The Evolution of Fractal Protein Modules in Multicellular Development

Harry Booth[1] and Peter J Bentley[1]

[1] Department of Computer Science, University College London
Gower Street, London WC1E 6BT, UK
harry.booth.19@ucl.ac.uk, p.bentley@cs.ucl.ac.uk

Abstract

Regional specification, or pattern formation, is the process by which developing cells in different regions are switched into different developmental pathways. We investigate this process through an ALife model of multicellular development using fractal proteins, where genes are expressed into proteins comprised of subsets of the Mandelbrot Set. The resulting network of gene and protein interactions can be designed by evolution to produce specific patterns, that in turn can be used to solve problems. Here fractal gene regulatory networks are incorporated into a multicellular model of development, and tested on the morphological problem of regional specification, using Map-Elites to explore the space of solutions. The results indicate the ability of this system to learn regularities in solutions and automatically create and use developmental modules, illustrating how an artificial system can replicate some of the fundamental processes of development.

Introduction

The phenotypes of organisms throughout the biological world are incredibly complex. As an example, a fully formed adult is made up of approximately 10^{13} cells, which form an all manner of specialised tissues and organs. However, the complete instruction set for the development of a human being is found in every cell, and contains a relatively small amount of information compared to the complexity of the phenotype previously described. There are approximately 3^9 base pairs within the human genome (Brown, 2002), and since there are 4 types of nucleotide bases (adenine, cytosine, guanine and thymine) the identity of each base pair can be specified using only 2 bits. This means the entire genome contains around 750 megabytes of data required for development - for reference, this is slightly larger than an audio CD. The crucial components in this journey from genotype to phenotype are *development processes*, which are coordinated through the expression of genes in specific temporal and spatial patterns (Levine and Davidson, 2005). There is growing evidence that many of these patterns are highly preserved and recombined during evolution as developmental modules (Lacquaniti et al., 2013).

The inclusion of development processes into artificial evolutionary systems has many advantages (Eggenberger et al., 1997). This work describes an artificial model of development, in which fractal proteins (Bentley, 2003b) are used for the first time to both regulate gene expression and determine other developmental parameters within a spatially extended multicellular environment. The system is tested on a fundamental morphological problem known as regional specification, with the aim of understanding how evolution is able to coordinate temporal and spatial gene expression in a multicellular assembly, paying specific attention to the evolution of developmental modules in the gene regulatory networks evolved by Map-Elites.

Background

The mapping from genotype to phenotype in natural evolution is a development growth process (Bowers, 2005). The study of such artificial mappings is sometimes referred to as computational embryology (CE) (Kumar and Bentley, 2003). Cell chemistry approaches to CE attempt to mimic how physical structures emerge in biology (Stanley and Miikkulainen, 2003). Here, interacting systems inspired by nature generate complex and indirect development programs which result in an emergent phenotype (Bentley et al., 1999). Often these systems consist of multicellular assemblies, where the regulation and subsequent expression of certain genes within each cell can trigger events such as mitosis, cell differentiation and apoptosis (Eggenberger et al., 1997). A fundamental process in the controlled formation and development of plant and animal phenotypes is regional specification, also known as pattern formation (Kumar and Bentley, 2003). For a multicellular assembly, the problem is one of self-organization - cells must differentiate in the correct spatial locations, with no explicit knowledge of their position within the final assembly (Friston et al., 2015). The essence of spatial patterning has been distilled into a metaphor which was first described by Wolpert (1969) and which subsequently became known as the French Flag (FF) Problem. Due to its self-contained nature and clear problem statement, the FF Problem has become a test bed for artificial development systems.

Some of the earliest work in testing an artificial develop-

ment system on the FF Problem was by Miller (2004), who used Cartesian Genetic Programming to evolve a program which controlled the actions of cells on a grid. This organism was able to successfully grow into a FF pattern, and in addition the development program was robust to damage applied to the developing phenotype. Chavoya and Duthen (2008) used a gene regulatory network (GRN) based on a bit-string representation (Banzhaf, 2003) within a cellular automata (CA) framework. Specific regulatory proteins determined the activation of structural genes, which in turn led to the adoption of an associated CA look-up table for cell reproduction. Knabe et al. (2008) integrated a GRN into the Cellular Potts computational model of cells and tissues. Protein concentrations determined individual cell parameters such as size, shape, adhesion, morphogen secretion and orientation. Evolved organisms were able to autonomously set up an asymmetric morphogen gradient and ultimately organize into a close match of the FF pattern. Joachimczak and Wróbel (2009) extended the the FF Problem into 3D by designing a development system consisting of spherical cells within a simulated physics environment.

