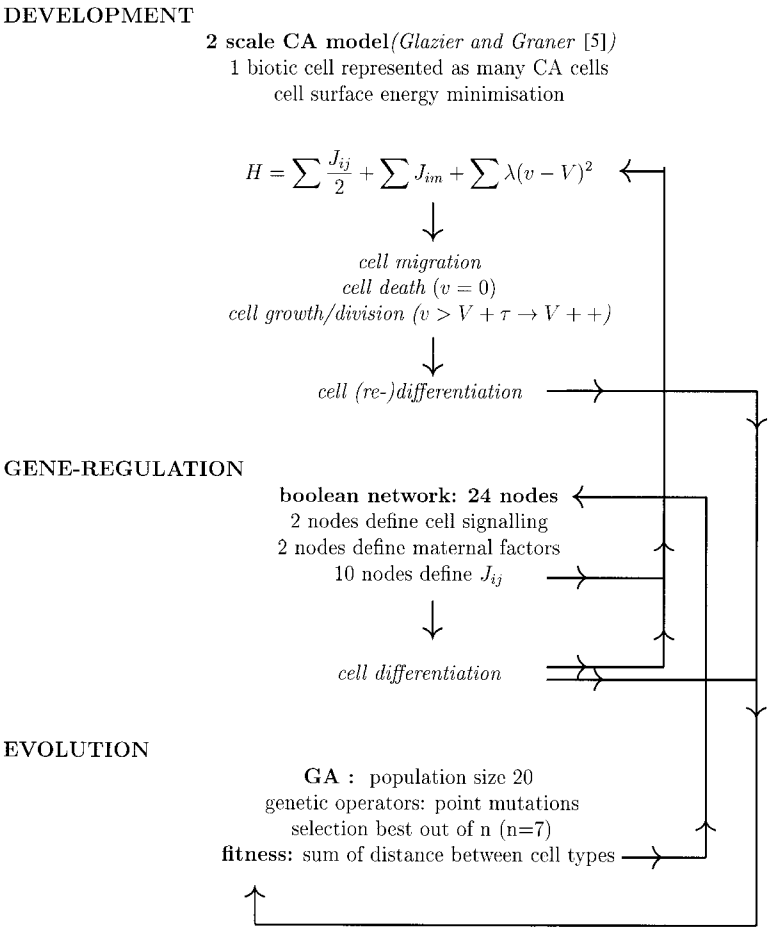

Figure 1. Schematic overview of the model. The entanglements between the within cell dynamics, morphogenesis, and evolution are shown.

contrast to the uniqueness of each evolutionary history which is started from random initial gene-regulation networks.

2 The Model

The model includes evolution, development, and gene regulation, which mutually define each other (see Figure 1). In evolutionary time gene regulation networks evolve. In developmental time cell divisions take place, and the evolved gene regulation networks plus "maternal signals" lead to differential gene expression and therewith to differential adhesion between cells.

Differential cell adhesion plays a central role: it is regulated by the gene regulation network, and regulates cell movement, cell growth/division, and cell death, and therewith influences gene expression through changing cell contacts. Thus, by surface energy minimization all these different processes are automatically coordinated. We use the model formalism proposed by Glazier and Graner [5] to model these processes and their coordination. We use a 2D implementation for convenience, but extensions to 3D are straightforward given computational resources. In this model formalism each

"biotic" cell is represented by many Cellular Automata cells; their common state is the cell identification number. The CA updating rules copy the state of a neighboring cell so as to minimize surface energy under the constraint of (approximate) volume conservation of the "biotic" cells (see Figure 1). The probability of copying is given by the Boltzmann distribution of *DeltaH* (i.e., the change in surface energy if the copying were to take place; see the equation in Figure 1.). This energy minimization process leads to cell movement, that is, cell sorting of cells with mutually different adhesion [5]. In our model we choose a relatively low degree of volume conservation ($\lambda = 5$, see Figure 1) so that, dependent on local conditions, a cell can "die" because its volume goes to zero. Interestingly, it has been shown in in vitro experiments [2,22] that squeezing of cells can indeed initiate apoptosis (programmed cell death) in biotic cells. It has also been shown in in vitro experiments that stretching of cells can lead to cell growth and cell division. A very simple extension of the model implements an analogue of the latter: When, by surface-tension-induced stretching, a cell exceeds its target volume by τ, we update the target volume by 1. In the experiments reported here $\tau = 3$. When the target volume is twice the reference target volume, the cell divides (perpendicular to its longest axis). Thus, cell growth/division is also entirely governed by differential adhesion.

The development is initiated with one cell (the "zygote"), which undergoes seven pre-scheduled cleavage divisions. The first and second cleavages are asymmetric in the sense that one "gene" is switched (on-off or off-on) one time step in one of the daughter cells. Herewith we model "maternal factors" which cause the initial differentiation in biotic early development. Note, however, that whether such a switch will lead to persistent differentiation depends on the gene regulation network. Note also that by including these initial differentiation signals we conform to biological observations, but depart from the main trend in modeling morphogenesis which emphasize symmetry breaking (e.g., Turing patterns (see [18]), and isologous differentiation [14,4].

Gene regulation is modeled by a simple synchronously updated Boolean network. The genome consists of 24 nodes and is coded as a list of the Boolean function of each node and its two inputs. The inputs are chosen from a range -24 to 24, indicating indices of nodes of the gene regulation network, whose states are the input for the node under consideration. The positive numbers indicate nodes (genes) of the cell under consideration, and the negative numbers indicate nodes from neighboring cells. (The states of neighboring cells are combined by an OR function in an environmental factor, which provides the input for the cell under consideration.) However in the experiments reported here, only two nodes provide intercellular signaling, that is, -1 and -2 connect to nodes 1 and 2 of the neighbors; the other negative numbers provide an invariant input value of "0". Thus, a variable functional connectivity is achieved. This coding, moreover, provides a basic level of redundancy at the genotype level. 10 nodes map to "surface receptors" which determine the J_{ij} (i.e., surface energy parameters; see Figure 1) which determine the intercellular adhesion and the adhesion to the substrate. Two nodes represent the above mentioned maternal factors.

