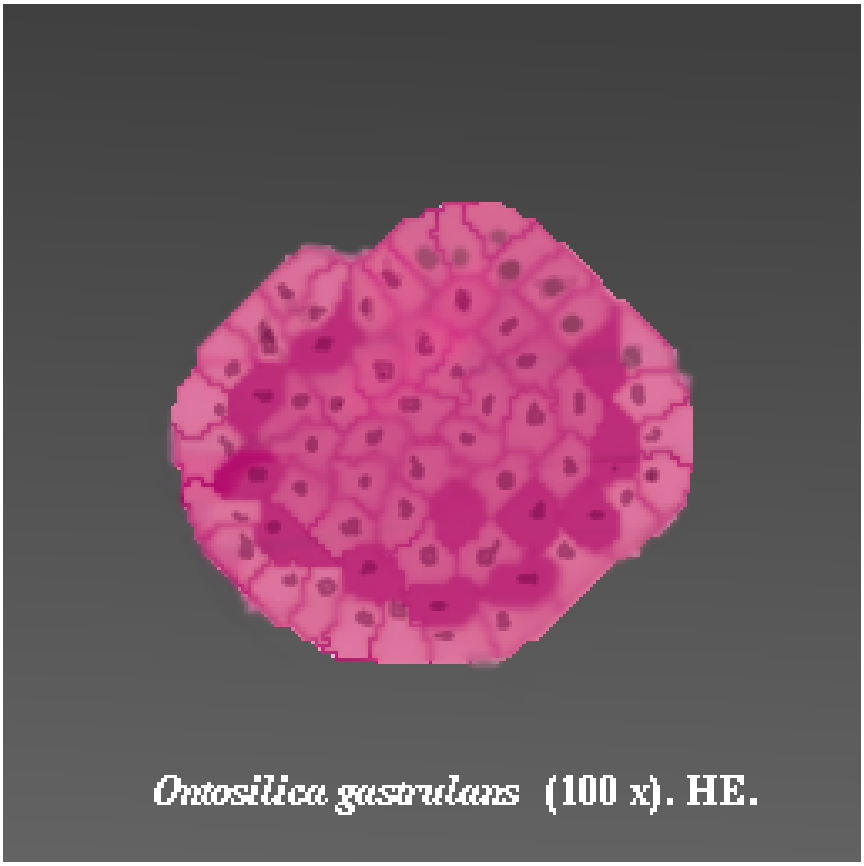
Evolving “Metazoan” Development

Roeland Merks
Group of Theoretical Biology and Bioinformatics
Department of Biology
Utrecht University
The Netherlands



September 1997

Front page illustration: Manipulated image of a “metazoan” evolved in a preliminary version of the system described in this report. Nuclei have been drawn and the colours have been adjusted to get a more lifelike appearance. The text “(100 x). HE.” is fictitious. It was added to the caption in order to improve the illusion of a hematoxilin-eosin coloured microscopic coupe of a real animal. The “metazoan” was called *Ontosilica gastrulans* since its development, during which three tissue layers are formed, resembles a gastrulation process.

Abstract

A paradigm system for the evolution of multicellular animals is constructed. In many evolutionary models, the non-linearity of the genome-phenome mapping is ignored. However, the results of evolutionary paradigm systems that did include a non-trivial, complex genome-phenome mapping have suggested a framework joining seemingly conflicting evolutionary "points of view" like neutral evolution, Punctuated evolution and "gradualism".

The embryonal development of multicellular animals adds many new levels of complexity to the genome-phenome mapping. (1) Genes interact resulting in a differentiated pattern of gene expression: a cell type. (2) Cells interact, generating cellular diversity and pattern. (3) Cells sort out to form tissues and organs. (4) Tissues and organs interact.

The first three levels of complexity were included in the paradigm system. Artificial evolution using a trivial fitness criterion resulted in a metastable sequence of epoches, each characterised by a predominant type of development. The evolved "metazoans" successively include in their developmental programs cell polarity, cell-cell communication, cell movement, positional information and genetic redundancy.

Prelimary results suggest that the evolved genomes are hierarchically structured. "Regulatory genes", being highly sensitive to mutation, control the expression of mutationally less sensitive "downstream" genes. It is hypothesised that evolutionary innovations result from mutations in the regulatory genes, whereas small scale changes within epochs result from mutations in the downstream genes.

Contents

1 Introduction **4**
1.1 Non-linear mapping from genotype to phenotype 4
1.2 Generation and maintenance of cellular diversity 6
1.3 Differential adhesion driven morphogenesis 7
1.4 Putting the lines of reasoning together 7

2 The Model **9**
2.1 A general overview of “metazoan” development 9
2.2 Intracellular dynamics 10
2.3 Cell cell communication 11
2.4 Differential adhesion driven cell movement 12
2.5 Cell division and maternal “genes” 15
2.6 Cell death: “Lonelyness penalty” 15
2.7 A detailed overview of the implementation of a *metazoan*’s development 16

3 The ontogeny of a simple organism **17**
3.1 Analysis of the network: housekeeping and dynamic *gene*s 17
3.2 Early pattern formation 20
3.3 Morphogenetic cell movements 20
3.4 Final pattern formation 23

4 Evolution of the metazoans **24**
4.1 The genetic algorithm 24
4.2 Fitness criterion 25
4.3 Computation of cell type diversity 26

5 An evolutionary run **27**
5.1 Zygotic cell polarity induces two cell lines 37
5.2 Discovery of cell cell communication 37
5.3 Fixing obtained information 39
5.4 The evolution of positional information 40
5.5 “Epiboly”: NB cells engulf B cells and induce a new cell type 41
5.6 More “tickling”: pattern destabilisation 42

6 Discussion and conclusions **50**
6.1 Summary of the results 50
6.2 Discussion and speculation on the results 51
6.3 Mutational robustness 52
6.4 Caveats of the fitness criterion: 53

7 Future work **54**
7.1 Buggy “atol” suggests to include chemotaxis 54

Chapter 1

Introduction

Three lines of reasoning led to the research described in this paper.

1.1 Non-linear mapping from genotype to phenotype

For many years the study of evolution has been shown to be fruitful without taking into account that there is no such thing as direct mapping from the coding of an organism to its fitness evaluation. In such studies on evolutionary dynamics a change in the genotype of an organism results in an equivalent change in its phenotype.

The last decennia however, a number of paradigm systems have been developed that do include a non-linear genotype to phenotype transition. These systems include NK-landscapes [29] and models on RNA-evolution [10, 23, 22], but also genetic algorithms (GA) [20] and genetic programming (GP) [31]. The study of these paradigm systems has, or should have, profoundly reformed thinking on evolutionary change.

Although genetic algorithms were designed not primarily for the study of evolutionary dynamics, they offer important insight into the behaviour of evolutionary systems having a non-linear genotype-phenotype mapping. In genetic algorithms, a solution to a predefined computational problem is "evolved" by selecting possible solutions from a population. The solutions that are more able to cope with the problem than their brothers and sisters reproduce and form a new population. During this reproduction small changes are made to the solutions by means of genetic operators such as point mutations and cross-overs.

In many of these genetic algorithms the problems comprise the setting of parameters in a predefined system. In some genetic algorithms however the solution to a predefined computational problem is coded in a representation that is non-linearly related to the actual solution. For example, the parameter setting of a system could be encoded in a bitstring. If each bit in this would have an equal chance of being mutated a change from, say, 127 to 255 would be as probable as a change from 254 to 255. The performance of the solution to the problem that this bitstring represents can be seen as its phenotype. This phenotype is now non-linearly related to the coding, or *genotype* of the problem.

Typically, genetic algorithms show metastable behaviour. The fitness of an evolving popula-

tion remains rather stable for a while, then increases rapidly as if a new discovery has been made, followed by another period of evolutionary stasis. Mathematical analysis of a simple genetic algorithm suggests that this behaviour might be very common genetic algorithms [41]. The genetic algorithm that was analysed (the "Royal Road" GA) did not have a complex genotype-phenotype mapping. Still, the genotype-*fitness* mapping was not linear in the sense that a mutation does not directly result in an equivalent fitness-change. Unlike fitness changes in classical evolutionary analysis (see f.i. [33]), in the "Royal Road" GA that was analysed several mutations need to be "collected" for a jump in fitness. This study suggests that even a slightly non-linear genotype-phenotype mapping can result in a "punctualist" mode of evolutionary change.

Studies on RNA-evolution have deepened the understanding of how genotypes may map to phenotypes, and how the structure of a phenotype landscape resulting from a given genotype-phenotype mapping reflects the evolutionary dynamics. These studies have led to a framework connecting the seemingly conflicting types of evolutionary dynamics neutral evolution, neutral evolution and "gradualism" using a "selectionist" point of view.

Using energy minimisation algorithms [19, 13] the secondary structure of RNA-strings is now reasonably well predicted. In this way the concept of "phenotype-space" was constructed, describing the phenotype (in the case of RNA strings the secondary structure) for every possible genotype (in the RNA case the sequence of an RNA string). It appears that a given secondary structure (phenotype) of an RNA-string is not at all linked to a certain region in sequence space (genotype-space). Sequences folding into a secondary structure come from all over sequence space. The same is true for sequences. A cloud of sequences in a small region of sequence space folds into secondary structures from all over phenotype space.

Evolution to a target in phenotype space (f.i. a double loop) again resulted in periods of stasis intermitted by sudden fitness jumps. For example, it has been demonstrated that RNA-strings can evolve "neutrally", i.e. without a gain in fitness, to a "smooth" part of a phenotype-landscape if stability is a beneficial property, such as in evolution towards a predefined goal. A "neutral walk" through sequence space brings the population to a region where it is less sensitive to mutation [23]. In contrast, if RNA strings are selected to change very fast, such as in a predator-prey setting, RNA strings evolve to "rugged" parts of the phenotype landscape [22].

The question now is, whether the ideas of evolutionary change emerging from the research on RNA-evolution can be projected onto the evolution of multicellular animals. Some data suggests that metazoans evolve "metastabily". Paleontological studies at least do claim to observe metastability in the fossil record (discussed in [18]).

The main reason to believe that the concepts developed in the context of RNA evolution can be used to mould our thoughts on the evolution of multicellularity is the exorbitantly non-linear genotype-phenotype mapping of multicellular organisms in the sense of the embryonic development. The development of multicellular organisms is non-linear at many structural levels. The mapping of the information contained in a gene to the tertiary structure of its gene-product is highly

non-linear. Also, there is no linear relation between the genetic information and the set of different cell types that can be attained by the cell containing this information. This problem, the dynamics of genetic regulatory networks, has been studied in great detail in the modelling framework of the Boolean network [29].

A third level of complexity of multicellular organisms is the property of the cells to differentiate in reaction to an external signal, in general imposed by the other cells. This problem is not as simple as it may seem. The cells of a multicellular organism all contain the same hereditary information stored in the DNA, where these cells should differentiate into a large set of quite distinct phenotypes. The evolution of the ability of genetic regulatory networks to react on external signals and to be robust to environmental noise has been studied in the framework of continuous, Hopfield like, regulatory networks by [5, 4].

A fourth level of complexity, adding to the non-linearity of the genotype-phenotype mapping, is the level of the organism. The information stored in a genome should not only be able to generate a set of distinct cell types, each having a different "function". These cell types should also be patterned in a particular way, such that the different cell types act as a coherent whole.

In order to explore the effect of the non-linear genotype-phenotype mapping in multicellular organisms we included the second (genome to protein sets), the third (cellular interactions) and part of the fourth level (pattern formation) of complexity in the model described in this report.

Using this system it may be possible to make a start in understanding the structure of the phenotype landscape of multicellular animals.

1.2 Generation and maintenance of cellular diversity

Another problem that is addressed here, is the generation and maintenance of cellular diversity. One of the central problems in developmental biology is, how it is possible that a single zygote gives rise to cell types as different as red blood cells and brain cells. The question is not only how it is possible that cells having identical genetic information are able to differentiate, but also how such differences are generated and patterned.

Prior to and during the evolution of multicellularity, mechanisms must have evolved that initiate, amplify and stabilise differences between cells during ontogeny. Cellular diversity can be generated with roughly two distinct mechanisms. A so-called "pre-pattern", laid down by maternal determinants, may be present in the zygote. Conversily, dynamical processes may generate a pattern during the development.

This second problem has been addressed by several lines of research. One of the approaches has been inspired by the phenomenon of isologous diversification and dynamic clustering in continuously stirred bacterial cultures [30]. In this approach, cell differentiation patterns arise due to the dynamical interaction of — initially identical — cells.

In this model, no maternal "pre-pattern" is applied. Still, several clusters of bacterial cells differentiate. Abstract model studies [27] have suggested that this type of behaviour can be understood in the following way. Consider a system consisting of chemical networks globally coupled via an external medium. The number of chemical networks, and in this way the dimension and the degree of freedom of the whole system, increases as a result of "division" of the networks. The dynamics of the networks is chaotic, i.e. the networks are very sensitive to initial conditions. Hence, tiny differences between two networks grow in time. At the same time, however, the global coupling synchronises the networks' dynamics.

So, high dimensional chaos on the one hand, and cell cell communication on the other hand is able to generate a clustered diversification of cellular phenotypes.

Another approach to the problem of cellular diversification was initiated by studies on arrays — growing in size by "division" — of locallly coupled Boolean networks [24, 7, 6]. In the model of Jackson et al. random one-dimensional arrays of Boolean networks were studied in which a fixed number of genes communicate with genes of neighbouring cells. Two important results emerged from this research. Cell diversification is maximised if approximately 20 % of the genes communicate with neighbouring cells. Secondly, solely as a result of cell-cell interactions, two simple patterns were generated in a simple model organism. One of these patterns was repetitive, a gene was turned on in every third cell. Another pattern consisted of a block of genes expressed in two regions of the cell, whereas these genes were unexpressed in the rest of the organism.

The evolvability of systems of locally coupled Boolean networks was first explored in work of [7]. In their model, "creatures" consisting of two-dimensional arrays of Boolean networks, were evolved using genetic algorithms according to fitness function maximising the number of cell types.

1.3 Differential adhesion driven morphogenesis

Evidence has been found that differential cell adhesion is important in morphogenetic processes. Dissociated animal tissues and organs have been reported to sort out into anatomically correct structures in a number of experimental systems (reviewed in: [39, 2], early amphibian embryos: [40]; whole sea urchin embryos: [16]; amphibian limbs: [35]; chicken retinas: [2]).

This cell sorting behaviour has been interpreted to be caused by differential cell adhesion. For example, ectodermal cells are thought to adhere more strongly to ectodermal cells than to endodermal cells [40].

Experimental results suggest that intercellular affinities may change during a morphogenetic process. It has been demonstrated that at the onset of sea urchin gastrulation, the ingressing mesenchyme cells decrease the affinity to their neighbours and to the extracellular matrix [12, 34].These experimental results indicate that differential adhesion may be an important driving force behind morphogenetic processes. Differential adhesion has even been called in a textbook "the dominant paradigm of morphogenesis" [15].

A recently developed cellular automata based algorithm has resulted in biologically defensible models of differential adhestion driven cell sorting [17, 36]. Using this algorithm, the aggregation and morphogenesis of *Dictyostelium discoideum* up to the crawling slug stage was modelled [37].

1.4 Putting the lines of reasoning together

The three lines of reasoning built up above have resulted in the following model system of "metazoan" evolution. In the system articifial "metazoans" are selected for the ability to develop and to maintain cellular diversity. The "metazoans" are allowed to make use of differential adhesion

driven cell rearrangement and contact signalling in the development and maintenance of cellular diversity. It is examined how the complex genotype-phenotype mapping of these “metazoans”, being a model of the complex genotype-phenotype mapping of multicellular animals, influences evolutionary dynamics. The mechanisms the “metazoans” use to build up cellular diversity are studied.

In this report the following style convention is used. In the model several entities have names like “gene” or “metazoan”. In order to indicate that these names refer to model entities rather than to biological concepts, they are printed in *slanted* style.

Chapter 2

The Model

Basically, the paradigm system of the evolution of multicellular animals, as well as the evolution of "real" *ex silico* metazoans, consists of two parts: development and selection.

Genomes are "executed" during development, resulting in a phenotype. This phenotype is "evaluated" in its performance against the living and the non-living environment[1].

The developmental part of the paradigm system is described in the present chapter. Chapter 3.2 describes the development of a simple organism. In chapter 4 the evolutionary part of the system is described and discussed. Additionally in chapter 5 the results of an evolutionary run are discussed.

2.1 A general overview of "metazoan" development

During the development of a multicellular animal, the zygote gives rise to hundreds and often milliards of cells building the adult body. During this process, cells differentiate into different cell types, they interact, and they sort out to form tissues and organs. These processes are interdependent. Cell sorting changes the interaction structure between cells. Different interactions result for a cell in different signals. Different input signals may result in a new differentiated state. This change in gene expression can result in different affinities for other cells, which can result in a different cell sorting.

The number of cells increases over development by division. Cells slowly die if they get isolated from the other cells. The processes taking part in the development of a "metazoan" are described in the following sections.

The cells' intracellular dynamics is described and discussed in subsection 2.2. The interaction between cells is described and discussed in subsection 2.3. Cell sorting is described in section 2.4. A discussion on cell division and inequal cell divisions is given in section 2.5. Finally, a quick overview of a "metazoan's" development is given in section 2.7.

[1] For the moment it is ignored that embryos face selectional forces as well.

2.2 Intracellular dynamics

The first part of our model consists of the genetic network of the cells. Gene regulation takes place at several levels of transcription. On the level of the DNA, transcription factors bind to promoters and enhancers, in this way initiating the transcription of a gene. Additionally a gene product may need to modified after transcription, before it is active [1].