This work introduces a system of artificial development which incorporates fractal proteins. Fractal proteins were first introduced by Bentley (2003b), and are an example of an artificial chemistry. Fundamental to the interaction of fractal proteins with any system is their shape (Bentley, 2009). This shape is genetically specified by a triple (x, y, w), which points to a finite region of the Mandelbrot set. As well as affording a compact representation, the geometry of the Mandelbrot set makes the 'fractal genetic space' highly evolvable, due to its continuity and self-similarity. For tasks such as gene regulation - which involves the interaction of fractal proteins and the genome - the shape of fractal proteins have been exploited very successfully to produce desirable concentration dynamics (Bentley, 2004b). In addition, gene regulation with fractal proteins has displayed desirable properties such as graceful degradation (Bentley, 2004a) and the emergence of modules - sub-routines which are reused and used with minor modifications to build more complex solutions (Bentley, 2003a). Modularity is a key feature of biological systems, and is seen as a desirable feature of evolutionary design. Consequently, there have been several studies examining modularity in simulated evolution. Garibay et al. (2003) introduced the modular genetic algorithm (MGA), which was explicitly designed to exploit regularity within the problem space using modularity. The MGA was found to both outperform a standard genetic algorithm and scale better for increasing problem complexity. Pollack and Lowell (2016) investigated the concept of hierarchical modularity. They found that an evolved gene regulatory network (GRN) displayed greater modularity than a neural network, due to the use of developmental encodings.

Methodology

This section describes an artificial development system which combines fractal proteins with a Cellular Potts model of cells. This system is referred to as Fractal-Potts. As in other works connecting gene regulatory networks with multicellular models of development, cellular behaviours are derived from regulated concentration dynamics to construct a development program. However, Fractal-Potts additionally exploits the underlying spatial nature of fractal proteins to define other interactions within the multicellular environment, thereby increasing the number of evolvable developmental parameters.

Fractal proteins and gene regulation

Fractal proteins have the following attributes, which determine their interactions and behaviour within the context of various systems.

- *Shape* - The shape of a fractal protein is represented by a subset of the Mandelbrot set
- *Environment concentration* - Fractal proteins exist in varying concentrations within an environment, such as a cell cytoplasm or an extra-cellular medium
- *Functional role(s)* - Fractal proteins can: regulate gene expression (F_r-proteins); be used within the cell for behavioural and structural purposes (F_b-proteins); move outside of the cell (F_e-proteins); behave as receptors for extracellular signals (F_s-proteins)
- *Chemical meta-properties* - Protein degradation rates specify the rate at which their concentration decreases, in absence of their production within an environment

The regulation of gene expression within a cell is achieved by the production of appropriate transcription factors. Transcription factors are formed by 'merging' the fractal proteins found within a cells cytoplasm at each developmental time step. This creates a single fractal protein through a kind of 'fractal chemistry', which considers both the shape and concentration of each reactant protein in determining the final product. These transcription factors then interact with the genome, activating specific genes. Each gene within the genome has the following attributes, which determine their interaction with transcription factors and the subsequent regulatory dynamics:

- *Promoter region* - This specifies a subset of the Mandelbrot set.
- *Affinity threshold* - This along with the promoter region creates a precondition for gene expression, by specifying a minimum shape similarity that must exist between the promoter region and transcription factor for gene activation

- *Coding region* - This specifies the shape of the fractal protein produced given gene activation

- *Gene type* - This specifies the functional role(s) of the fractal protein associated with the coding region

- *Transcription meta-properties* - Gene activation results in the concentration of the coding protein increasing. The amount produced is dependent on the concentration of the transcription factor, a concentration threshold specific to each gene, and a number of global parameters (i.e. constant across all genes).

For full details of the regulatory dynamics, interested readers should consult Bentley (2004a).

This work introduces an additional fractal protein attribute. It is a real valued parameter, and is encoded within each gene for the associated coding protein. It has different purposes, depending on the proteins functional role:

- *Regulatory (F_r) and environmental (F_e) proteins* - Specifies a minimum concentration, below which the protein is not considered part of the 'merge' operation for the determination of transcription factors. Whilst Bentley (2003b) does specify a minimum concentration below which a fractal protein is no longer considered present in the cell, it is not specific to each protein and furthermore is not evolved. This is likely to make the transcription factors seen through development increasingly varied and dynamic.