The gene regulation networks are selected by the evolutionary process. Only point mutations are used: They change connections and/or Boolean functions of the gene regulation network. The fitness criterion is the number of different gene expression patterns and the amount of difference between them after a fixed developmental time (after 10,000 time steps the minimum diversity in 500 time steps is taken as the fitness). During the evolutionary run cell growth by stretching was not implemented, as it was when we studied the evolved critters in detail, and for longer developmental times later on. The selection criterion is the best out of N samples from the population ($N = 3 * NPOP$, where $NPOP$ is the fixed population size). In the experiments reported here we select for fixed point attractors in the gene regulation network or very short

(≤ 2) cycles: only those add to the fitness function. Because of computational resources (20 PCs running PVM) our population is small (20).

We would like to stress the following features of the model, which in our opinion help to capture some essential features of living systems:

- The model focuses on the entanglement between levels of organization in biological development: Instead of abstracting to a minimum process at one level, we abstract to a minimum process connecting different levels of developmental systems.
- The model focuses on the side-effects of an evolutionary optimization, instead of on evolutionary adaptation and/or neutral drift.
- The model focuses on the interplay between "self-organization" and genetically encoded information.

Moreover:

- All implemented processes represent processes which are ubiquitous in biological systems (in contrast with the otherwise fascinating studies on the evolution of morphology by Sims [26])
- There are very few parameters specifying the development: λ, the volume conservation parameter, τ, the growth threshold, and T, the temperature in the Boltzmann equation for energy minimization; all others are generated by the evolutionary process. The variation we find arises as a priori equivalent alternative evolutionary paths subjected to the same fitness criterion.
- In our model all processes interact through surface energy minimization. In biological systems cell growth and cell death can also be regulated "from within," that is, directly governed by gene expression. We exclude these from our model to be able to study the morphogenetic potential of the interplay between cell surface energy minimization and cell differentiation alone.

3 Results

About one-third of the evolutionary runs lead to extensive cell differentiation and morphogenesis. Taking one such evolutionary history as an example, we discuss first the evolutionary dynamics at several levels of description; next we discuss the "discovered" mechanisms of morphogenesis; and finally we discuss the peculiar mosaic-like, self-repeating evolutionary patterns at the morphological level which are the focus of this article.

3.1 Evolutionary Dynamics

Our model allows for four levels of description, that is, the level of the genome, the level of functional gene-regulation, the level of cell differentiation (which defines the fitness), and the level of morphogenesis sensu stricto which involves cell movement, cell growth, and cell death. The evolutionary dynamics of each of these levels has a quite distinct signature. The evolutionary dynamics of these levels will be discussed here, and are depicted in Figure 2.

The morphogenesis level is the focus of this article and will be discussed further in the next section.

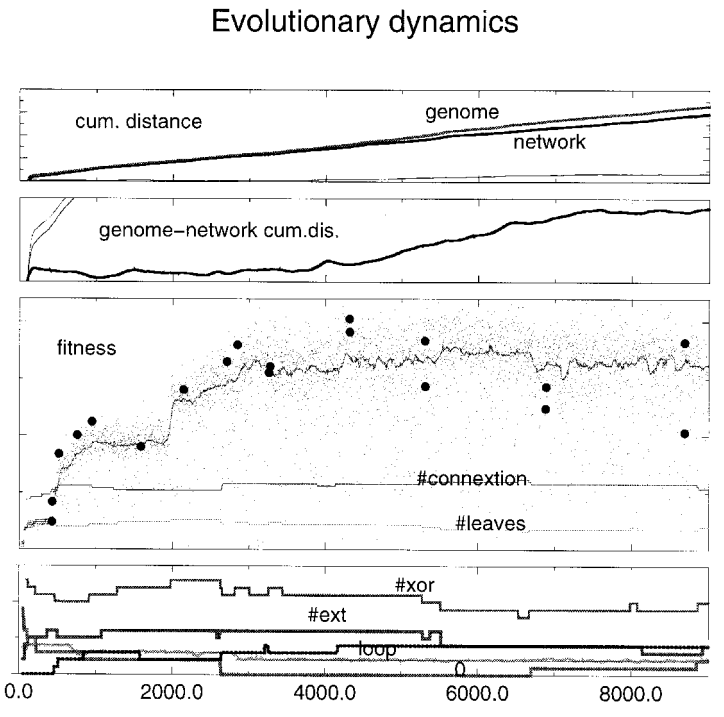

Figure 2. Evolutionary dynamics at different levels of description. The evolutionary dynamics through time. Time is measured in terms of the number of the critters generated, and thus uniquely identifies a critter. We show:
Upper panel: Length of trajectory traveled through genome space. This is calculated as the cumulative distance between the centroids of a population of 10 critters, 100 time steps apart. The upper line is the distance between genomes, that is, the number of differences in the genomes. In the lower line the distance is measured as the number of differences in the functional gene regulation network. Here the network is reduced to functional links (i.e., those which influence the gene expression). Note that one mutation can cause multiple changes in the functional network.
Second panel from above: The difference between the length of the trajectory traveled through genome space and through network space. Initially mutations cause, on average, the same amount of change in the genome and in the functional network. Once the full cell differentiation pattern is reached, the change in the functional network slows down.
Middle: Fitness over time. The dots give the fitness of each evolved critter: Single mutations cause major changes in fitness. The thick line is the running median fitness over 100 time steps. The large dots are the samples depicted in Plate 1. The lower lines depict some network properties: The upper line gives the number of network connections, the lower line the number of downstream genes, which do not regulate any other gene(s).
Bottom: Network properties (shown as running median (200 time steps)):
XOR: The number of non-forcing functions in the network, that is, the Boolean functions which depend on all inputs, regardless of the value of the inputs. For 2-connected networks these are "exclusive or" and its negation. Their number declines and therewith tends to increase the mutational stability of the network.
Loop: The number of nodes in network loops. Loops are important for stable cell-based memory. Here it is not functional and declines through drift.
#ext: The number of nodes connected to external signals. Initially this increases, and later drifts, but initial neutral changes become "functional" later on (i.e., they cause phenotypic changes only after other mutations have occurred).
0: Number of genes which are never expressed. At first this declines sharply, and becomes zero, and later it increases again.