We chose to simplify the genetic regulatory network of our cells, because we want to focus not on the particular properties of transcriptional regulation, but moreover on interactions between cells. For this simplification the formalism of the Boolean network was used. In Boolean networks, the activity of a *gene*s is considered to be binary. They can be either active or inactive. The transcriptional activity of a *gene* is regulated by a fixed number of other *gene*s. The decision of a *gene* to be either active or inactive upon a certain combination of inputs from other *genes*, is made by a Boolean function.

Several molecular biological examples suggest that it is defensible to model gene regulation with Boolean functions. A well known example is the activation of the *lactose*-operon in *E. coli* [25, 26], whose regulation reminds of a *not (exclusive repressor)* Boolean function. The genes regulated by the *lac*-operon are only transcribed if the bacterium meets lactose. If no lactose is present, a repressor protein prohibits binding of the RNA-polymerase complex to the operon. However, if a lactose molecule binds to the repressor, the repressor is released from the operon and transcription is initiated.

We have chosen to fix the number of inputs per *gene* in the Boolean network on two. In other words, each *gene* in the Boolean network is regulated by exactly two other *genes* or *receptors*.

The properties of Boolean networks have been extensively studied by Kauffman [28, 29]. These studies show that the dynamics of Boolean networks, especially those with two inputs per *gene* has a number of properties reminding of differentiating cells. First, in a K=2 Boolean network, differentiation is persistent. A stable state or state cycle is stable to 80 to 90 percent of small perturbations like transiently flipping a bit. Second, induction can push a differentiated Boolean network to another attractor. In the remaining 10 to 20 percent, the network falls into another attractor after a small perturbation. These properties make them well suited for our aim: studying the behaviour of interacting cells.

Kauffman [28, 29] suggested that a state cycle or a stable state in a regulatory network can be interpreted as a cell type. In the model described in this report this suggestion is followed. During the updates of the boolean networks the previous states are scanned in order to detect state cycles and stable states. Whenever a state cycle or a stable state is detected, a cell is said to be *differentiated* and a colour is assigned to it. This colour depends on the state cycle via a hash function. In this way a particular state cycle is always assigned to the same colour in all the "metazoans".

In total, there are 16 Boolean functions having two inputs. They are listed in table 2.1, together with their names used throughout this paper.

Name	outputs			
	A=1,B=1	A=0, B=1	A=1, B=0	A=0, B=0
ALL0	0	0	0	0
0 ON	0	0	0	1
xA	0	0	1	0
!B	0	0	1	1
xB	0	1	0	0
!A	0	1	0	1
XOR	0	1	1	0
¡2	0	1	1	1
AND	1	0	0	0
!XOR	1	0	0	1
A	1	0	1	0
!xB	1	0	1	1
B	1	1	0	0
!xA	1	1	0	1
OR	1	1	1	0
ALL1	1	1	1	1

Table 2.1: All possible Boolean functions with two inputs. Note that the two inputs are not equivalent.

2.3 Cell cell communication

Cells communicate in many different ways. Roughly, two kinds of intercellular communication can be distinguished. Signals can be transmitted by means of diffusible microhormones, excreted by the cell in the extracellular medium. These diffusing signals possibly have a reach of several cell perimeters. Many intercellular signalling mechanisms are known however, where intimate contact between the cells is prerequisite. Biological examples of such contact dependent signals are reviewed in [11].

On the one hand for ease of implementation, on the other hand to explore the role of contact signalling in morphogenesis, in our model cells communicate only via contact dependent signals.

Contact dependent cell cell signalling in our model was implemented in the following way. In the Boolean networks, a number of *genes* receive inputs from the networks of neighbouring *cells*.

The *cells* are assumed to express a number of proteins, that are "presented" at the cell surface. These *ligands* bind to the cell surface *receptors* of the surrounding cells. For every possible type of *ligand* a receptor is present on every cell. The output of such a *receptor* is `1` whenever the ligand matching it is presented by at least one of the surrounding cells.

In the simulations presented in this paper, the *genes* 1-6 code for *ligands*. The matching *receptors* of these *ligand* are indicated with negative numbers. -1, for instance, indicates the *receptor* for the *ligand* expressed by *gene* 1.

A consequence of this intercellular communication may be that many more different state

cycles are possible in a system of coupled Boolean networks than in an isolated network. The networks may "drive" eachother's dynamics.[2]. It is easy to see that in such a coupled system the extra state cycles are "driven" by neighbouring cycles.

2.4 Differential adhesion driven cell movement

As it was discussed in section 1.3, differential adhesion driven cell rearrangement is assumed to be an important process in the morphogenesis of multicellular animals. Differential adhesion driven cell sorting is easily and beautifully modelled using an energy minimisation algorithm devised by [17]. This algorithm has been further developed for use in models of morphogenetic processes, such as the development of *Dictyostelium discoideum*, by [37, 36].

The Glazier and Graner algorithm is a cellular automaton, in which cells are represented as pathes of CA cells in the same state. This state uniquely identifies a cell. To avoid confusion, following Savill and Hogeweg, in the rest of the text these patches of CA cells will be called "cells". We will refer to CA cells as "sites".

In "real" embryos, cell cell adhesion is a complicated process. Cell adhesion is mediated by cell surface molecules such as N-CAMs and cadherins and by adhesion to the extracellular matrix. In this algorithm all the different processes playing a role in cell cell adhesion have been lumped together in the concept of "surface energy". The lower the surface energy between two cells, the stronger they adhere.

Sites of unequal state, i.e. sites belonging to different "cells", are connected via dimensionless energy bonds. The strength of these energy bonds depends on the "*gene* expression pattern" of the cells.

The surface energy of cell i is defined by:

$$H_i = \sum_j J_{i,j} + 2 \sum_j J_{i,medium}$$

where H_i is the surface energy of cell i, and $J_{i,j}$ represents the strength of the surface energy between cell i and cell j.

In each iteration of the CA, a random site at the border of two cells is chosen. It is checked whether copying the state of a random neighbour into this site (one could see this as the extension of a "phylopodium") *would* free any local surface energy (i.e. whether this operation would result in a surface energy drop). If, and only if this is so, the state of this random neighbour is really copied into the site. In addition, some extra copying steps due to "thermal noise" are allowed to prevent the algorithm from getting stuck into local minima. These "thermal noise" copying steps, leading to a surface energy rise, are accepted according to the Boltzmann probability function $P = e^{\frac{-\Delta H}{T}}$ (fig. 2.1)

The best way for a cell to minimise its surface energy, is to decrease its surface. Therefore, it is assumed that a cell has an optimal size. If is the cell is smaller or larger than this optimal size,

[2]Unreported data by Kauffman [29], pp. 547, would suggest that the number of state cycles in system of spatially coupled Boolean networks is twice the number of state cycles in an isolated network

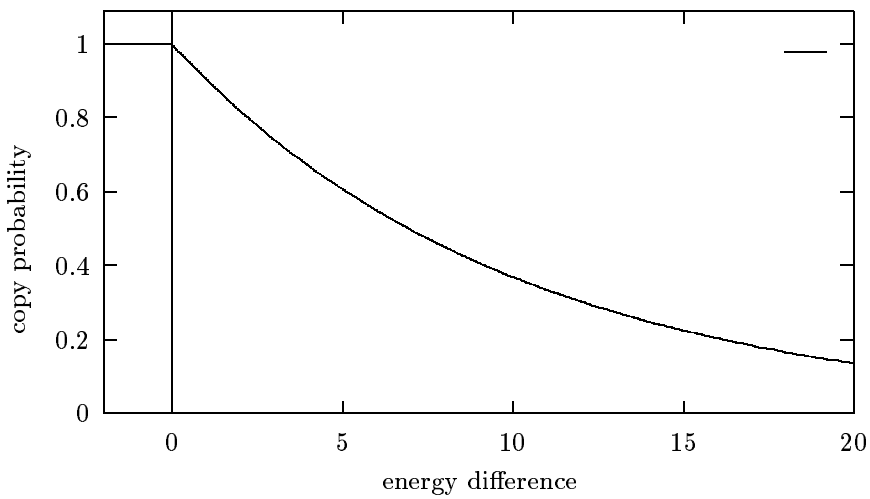


Figure 2.1: The Boltzmann probability function

an opposing "elastic" force will bring it back to its original size. The elastic force is implemented by an extra term in the energy function of the cells, which is now defined by:

$$H_i = \sum_j J_{i,j} + 2\sum_j J_{i,medium} + \lambda(V_i - v_i)^2$$

where V_i is the optimal size, and v_i is the actual size of cell i. λ represents the elasticity of the cells. The higher this parameter, the more energy is needed to deform the cell's membrane.

In this model, several behaviours can be distinguished that will be important for understanding the morphogenetic processes in the evolved beasts. In table 2.2 some examples of these behaviours are shown. In each of these examples, two cell types are used.

If the cells bind more tightly to their own type than to cells of the other type, the cells sort out. Contrary, if the cells bind more tightly to the other type than to their own type, the cells mix. Finally, engulfment occurs if one of the cell types has a higher surface tension to the medium than to the other type.

The adhesion strength between two cells is determined by the Boolean networks. Ten of the bits in the Boolean networks have been assigned the function of "cell surface protein". Five of these adhesion bits act as receptors, or "locks", while the other five bits act as donors, or "keys". Some of the receptor-donor pairs bind more strongly than other pairs, in order to allow the cells to fine-tune their adhesion strength to the other cells.

To compute the adhesion strength between two cells, the following procedure is followed. The adhesion bits (bits 2-11)[1] are extracted from the state vectors of both cells. One of these vectors is mirrored, such that the receptor bits of the first cell are aligned with the donor bits of the other cell. Then the "match vector" is computed, which is the logical AND of the adhesion vector of the first and the mirrored adhesion vector of the second cell. The strength of the surface energy

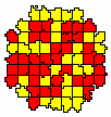
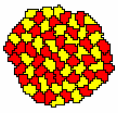
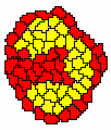
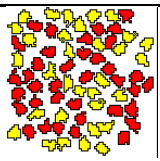
Initial configuration	
Cell Sorting $J_{white,white} = J_{grey,grey} < J_{white,grey}$	
Cell Mixing $J_{white,white} = J_{grey,grey} > J_{white,grey}$	
Engulfment $J_{white,grey} < J_{white,medium}$ $J_{grey,medium} < J_{white,medium}$	
No cell cell adhesion $J_{cell,cell} > 2J_{cell,medium}$	

Table 2.2: A list of cell sorting behaviours in the Glazier and Graner model

bond between the two cells is equal the number that is represented by the logical OR between the lower five bits and the mirrored higher five bits of the match vector to which an small "energy-offset" is added. The advantage of this definition of the surface energy between two cells relative to alternative definitions is, that the binding strength between two cells is directly determined by the "gene expression patterns" of the cells. Also, high and low adhesion strength are relatively well distributed over the possible gene expression patterns. The somewhat complicated mirror operations were necessary to assure symmetric binding. Cell A should bind as strongly to cell B as cell B binds to cell A.

For determining the adhesion strength between cells and medium, it is assumed that half of the "surface proteins" encoded by the adhesion bits are relatively fatty, or apolar. In other words, they repel the medium. Each of these hydrophobic surface proteins adds more or less to the surface tension between the cell and the medium. Just as in the computation of the adhesion strength between two cells, there are strong and less strong hydrophobic "surface proteins". The energy bond between a cell and the medium is equal to the value of the five bit number formed by the even adhesion bits.

2.5 Cell division and maternal "genes"

During the development of a *metazoan* the cells divide 7 times, upon a global signal. Cells divide over their shortest axis. The Boolean networks are duplicated, including their states. Duplication of the the Boolean network can be seen as duplication of the DNA. Copying the state vector resembles copying the cytoplasm of the cells, in this way introducing a simple form of "cell memory".

It was necessary to introduce a minimal number of unequal divisions in order to "prime" the diversificating process. After the first division in one of the cells bit 21 of the Boolean network is flipped. This signal was called the "bicoid" signal, as it resembles the *Drosophila* maternal polarising bicoid signal [8, 9]. In some runs, such as the run within which the organism described in chapter 3.2 was bred, the second division is unequal as well. After the second division bit 22 is flipped in one of the four cells. This signal was called "activin", inspired by the possible role of activin in *Xenopus* dorsoventral patterning [3].

2.6 Cell death: "Lonelyness penalty"

In many preliminary runs, the *metazoans* were pulled towards a quite uninteresting mechanism to generate cellular diversity. Non-adhering cells with very long cycles were developed. During development more and more cells disattached from the "embryo". Dependent on the phase of their cycle at which they disattached many different cell types were produced. Even more cell types were generated by a mechanism in which the cells "tickled" eachother during a few time step, in this way "pushing" them out of their cycle.

In order to prevent this behaviour a penalty was given to cells that didn't touch any other cell. Every time step a cell was "lonely", the the cell's target size was decreased by one site with a

probabiliy that was dependent on the size of the cells. Large cells shrunk faster than small cells.

2.7 A detailed overview of the implementation of a *metazoan*'s development

1. The CA plane is primed with an elipsoid zygote. The zygote's Boolean network is primed with zeros.
2. The cell divides, the Boolean networks is duplicated, together with the state of the Boolean network.
3. If the cells have divided for the first time and if the bicoid signal has been enabled, the bicoid bit (bit 21) is flipped.
4. If the cells have divided for the second time and if the activin signal has been enabled, the activin bit (bit 22) is flipped.
5. The cell's neighbours are determined and the receptor vector is computed.
6. The Boolean networks are updated. Closed cycles are detected.
7. The adhesion between the cells is computed from the network state vectors of the cells.
8. The cellular automaton is updated, resulting in cell movement.
9. Iterate steps 5-8 for 100 steps after the first, the second and the third division, for 1000 steps after the next divisions and for 5000 steps after the eighth division.
10. Iterate steps 5-8 for another 100 steps and compute lowest cellular diversity during these iterations.

Chapter 3

The ontogeny of a simple organism

In the next chapter it will be explained how the *metazoans* are evolved. First, however, we will follow the development of a simple *metazoan*. This will clarify how the different parts of the model are put together to result in a the simulated development of a multicellular "organism". Additionally, it may evoke some intuition about how the *metazoan* developmental program is coded in its *genome*. This may help the reader in understanding what happens during the *metazoan* evolution.

In order to obtain a simple example organism, an evolutionary search was set up selecting for *metazoans* that developed stable cell types[1]. From this run a nice example was chosen. The development of this example is described in detail below.

3.1 Analysis of the network: housekeeping and dynamic *gene*s

The network of our organism is shown in fig. 3.1. The first thing to notice is, that many of the *gene*s are connected to non-communicating *gene*s of the neighbourhood vector. The inputs coming from these *gene*s are set to zero.

Second, many *gene*s have superfluous inputs. The functions `ALL1` and `ALL0` do not need any input at all, while for the functions `A`, `B`, `!A` and `!B` only one of the inputs is functional. For simplicity, we could remove from the figure all these non-functional nodes and connections.

If we take a closer look at the network however, many of the *gene*s, for instance the *gene*s 12, 5 and 19, turn out to be continuously turned on or off. These *gene*s are called *housekeeping gene*s, because they are always expressed or turned off in all the cells.

On the other hand, there is a core of *gene*s that is sensible to inputs from neighbouring cells. The cells diversify by changing the expression of these *gene*s. Let's call them *dynamic gene*s. In figure 3.2, all the non-communication connections and *housekeeping gene*s have been removed from the graph. Additionally, the names of the Boolean functions have been simplified where possible.

[1]The reader may be interested in the complete results of the evolutionary process selecting for stable cell types. However, due to time- and software-related problems, the run was interrupted at a premature stage. Complete results will be available later.