- *Behavioural (F_b) proteins* - Allows the re-interpretation of continuous protein concentrations as Boolean values, for all-or-nothing structural changes or decision-like behaviours. This is referred to as a 'switching threshold'.

This additional attribute can be thought of as a function (and therefore context) dependent expansion of the chemical meta-properties associated with general fractal proteins.

Multicellular fractal proteins

The Cellular Potts (CP) model (also known as the Glazier-Graner-Hogeweg model) is a computational model of cells and tissues. Introduced by Graner and Glazier (1992), it was used to simulate the sorting of a mixture of two types of biological cells. The fundamental components of the basic CP model are a lattice $L \subset \mathbb{Z}^2$ representing a spatial environment and two functions, the cell identity function $\sigma : L \to \{1, \cdots, N_{\text{cells}}\}$ and the cell type function $\tau : \{1, \cdots, N_{\text{cells}}\} \to \{1, \cdots, N_{\text{types}}\}$. In addition, there is a matrix containing the differential surface energies between cell types, $J \in \mathbb{R}^{N_{\text{types}} \times N_{\text{types}}}$. A hamiltonian $H(t) : L \to \mathbb{R}$ describes the energy of a given configuration of cells on the lattice:

$$H(t) = \sum_{\substack{(i,j),(i',j') \\ \text{neighbors}}} J\Big[\tau(\sigma(i,j)), \tau(\sigma(i',j'))\Big]\Big[1 - \delta_{\sigma(i,j),\sigma(i',j')}\Big] + \lambda \sum_{\substack{\text{cells} \\ k}} \Big[a_k(t) - A_k(t)\Big]^2 \mathbb{1}(A_k(t) > 0) \tag{1}$$

Here $a_k(t)$ is the size of cell k, and $A_k(t)$ is its target size at time t. λ is a constant which determines the elasticity of the cell wall. The dynamics and evolution of the system can be simulated through a metropolis-style update (Graner and Glazier, 1992). Fractal proteins are introduced quite naturally to this model - at each point $(i,j) \in L$ in the environment and at a given time t there is a concentration of each fractal protein (which can of course be zero). The concentration of fractal protein F_a at site $(i,j) \in L$ at a time t is denoted $C_{i,j}(F_a, t)$. At each development time step, each of the N_{cells} completes its own gene regulation cycle independently based on the concentrations of proteins found within its interior. The total concentration of protein F_a found within cell $k \in \{1, \cdots, N_{\text{cells}}\}$ at time t is

$$C_k(F_a, t) = \sum_{(i,j) \in \text{cyto}_k} C_{i,j}(F_a, t) \tag{2}$$

where $\text{cyto}_k = \{(i,j) \in L \text{ s.t. } \sigma(i,j) = k\}$. These aggregated concentrations determine the set of fractal proteins which react to form the transcription factor at each development time step. If a gene whose coding region specifies F_a is activated, resultant concentration updates $\Delta C_k(F_a, t)$ are distributed across the cells cytoplasm uniformly - for each $(i,j) \in \text{cyto}_k \subset L$,

$$C_{i,j}(F_a, t) = C_{i,j}(F_a, t) + \frac{\Delta C_k(F_a, t)}{|\text{cyto}_k|} \tag{3}$$

Environmental proteins have the ability to move freely throughout the multicellular environment. This is achieved by introducing a diffusion process on the lattice L. Given an environmental protein F_e a diffusion term is added to the usual update rule, which then gives the following: For each $(i,j) \in \text{cyto}_k \subset L$,

$$C_{i,j}(F_e, t) = C_{i,j}(F_e, t) + \frac{\Delta C_k(F_e, t)}{|\text{cyto}_k|} + \alpha_e \nabla_L^2 C_{i,j}(F_e, t) \tag{4}$$

where ∇_L^2 is the appropriate lattice Laplacian and $\alpha_e \in \mathbb{R}$ is a diffusion constant specific to each environmental protein. The complete dynamics of environmental proteins can therefore be considered as a reaction-diffusion system.

Cellular behaviours

In this work, both protein concentrations and their fractal shape determine their role within development mech-

anisms. A key tool is the fractal protein similarity measure, which is defined as follows for two $n \times n$ proteins $F_1, F_2 \in [0, \cdots, 255]^2$:

$$S(F_1, F_2) = 1 - \frac{1}{255 \times n \times n} \sum_{i,j \in n \times n} |F_1[i,j] - F_2[i,j]| \quad (5)$$

Note that $0 \leq S(F_1, F_2) \leq 1$ for any two fractal proteins.