3.1.1 Genome

Figure 2, upper panel shows that the cumulative change at the genome level is linear: The amount of change per time unit is the same early and late in evolution. In other words, it behaves as a "molecular clock," as is also the case for genomes of biotic systems. Important for this result is the redundancy at the genome level. Over and above this built-in redundancy mentioned above, there is also less trivial redundancy

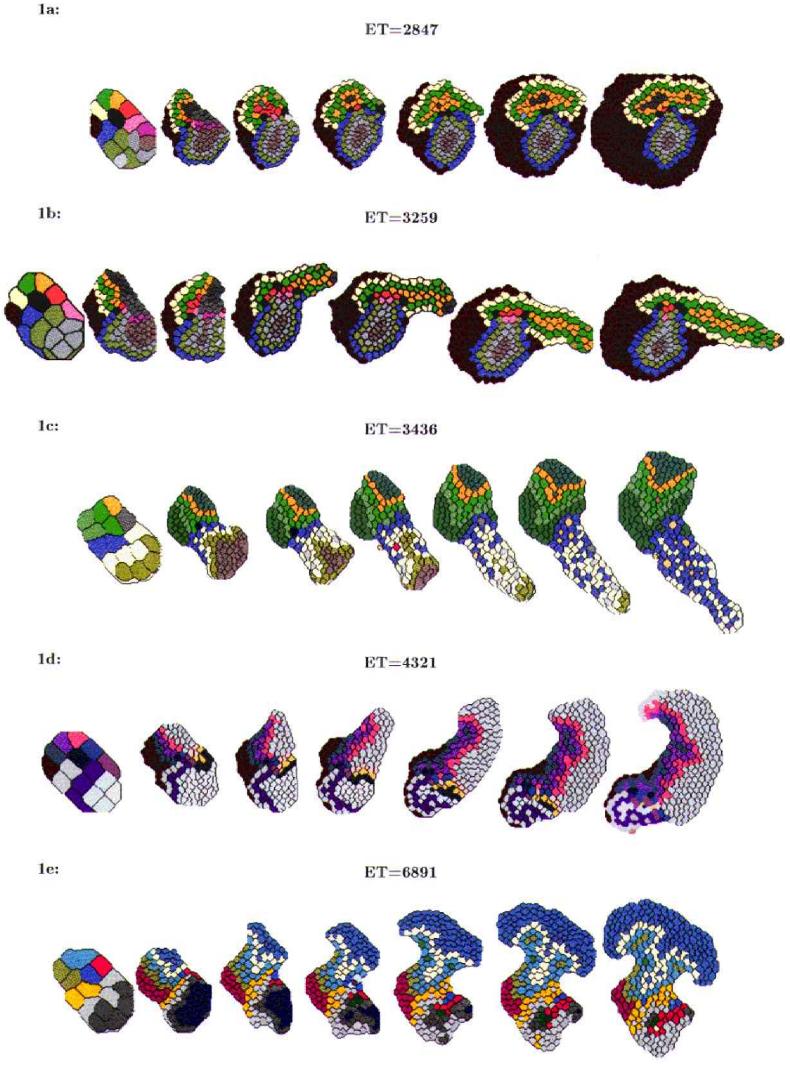

Plate 1. Developmental dynamics of morphogenesis. The depicted developmental sequences (horizontal rows) all occur in the same evolutionary history, at the evolutionary time (ET) indicated. The following stages are shown (DT is the number of developmental time steps):

1a: DT =2500,5000,12500,18750,25000,37500,5000 0;
1b: DT =2500,5000,12500,25000,37500,50000,625 00;
1c: DT =2500,5000,7500,12500,18750,25000,3750 0;
1d: DT =2500,5000,18750,37500,65000,83000,100 000;
1e: DT =2500,5000,21500,30000,48000,60000,700 00.

Note the different rate of development. Note also that especially just after the 7th cleavage (second stage shown, after 5000 developmental time steps) the similarity of the cell differentiation pattern can be seen This is reminiscent of the "zootype," the highly conserved developmental stage in animals (for a relatively recent discussion see Slack et al. (1993)). For further explanation see Section 3.2.