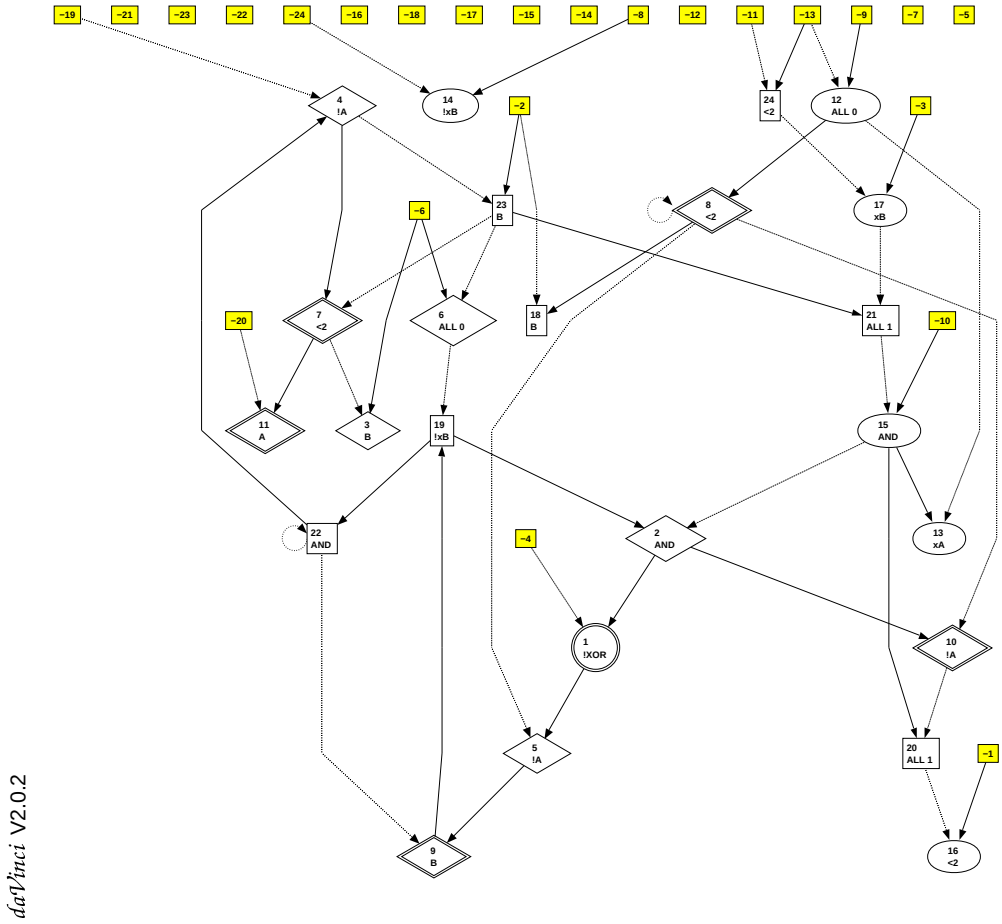


Figure 3.1: The Boolean network, or “genome” of a simple organism

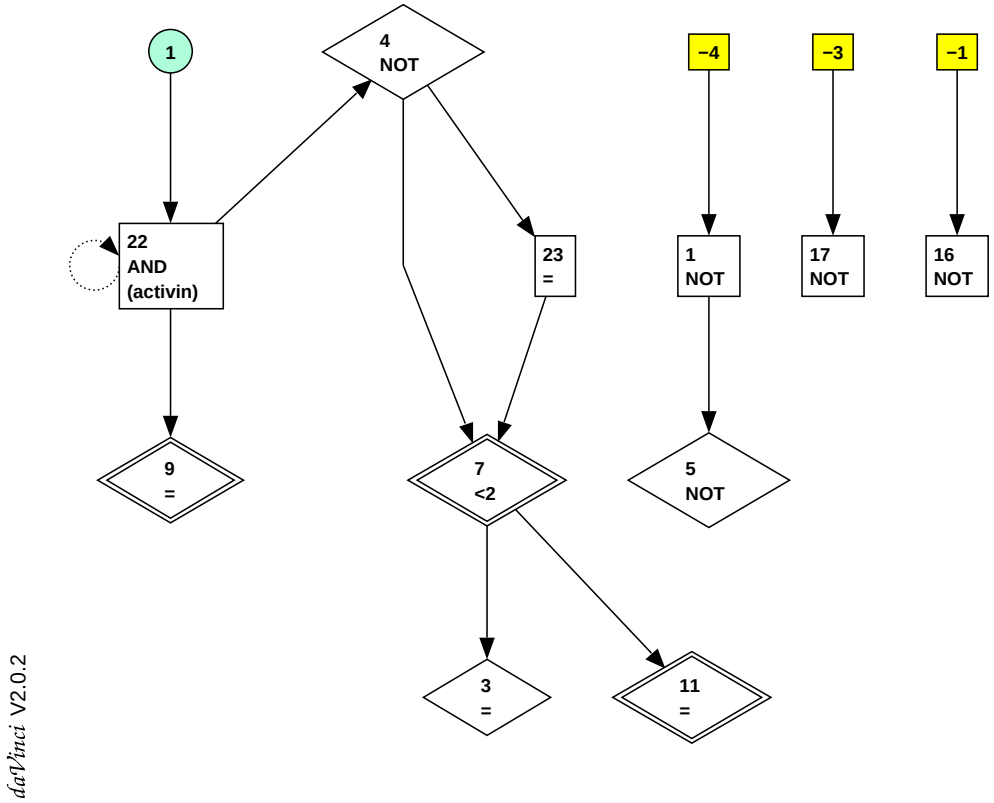


Figure 3.2: The dynamic part of the network shown in fig. 3.1. All non-communicating connections and housekeeping *gene*s have been removed. Diamonds indicate "sticky *gene*s", gray squares indicate neighbourhood connections and gray circles indicate stable inputs.

3.2 Early pattern formation

In figure 3.3(a) the initial condition, or "zygote" of the organism is shown. The state vector and the neighbourhood vector of the zygote are primed with zeros. After two iterations, the housekeeping *gene*s have stabilised. After the first division, the bicoid *gene* (`# 21`) will be flipped. In our case it was turned on, so we switch it off. Obviously, after one iteration the bicoid bit is turned back on, because of its Boolean function `ALL1`. In spite of the network's insensitivity to the bicoid bit, the daughter cells do differentiate after the first division, because the cells "sense" each other's neighbourhood vector. Whereas all the bits of the neighbourhood vector were turned off in the zygote stage, the neighbourhood vector of the two cell is now equal to `001001`. It is easy to see, that bit 16 will be turned off, because bit 1 of the neighbour vector is turned on. Bit 1 of the state vector is switched off because bit 4 is turned on. In the next iteration, bit 1 will be turned off in the neighbourhood vector, causing bit 16 to flip back on.

After the second division (figure 3.3(b)) in one of the cells the "activin" bit is flipped. The network needs a mechanism to store this signal, because it is only applied during one iteration. This is nicely achieved by a feedback loop on the activin *gene*. If the activin bit is turned off, it inputs a `0` to itself. The next state will be a 0 again, and so forth. If, however, the activin bit is turned on, it will keep itself turned on.

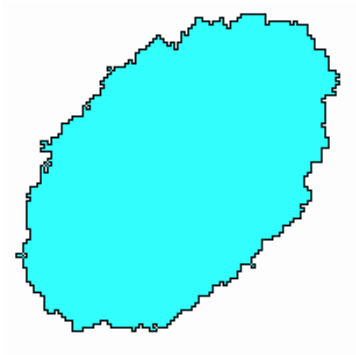
Interestingly, the state of the activin bit is propagated to a subnet of the bits `4`, `23`, `7`, `11` and `3`. To understand the behaviour of this subnet, let us first return to the zygote. In the zygote, all the bits of state vector and the neighbourhood vector were turned off. After the first iteration, bit 4 was switched on. Bit 7 was turned on, because one of its inputs, bit `23` was 0. After the second iteration, bit `23` was turned on, because its input is 1. This caused bit 7 to be turned on. As a result, the bits `11` and `3` were switched off after the following iteration of the network. Now that activin has been switched on, the expression of the subnet is inverted. First, bit `4` is switched off. As a result, the bit `23` is switched off, whereas bit `7` is turned on. Bit `7` finally, switches on the bits `11` and `3`.

Now that the activin cell has differentiated, the neighbourhood vector of the surrounding cells changes. In the activin cell, *gene* `3` is expressed. By these means, in all the surrounding cells *gene* `17` is suppressed by bit `3` of the neighbourhood vector.

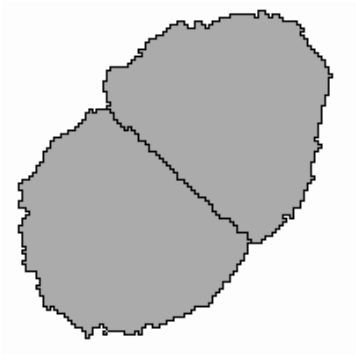
After the third division, the main pattern of the creature has been laid down. In figure 3.3(d) two changes are apparent. First, some pink cells have dedifferentiated into gray cells. They do not touch the "activin" cell any more, so that *gene* `17` is expressed again. Second, the black "activin" cell has divided. It differentiated into the greenish brown cells in about the same way the gray cells appeared after the first division. Now that it is connected to another black cell, bit `3` of the neighbourhood vector is expressed, by which means *gene* `17` is suppressed.

3.3 Morphogenetic cell movements

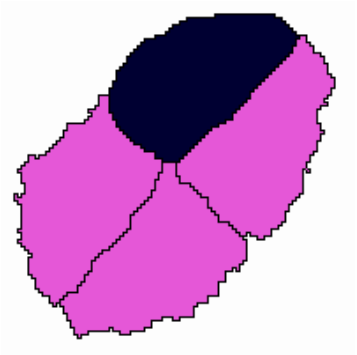
Now that the basic pattern of the organism has been laid down, after the third division cellular movements become apparent. First, the pink cells engulf the greenish brown cells (see figure 3.3(d)). Finally, the gray cells dissociate (figures 3.3(g)-3.3(k)).



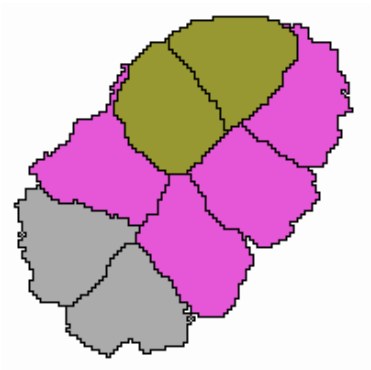
(a) Zygote



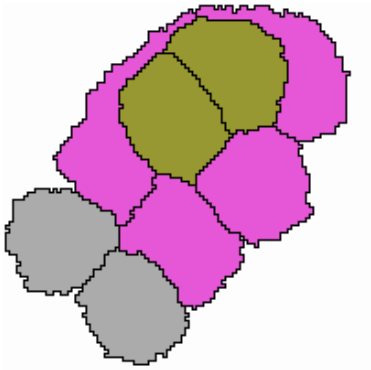
(b) Two cell stage



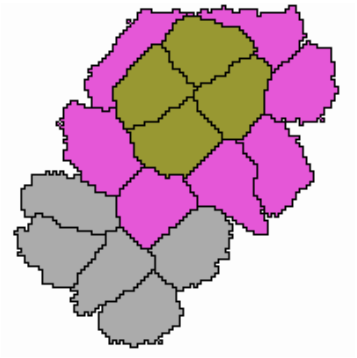
(c) Four cell stage



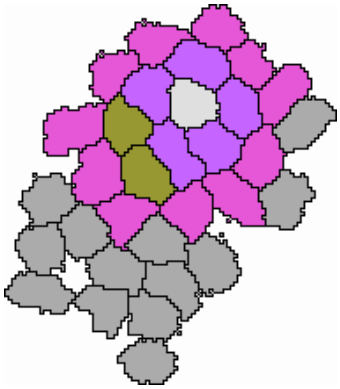
(d) Eight cell stage, early



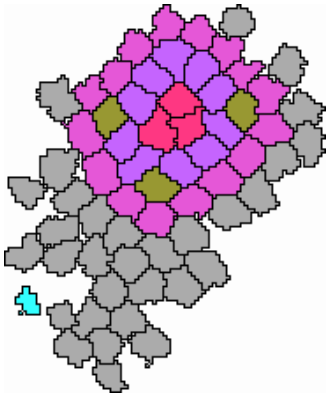
(e) Eight cell stage, later



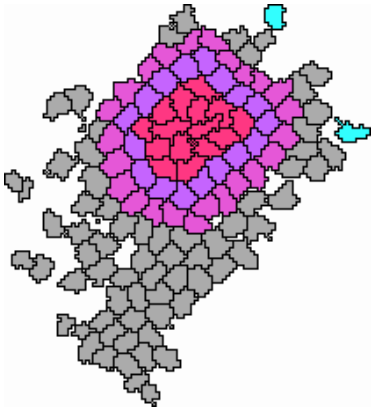
(f) Sixteen cell stage



(g) Thirty-two cell stage

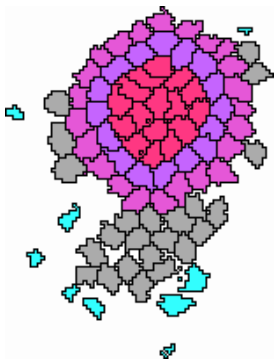


(h) Sixty-four cell stage

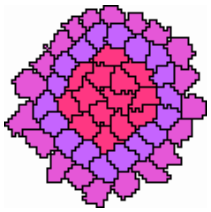


(i) 128 cell stage, early

Figure 3.3: The ontogeny of *Ontosilica gastrulans segundo*



(j) 128 cell stage, later



(k) 128 cell stage, final

Figure 3.3: continued.

Interestingly, these cellular movements can be traced back to a striking difference between the cell-medium adhesion of the cells descending from the "activin" cell, and the cells that do not descend from the "activin" cell. In the last paragraph it was shown that in the *activin cell line* bit 22 is constantly set. Bit 22 in turn keeps the bits `11`, `9`, `7`, and `3` turned on.

Remember that the strength of the energy bond between a cell and the medium is computed by taking the even bits of the *sticky vector*, i.e. the bits `11`, `9`, `7`, `5` and `3`. For the activin lineage this gives `111*1`. `*` is equal to 1 if the cell touches a cell that does not express the activin bit. Adding an `energyoffset` of 3 gives an energy bond of 32 or 34. In the other cells the activin bit is reset, giving a energy bond of 5.

The adhesion strength between the *activin* cells is computed as described in the last chapter. First, one of the *sticky vectors* is mirrored, such that the `key` *gene*`s` are aligned with the `lock` *gene*`s`. The `AND` of these vectors gives us the `match vector`, describing which `key-lock` pairs match between the cells:

```
111110*010
010*011111
———————————  AND
010*00*010
```

Again, `*` represents a `1` for the cells that touch a non-activin cell.

The energy bond between the cells is computed from the `OR` between the left five bits of the match vector and the mirrored right five bits:

```
0*010
0*010
———————  OR
0*010
```

So, between two *activin cells* the strength of the energy bond is 5 if neither of the cells touches a non-activin cell, otherwise it is 13. Following the same procedure, we find that the adhesion strength between two *non-activin* cells is always 11. The adhesion strength between activin cells and non-activin cells is 17.

In summary, the following conditions are valid:

$$J_{act,act} < 2J_{act,medium} \tag{3.1}$$
$$J_{na,na} > 2J_{na,medium} \tag{3.2}$$
$$J_{na,na} < J_{na,act} \tag{3.3}$$
$$J_{act,act} < J_{na,act} \tag{3.4}$$
$$J_{act,na}(+J_{na,medium})? < J_{act,medium} + J_{na,medium} \tag{3.5}$$

$J_{A,B}$ represents the energy bond between cell type *A* and cell type *B*, *act* stands for cells of the act*ctivin* lineage and *na* stands for n*on*-a*ctivin* induced cells.

From the conditions 3.3 and 3.4 it is easy to see that the activin lineage and the non-activin lineage will remain sorted out (see also table 2.2). It is also clear from condition 3.1 that the activin cells will remain adhered to each other. In contrast, the *non-activin* cells dissociate, because they adhere more strongly to the medium than to each other. Finally, condition 3.5 shows that the non-activin cells will engulf the activin cells. This can be understood intuitively in the following way. In figure 3.3(e) the greenish brown , activin derived cells repel the medium. At the same time, the pink non-activin cells adhere strongly to the medium, whereas they adhere, albeit weakly, to the greenish brown cells. This causes the pink cells to cover the “hydrophobic” surface of the greenish brown cells.

3.4 Final pattern formation

After two more divisions, in figure 3.3(g) two new cell types appear, a white one and a purple one. These cell types arise in an interesting cascade of differentiations. First, one of the greenish brown cells differentiates into the white cell type. When this cell is isolated from the pink cells, bit `4` of its neighbourhood vector is set to 0. This causes bit 1 to be turned on. In the next iteration, bit 1 switches off bit 5.

As a result of the expression of bit 1 in the white cell, bit 16 is switched off in the surrounding greenish brown cells. This results in a new — purple — cell type.

In figure 3.3(h) the white cell has divided. In the resulting red pink cell bit `16` is suppressed by bit `1` of the neighbouring cells.

In parallel, the gray cells dissociate. Finally they vanish (fig. 3.3(k)), because of the “LONELYNESS PENALTY” that was given to dissociated cells. Also note, that the gray cells differentiate in the cyan “zygote” cell type if isolated from the other cells.

Chapter 4

Evolution of the metazoans

4.1 The genetic algorithm

The evolutionary process of the *metazoans* is simulated using a genetic algorithm [20] (GA) on a *parallel virtual machine*[1](PVM) configuration. This configuration consists of an array of 16 *Pentium* 166 Mhz personal computers running Linux 1.2.13, mastered by a Silicon Graphics machine.

The developmental process of the *metazoans* is simulated on the Linux machines. A master process, running on the Silicon Graphics machine, performs the genetic algorithm. The evolutionary process is initiated by generating 16 random *genomes*. These *genomes* are sent to the slave machines where they are developed. The master process keeps track of the 16 genomes currently being developed in the socalled *beast buffer*. The *beast store* contains the last 16 genomes whose development has been completed, together with their fitnesses. Whenever a slave machine has finished developing a beast, it sends the fitness to the master process, together with a dump of the last stage of the *metazoan*'s ontogeny. This dump consists of a "snap shot" of the CA, a dump of the cell cycles and a dump of the states of the Boolean networks. The master process writes a *fossil record*, consisting of the *genomes* of all the *metazoans* evolved, together with their fitness, their *beast IDs* (a number ranking the point in evolutionary time at which the *metazoan* evolved), their parent's *beast ID* and the dump of the last stage of their ontogeny.

A tournament selection scheme is followed. Whenever a machine has finished developing a beast, the beast's *genome* is shifted from the beast buffer to the beast store, in this way removing the beast that was present at this entry of the beast store from the reproductive process. A sample of eight individuals is extracted from the *beast store*, consisting of the 16 most recent genomes whose fitness was known. Out of these eight individuals the genome of the best performing *metazoan* was mutated in 50% of the cases and it was developed on one of the slave machines. If the best performing beast appeared to have developed only a single cell type, a random genome was generated. This initial "bootstrap" procedure saved the algorithm from initial searches to a *metazoans* having more than one cell type. Such an initial search would last very long, due to

[1]A recent version of the Parallel Virtual Machine can be obtained from:
`http://www.epm.ornl.gov/pvm/pvm_home.html`

the small population diversity.