Inter-cellular communication Environmental proteins can be absorbed into one cells cytoplasm after being produced within the cytoplasm of a neighbour through diffusion, establishing a form of inter-cellular communication. In this work, diffusion speed is derived from the shape of the diffusing protein itself. We model a 'cellular substrate' as the diffusing medium. This medium contains a specific fractal protein F_{sub} in which all entries are zero, i.e. $F_{\text{sub}}[i,j] = 0$. The speed α_e at which an environmental protein F_e diffuses through the cellular substrate is then assumed to be its similarity in shape to F_{sub}:

$$\alpha_e = \beta_{\text{diff}} \cdot S(F_e, F_{\text{sub}}) \quad (6)$$

where β_{diff} is a user set parameter that ensures the numerical stability of Eq. 4.

Mitosis and apoptosis Cell size scales with the concentration of a specific behavioural protein F_g within the cell. Computationally this is achieved by linking the target cell size in Eq. 1 with this concentration: $A_k(t) = \beta_{\text{size}} \cdot C_k(F_g, t)$, where β_{size} is a user-set parameter based on lattice size. Mitosis occurs when $a_k > 0.25 * \beta_{\text{size}}$ - the cell is split into two, perpendicular to the longest axis associated with its 2D grid shape. After such an event, the concentration of all proteins are split randomly between the two cells. In this work, a cell can only undergo mitosis once. Apoptosis occurs when the concentration $C_k(F_g, t)$ falls below a user set threshold β_{death} - all proteins within the cell are destroyed, except from environmental proteins which remain.

Cell type differentiation Cell type is represented by one of a finite number of pre-determined colours - red, green, blue or white. Differentiation is interpreted to be an all-or-nothing event. The approach taken is to re-interpret continuous protein concentrations as Boolean values using the switching threshold values specified in the genome for each coding protein. Using a binary mapping, the cellular concentrations of two behavioural proteins can specify each of the four types, as shown in Table 1.

Inter-cellular adhesion The importance of the interplay between cell type and inter-cellular adhesion in producing complex morphologies has been demonstrated by Hogeweg (2000). In the CP model, cellular adhesion is determined energetically - in Eq. 1, the matrix $J \in \mathbb{R}^{N_{\text{types}} \times N_{\text{types}}}$ specifies the surface energy between each possible pairing of the

State	$C(F_{b1}, t) > T_{b1}$	$C(F_{b1}, t) \leq T_{b1}$
$C(F_{b2}, t) > T_{b2}$	red	green
$C(F_{b2}, t) \leq T_{b2}$	blue	white

Table 1: Specification of cell type through behavioural protein concentrations

N_{types} cell types via an appropriate matrix entry. In addition, an energy is specified between each cell type and the extra-cellular medium, which in this work is made equal to 1. Similar cell types are expected to adhere strongly to each other, and hence cells of the same type are not energetically penalised (the diagonal of J is zero). In this work the approach is to specify the surface energies between different cell types through a cellular fractal chemistry - the matrix J is constructed by associating a new fractal protein with each distinct cell type. The proteins associated with the four cell types are derived from the two behavioral proteins responsible for cell type determination. This is done by considering four possible reactions between the proteins, as shown in Table 2.

State	$C(F_{b1}, t) > T_{b1}$	$C(F_{b1}, t) \leq T_{b1}$
$C(F_{b2}, t) > T_{b2}$	$R_{\hat{F}} = \{F_{b1}, F_{b2}\}$	$R_{\hat{F}} = \{F_{b2}\}$
$C(F_{b2}, t) \leq T_{b2}$	$R_{\hat{F}} = \{F_{b1}\}$	$R_{\hat{F}} = \emptyset$

Table 2: Protein metabolism products associated with type-defining behavioural proteins

These four combinations of reactants $R_{\hat{F}}$ define four protein products through the same fractal chemistry used to determine transcription factors within the cell. In this work, the empty set of reactants is associated with the substrate protein F_{sub}. The entries of the matrix J are then calculated as one-minus the similarity measure between the different pairs of products. For example, given the products P_{red} and P_{blue}

$$J[1,3] = J[\text{red}, \text{blue}] = 1 - S(P_{\text{red}}, P_{\text{blue}}) \quad (7)$$

Chemotaxis Chemotaxis can be incorporated into the CP model by introducing a chemokine which diffuses across the lattice, and increasing the likelihood that a lattice site $(i,j) \in L$ will be changed to its neighbour $(i',j') \in L$ if the chemokine concentration is higher at (i,j). In our model an environmental protein F_c acts as a chemokine. Following Savill and Hogeweg (1997), the modification to the energy in the metropolis update is therefore

$$\Delta H' = \Delta H - \mu\Big(C_{i,j}(F_c, t) - C_{i',j'}(F_c, t)\Big) \quad (8)$$