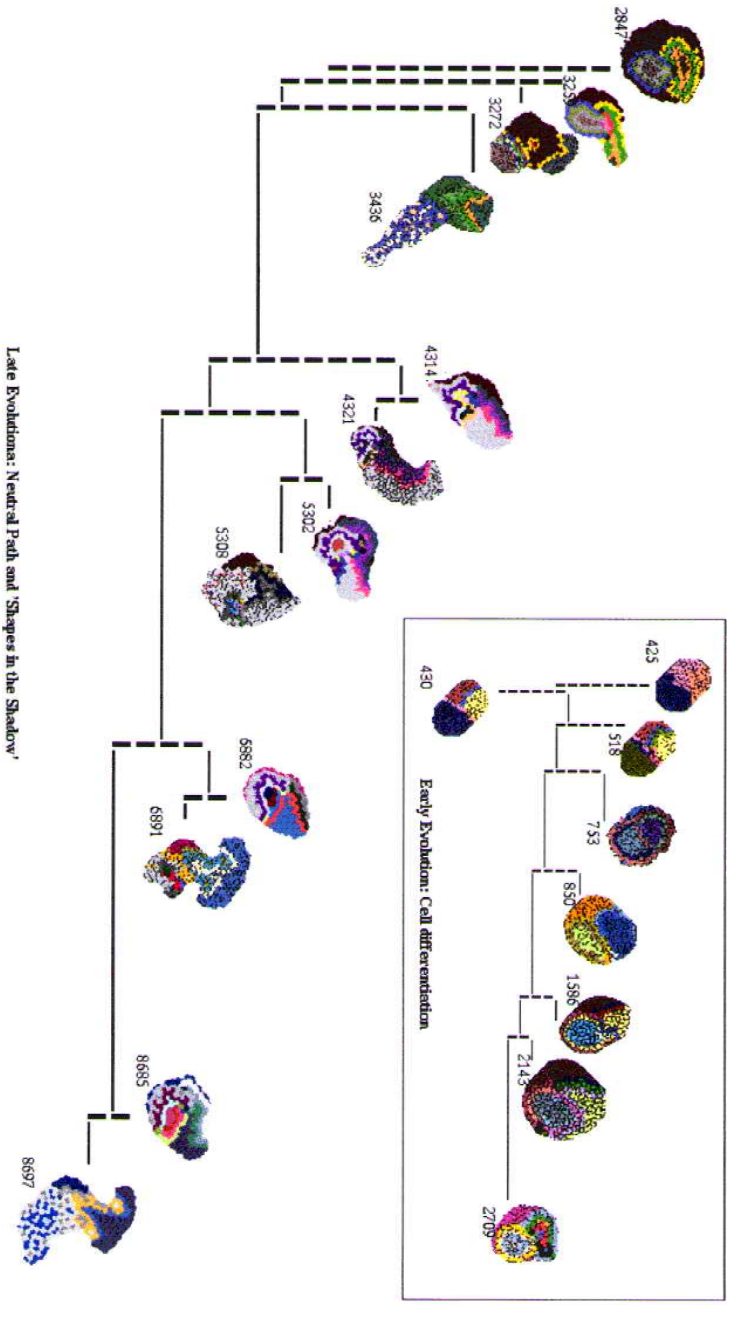

Plate 2. Molecular clock and mosaic evolution. The phylogenetic trees are calculated from the genomic distances using the Neighbor Joining method (Saitou and Nei [23]). The numbers indicate rank order in evolutionary time (ET notation omitted). The inset shows the early stages of evolution of cell differentiation. The large tree shows the evolution along the neutral path once maximum cell differentiation has been established. The length of the branches correlates with evolutionary distance: genomes diverge linearly with time (which is referred to in biological evolution as the "molecular clock"). However the similarity of the morphologies does not correspond to the position in the tree. Along the main stem overall morphology remains similar from ET4314 onwards. However the shapes that arise as close mutants combine various morphological features in various ways, and therefore appear to have a mosaic-like evolution. Also similar morphologies are reinvented at large evolutionary distances. For further explanation see Section 3.3.

because a second input to a Boolean function may have no influence on the output of that function (see below).

3.1.2 Functional Gene Regulation

We extracted a description of the functional gene regulation network by iteratively removing noninformative connections. These include the nonfunctional intercellular signals (see above), and all connections to nodes which are "forced" [15] through the other (variable or constant) input. Although the behaviors of two different functional networks can be identical when the states of nodes are fully correlated, much redundancy of the full networks is removed. The functional networks provide new observables as they have, for example, variable size and variable degree of connectivity. Note that one mutation at the genome level may cause no change in the functional network, but may also cause an avalanche of changes. The top two panels of Figure 2 show that during the initial phase of evolution the rates of change of the genome and the functional gene regulation network are the same, but once high fitnesses are obtained, the rate of change of the so defined functional networks slows down: conservation of cell differentiation reduces the number of accepted changes at this level. Note, however, that progressive change still does take place.

3.1.3 Cell Differentiation and Fitness

In contrast to the regular rate of change at the lower levels, the amount of cell differentiation, that is, the fitness criterion used for the evolution, changes stepwise in a manner that is well known for evolution with a nonlinear genotype fitness transition, for example, RNA evolution [12]. "Epochs" [19] of "stasis" are "punctuated" by sudden changes in fitness. Moreover, as has been analytically demonstrated by van Nimwegen [19] the early epochs have a short duration, whereas the later ones last ever longer.

In the middle panel of Figure 2 these properties are clearly seen in the median fitness of the population (line) or the entire fitness distribution of the population (dots). The dots show also the large variation in fitness in the population. This variation is maintained because single point mutations can partially or completely eradicate the differentiation, and therewith the fitness of the critter, and because the fitness of one and the same gene regulation network can vary due to random differences in cell movement and therewith cell differentiation.

The large dots in the middle panel of Figure 2 represent samples which are depicted in Plate 1 and Plate 2. The samples are labeled by numbers giving the rank order of their appearance in evolutionary time (ET). The inset of Plate 2 shows the successive increase in cell differentiation in early evolution. In the first three samples shown (in the inset of Plate 2), cell movement does not yet occur. At first there is only differentiation in three types due to the maternal signals. These are stably differentiated and keep their signature all through development, and in fact all through the subsequent evolution, but later on they do give rise to many different cell types through induction by neighboring cells. This later differentiation is neighborhood dependent, and so it is reversible when the neighborhood changes. A mutation in a Boolean function makes a (pre-existing) external link functional in the offspring of ET425. This produces ET430 in which the cell layer between the maternal lineages is differentiated. A second external signal causes the double layer of differentiated cells from ET518 onwards. The next "invention" is cell movement. In ET753 the upper-left maternally induced cell lineage engulfs the entire critter. In the next two stages the extent of the engulfing is modified, being slower, and involves squeezing between the lineages in ET950, whereas it does not engulf the upper-right part at all in ET1586. These changes occur on the neutral path. The next stage has jumped up to a higher fitness plateau. Curiously this major evolutionary step is not accompanied by obvious changes in the structural properties of the gene

regulation network (Figure 2, lowest panel, shows the addition of an XOR node, but this does not cause the fitness increase and disappears later). ET2143 combines an engulfing pattern similar to ET1586, with engulfing of the (non-engulfed) green cell layer; new cell types are induced when the two engulfing layers meet. Although in ET753 cell division occurs in the outer layer, and diffuse cell division occurs in ET850, this is the first stage in which cell division causes further cell differentiation. The rotational cell movement will persist all through evolution, and will give rise to quite different morphologies later on, but no major increase in cell differentiation will occur. Thus, subsequent evolutionary change is largely due to diffusion over a neutral network.