4.2 Fitness criterion

As it was discussed in chapter 1, the aim of the paradigm system of the evolution of multicellular creatures is twofold. The first aim is to make a start in an understanding of the phenotype landscape and the evolutionary dynamics caused by the highly complex metazoan genotype-phenotype mapping. The second aim is to "breed" hypotheses on how cellular diversity can be produced and maintained during the development of multicellular animals using cell communication and cell movement.

Consequently, the fitness criterion that was constructed had to face these two aims. For the understanding of the "metazoan" fitness landscape we could have set up and evolutionary search towards a predefined morphology of cell type pattern. For example, we could have selected for "metazoans" exhibiting bilateral or radial symmetry. This strategy has proven to be fruitful in the study of RNA phenotype landscape [21]. In these studies, an evolutionary search was set up towards an secondary structure with two loops.

Still, the choice was not to set up an evolutionary algorithm to search for a predefined morphology. As a fitness criterion, a measure for cell type diversity was used. This criterion satisfied our second aim: the generation and maintenance of cellular diversity. However, we consider it more important that this fitness criterion is trivial with respect to our evolutionary search image. A *metazoan* is considered more complex if it has more different cell types, irrespective of how it produces them. An evolutionary scheme selecting for the number of cell types developed was thought to be sufficiently "undefined" to allow the developing *metazoans* in an unpredefined way.

Selection for the number of cell types was done in two ways. As a first try, each different state cycle was simply called a different cell type, regardless of how different the cell types actually were. In this method, two stable states having differences as tiny as a single flipped bit, gave as much fitness to an organism as, say, a stable state and a state cycle.

In a way, this made diversification to easy. The organisms very quickly evolved long state cycles, in order of 64 steps, that diversified by means of two different mechanisms. The first mechanism was, to propagate a desynchronisation between the cell types induced by the bicoid and the activin signals on the one side and the other cells on the other side. This desynchronisation would induce differences between the cell types on the order of several bits. The second mechanism was to evolve very "fluent" cell types (i.e. with a low energy difference to the surrounding cells). These cell types would squeeze out philopodia between their surrounding cells. Once a philopodium touched a new cell. it differentiated, an retracted the philopodium.

The actual fitness criterium was based on the diversity of cell types a "beast" was able to develop, rather than its *number* of cell types. For each cell type that was present at the moment of fitness evaluation, the difference to the other cell types was computed. This was done irrespective of the ubiquity of the cell types. In other words, a single cell of a new cell type contributes as much to the fitness of a `beast` as ten cells of this new cell type. In order to select for *metazoans* that were able to maintain their cellular diversity for a period of time, the fitness of a *metazoan* was defined to be lowest cellular diversity in the last one hundred time steps of its development.

The algorithm computating the fitness is explained in box 4.3.

4.3 Computation of cell type diversity

First the mean expression over time of each *gene* in the state cycles is computed

$$m_j = \frac{1}{T} \sum_{\phi=1}^{\phi=T} x_{\phi,j}$$

where m_j is the mean expression in the state cycle of *gene* j, $x_{\phi,j}$ is the expression of *gene* j in phase ϕ of the cycle and T is the period of the state cycle.

Then, for each *gene* of the state cycles, the difference in mean expression to the other state cycles is computed.

$$D_{a,b} = \sum_{j=1}^{j=N} |m_{a,j} - m_{b,j}|$$

$D_{a,b}$ is the *gene* expression difference between state cycle a and state cycle b. N is the number of *gene*s in the network and $m_{a,j}$ is the mean expression of *gene* i in state cycle a.

The cellular diversity of the organism is defined as the mean expression difference between the state cycles

$$diversity \doteq \frac{1}{C} \sum_{a=1}^{a=C-1} \sum_{b=a+1}^{b=C} D_{a,b}$$

a, b, and $D_{a,b}$ are defined as above. C represents the number of cell types of the organism.

The *fitness* of a *metazoan* is the lowest diversity it reaches during the last 100 time steps of its ontogeny.

Chapter 5

An evolutionary run

In figure 5.1 the cumulative genetic distance of the genomes relative to a 90 time steps younger individual is shown. It is clear that the evolutionary change at the genetic level progresses at a fairly constant rate. The molecular clock is ticking at its maximum rate, as if there would be no selection at all. This suggests that no strong selection is acting on most of the possible mutations.

However, at the phenotypic level, the fitness of the individuals increases stepwise. Periods of stasis, where the fitness of the creatures remains constant, are intermitted by short periods of evolutionary change.

Interestingly, the fitness jumps or *evolutionary innovations*, are correlated to key innovations in the structure of the Boolean networks, and as a result to changes both in the development and in the adult phenome[1] of the creatures. Table 5.5 lists the epochs together with the phenotypic and genotypic innovation that have most probably caused the transient from one epoch to the next.

In order to understand which structural changes in the genotype have caused a transient from one epoch to the next, the same procedure as in the example of chapter 3.2 was followed. The *gene*s whose state was stable throughout development were stripped of the network and the Boolean functions were adjusted. For example, from a `NOT A` function, the B connection was removed and the function's name was changed into `NOT`. In the same way, an `AND` function which had one of its inputs constantly set to one, while the other was dynamic, was changed into `=`.

For each of the epoches, a typical organism was chosen. As the morphology of these beasts was considered typical for a particular epoch, these beasts were called the *morphotype* of an epoch.

In figure 5.3 for each epoch the morphotype is shown. For each of these morphotypes the network was simplified as described above, such that only the dynamic *gene*s remain. These simplified networks are shown in figure 5.4.

Interestingly, the dynamic part of the networks gets larger and larger over evolutionary time. Figure 5.5 shows the number of dynamic *gene*s for each of the morphotypes.

[1]The word *phenome* is used here to indicate morphological and cybernetic properties of the organism that may but need not necessarily determine the fitness of the organism, as opposed to *phenotype* that implies an effect on fitness. See [32] for a discussion on the words genotype, genome, phenotype and phenome.

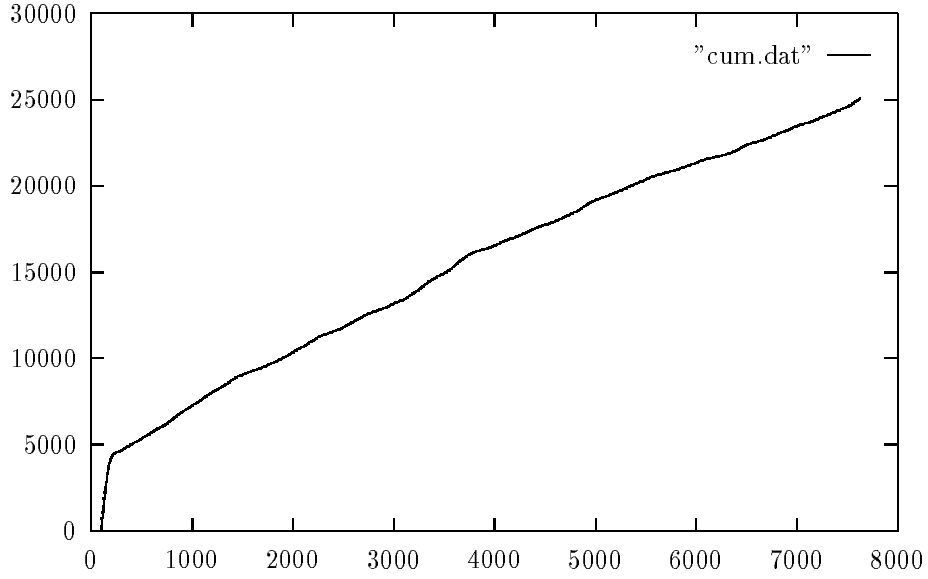


Figure 5.1: The cumulative running mean genetic distance between a population of ten individuals and ninety time steps earlier population. This plot shows that the genetic "walking" speed remains rather constant over evolutionary time. The increased walking speed at the start of the plot is due to the initial "bootstrap" procedure. During this procedure random genomes are generated until a creature having more than one cell type is found.

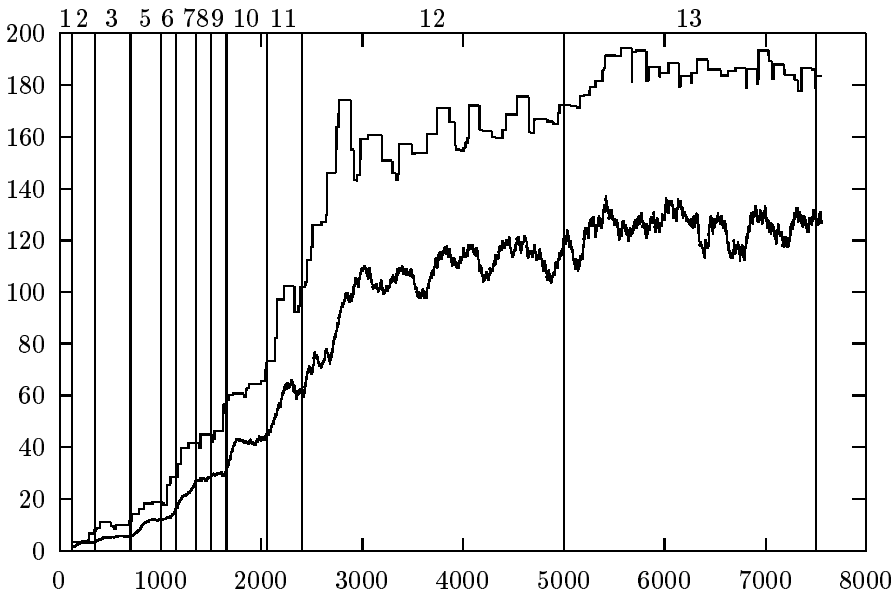
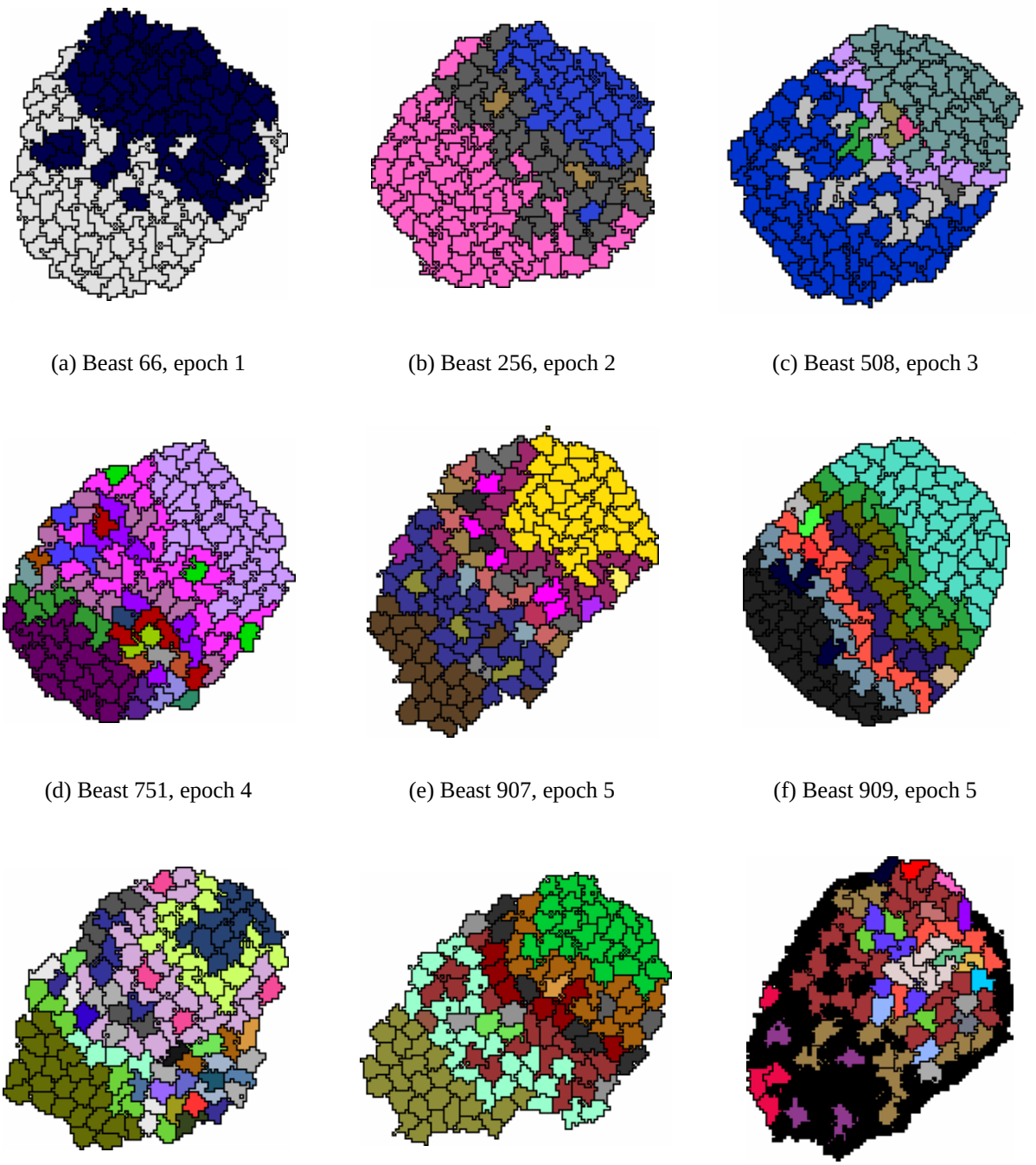


Figure 5.2: Running average and running maximum of the fitness over time. Length of the running average and maximum is 100 time steps. The numbers 1-13 indicate the epoches. The epoches are delimited by the vertical lines. The epoches are characterised by a particular mode of development. See text for explanation.

Early in evolution, in beast 66, only 50% of the *gene*s is dynamic. Hence, half of the *gene*s is able to diversify. Later in evolution, from beast 4000 onwards, 83% of the *gene*s is able to change.

Additionally, the number of *gene*s being part of a regulatory cycle increases, albeit less quickly. Beast 66 only has one cycle containing six *gene*s. In epoch three ten *gene*s are embedded in a cycle. A new cycle arose, consisting of the *gene*s 13, 23, 18 and 3. In both of the two coexisting morphotypes of epoch 5, ten genes are embedded om a cycle. In addition, two genes act as "intermediaries" between the two cycles. In epoch 6, finally, eleven "cyclic" genes are present.

Interestingly, for almost all the transitions between epochs, a change at the genetic level accounting for the rise in fitness was identified. Because it was possible to make "loss-of-function knock-outs" on these *gene*s such that the phenotype of the organism would "fall back" to the morphotype of the preceding epoch, these structural changes were called *key mutations* resulting in a new morphotype. The change at the phenotypic and developmental level between two epoches is called a *key innovation*.



(a) Beast 66, epoch 1

(b) Beast 256, epoch 2

(c) Beast 508, epoch 3

(d) Beast 751, epoch 4

(e) Beast 907, epoch 5

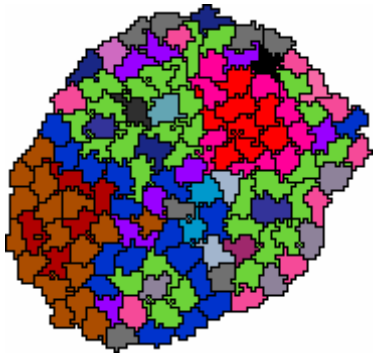
(f) Beast 909, epoch 5

(g) Beast 1124, epoch 6

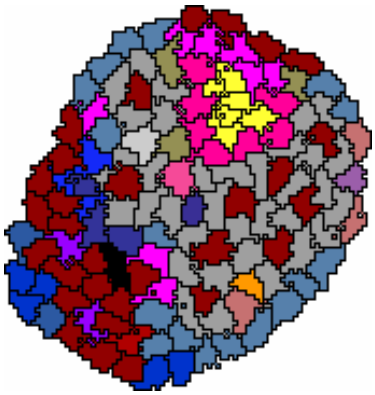
(h) Beast 1130, epoch 7(?)

(i) Beast 1389, epoch 8

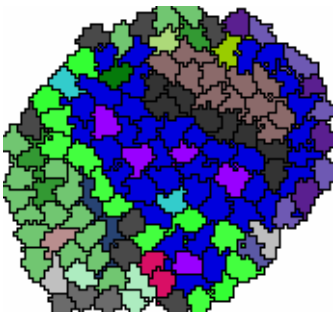
Figure 5.3: Morphotypes of the epoches. Morphotypes are morphologies considered typical for an epoch. The morphotypes were selected by a human observer scanning through a movie file containing the adult stages of the evolved creatures.