Receptor clustering is a metabolic process that results in grouping of a set of receptors at a cellular location, often

to amplify the sensitivity of a signaling response. In some cellular systems, clusters form dynamically in response to activation by an extracellular ligand (Duke and Graham, 2009). In this model, clustering is imagined to occur when the chemokine binds to a specific receptor protein. The similarity between these proteins hence determines the degree of clustering, and as a result the baseline chemotaxis sensitivity μ:

$$\mu = \beta_{\text{chemo}} * S(R_{\text{prom}}, F_c) \tag{9}$$

Here β_{chemo} is a user-specified parameter which determines the maximum possible sensitivity - it is chosen in consideration with the other CP meta-parameters such as T and λ.

Developmental modules

In this work a module is a sub-routine of the development process, and is identified by a characteristic pattern of gene expression occurring at a number of points throughout development. Within the Fractal-Potts model, the activation of a gene at a specific point in development occurs when a suitable transcription factor is created within the cell. In order to be activated, the fractal similarity between the genes promoter region and the transcription factor must reach a threshold. The full set of unique transcription factors created through the cells history therefore characterises the full set of gene activation patterns used throughout development. In the case of the multicellular model, these unique transcription factors can be collected for each cell and combined into a super-set, which then characterises the full set of gene activation patterns for the complete development of the organism.

Evolutionary strategies

In this work the MAP-Elites algorithm (Mouret and Clune, 2015) is used for evolutionary search. The MAP-Elites algorithm is designed to deliver a large set of diverse, high-performing individuals, embedded in an archive that describes where they are located in the phenotype space. This archive is a collection of 'bins', a discretization of the phenotype space using some dimension of variation which is of interest. In this work, the selected dimension of variation is final colour composition. In the Fractal-Potts model there are four possible cell colours - red, green, blue and white. This therefore defines an archive of size $2^4 - 1 = 15$ (minus one since we do not search for organisms with no colour - i.e. dead). In order to create offspring from fit individuals the genetic operators described by Bentley (2003b) are used. Crossover allows for variable sized genomes, with appropriate mutation operators responsible for enlarging or shrinking the genomes within a population. Individuals develop for T_{dev} time steps. To improve robustness, development is run 20 times with different random seeds. From this set of trials the most common phenotype classification is identified, which then becomes the archive bin to which it is assigned. From the set of trials corresponding to this assignment, the average fitness is taken. Details of the fitness function are given in the next section.

Experiments

The aim of the experiments completed in this work were twofold. The first aim was to investigate the use of fractal proteins in coordinating temporal and spatial gene expression within a multicellular development system, and determine whether the system inherited any desirable properties such as stability and robustness. The second aim was to identify evidence of module formation in the regulatory networks evolved for the development task.

To this end, Fractal-Potts was tested on an expanded formulation of the FF Problem, referred to in this work as the General Flag (GF) Problem. The GF Problem arises naturally from the novelty search algorithm and involves the search for organisms which can develop flag-like morphologies. This abstraction of the development target creates a related set of problems, and therefore it might be expected that evolution preserves specific developmental modules. This was investigated in the second part of the experiment, where a cross-sectional study across evolved organisms was completed to search for evidence of module formation.

Fitness evaluation In the GF Problem, an n-colour flag-like morphology has n equally sized cellular regions of homogeneous cell type (colour) which are organised sequentially from left to right. The MAP-Elites archive bins only consider which colours exist in the final phenotype, and not the order in which they are sequenced across the grid. The fitness function is therefore designed to be flexible to alternative orderings - the pattern B-W-R (blue,white,red) would be assigned the same fitness as the pattern R-B-W within the [W,B,R] archive bin. This is achieved by assessing the number of pixel matches given each possible colour ordering, and selecting the one which provides the maximum fitness. Hence for n-colour phenotypes there are $n!$ assessments. Fitness is ultimately measured as the percentage of correct pixel matches with what would be considered the perfect flag.

Measuring stability and robustness Evolved organisms are assessed for stability by running the development time for twice that used in evolution. Organisms are assessed for robustness by randomly perturbing the internal concentration levels of a subset of cells at a random point through development. Each cellular perturbation is applied as a random increase or decrease in its protein concentrations, the magnitude of which can be up to 10% of the original. In each trial, 30% of cells are randomly targeted. In both stability and robustness, results are reported as a percentage of the original fitness.