In conclusion: Gradual change at the genome level and the functional network level gives rise to a stepwise change in cell differentiation/ fitness. The fitness steps do not always co-occur with major changes in the shape of the functional networks. At the morphological level the fitness steps can be traced ex post facto, but interesting morphological "inventions" often occur without change in fitness. This is discussed in the next sections.

3.2 Morphogenesis

In a previous paper [10] we identified morphogenetic mechanisms and presented examples of each selected from different evolutionary runs. Four of these mechanisms occur in various "flavors" in the evolutionary run we describe here. We have labeled the mechanisms *Engulfing, Budding and elongation, Intercalation and elongation*, and *Meristematic growth and differentiation*. Examples of development governed by these processes are shown in Plate 1.

3.2.1 Engulfing

This is a basic cell sorting mechanism studied by Glazier and Graner [5] when they introduced the two-scale CA we use here. It occurs by differential adhesion alone, independent of dynamic cell redifferentiation. In combination with cell death and dynamic cell redifferentiation, engulfing by energy minimization is sufficient to produce gastrulation-like cell movement [10].

Engulfing plays an important role in the morphogenesis of all (fit) critters from ET753 onwards. Interestingly, the maternally differentiated upper-left cell lineage (pink in the earliest critter depicted in the Plate 2 inset) remains prone to engulf the critter all through the particular evolutionary history discussed here. It engulfs both of the other maternally derived cell layers, and induces cell differentiation in itself and the layers it engulfs. Variation in the extent and the speed of engulfing produces much morphological variation, as discussed already above. Engulfing of cell layers around other cell layers of the same maternally derived cell lineage occurs to various extents from ET2143 onwards. In this case cells can redifferentiate into each other, which gives rise to additional morphogenetic mechanisms.

3.2.2 Budding and Elongation

This mechanism is the first one invented in the run under consideration which gives rise to non-blob-like structures. The critter ET3259 (Plate 1b shows its shape and cell differentiation pattern in a number of developmental stages) produces an elongated structure by this mechanism. It differs in only two point mutations from ET2847. The crucial difference is in the dark-gray upper-right cell layer, which in the elongating critter has much smaller surface energy with the medium than the equivalent cells in the earlier critter (see the second developmental stage in Plate 1a,b). These cells form the "bud". The orange cells, on the one hand, try to engulf it, and on the other hand stick more strongly to each other than to the medium. They therefore push the bud

outwards. If they do engulf it, they themselves differentiate into dark-grey, because the orange cells are induced as such by green cells, otherwise they are "bud-cells." "New" green and orange cells are produced by growth and redifferentiation caused by the engulfing maternally differentiated upper-right cell lineages (which makes the black, the light, and the blue cells). This mechanism, that is, a "frustrated" engulfing of a small bud, plus a supply of "new" cells by growth and redifferentiation has often been observed in various forms in our evolutionary runs (see [10]).

3.2.3 Intercalation and Elongation

Another mechanism of elongation is driven by intercalation, that is, the squeezing of one cell type between the cells of a homogeneous layer of another cell type. An example is depicted in Plate 1c. It is primarily driven by the intercalation of the upper-left maternally derived cell lineage (blue) into the rear (white). Because of the decreased surface energy the blue cells experience once they are surrounded by white cells they grow, divide, and continue to invade the white area. In addition, a bud-like mechanism as described above plays a role in the elongation in the rear, and a meristematic growth, as to be described below, enlarges the front part. Note that the critters in Plate 1b and 1c are quite close in evolutionary time and differ by only two functional point mutations, one involving an extra external link. Although both elongate, they do so in different cell lineages and by a different mechanism.

3.2.4 Meristematic Growth and Differentiation

This powerful morphogenetic mechanism is most clearly exemplified in critter ET 6891 (Plate 1e) forming the upper cap, but also occurs in ET3436 and ET4321 (Plate 1c,d). It involves two or more neighborhood-dependent cell layers, which can redifferentiate into each other. The more distal layers are invaded and/or partially engulfed by the more proximal layers because this reduces surface tension. The reduction in surface tension causes growth and eventually cell division (growth is self-reinforcing because of curvature effects). The daughter cells differentiate according to the neighborhood in which they find themselves. In this way, in Plate 1e, the white cells invade the upper dark blue cells, and divide and differentiate into dark blue cells when not in contact with the brighter blue cells, while the latter invade and engulf the white cells, differentiating into white and dark blue cells as the neighborhood dictates. On both sides, initially one, later a few cells, derived from the other maternal lineage, which induces the bright blue cell type, cause the curl to form around them. Similarly, in Plate 1d the outgrowth of light gray cells is derived from the partially intercalating and differentiating red cell layers, and the outgrowth of green cells in Plate 1c by intercalating and differentiating blue cells. The redifferentiation is necessary to maintain the integrity of the layers. Similar growth zones occur in plants where cell fate is largely neighborhood-dependent. They are known as meristemes. Some stem cell layers in animals may be regulated similarly, but many of those also use other mechanisms than neighborhood-dependent determination of cell fate.