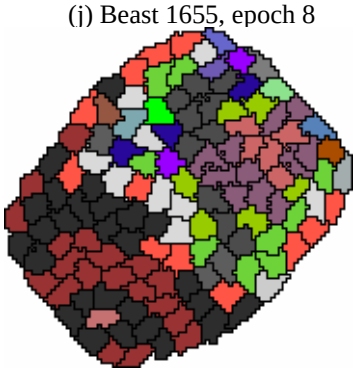
(j) Beast 1655, epoch 8



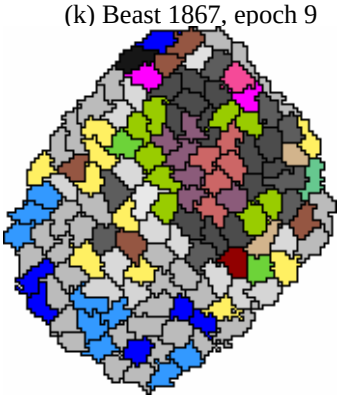
(k) Beast 1867, epoch 9



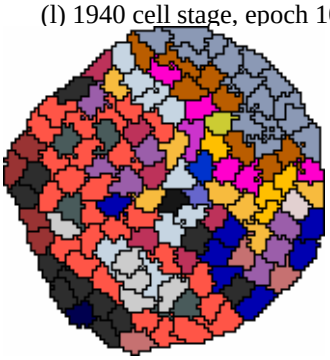
(l) 1940 cell stage, epoch 10



(m) Beast 2126, epoch 10

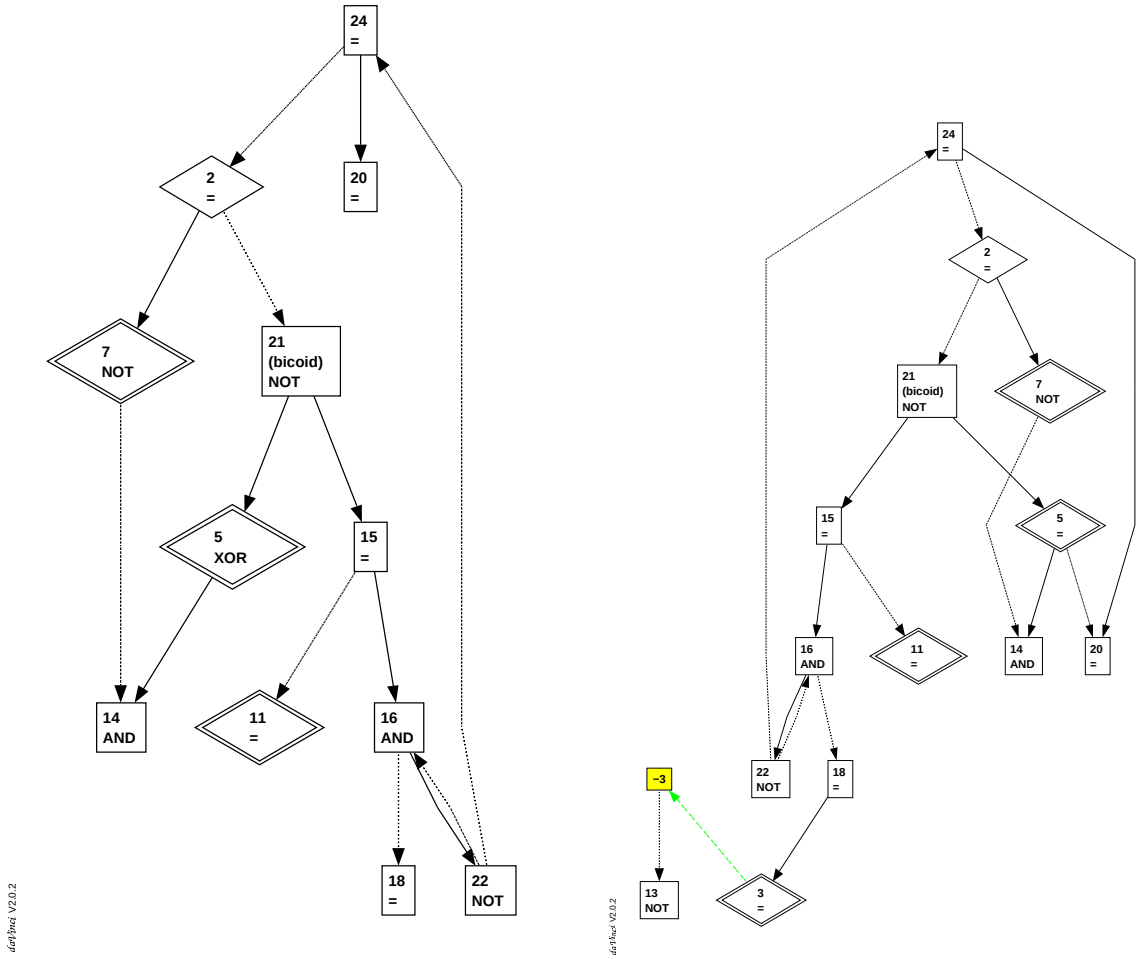


(n) Beast 2162, epoch 11



(o) Beast 2309, epoch 12

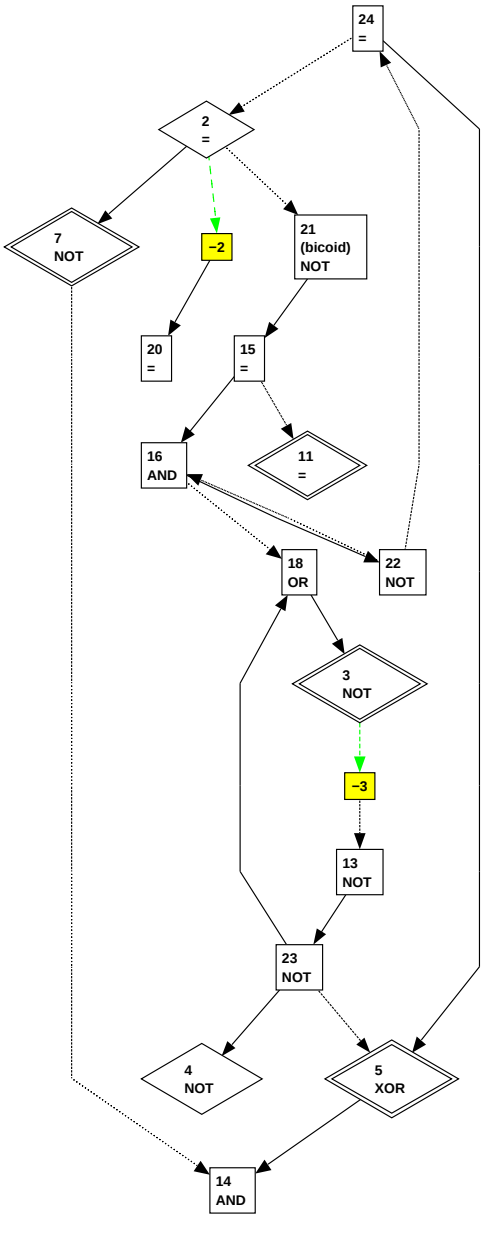
Figure 5.3: continued



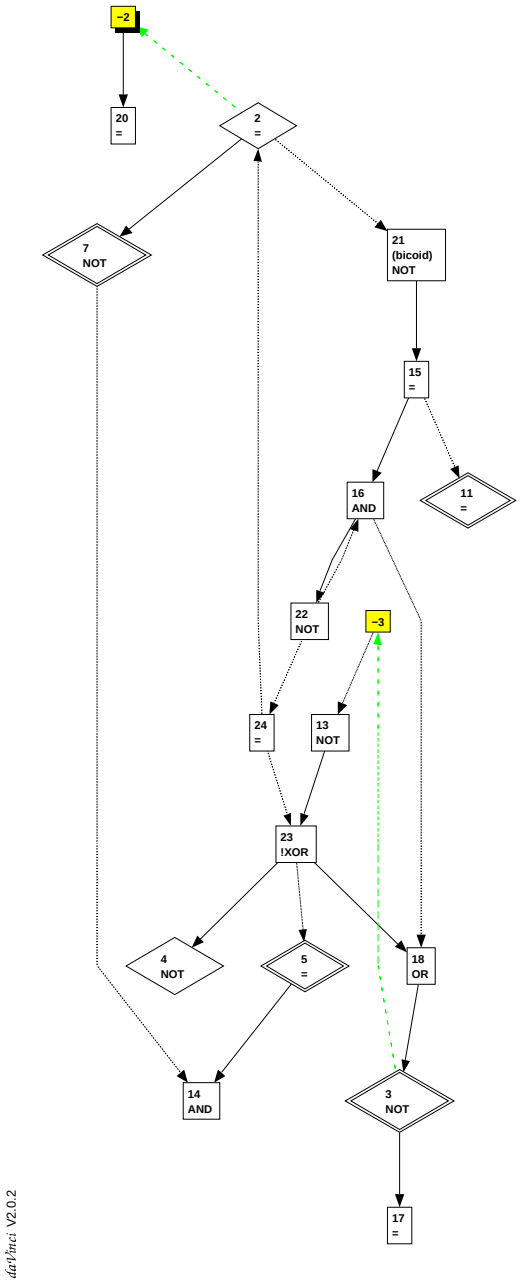
(a) Network of beast 66, epoch 1

(b) Network of 256, epoch 2

Figure 5.4: The Boolean networks of the beasts shown in. figure 5.3. All the *genes* that are stably expressed in all the cell types and in all the stages have been omitted, in order to simplify the networks. The names of the Boolean functions have been simplified where possible. Gray squares indicate *receptors.* Gray dashed arrows indicate *receptor-ligand* interactions.

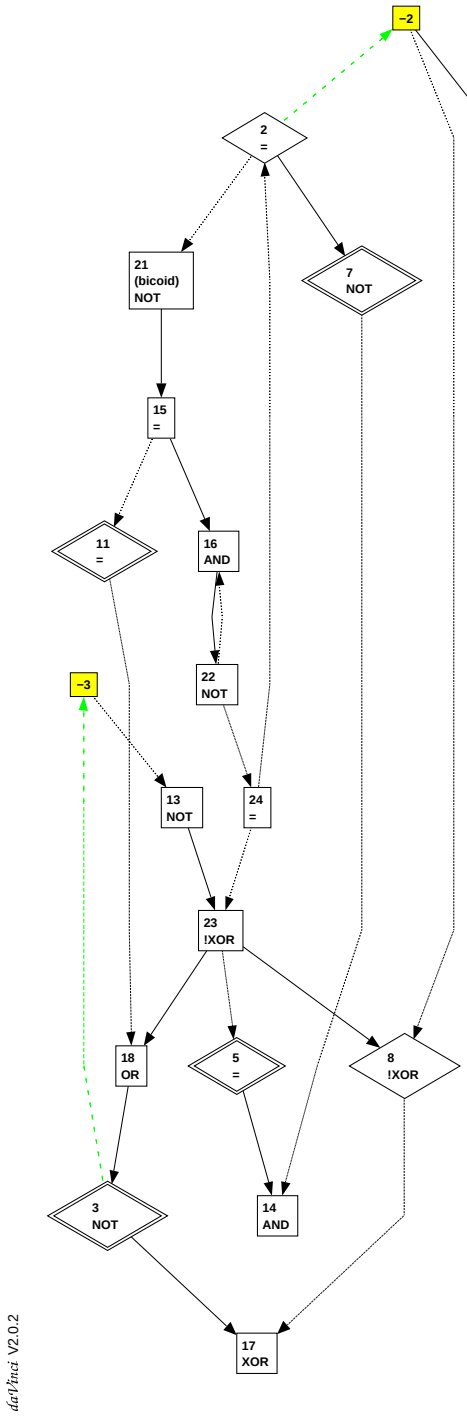


(c) Network of beast 508, epoch 3

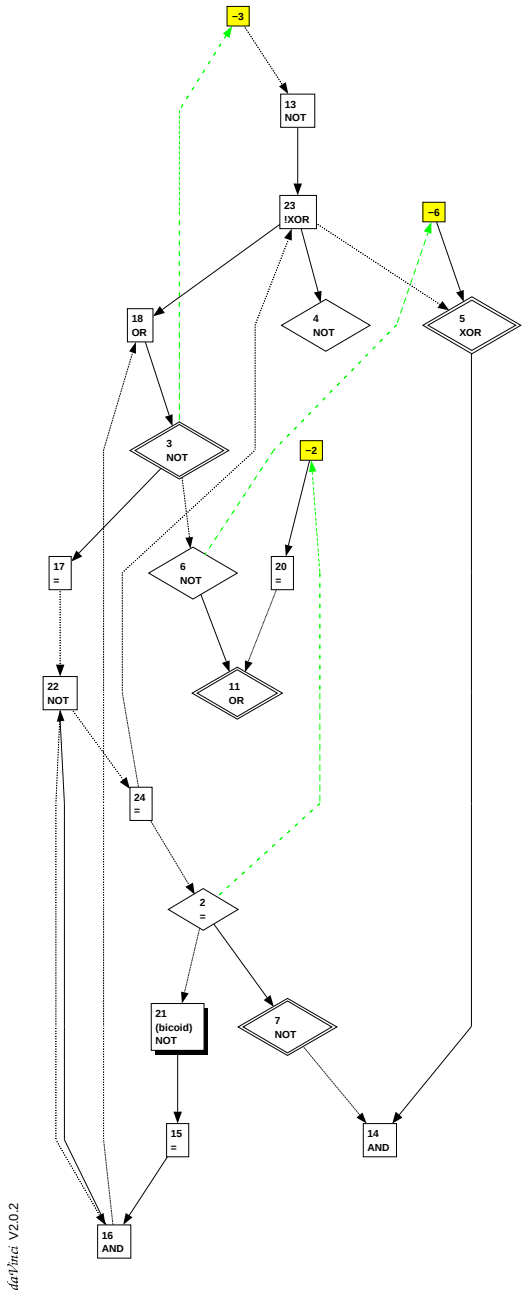


(d) Beast 751, epoch 4

Figure 5.4: continued

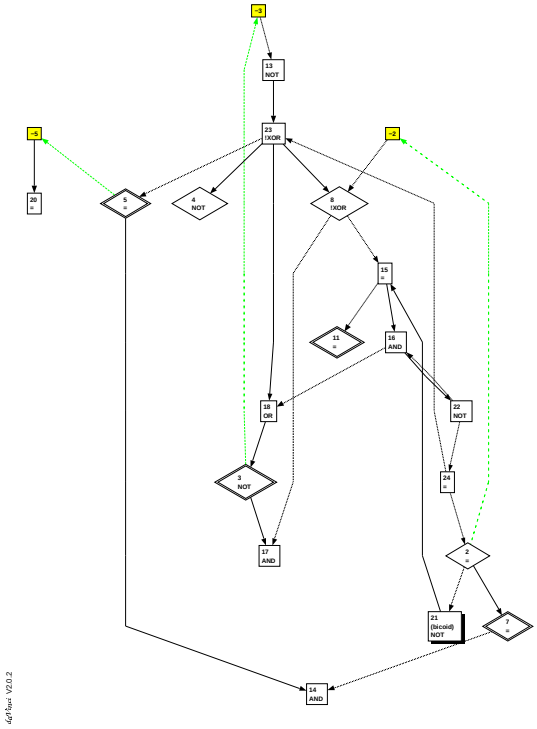


(e) Beast 907, epoch 5



(f) Beast 909, epoch 5

Figure 5.4: continued



(g) Beast 1124, epoch 6

Figure 5.4: continued

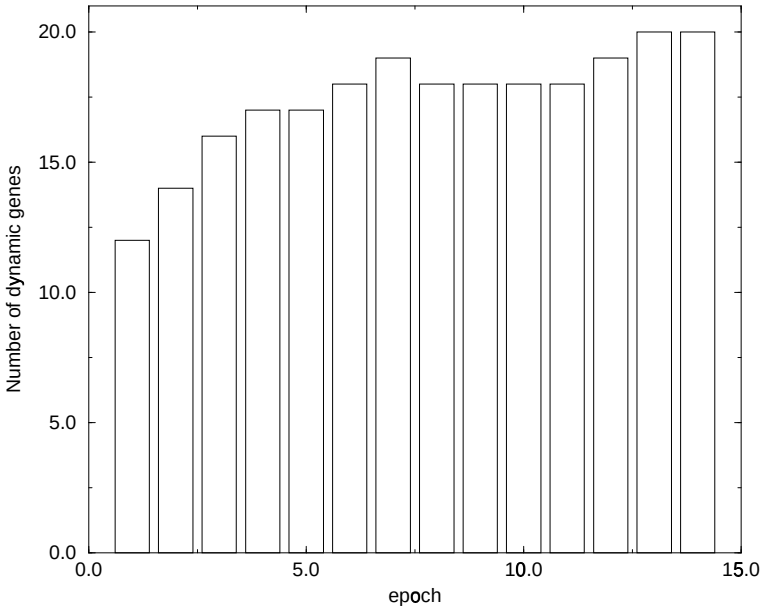


Figure 5.5: The number of "dynamic *gene*s" for the analysed genomes. The number of dynamic genesincreases over evolutionary time. In this way, the part of the genome potentially taking part in cellular diversification increases.

Gene:	2 =	21(bic) NOT	15 =	16 AND	22 NOT	24 =	
1	0	0	0	0	0	0	
2	1	1	0	0	1	0	
3	0	0	1	0	1	1	
4	1	1	0	1	1	1	
5	1	0	1	0	0	1	
6	1	0	0	0	1	0	
7	0	0	0	0	1	1	
8	1	1	0	0	1	1	
9	1	0	1	0	1	1	
10	1	0	0	1	1	1	
11	1	0	0	0	0	1	
12	1	0	0	0	1	0	(converged on step 6)

Table 5.1: Convergence to the period 6 state cycle of the non-bicoid cell

5.1 Zygotic cell polarity induces two cell lines

In the first epoch the organisms have discovered the use of the *bicoid gene*. A cycle of six *gene*s falls into one of two alternative period six state cycles, depending on the state of the *bicoid gene* during the first iteration.

The transient towards these two state cycles is shown in the tables 5.1 and 5.2.

The state of being trapped in either the one or the other of these two state cycles is stably inherited over cell division. It is easy to see that this will happen, because the state of the network is inherited.

In this way two cell lines arise, the cells descending from the *bicoid* induced and the cells descending from the non-*bicoid gene*. Interestingly, the non-*bicoid* cells have the slight tendency to engulf the *bicoid* cells. This can be understood from the fact that the *bicoid* cells adhere less strongly to the medium than the non-*bicoid* cells. (see table 5.3) This difference in hydrophoby is a result of the difference between the state cycles of the *bicoid* and the non-*bicoid* cells.