Cross-sectional study of transcription factors The cross-sectional study of unique transcription factors is completed across all of the evolved organisms. The fittest organism from each of the 15 flag species is selected and developed. Throughout development the unique set of transcription factors is then collected, as well as information relating to each transcription factor's time of appearance and its duration within the organism.

Experimental setup Evolution ran for 500 generations. Evolutionary runs were seeded with a population of 10,000 genomes, each of which were initialized with 35 genes - 1 receptor, 3 behavioural, 6 regulatory and 1 environmental, with the remaining assigned random functional roles. Organisms were allowed to develop for 800 time steps, starting from a single cell with concentrations set at zero. Protein size was 15x15. For determining transcription dynamics the following parameters were used: $C_t = 0.2$, $C_s = 0.1$, $C_w = -0.1$, $C_p = 20$, $C_i = 1.5$. For more details see Bentley (2003b). The CP lattice size was 30x30, with temperature $T = 1$ and cell elasticity $\lambda = 10000$. For the cellular behaviours, $\beta_{\text{diff}} = 0.25$, $\beta_{\text{size}} = 100$, $\beta_{\text{death}} = 0.05$, $\beta_{\text{chemo}} = 5000$.

Results

In the experiments all of the possible colour combinations were discovered in a single evolutionary run, resulting in single, bi, tri and quad-colour flags. Organisms of high fitness were obtained, with excellent degrees of regional colour specification. In addition some phenotypes demonstrated stability when allowed to develop for a longer amount of time, with the average phenotype maintaining 80% of its original fitness. The fitness of some however did deteriorate quite significantly. Some organisms additionally demonstrated robustness, being able to recover from random chemical perturbations applied through development and still obtain high fitness solutions. The full set of quantitative assessments is shown in Table 3, along with examples of developed phenotypes in Fig. 1.

In the cross-sectional study of transcription factors, a total of 199 unique transcription factors were collected through development across the 15 species. Interestingly but perhaps expected, the distribution of this total was not uniform across the different flag species. Single colour organisms had an average of 9 unique transcription factors created through development. Bi-colour and Tri-colour organisms on average had 19 and 24 respectively, with the quad-colour organism demonstrating 21. This is suggestive that the number of transcription factors observed is related to the complexity of the development task. The full history of unique transcription factors for each species is shown in Fig. 2. The colour-bar indicates the time at which the transcription factor first appears in the organism. A number of the transcription factors (36) were found to be present in multiple organisms at some point through their development. This set of common transcription factors is shown in Fig. 3. It was found that every organism had at least one transcription factor which was identifiable in another organism at some point.

Species	Average fitness (50 trials)	Fitness standard deviation	Average stability (50 trials)	Average robustness (50 trials)
d-R	100%	0%	100%	99.7%
d-B	100%	2.6%	99.9%	99.6%
d-G	100%	0%	100%	99.3%
d-W	100%	0.7%	98.3%	99.3%
d-BR	86.6%	1.8%	98.3%	83.3%
d-RG	92.5%	6.2%	56.6%	71.5%
d-GB	67%	3.6%	74.2%	82.1%
d-WR	75%	6.1%	66.8%	78.2%
d-BW	92.1%	10.8%	72.5%	61.6%
d-GW	92.2%	5.6%	99.9%	91%
d-RGB	72.4%	2.5%	61%	71.8%
d-BWR	77.3%	7.4%	62.2%	68.4%
d-RGW	60.3%	7.4%	59.2%	76.2%
d-GWB	76.7%	2.4%	78.3%	91.4%
d-RGWB	46.9%	3.7%	92.3%	83.3%

Table 3: Fitness, Stability and Robustness of fittest organism per species. The species notation indicates the final colours and orderings, e.g. d-RGB develops red-green-blue, patterned left to right.

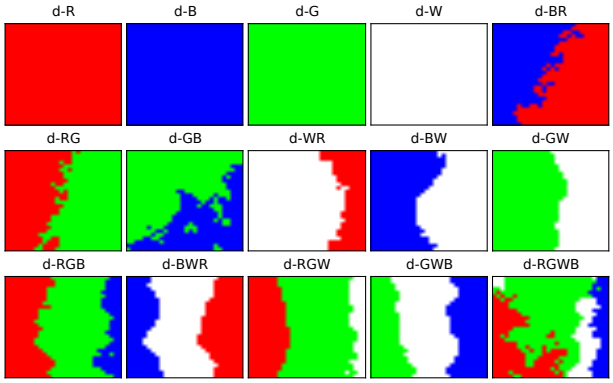

Figure 1: Example development of the fittest evolved organisms

Discussion

The first part of the results shows that Fractal-Potts is well specified, and can achieve success on the classic problem of flag development. In addition some evolved organisms demonstrate desirable properties, such as robustness to environmental perturbations and the ability to maintain stable phenotypes. MAP-Elites was beneficial in the search for flag-like organisms. In particular no penalization terms were added to the fitness function in order to encourage organisms early on in evolution to use all colours - the novelty search naturally resulted in this behaviour.