3.2.5 Long-Range Interactions

Local cell surface energy minimization by itself, and in particular in combination with cell movement, cell growth/division, and cell death, leads to long-range interactions. Therefore new morphological features arise when two or more of the above described mechanisms arise and interact. A case in point is shown in Plate 1d,e. Similar meristematic growth occurs in homologous cell layers in both cases. The critter of Plate 1e combines this with intercalation of cells of the upper-left cell lineage into the rear (bottom). This causes the different shape of the whole critter, in particular the front (top).

Such long-range interactions between local morphogenetic features cause a more than combinatorial variation in critters.

3.2.6 Conclusion

Intricate shapes can develop from single "eggs" by the combination of cell surface minimization and cell differentiation. Striking is the pseudo-isomorphic "life-like" outgrowth of the critters. (We use the term pseudo-isomorphic to stress the fact that although the shape and relative sizes change over developmental time (as they do in all biotic developing critters), the overall shape remains sufficiently similar to recognize the shapes as stages of the same developmental process). Initially shape changes arise as transient from the nonequilibrium state produced by cell differentiation. The transients are dynamically maintained by intrinsic frustration between surface energy minimization and cell growth/division and cell differentiation.

3.3 Shapes in the Shadow

Plate 2 shows the phylogenetic tree of the evolutionary history under discussion. It is constructed from a selected set of genomes using the neighbor-joining method [23] on genomic (Hamming) distance. It reflects the "true" phylogeny quite well. It is clear from the picture that it does *not* reflect the similarity of the critters at the morphological level. Along the main stem of the tree from ET4314 onwards (here represented by ET5302, ET6882, ET8685) the morphology remains fairly similar: The critter rotates by engulfing layers of cells, and there is a slow outward expansion of a homologous (initially outmost upper-right) cell layer. Thus, although the diffusion along the neutral path here changes the (functional) gene regulation network as much as it is changed along the adaptive evolutionary path leading up to this stage, similar morphologies occur all along the path. When branching off from the stem, however, quite different morphologies occur as close kin of the depicted main stem morphologies. Using the terminology of Huynen [13], these occur in the "shadow" of the neutral path. We note the following features of these "shapes in the shadow":

- Extensive cell movement and cell division occur in many of the overtly non-bloblike critters. This leads often to a net loss of cell types during development, although new cell types can also arise late in development, so that over the entire development they show the most cell differentiation. Because cellular diversity is mapped to the fitness over a small period only, however, these critters have relatively low fitness and are therefore out-competed (i.e., occur only in the shadow). Thus, the evolution of interesting shapes is not merely a side-effect of the evolution of cell differentiation, but these shapes are actually negatively selected.
- Similar shapes reoccur at quite distinct evolutionary times (e.g., ET6891 and ET8697; we found this shape 6 times in the shadow of the neutral path). Notwithstanding the genotypic change, these shapes remain within a distance of one or two point mutations from the main stem.
- Morphogenetic "inventions" occur in different combinations, and thus give rise to a combinatorial set of shapes (of which only a few are shown). For example, extensive intercalation and growth of the upper-left cell lineage into the rear occurs in ET3436 and ET8697. This does not occur in intermediate critters, and in ET8697 it is combined with a "neck" and strong meristematic growth like in ET6891, which do not occur in ET3436.

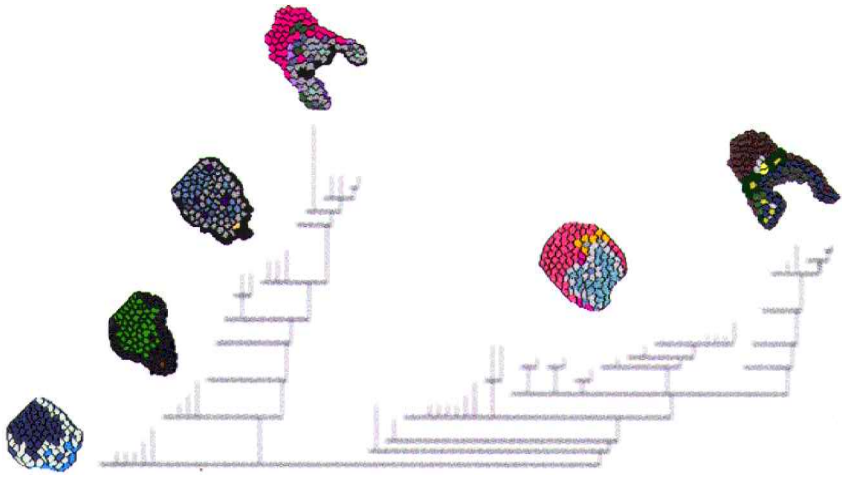

Plate 3. "Shapes in the shadow" reoccur "when the tape is played twice." At the stage of the leftmost picture the evolutionary run was restarted. Genotypically the 2 runs diverged, giving rise to the 2 separate branches of the reconstructed phylogenetic tree (using the Neighbor Joining method (Saitou and Nei [23])). The morphology on the main stem also diverges: in the left branch the bottom engulfs the top, whereas in the right branch the top engulfs the bottom. Nevertheless, a very characteristic 2-armed shape arises in the "shadow" of the neutral path in both evolutionary histories. For further explanation see Section 3.3.

- Thus, evolution appears to be mosaic-like on the morphological level, although recombination is not implemented.
- Notwithstanding the diversity of shapes which appear in the shadow of one evolutionary run, they do share a basic differentiation pattern, and are thus recognizable as family members. Other evolutionary histories produce quite different critters.
- Thus, constraints on evolution occur not only for the selected critters, but also for their mutational shadow.