5.2 Discovery of cell cell communication

In epoch 2, the beasts have discovered cell-cell communication as a way to generate new cell types. In the *bicoid* lineage, *gene* 3 is turned ON during one of the six states of the state cycles, whereas it is turned ON during two steps in the non-*bicoid* cells. As a result, *gene* 13 is suppressed during two out of six states in cells touching a non-*bicoid* cell. Therefore, apart from non-*bicoid* cells touching other non-*bicoid* cells and *bicoid* cells that do not touch a *bicoid* cell, two new cell types are developed: *bicoid* cells touching non-*bicoid* cells (the gray cells at the border of the blue and the pink cells) and non-*bicoid* cells isolated from the other non-*bicoid* cells (the brown cells embedded in gray cells).

Gene:	2 =	21(bic) NOT	15 =	16 AND	22 NOT	24 =	
1	0	1	0	0	0	0	(bicoid bit set)
2	1	1	1	0	1	0	
3	0	0	1	1	1	1	
4	1	1	0	1	0	1	
5	1	0	1	0	0	0	
6	0	0	0	0	1	0	
7	0	1	0	0	1	1	
8	1	1	1	0	1	1	
9	1	0	1	1	1	1	
10	1	0	0	1	0	1	
11	1	0	0	0	0	0	
12	0	0	0	0	1	0	(converged on step 6)

Table 5.2: Convergence to the period 6 state-cycle of the bicoid cell

cycle phase	**J[B][med]**	**J[NB][med]**	**J[B][NB]**
1	3	3	3
2	7	3	3
3	9	7	3
4	21	5	4
5	19	19	4
6	3	3	3

Table 5.3: Energy bonds between the different lineages of beast 66. *B* denotes the cells from the *bicoid*-induced embryonal cell, *NB* denotes the cells descending from the other embryonal cell, not induced by *bicoid*. The energy difference between the medium and the non-bicoid cells is slightly lower than the energy difference between the bicoid cells and the medium. This results in a slight engulfment of the bicoid cells by the bicoid cells.

The *key mutation* for epoch two is the mutation coupling *gene* 3 to the 6 *gene* regulatory loop via the newly evolved connection from *gene* 18 to *gene* 3. In this way the period 6 cycle generate by the 6 *gene* loop drives the expresision of *gene* 13 — which is coupled to receptor 3 — in the surrounding cells. This was tested by making a knock-out mutation. The Boolean function of *gene* 3 was mutated to `ALL 0`. In this way, one could say that the expression *ligand* 3 is prohibited, preventing cell cell communication. In figure 5.11(a) the resulting phenotype is shown. It closely resembles the morphotype of epoch 1. The two new cell types have disappeared.

5.3 Fixing obtained information

In epoch 3, the information coming in from receptor 3 is used more intensively. A new connection from *gene* 23 to *gene* 18 has closed a new cycle of four genes, generating a new period 8 cycle. In association with the period 6 cycle a new 24 cycle is generated in the non-*bicoid* cell line, after the second division, if two non-*bicoid* cells are present. The period 24 cycle needs two non-*bicoid* cells because an important connection of the period 4 cycle is running through the cell-cell communication canal generated by *gene*-receptor pair 3.

The *bicoid* cells iterate through a period 6 state cycle, which is "phase locked" by *gene* 16 of the period 6 state cycle, being turned on 2 out of 6 states. The *bicoid* cells, finally, touching the period 24 non-*bicoid* cells, cycle around in a succession of four period 6 cell types. The generation of new cycles and the addition of alternative routes to existing cycles turn out to be important mechanisms for the evolution of new cell types.

In epoch 3, *gene* 20 is coupled to receptor 2. *Gene* 2 is part of the period 6 cycle. This new connection results in an extra diversification between the four cell types in epoch 2.

At the phenotypic level, the closure of the period 4 cycle has led to a new mechanism of differentiation. In beast 258 some of the gray cells diffuse into the blue, *bicoid* descended part of the organism. In beast 508 the same thing happens. However, here the diffusing cells differentiate into a new type whenever they loose contact to the non-*bicoid* cells.

In order to test whether the closed 4 cycle could account for this mechanism, the connection from *gene* 13 to *gene* 23 was blocked. This was done by changing the Boolean function of *gene* 23 from `!XOR` to `ALL 0`. By this mutation, the connection 13 $\rightarrow$ 23 is blocked, while *gene* 18 still propagates the signals coming in from *gene* 16.

In figure 5.11(b) the resulting phenotype is shown. It resembles the morphotype of epoch 2. The cells from the "mesoderm" layer between the *bicoid* and the non-*bicoid* layer do not differentiate if they diffuse into the *bicoid* part of the organism. A second change in epoch 3 is the connection of *gene* 20 to receptor 2. In contrast to the mutation described above, this change does not result in a "key innovation" at the phenotypic level. It strengthens, however, the diversificating effect of the mechanism of epoch 2. Using this extra receptor *gene*, both *gene* 13 and *gene* 20 diversify at the border of the *bicoid* and the non-*bicoid* lineage. The mutation 20:A$\rightarrow$`ALL 0` (figure 5.11(c)) results in a fitness drop from 4.75 to 3.25.

Two fundamentally different types of mutations were described in this paragraph. The first one, the mutation that formed a cyclic interaction betweeen genes, results in a "morphological" change. The second type of mutation does not affect any cyclic interaction. Instead, it recruits a

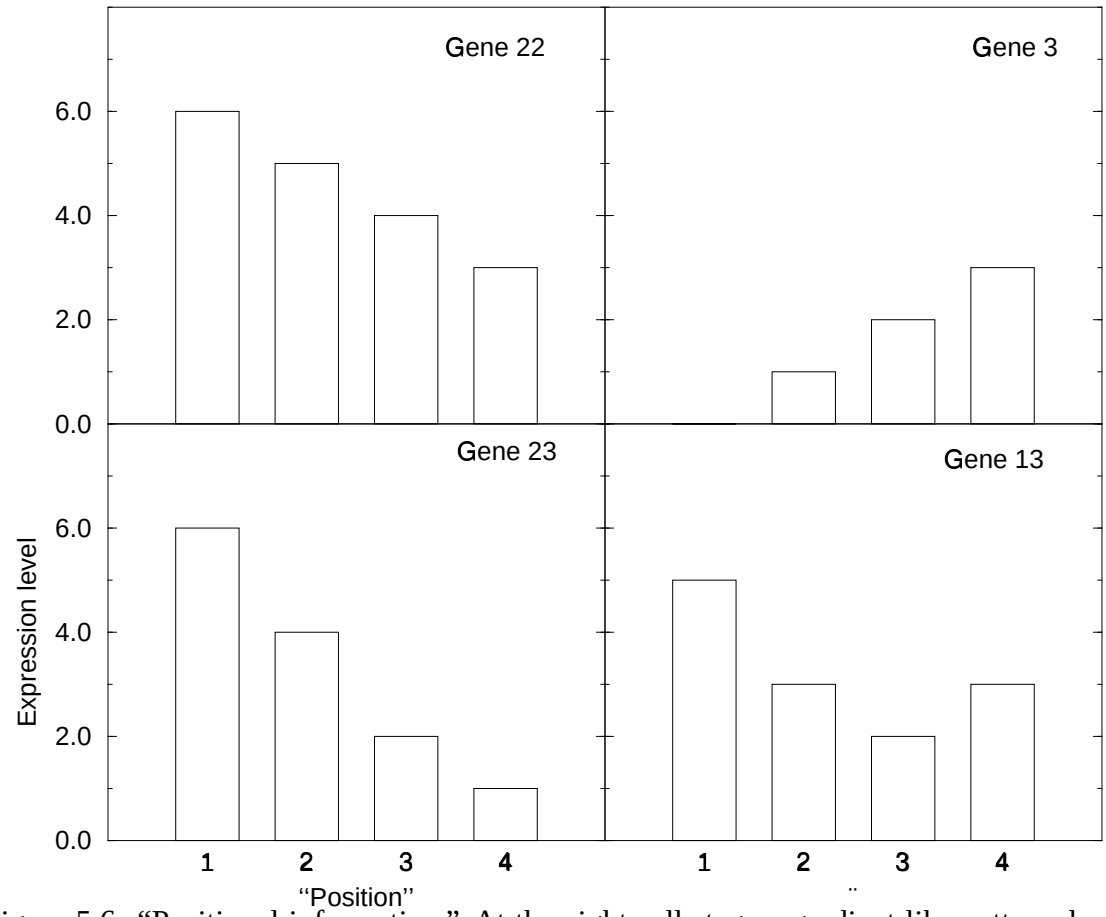


Figure 5.6: "Positional information." At the eight cell stage a gradient like pattern develops in *metazoan* 751 (shown here), but also in the *metazoans* until epoch 8. The pattern consists of four "bands", number from 1 (bicoid side) until 4 (non bicoid side). The expression level indicate the number of steps of the period 6 cycle the *gene* is expressed.

second gene in the differentiation process, thus generating a fitness rise.

5.4 The evolution of positional information

Interestingly, a simple form of positional information has evolved in epoch 4. In the 8 cell stage a gradient like pattern in the expression of *gene* 3, 13, 22 and 23 develops. The expression of these *gene*s is plotted in figure 5.6. It is not yet completely understood how this pattern is formed. However, it was possible to identify the *key mutation* changing morphotype 3 into morphotype 4.

In morphotype 4 a second connection from the 6 *gene* regulatory loop onto the 4 *gene* regulatory loop has evolved. This connection seems to "phase lock" the 4 *gene* regulatory loop together

with the already existing connection, resulting in a set of four different period 6 cycles.

In order to test whether both connections from the four *gene* and the six *gene* cycle were needed for the formation of the four band pattern, the following knock out experiment was performed. The connection from *gene* 16 to *gene* 18 was blocked by changing *gene* 18's `OR` into `A`. The phenotype resulting from this knock out experiment is shown in figure 5.11(d). It clearly shows that the interaction 16→18 is important for the early patterning of morphotype 4. Moreover, this knock out experiment clearly shows the role of interaction 16→18 in phase locking the period 8 cycle generated by the *gene*s 18, 3, 13 and 23. In the knock out phenotype the period 8 cycle and the period 6 cycle are completely decoupled resulting in a period 24 cycle.

After the four "bands" of *gene* expression have developed in the remaining part of the development of beast 751, the "bands" diffuse into each other like in epoch 4. During this process many additional cell types are generated. The importance of this process is easily shown in a simple experiment. If cell movement is prohibited, simply by omitting the CA steps during the developmental process, none of these additional cell types are produced, resulting in a fitness drop from 11.51 to 7.73.

It is interesting to speculate that selection may have favoured the phenotype of beast 751 because the fitness function turns out to select for stable phenotypes. In contrast to the fitness of morphotype 4, the fitness of the beasts in the preceding epoch is not stable all the time. The period 24 cycle in B cell line induces a sequence of four period 6 cycles in the NB cells. After the cell "bands" have diffused into each other, once in every four time steps two "bands" having a different cell type in the other steps, now have the same type resulting in a temporal diversity drop. Because the fitness function is defined as the lowest cellular diversity over the last one hundred time steps of a beast's life a considerable penalty is given to this type of behaviour.

5.5 "Epiboly": NB cells engulf B cells and induce a new cell type

A very interesting innovation can be observed in epoch 7. Like in epoch 6, a pattern of four cell types develops in the eight cell stage. In contrast to epoch 6 however, these cell types have a period 8 cycle instead of a period 6 cycle.

As is shown in figure 5.7 one of the cell types (the black one) of the NB-lineage engulfs the orange cells of the B-lineage. The mechanism of this engulfment is easily understood. As it was discussed in section 5.1, in beast 66, the NB-lineage engulfs the B-lineage, because (i) the B-lineage adheres less strongly to the medium than the NB-lineage does, and (ii) the B-lineage adheres more strongly to the NB-lineage than to the medium (see table 5.4).

In epoch 7, this difference in "hydrophoby" between the NB and the B-lineage has increased considerably. Whereas the mean difference in the hydrophoby between the B-cells and the NB-cells was 3.7 in beast 66 (see table 5.3), in beast 1389 this difference has grown to 9.625 (table 5.4). The stronger engulfment resulting from this increased adhesion difference results in the following mechanism of cell diversification.

In figure 5.7 one of the black NB cells has almost lost contact to the other black cells. As

cycle phase	J[B][med]	J[NB][med]	J[B][NB]
1	25	26	4
2	23	5	3
3	23	23	4
4	26	4	3
5	23	24	4
6	23	6	3
7	25	25	4
8	25	3	3

Table 5.4: Strength of the energy bonds between the different lineages of beast 1389. *B* denotes the cells from the *bicoid*-induced embryonal cell, *NB* denotes the cells descending from the other embryonal cell, not induced by *bicoid.*

soon as it has crawled so far over the orange B cells that is completely disattached from the black cells, it differentiates. As a result, the NB cell differentiates, inducing a new cell type in the underlying B cells (figure 5.8). This behaviour reminds of the mechanism of neural induction in amphibian development like it was first shown in the newt *Ambystoma mexicanum* [38]. In this process neural tissue is induced in ectodermal tissue by a mesodermal tissue layer covering it at the internal side.

Using this mechanism additional cell types are created by keeping the pattern in a transient state. If a NB cell disattaches and differentiates the surrounding cells keep on switching between a number of cell types before they have reached a stable pattern like that in figure 5.7. Especially during late development the NB cells keep on attaching and disattaching the other NB cells keeping the creature in the transient, diverse state.

5.6 More “tickling”: pattern destabilisation

In epoch 8 the “positional information” mechanism is disrupted. As a result the pattern of cell types becomes less stable. Due to a mutation in the old 6 *gene* regulatory loop, a new 6 *gene* regulatory loop is formed, running through one of the shortcuts evolved in epoch 7 (see table 5.5). This new 6 *gene* loop appears to be unable to generate a stable banding pattern. The “embryonal” pattern that is formed resembles the pattern of beast 508 (epoch 3): a period 8 cycle in the NB cell and the B cells touching the NB cells. These cells in turn induce a periodic sequence of a stable state and a period 2 cycle in the remaining B cells.

In contrast to what we would expect — a severe drop in cellular diversity — extra diversity is produced in collaboration with the epiboly mechanism described in chapter 5.5.

It was discussed above how epiboly of the NB cells is able to keep the pattern in a continuous transient towards a stable cell type pattern. The epibolic NB cell engulfs the B cells, disattaches from the other NB cells, differentiates, induces its surrounding cells, reattaches to another NB cell and dedifferentiates.

In beast 1655 (epoch 8), this “tickling” mechanism is able to generate a much higher diversity.

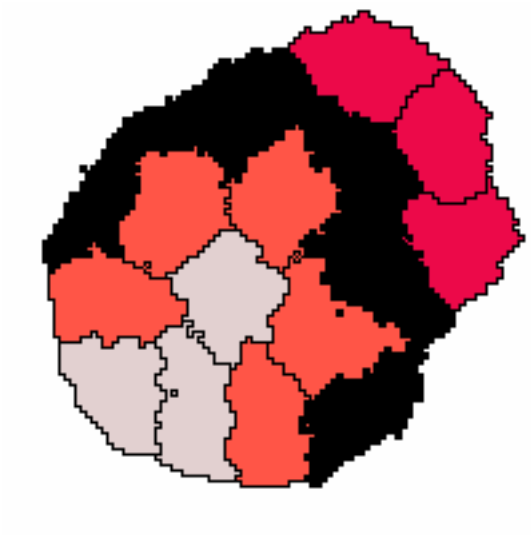


Figure 5.7: The black NB cells engulf the orange B cells, resembling "epiboly", expansion of ectodermal precursors over the embryo, taking place during the gastrulation of many metazoans (reviewed in (Gilbert, 1991)). The lower right black cell is about to disattach from the other black cells. It will differentiate and induce a number of new cell types (figure 5.8).

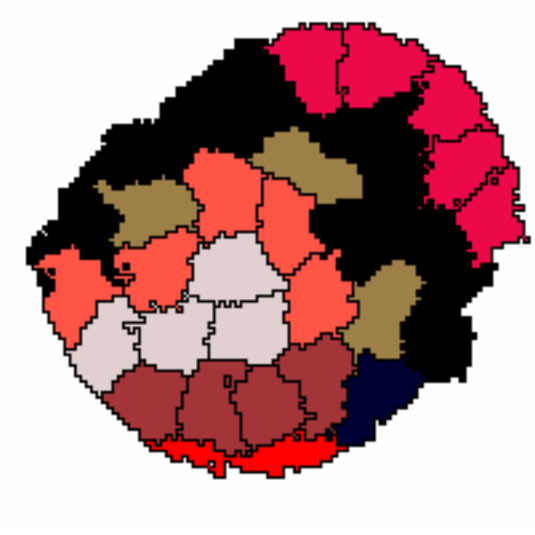
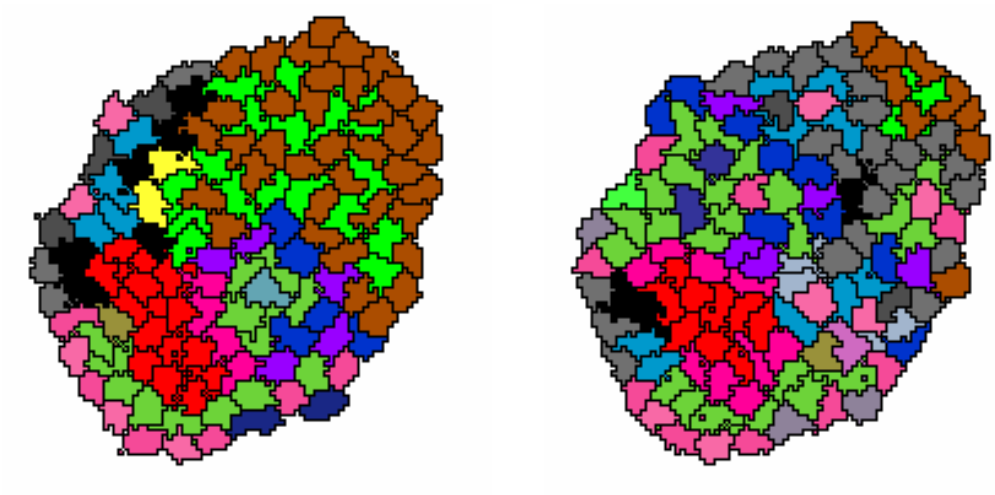


Figure 5.8: "Induction" in beast 1389: the lower red cell is a differentiated black cell, disattached from the other black cells. It has induced the four brown cells.