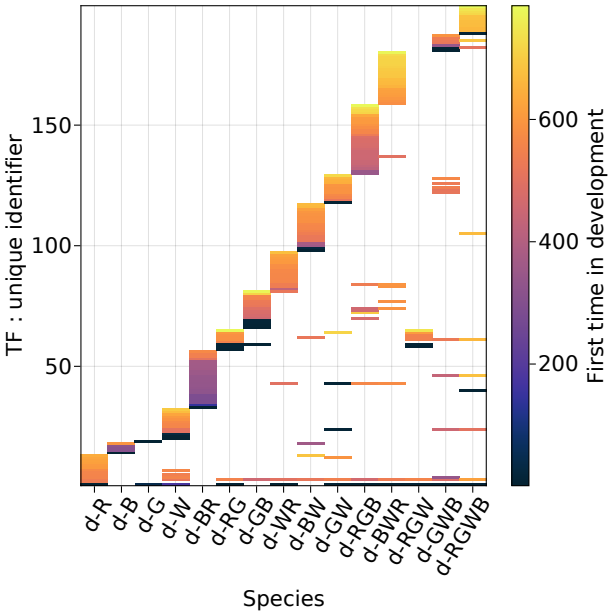

Figure 2: Transcription factors (TF) seen across organisms through development

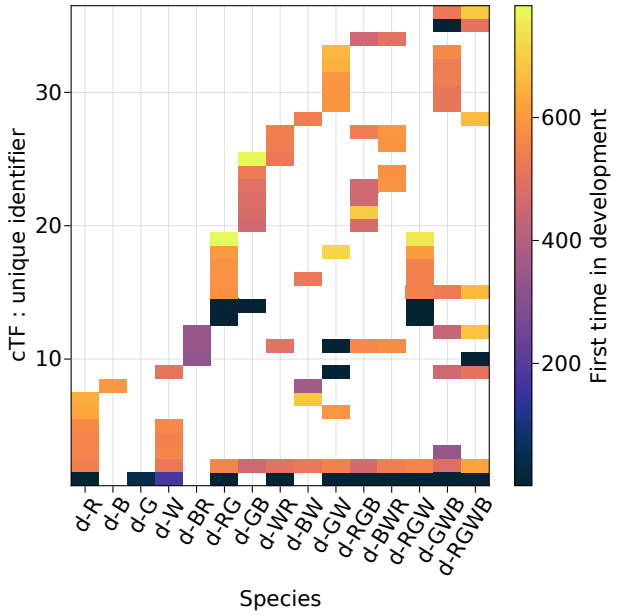

Figure 3: Common transcription factors (cTF) seen across organisms through development

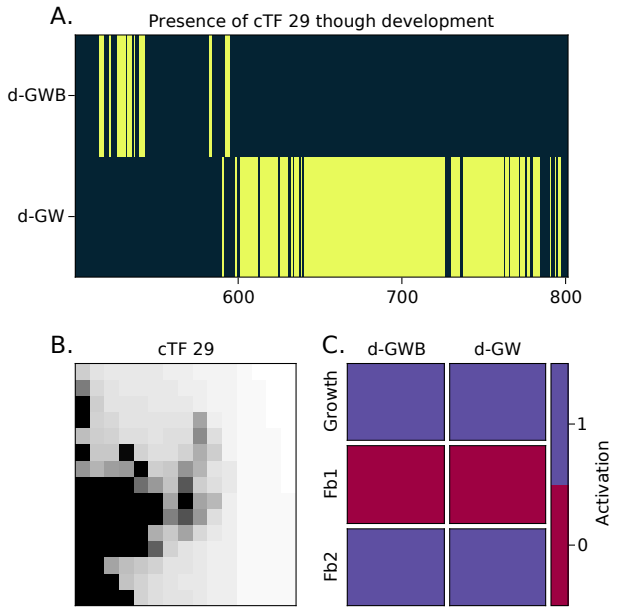

Figure 4: Presence of a common transcription factor (cTF-29) and the corresponding behavioural gene activation pattern in two different species