The latter points are illustrated in Plate 3 which shows a phylogenetic tree of some other evolutionary run. Here we tested "what would be conserved when the tape was played twice" by reinitiating the run at a point after the main cell differentiation was established (leftmost critter). The two branches of the reconstructed phylogenetic tree indeed represent the two runs, which diverge genotypically. Surprisingly a quite distinct morphology with two "arms" appears in the shadow of both branches amid the rather nondistinct blobs of the main stem. They are formed by an "intercalate and stretch" mechanism in which the length of the interface between two cell types is maximized by intercalation of outer cell layers which differentiate into the boundary cell types; engulfing and redifferentiation also lead to growth of the tips of the branches. In both runs this two-armed morphology appears repeatedly in the shadow, while the main stem critters of both branches are both uninteresting, but differ profoundly with respect to the engulfing cell layer.

In conclusion: The outcome of evolution is very sensitive to the initial evolutionary stages. Every run we have studied has its unique signature. This signature includes not only the fittest critters on the highest neutral path but also the ones in the "shadow" of the neutral path. These "shapes in the shadow" combine a number of morphological features which are repeatedly "re-invented" in a combinatorial way. This leads, in combination with long-range interactions, to much morphological variation, mosaic-like

evolution, and "determinism" in evolution, that is, shapes reoccurring in independent branches of the evolutionary tree.

4 Discussion and Conclusions

In this paper we have demonstrated the morphogenetic potential of the interplay between differential adhesion and dynamical redifferentiation of cells. We did this by evolving shapes as a side-effect of maximizing cell differentiation. We found a mosaic-like evolution at the morphological level. Here we discuss our results in terms of long-term information integration in evolution and in terms of the feasibility to study "generic, nongeneric phenomena" (as life appears to be), that is, phenomena which are nongeneric in the sense that they are rare for arbitrary initial conditions, but generic in the sense that they do reoccur in many different settings.

4.1 Information Integration in Evolution

Most research on biological evolution is about either immediate fitness benefits or neutral drift, or the combination of both. Long-term effects have been largely ignored. For example, Maynard Smith and Szathmáry [17] reconstruct major transitions in evolution that demand immediate benefits for all intermediate steps. Although this is indeed a necessary and sound research methodology for such a reconstruction attempt, long-term information integration under a Darwinian mutation selection regime has been demonstrated to occur in models, and might operate in biotic systems as well.

As mentioned above, long-term evolution in a (relatively) constant environment leads to a decrease in phenotypic variation due to less phenotypic effect of mutations or to direct "physiological" adaptation to (relatively slight) environmental variations. This has been observed in biotic systems [24], studied from a population genetic point of view as "canalization" [28], and in genetic programming [1] and is a property of (e.g., RNA) evolution on a neutral network [11,20]. Environmental variation on the time scale of many generations (which may occur by spatial pattern formation and/or coevolution), can be dealt with by evolution in two different ways: either a genetic predisposition to track the changing environment (leading to red-queen-like coevolution), or information integration in individual genomes of adaptive demands of which they experience only a very sparse subset during their lifetime. The latter is a powerful scheme to obtain general-purpose "solutions," while evaluating only a sparse subset of "problems" [6]. Pagie and Hogeweg [21] have shown that these general-purpose solutions do bear a signature of the former "strategy" as well: The phenotype is very sensitive to mutations.

In the present article we have studied the evolution of morphogenesis by an interplay of differential adhesion and gene regulation. We have demonstrated that interesting shapes appeared relatively late in evolution as a side-effect of cell differentiation. The interesting shapes appear mostly in the "shadow" of the neutral path along which the population diffuses once it has obtained a high degree of cell differentiation, that is, they are less-fit mutants of the fittest individuals (as mentioned, our fitness criterion unintentionally disfavored cell movement and therewith morphogenesis). The evolutionary dynamics of these "shapes in the shadow" bears a distinct signature: it is mosaic-like because morphological features are repeatedly reinvented and combined in different ways, and it is "deterministic" in that similar shapes arise in independent parallel branches of the phylogeny.

Rapid and repeated morphological adaptations are described for several biological systems. For example, *Anolis* lizards on the Caribbean islands show a similar niche differentiation and accompanying morphological differentiation on different islands. Molecular studies have shown that the differentiation evolved anew on each of the islands, starting from a single type that happened to colonize that island [16]. Another

example is the *Cichlid* fishes in the African great lakes. Similar species flocks diverged in different lakes, showing similar morphological adaptations to similar niches. In both cases classical taxonomy mistakenly grouped the similar adaptation patterns from different locations into taxa, but molecular data show that phylogeneticly they should be grouped in location-specific morphologically diverse groups.

Both of these examples concern species which have repeatedly invaded "new" environments in which they could occupy a range of habitats. Our experiments show that a set of potential morphemes can reside close to a specific cell differentiation pattern, and therefore can rapidly evolve whenever "needed." Like the biological examples just mentioned, they combine "uniqueness" (*Anolis*, *Cichlids*, a particular evolutionary run) with "determinism" (repeated evolution of similar adaptations). Previous studies which recognized the existence of a mutational "shadow" along a neutral path [13,20] considered such a shadow as a constraint on "true" innovation, that is, entirely novel shapes encountered in the vicinity of the neutral path. Indeed in the RNA evolution studied by them the shadow is occupied by rather uninteresting variations of the secondary structure selected. In contrast, our experiments, together with the biological examples mentioned, suggest a mode of long-term information integration in evolution which involves a mutational shadow of "useful" shapes, instead of simply sensitivity to mutations as previously demonstrated in models. To demonstrate this, artificial worlds allowing repeated invasions of "virgin" environments with similar niches with complex demands on fitness should be studied.