(a) (b)

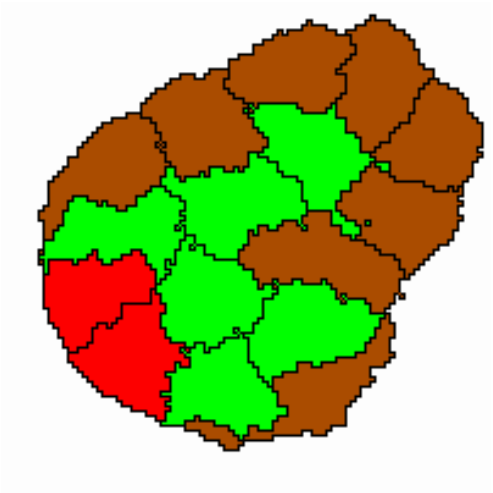
Figure 5.9: A second “tickling” mechanism can be observed in beast 1655. The green cells in 5.9(a) are derived from the *bicoid* cell and have mixed with the brown non-*bicoid* cells. In 5.9(a) the green cells all touch another green cell. In 5.9(b) some of the green cells have lost contact to the the other green cells. As a result a higly percolating pattern of other cell types is induced.

Because of the decreased stability, the “tickling” results in a pattern destabilisation, percolating into a considerable part of the organism (see figure 5.9(b)).

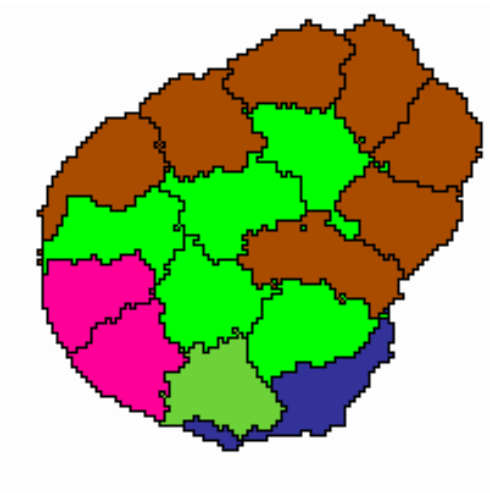
Whereas in epoch 3 the lack of stability led to a periodic diversity drop, causing a decreased fitness, here a comparable lack of stability results in an increased cellular diversity as long as the epibolic NB cells keep on “tickling” the B cells.

A second “tickling”-like mechanism appears later in the ontogeny of beast 1655. Most apparently after the last division, a number of B cells mixes with the NB cells (see figure 5.9(a)). If, as a result of the cell mixing process, one of the B cells looses the contact to another B cell, it differentiates, and, as a result, a new cell type is induced in the surrounding NB cells (figure 5.9(b).

These diversifying mechanisms highly depend on the cells’ ability to move. In figure 5.10, one of the brown surface cells attaches to a red cell. As a result, the surrounding cells differentiate.

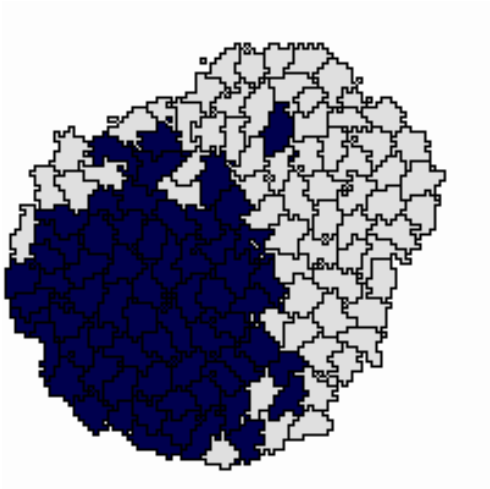


(a) One of the brown surface cells is about to loose contact to the other brown cells at time step 1999

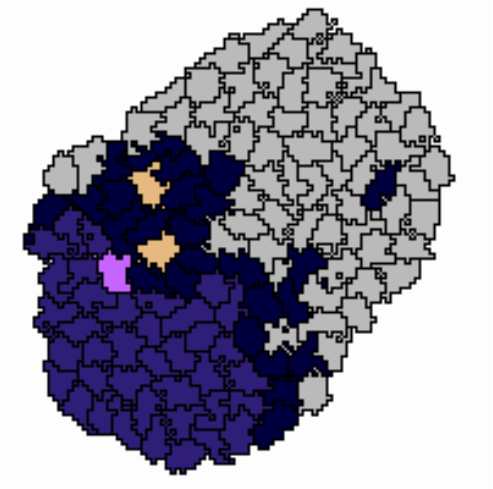


(b) At time step 2000 the red cells, the surface brown cell and one of the green cells has differentiated.

Figure 5.10: “Tickling mechanism” at an earlier stage of the development.

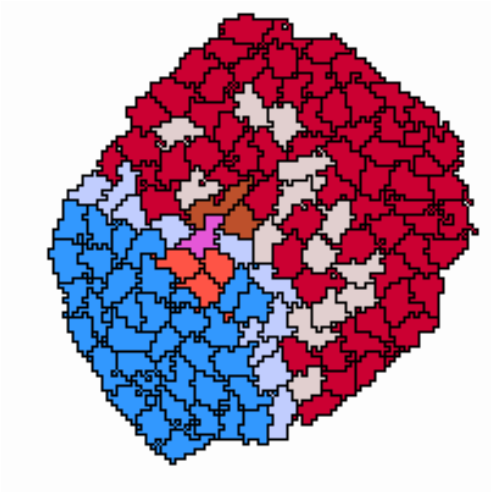


(a) A knock out phenome of beast 258. Cell cell communication via *receptor 3* was prevented by a substitution of the Boolean function = of gene 3 for ALL0. One could understand this as a knock out of the expression of *ligand* 3. The resulting phenome lacks the two newly evolved cell types of epoch 3

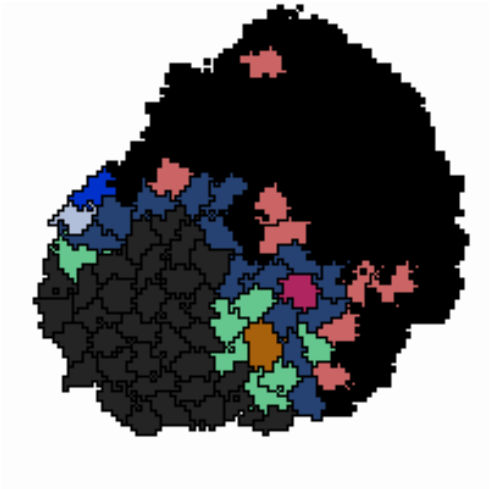


(b) A knock out phenome of beast 508. The four *gene* regulatory loop running through *receptor* 3 was knocked out by blocking connection: 13 to 23. The resulting phenome resembles the phenome of epoch 2. The black non-*bicoid* cells do not differentiate if they diffuse into the *bicoid* part of the organism.

Figure 5.11: A zoo of knock out phenomes. In most of the knock-out's shown here the genomic "key innovation" (see table 5.5) was knocked out, resulting in a regression to the morphotype of an earlier epoch.



(c) A second knock out phenome of beast 508. The mutation 20:A to ALLO blocks the connection from receptor 3 to *gene* 20, resulting in a fitness drop from 4.75 to 3.25. Morphotypically the *metazoan* is identical to the "wildtype". The connection of *gene* 20 to *receptor 3* is an example of a small scale evolutionary change, just affecting the cell types from which a morphotype is "build".



(d) A knock out phenome of beast 751. The new connection 16 to 18 from the four *gene* regulatory loop to the six *gene* loop plays a role in the formation of a "positional" banding pattern of four cell types in the eight cell stage. This is shown by the knock-out mutant 18:OR to A. The resulting phenome closely resembles morphotype 3.

Figure 5.11: continued

Table 5.5: Evolutionary innovations

Evolutionary innovations		
Epoch	**Phenomic innovations**	**Genomic innovations**
1	Zygotic cell polarity is used. Two cell lines arise. The cells derived from the bicoid induced cell (B) and the cells from the non-bicoid induced the cell (NB).	Regulatory loop of 6 *gene*s containing the bicoid bit
2	Two new cell types have arisen. Four cell types are present now: B (bicoid lineage) cells not touching NB cells, B cells touching B cells. NB cells touching NB cells and NB that do not touch an NB cell.	*Gene* 18 is coupled to receptor 3.
3	Temporal sequence of 4 period 6 cycles in the B cells is driven by a period 24 cycle in the NB cells. The differentiated B cells differentiate into a new cell type if they diffuse into the other B cells and, as a result, disattach from the NB cells.	*Gene* 18 is coupled to *gene* 24, in this way closing the new four *gene* regulatory loop 3$\longrightarrow$[2] 13$\longrightarrow$ 23$\longrightarrow$ 18$\longrightarrow$ 3.
4	"Positional information" has evolved. An embryonal banding pattern at the eight cell stage consisting of period six state cycles has evolved, resulting in a more stable diversity (see text).	Second connection from the six *gene* regulatory loop onto the four *gene* regulatory loop seems to "phase lock" the cycle generated by the four *gene* regulatory loop. As a result, the cells are in a period six cycle in stead of a period 24 cycle.
5A	(i) NB cells are slightly better engulfed by the B cells. (ii) Mechanisms of epoch 4 are still used, but produce increased diversity.	(i) New connection 2$\Longrightarrow$ 20 puts the two *gene*s 8 and 17 under control of receptor 2. (ii) Connection 16$\longrightarrow$ 18 is substituted by 11$\longrightarrow$ 18.
5B	In the adult stage five cell type "bands" are present at the border between the B and the NB cells. The pattern is preserved even if the cell shift in position relative to each other.	(i) New connection 20$\longrightarrow$ 11 puts *gene* 11 under control of receptor 2. (ii) New connection 6$\Longrightarrow$ 5 generates new cell-cell interaction controlling *gene*s 5 and 14.

continued on next page

[2] $\Longrightarrow$ indicates a receptor-ligand interaction

continued from previous page

Epoch	**Phenomic innovations**	**Genomic innovations**
6	(i) Morphotype 5A "won" over morphotype 5B in epoch 5. Using another mechanism, the "five band pattern" like that of morphotype 5B is again generated. The "band remixing" mechanism of epoch 4 is superimposed on this "five band pattern. (ii) Increased engulfment of NB-cells by B cells.	(i) The mutation 2$\Longrightarrow$ 20 to 2$\Longrightarrow$ 8 puts the six *gene* regulatory loop under control of receptor 2. Additionally an alternative route for the six *gene* regulatory loop is generated: 2$\Longrightarrow$ 8 $\longrightarrow$ 15. (ii) The new connection 8$\longrightarrow$ 15 adds a second alternative route to the six *gene* regulatory loop: 24$\longrightarrow$ 23 $\longrightarrow$ 8 $\longrightarrow$ 15. Moreover, this connection puts the six *gene* loop under control of the four *gene* loop.
7	(i) engulfment of B cells by NB cells is dramatically improved. (fig) As a result NB cells keep on "tickling" the outer B cells that did not the NB cells before, resulting in continuous destabilisation of the pattern. (ii) Embryonic "banding pattern" of period 6 cycles changed in pattern of period 8 cycles.	(i) The two short cuts on the six *gene* loop (2$\Longrightarrow$ 8 $\longrightarrow$ 15 and 24 $\longrightarrow$ 23 $\longrightarrow$ 8 $\longrightarrow$ 15) have disappeared as a result of the mutation: 8 $\longrightarrow$ 15 to 19$\longrightarrow$ 15. (ii) Mutation 19:`ALL1` $\longrightarrow$ `0ON` generates two short cuts to the six *gene* loop: 16$\longrightarrow$ 19 $\longrightarrow$ 24 and 21 $\longrightarrow$ 19 $\longrightarrow$ 24.
8	6 gene regulatory loop has changed as a result of the broken connection 16$\longrightarrow$ 22. Connection was broken because of the mutation 16$\longrightarrow$ 22 : 12$\longrightarrow$ 22, causing *gene* 22 to be constantly turned on.	Four band "positional" pattern has disappeared and has been substituted for embryonic pattern of three cell types: period 8 cycle in the NB cells and in the B cells touching the NB cells, a periodic sequence of a period 1 cycle and a period 2 cycle in the remaining B cells. Cell types are formed as a result of the "tickling" mechanism. (i) The NB cells engulf the B cells. If one of the NB cells looses contact to the other NB cells it differentiates and induces a pattern destabilisation (figure 5.10(a) and figure 5.10(b)). (ii) B cells mix with NB cells. Isolated B cells induce an extra cell type in the NB cells (figure 5.9(a) and figure 5.9(b)).

Chapter 6

Discussion and conclusions

A paradigm system for the evolution of multicellular animals has been constructed.

The aim of the paradigm system was twofold. The first aim was to understand how the highly complex genotype-phenotype mapping generated by the development zygote to adult may effect evolutionary dynamics. The second aim was to "breed" hypotheses on how cellular diversity can be generated and maintained in multicellular animals.

6.1 Summary of the results

Morphotypic metastabiltity

The evolution of our *metazoans* showed very characteristic behaviour. Although there was a constant speed in the evolution at the genetic level, evolution at the fitness level and at the morphological level progresses stepwise.

Most importantly, most of the time a fitness change was correlated to a large scale morphological change, as if new "Bauplan" emerge. It is proposed to call this behaviour *morphotypic metastability*.

Networks are hierarchically structured

The analisis of the networks described in chapter 4 has led to the distiction of three types of *gene*s: *regulatory genes* — embedded within a regulatory loop — controlling the expression of *downstream genes*. Finally, *housekeeping genes* that are constantly expressed can be distinguished.

Key mutations are mutations in the regulatory part of the network

Key mutations, the mutations marking the transition from one epoch to another, are in general mutations in the regulatory part of the network.

6.2 Discussion and speculation on the results

Evolution of regulatory *gene*s and downstream *gene*s may have resulted in morphotypic metastability

In this section it is hypothesised that the evolved distinction between regulatory loops and downstream *gene*s may have resulted in the characteristic morphotypic metastability that was observed in our model.

In the *fossil record* of the evolutionary run, two types of phenotypic change can be observed. On the one hand, large scale phenotypic changes, *phenotypic innovations*, are apparent. Such a phenotypic change is in general caused by a fundamental change in the *metazoan*'s developmental pathway. In this report, I have focused on these large scale evolutionary changes. While observing the beasts within a single epoch, the second kind of evolutionary changes becomes apparent. Although the morphotype remains by definition unchanged within an epoch, the colours in which the morphotypes are "painted" change a number of times. Hence, although the "developmental program" of the *metazoan*s does not evolve within an epoch, the set of cell types "building" them keeps changing.

The analysis of the morphotypic changes in chapter 4 has shown that it is possible to find for each epoch a "key mutation" that was most probably responsible for the morphotypic change. The greater part of these "key mutations" were changes in one of the regulatory loops of the *genome*. In epoch 3, a new regulatory loop was formed, in epoch 4 this new regulatory loop is put under the control of the first regulatory loop, in epoch 6 the first *gene* loop is connected to a *receptor* (cell cell interaction) and two alternative routes to this loop are formed. In epoch 7 these two short cuts on the first regulatory loop have been substituted for two other short cuts. In epoch 8 the first *gene* loop has been remoulded.

Although this was not (yet) tested, it is highly probable that the small scale — within epoch — phenotypic changes were were driven by mutations in the "downstream" *gene*s and in the "housekeeping" *gene*s. For example, a mutation changing an `ALL0` function into an `ALL1` function would change the expression of one of the housekeeping *gene*s. Obviously however, the developmental program would remain identical because no regulatory *gene*s would be changed.

In this way one could imagine that the morphotypic metastability results from the evolved distinction between regulatory *gene*s and downstream *gene*s. Mutations in the downstream *gene*s and in the housekeeping *gene*s do not result very often in a fitness change, whereas a mutation in the regulatory part can result in a large scale morphotypic change.

The analysis of the genetic basis of development of the last decennia has shown that biological genetic networks are hierarchically structured. Relatively few regulatory genes control the expression of the downstream "structural" genes. It has already been hypothesised by [18] that the "punctuated" evolution often observed in the "real" fossil record may be the result of the hierarchical structure of biological regulatory networks. "We may find [...] that structural gene substitutions control most small scale, adaptive variation [...] while disruption of regulation lies behind most key innovations in macroevolution" [18].

6.3 Mutational robustness

In this section it is discussed how the evolved structural properties of the genetic networks may influence the mutational sensitivity of the *metazoans*. It is hypothesised that two structural properties of the networks may result in increased mutational stability. Firstly, the ratio between regulatory *gene*s and downstream/housekeeping *gene*s could be changed in the favour of the downstream *gene*s. Secondly, the number of interactions within a regulatory loop may be increased, in this way "spreading of the risk" of a mutation in the regulatory part of the network. The results available at the moment that support these ideas are discussed and additional experiments are suggested. Finally, a biological interpretation of this hypothesis is suggested.

Mutational stability as a result of the ratio between regulatory *gene*s and downstream/housekeeping *gene*s

It would be feasible that the mutational stability of a *metazoan* is increased by decreasing the amount of regulatory *gene*s. Imagine a fully connected *genome* in which there is no distinction between regulatory *gene*s and downstream/housekeeping *gene*s. In such a *genome*, all the *gene*s would be embedded in a regulatory loop. Further imagine that a highly sophisticated developmental process were implemented by this *genome*, resulting in a high fitness. Undoubtly, this *genome* would be very prone to mutation, since all the *gene*s and *gene* interaction would be part of one big complicated regulatory network.

However, if the same developmental programme would be implemented in only a few *gene*s, the other *gene*s would be "free" to be mutated.

Genetic redundancy: "Spreading the risk"

An almost contradictory strategy to increase the stability against mutations would be to increase the redundancy of the regulatory loops. Two events in the evolutionary run described in chapter 4 suggest that (i) this mechanism has evolved and (ii) that it increases the mutational stability. In epoch 6 two alternative routes to the six *gene* regulatory loop have evolved. The first one runs via a *receptor*, the second one runs via an internal *gene* interaction.

Already in the next epoch, these two connections have disappeared as a result of a single mutation. However, here two alternative short cuts in the six *gene* regulatory loop are present, both running via intracellular *gene* interactions. It still needs to be investigated whether one of the intermediate *metazoan*s possessed the four shortcuts at the same time. In epoch 8 a *gene* interaction that was conserved since epoch 1 has disappeared. However, the six *gene* loop is still present, because it now runs through one of the alternative routes evolved in epoch 7.

These preliminary data suggests that genetic redundancy has evolved in the *metazoan* evolution as described in chapter 4. However, certainly more analysis and a number of experiments need to be carried out in order to support this hypothesis.

First, more representatives of the epochs should be analysed. These analyses might answer the following questions: (i) Is the trend sketched above for the first eight epochs continued during

the rest of the evolutionary run? (ii) What happens during one evolutionary epoch? Are *genomes* at the onset of an epoch less redundant than later representatives of an epoch?

Second, I would propose a number of "bulk data" experiments. The first one would be to test the beasts in the evolutionary run on mutational stability. At regulatory intervals in evolutionary time, a *genome* could be extracted from the *fossil record*. Each of the *gene*s of this *genome* is than mutated a fixed number of times. A "mutational stability map" could be constructed. The series of mutational maps could show what happens to the mutational stability of the *metazoans* over evolutionary time and within one epoch. Together with the data obtained from the *genome* analysis these data could be specified with respect to regulatory *gene*s, downstream *gene*s and housekeeping *gene*s. It is expected that the mutational stability will increase presumably for the regulatory part of the network.

Additional data on the networks' structures could be obtained if the "network-stripping" procedure would be automated. Until now, in order to understand the network's structure, the housekeeping *gene*s were stripped off from the network manually. As this is a very time-consuming job only very sparse data could be obtained about the structure of the networks (see table 5.5).

The "stripping" procedure can be automated in the following way. First all the non-functional connections are removed, such as the B-connection of a function `A`. Then, the "obvious" housekeeping *gene*s `ALL1` and `ALL0` are searched. The connections fanning out from these functions are tracked down. A function such as `xB` with the A connection constantly turned on is set to 0. In this way a second level of housekeeping *gene*s has been found. The procedure is repeated until no more housekeeping *gene*s are found.

6.4 Caveats of the fitness criterion:

Optional division led to faster creatures

In a different version of the model the genetic networks could control whether they would divide or not. This was achieved as following: at each "division signal" only the cells divided that had their so called division bit set.

Initially to our astonishment repeatedly the metazoan cells evolved the habit of only dividing only a limited numner of times. The division bit was involved in the state cycle such that it was only turned on once in a few timesteps.

It is quite easy to see why this strategy evolves. In the evolutionary simulations a beast is able to reproduce as soon as it has finished its development. In the developmental simulations beasts with fewer cells consume less computer time. Hence, we were confronted with some "good old" evolutionary dynamics: faster reproduction leads to more offspring.

Chapter 7

Future work

7.1 Buggy “atol” suggests to include chemotaxis

On errors and ideas

In some cases the most interesting scientific results are produced by stupid mistakes. Famous is the example of the slobby Alexander Fleming who discovered the antibiotic penicillin on one of the dirty petridishes he had left rotting in the sink for several weeks.

In the construction of the paradigm system of *metazoan* development the most spectacular example of *metazoan* development appeared to be produced by a logical error in the algorithm determining the value of the energy bond between two cells having a different state vector.

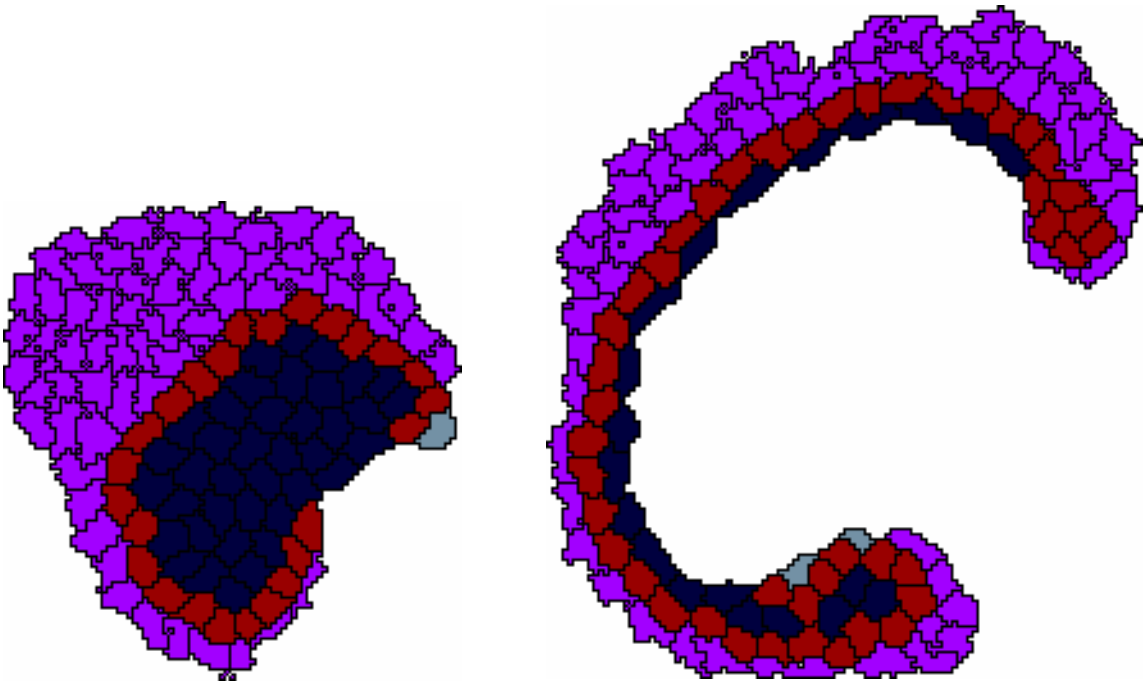
The bug

The phenome of this organism is shown in figure 7.1. Interestingly, it shows considerable morphogenetic movements resulting in an oblong shape, and this was all — according to our ideas at that moment — produced by energy minimisation.

The logical error that produced this interesting phenotype was the following. The algorithm computing the energy bond between two cells, takes the state vector of both cells, mirrors one of them and takes the AND of these two vectors. The resulting vector, the “match vector”, represents the matching “key-lock” pairs. The more key-lock pairs, the higher the energy bond. The logical error was that the right hand five bits where not mirrored. Instead, the bits were ORed pairwise, the 10th with the 9th, the 8th with the 7th, etcetera. It is easy to see now (but not before) that the consequence of this procedure is that the outcome depends on whether the state vector of the first cell or the state vector of the second cell is mirrored.

In the present model, the logical OR of the right hand five bits and the mirrored left hand five bits gives the energy bond between two cells.

During every cellular automata (CA) update only the needed energy bond values were computed. These were stored in a symmetrical matrix, so that the energy bond between two cells needed to be calculated only once a time step. Since it mattered whether the bond between A and B or the bond between B and A was calculated, in some CA updates the higher value was



(a) Age: 7000 time steps (128 cells)

(b) Age: 19000 time steps (128 cells)

Figure 7.1: The oblong shape of this organism was ascribed to energy minimization. However, it appeared to be due to a bug, that inadvertly introduced chemotaxis-like processes in the model.

used and in some CA updates the lower value was used. This would mean that the interesting morphology of the "atol" was produced by "trembling" energy bonds. Experiments however in which the energy bonds were chosen at random between the higher and the lower value, showed that in this case the oblong morphology completely disappeared.

The solution to the problem appeared to be quite simple. If two sites are picked up from the CA plane, and it is tested whether the first site's state should be copied in the second site, the energy bond between the two cells to which the sites belong, is calculated if it is unknown.

As a result, every first copuing step between two cells is biased. If $J_{A,B} = 20$ and $J_{B,A} = 5$ a first copy from cell B into A will always use the lower value, and vice versa. As a result, B is copied slightly more often into A than A is copied into B.

Interestingly, as a result of this bug, chemotaxis has been introduced. In this case cell B is attracted by cell A.

The idea: "chemotaxis"

After the first disappointment that the "atol" morphology was a result of the bug described above, Paulien Hogeweg came up with the following idea: include chemotaxis in the paradigm system.

Chemotaxis is nicely modelled in the Glazier and Graner model [17] using the extension constructed by [37]. In this extension, cells have a slightly higher copying probability in the direction of the gradient and a slightly lower copying probability against the gradient. The gradient is modelled in a PDE layer.

It is proposed here that the cells should be able to produce a metabolite. This metabolie is "excreted" by the cells and "diffuses" in the PDE layer. At the same time, cells are "allowed" to evolve the possibility to use this metabolite as a chemoattractant. They may also use it as a signalling molecule. In this way diffusing signals are included in the model as well.

The cells may produce chemoattractants and react to them in two ways.

As a first try, the metabolites will be excreted if one of the bits of the Boolean network is set. The cells will be attracted to the metabolites whenever another bit, the "receptor", is set.

If the results of these experiments are promising, it might be a good idea to proceed to a more dramatical change of the model. In this idea the work of professor Kaneko [27] on isologous diversification will be combined with the work presented in this paper. In this work the genetic networks of cells are modelled as metabolic continuous networks. As a first try, some metabolites may diffuse into the medium and from the medium into other cells through a selective membrane. This will allow the cells to communicate using diffusive signals. Later, the cells will be attracted along metabolite gradients from other cells, again only if the cells possess a "receptor" metabolite.

In this paradigm system, it will be not only possible to study how patterning and cellular diversification processes are guided by the genome and how they evolve, but also how the development of complicated morphologies is directed by the information stored in the genome.

Bibliography

[1] ALBERTS, B., BRAY, D., LEWIS, J., RAFF, M., ROBERTS, K., AND WATSON, J. D. *Molecular Biology of the Cell*, 4 ed. Garland Publishing, 1994.

[2] ARMSTRONG, P. Cell sorting out: the self-assembly of tissues in vitro. *CRC Critical Reviews of Biochem. Mol. Biol. 24* (1989), 119–149.

[3] ASASHIMA, M., AND ARIIZUMI, T. Control of the embryonic body plan by activin during amphibian development. *Zoological Science 12*, 5 (1995), 509–521.

[4] CHIVA, E., AND TARROUX, P. Studying genotype-phenotype interactions: a model of the evolution of the cell regulation network.

[5] CHIVA, E., AND TARROUX, P. Evolution of biological regulation networks under complex environmental constraints. *Biological Cybernetics 73* (1995), 323–333.

[6] DELLAERT, F. Towards a biologically defensible model of development. Master's thesis, Case Western Reserve University, January 1995.

[7] DELLAERT, F., AND BEER, R. D. Toward an evolvable model of development for autonomous agent synthesis. In *Artificial Life IV*, R. Brooks and P. Maas, Eds. MIT Press, Cambridge, 1994, pp. 246–257.

[8] DRIEVER, W., AND NÜSSLEIN-VOLHARD, C. The bicoid protein determines position in the *drosophila* embryo in a concentration dependent manner. *Cell 54* (1988), 95–104.

[9] DRIEVER, W., AND NÜSSLEIN-VOLHARD, C. Determination of spatial domains of zygotic expression in the *drosophila* embryo by the affinity of binding sites for the bicoid morphogen. *Nature 340* (1989), 363–367.

[10] EIGEN, M. Macromolecular evolution: dynamical ordering in sequence space. In *Emerging Synthesis in Science* (Reading MA, 1985), D. Peins, Ed., vol. I of *Santa Fe Institute Studies in the Sciences of Complexity*, Addison Wesley, pp. 25–69.

[11] FAGOTTO, F., AND GUMBINER, B. M. Cell contact-dependent signalling. *Developmental Biology 180* (1996), 445–454.

[12] FINK, R., AND MCCLAY, D. Three cell recognition changes accompany the ingression of sea urchin primary mesenchyme cells. *Developmental Biology 107* (1985), 66–74.

[13] FONTANA, W., KONINGS, D. A., STADLER, P. F., AND SCHUSTER, P. Statistics of RNA secondary structures. *Biopolymers 33* (1993), 1389–1404.

[14] GILBERT, S. F. *Developmental Biology*. Sinauer Associates, 1991.

[15] GILBERT, S. F. *Developmental Biology*, fourth edition ed. Sinnauer Associates, Sunderland, Massachusetts, 1994.

[16] GIUDICE, G. Restitution of whole larvae from disaggregated sea urchin embryos. *Developmental Biology 5* (1962), 402–411.

[17] GLAZIER, J., AND GRANER, F. Simulation of the differential adhesion driven rearrangement of biological cells. *Physical Review E 47* (1993), 2128–2154.

[18] GOULD, S. J. Is a new and general theory of evolution emerging? *Paleobiology 6*, 1 (1980), 119–130.

[19] HOGEWEG, P., AND HESPER, B. Energy directed folding of RNA sequences. *Nucleic Acids Research 12*, 1 (1984), 67–74.

[20] HOLLAND, J. H. *Adaptation in natural and artificial systems*. University of Michigan Press, Ann Arbor, MI, 1975.

[21] HUYNEN, M., DANIËLLE KONINGS, AND HOGEWEG, P. Multiple coding and the evolutionary properties of RNA secondary structure. *Journal of Theoretical Biology*.

[22] HUYNEN, M., AND HOGEWEG, P. Selection for change; evolution of the potential to evolve in a predator-prey model.

[23] HUYNEN, M., AND HOGEWEG, P. Pattern generation in molecular evolution; exploitation of the variation in rna landscapes. *Journal of Molecular Evolution* (1993).

[24] JACKSON, E. R., JOHNSON, D., AND NASH, W. G. Gene networks in development. *Journal of Theoretical Biology 119* (1986), 379–396.

[25] JACOB, F., AND MONOD, J. Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology 3* (1961), 318–356.

[26] JACOB, F., AND MONOD, J. On the regulation of gene activity. In *Cold Spring Harbor symposia on Quantitative Biology* (1961), vol. 26, pp. 193–211.

[27] KANEKO, K., AND YOMO, T. Isologous diversification: a theory of cell differentiation. *Bulletin of Mathematical Biology 59* (1997), 139–196.

[28] KAUFFMAN, S. A. Metabolic stability and epigenesis in randomly connected nets. *Journal of Theoretical Biology 22* (1969), 437.

[29] KAUFFMAN, S. A. *The origins of order*. Oxford University Press, 1993.

[30] KO, E. P., YOMO, T., AND URABE, I. Dynamic clustering of bacterial population. *Physica D 75* (1994), 81–88.

[31] KOZA, J. R. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. MIT Press, Cambridge, MA, 1992.

[32] MAHNER, M., AND KARY, M. What exactly are genomes, genotypes and phenotypes? and what about phenomes? *Journal of Theoretical Biology 186* (1997), 55–63.

[33] MAYNARD SMITH, J. *Evolution and the theory of games*. Cambridge University Press, Cambridge, 1982.

[34] MCCLAY, D., AND ETTENSOHN, C. Cell recognition during sea urchin gastrulation. In *Genetic Regulation in Development* (New York, 1987), W. F. Loomis, Ed., Alan R. Liss, pp. 111–128.

[35] NARDI, J., AND STOCUM, D. Surface properties of regenerating limb cells: Evidence for gradation along the proximodistal axis. *Differentiation 25* (1983), 27–31.

[36] SAVILL, N. J., AND HOGEWEG, P. A cellular automata approach for modeling cell-cell interactions. In *Dynamics of Cell and Tissue Motion, in press* (Basel, 1997), A. Alt, A. Deutsch, and G. Dunn, Eds., Birkhäuser Verlag.

[37] SAVILL, N. J., AND HOGEWEG, P. Modelling morphogenesis: from single cells to crawling slugs. *Journal of theoretical biology 184* (February 1997), 229–235.

[38] SPEMANN, H., AND MANGOLD, H. Über induktion von embryonalanalagen durch implantation artfremder organisatoren. *Archiv für mikroskopische Anatomie und Entwicklungsmechanik 100* (1924), 599–638.

[39] STEINBERG, M. S. Adhesion in development: an historical overview. *Developmental Biology 180* (1996), 377–388.

[40] TOWNES, P. L., AND HOLTFRETER, J. Directed movements and selective adhesion of emryonic amphibian cells. *Journal of experimental zoology 128* (1955), 53–120.

[41] VAN NIMWEGEN, E., CRUTCHFIELD, J. P., AND MITCHELL, M. Finite populations induce metastability in evolutionary search. *Physics letters A 229* (1997), 144–150.