4.2 Generic-Nongeneric Phenomena and a Theory of (Artificial) Life

Morphogenesis in biotic systems is governed by an interplay of many processes. In this article we studied the morphogenetic potential of the interplay between some of these, that is, differential adhesion and gene regulation. We observed a rich variety of morphogenetic behavior. The observed morphogenetic mechanisms are generic in the sense that they occurred in many of the evolutionary runs, although they led to quite diverse morphologies due to different cell differentiation patterns. Moreover they appeared without being explicitly selected, that is, they appeared as a side-effect of evolution for cell differentiation. Similar mechanisms are observed in biotic systems as well. Notwithstanding the occurrence of these morphogenetic mechanisms in different settings, they are very rare in the space of gene regulation networks, that is, they are clearly nongeneric in that respect. We have demonstrated that we can discover and study such phenomena in evolving systems subject to a fitness criterion which can be seen as a prerequisite for the behavior of interest, but is not sufficient for it. In our opinion we have thereby demonstrated the potential feasibility of a theory of artificial life.

Acknowledgments

This work was initiated in collaboration with Roeland Merks. I am much indebted to him for the great amount of work he did in setting up and programming the combined developmental and evolutionary model. I also thank Nick Savill and Stan Marée for insightful discussions. They used the developmental model in combination with long-range signalling for modeling "real" morphogenesis of *Dictyostelium*. Finally I thank Ben Hesper for his long-term inspiration and support.

References

1. Altenberg, L. (1994). The evolution of evolvability in genetic programming. In J. K. E. Kinnear (Ed.), *Advances in genetic programming* (pp. 47–74). Cambridge, MA: MIT Press.
2. Chen, C. S., Mrksich, M., Huang, S., Whitesides, G. M., & Ingber, D. E. (1997). Geometric control of cell life and death. *Science, 276*, 1425–1428.

3. Fontana, W., & Schuster, P. (1998). Continuity in evolution: On the nature of transitions. *Science, 280*, 1451–1455.

4. Furusawa, C., & Kaneko, K. (1998). Emergence of rules in cell society: Differentiation, hierarchy, and stability. *Bulletin of Mathematical Biology, 60*(4), 659–687.

5. Glazier, J.A., & Graner, F. (1993). Simulation of the differential driven rearrangement of biological cells. *Physics Review E, 47*, 2128–2154.

6. Hillis, D. (1992). Coevolving parasites improve simulated evolution as an optimization process. In C. G. Langton et al. (Eds.), *Artificial Life II* (pp. 313–324) Redwood City, CA: Addison-Wesley.

7. Hogeweg, P. (1989). MIRROR beyond MIRROR, puddles of life. In C. G. Langton et al. (Eds.), *Artificial Life* (pp. 297–315) Redwood City, CA: Addison-Wesley.

8. Hogeweg, P., & Hesper, B. (1991). Volution as pattern processing: TODO as substrate for evolution. In J. A. Meyer & S. W. Wilson (Eds.), *From Animals to Animats* (pp. 492–497). Cambridge, MA: MIT Press/Bradford Books.

9. Hogeweg, P. (1998). On searching generic properties of non-generic phenomena: An approach to bioinformatic theory formation. In C. Adami, R. K. Belew, H. Kitano, and C. E. Taylor (Eds.), *Artificial Life VI* (pp. 285–294), Cambridge, MA: MIT Press.

10. Hogeweg, P. (2000). Evolving mechanisms of morphogenesis: On the interplay between surface energy minimization and cell differentiation. *Journal of Theoretical Biology (in press).*

11. Huynen, M. A., & Hogeweg, P. (1994). Pattern generation in molecular evolution: Exploitation of the variation in RNA landscapes. *Journal of Molecular Evolution, 39*, 71–79.

12. Huynen, M. A., Stadler, P. F., & Fontana, W. (1996). Smoothness within ruggedness: The role of neutrality in adaptation. *Proceedings of the National Academy of Sciences USA, 93*, 397–401.

13. Huynen, M. A. (1996). Exploring phenotype space through neutral evolution. *Journal of Molecular Evolution, 43*, 165–169.

14. Kaneko, K., & Yomo, T. (1997). Isologous diversification: A theory of cell differentiation. *Bulletin of Mathematical Biology, 59*, 139–196.

15. Kauffman, S.A. (1969). Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of Theoretical Biology, 22*, 437–467.

16. Losos, J. B., Jackman, T. R., Larson, A., de Queiroz, K., Rodriguez-Schettino, L. (1998). Contingency and determinism in replicated adaptive radiations of island lizards. *Science, 279*, 2115–2118.

17. Maynard Smith, J., & Szathmáry, E. (1995). *The major transitions in evolution*. Oxford, U.K.: Freeman.

18. Murray, J. D. (1989). *Mathematical biology*. Berlin: Springer-Verlag.

19. van Nimwegen, E., Crutchfield, J. P., & Mitchell, M. (1997). Finite populations induce metastability in evolutionary search. *Physics Letters A, 229*, 144–150.

20. van Nimwegen, E., Crutchfield, J. P., & Huynen, M. A. (1999). Neutral evolution of mutational robustness. *Proceedings of the National Academy of Science USA, 96*, 9716–9720.

21. Pagie, L., & Hogeweg, P. (1998). Evolving adaptability due to coevolving targets. *Evolutionary Computation, 5*, 401–418.

22. Ruoslahti, E. (1997). Stretching is good for a cell. *Science, 276*, 1345–1346.

23. Saitou, N., & Nei, M. (1987). The neighbour-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology of Evolution, 4*, 406–424.

24. Scharloo, W. (1991). Canalization: genetic and developmental aspects. *Annual Review of Ecology and Systematics, 22*, 65–93.

25. Schuster, P. K., Stadler, P. F., & Hofacker, I. L. (1994). From sequences to shapes and back: A case study in RNA secondary structure. *Proceedings of the Royal Society of London B, 255*, 279–235.

26. Sims, K. (1994). Evolving 3D morphology and behavior by competition. *Artificial Life, 1*, 353–372.

27. Slack, J. M., Holland, P. W., & Graham, C. F. (1993). The zootype and the phylotypic stage. *Nature, 361*, 490–492.

28. Wagner, G. P., Both, G., & Bagheri-Chaichian, H. (1996). A population genetic theory of canalization. *Evolution, 51*, 329–347.