The existence of common transcription factors across different organisms suggests that evolution has discovered and re-used patterns of gene activation, both within an organism's development (module re-use) and across different organisms (module inheritance) despite the selective pressure of MAP-Elites to find novel phenotypes. Evidence to support this hypothesis can be found by analysing the effects of individual transcription factors. For example, Fig. 4 illustrates one common transcription factor cTF 29. In Fig. 4.B the fractal shape of cTF 29 is displayed, and in Fig. 4.C the corresponding activation pattern of the 3 behavioral genes responsible for cell growth and colour differentiation is shown. It can be seen that the activation pattern is identical for the two organisms d-GWB and d-GW. Examination of Table 1 indicates that increased levels of the F_{b2} protein has a positive logical association with the transition to red or green cell types, depending on the concentration of F_{b1}. By observing the development process associated with both d-GW and d-GWB, it was confirmed that the occurrence of cTF 29 corresponds with the proliferation of green cells, in agreement with the activation pattern. In Fig. 4.A the presence (yellow bars) of cTF 29 is shown throughout development. The intermittency of the protein within both organisms indicates that the gene activation pattern occurs during different times through development, rather than continuously. Out of the 36 common transcription factors it was found that 17 displayed these characteristics, producing a common pattern of activation across the three behavioural genes in all organisms, but at different times. Of the 3! = 6 logical combinations of gene activation, 3 were present.

This is strong evidence that evolution preserved and reused developmental modules - common patterns of be-

havioural gene expression triggered by specific transcription factors - both across organisms and throughout the development process. It was also found that a number of the common transcription factors activated equivalent patterns amongst the set of regulatory genes shared across organisms. Whilst a full analysis is outside of the scope of this paper, this is an indication that evolution has additionally preserved important regulatory sub-routines.

Regional specification remains a relatively unexplored and rarely modeled process in ALife. Here we have shown that a combination of fractal proteins, Cellular Potts and MAP-Elites enables us to explore the space of developmental solutions and furthermore observe the emergence of biologically plausible phenomena, such as the learning of regularities in solutions and the automatic creation and reuse of modules. We anticipate that these methods will enable further insights into artificial developmental processes in the future.

References

Banzhaf, W. (2003). On the dynamics of an artificial regulatory network. In *European Conference on Artificial Life*, pages 217–227. Springer.

Bentley, P. J. (2003a). Evolving fractal gene regulatory networks for robot control. In *European Conference on Artificial Life*, pages 753–762. Springer.

Bentley, P. J. (2003b). Evolving fractal proteins. In *International Conference on Evolvable Systems*, pages 81–92. Springer.

Bentley, P. J. (2004a). Evolving beyond perfection: An investigation of the effects of long-term evolution on fractal gene regulatory networks. *Biosystems*, 76(1-3):291–301.

Bentley, P. J. (2004b). Fractal proteins. *Genetic Programming and Evolvable Machines*, 5(1):71–101.

Bentley, P. J. (2009). Methods for improving simulations of biological systems: systemic computation and fractal proteins. *Journal of The Royal Society Interface*, 6(suppl_4):S451–S466.

Bentley, P. J., Kumar, S., et al. (1999). Three ways to grow designs: A comparison of embryogenies for an evolutionary design problem. In *GECCO*, volume 99, pages 35–43.

Bowers, C. P. (2005). Formation of modules in a computational model of embryogeny. In *2005 IEEE Congress on Evolutionary Computation*, volume 1, pages 537–542. IEEE.

Brown, T. A. (2002). The human genome. In *Genomes. 2nd edition*. Wiley-Liss.

Chavoya, A. and Duthen, Y. (2008). A cell pattern generation model based on an extended artificial regulatory network. *Biosystems*, 94(1-2):95–101.

Duke, T. and Graham, I. (2009). Equilibrium mechanisms of receptor clustering. *Progress in biophysics and molecular biology*, 100(1-3):18–24.

Eggenberger, P. et al. (1997). Evolving morphologies of simulated 3d organisms based on differential gene expression. In *Proceedings of the fourth european conference on Artificial Life*, pages 205–213. Citeseer.

Friston, K., Levin, M., Sengupta, B., and Pezzulo, G. (2015). Knowing one's place: a free-energy approach to pattern regulation. *Journal of the Royal Society Interface*, 12(105):20141383.

Garibay, O. O., Garibay, I. I., and Wu, A. S. (2003). The modular genetic algorithm: Exploiting regularities in the problem space. In *International Symposium on Computer and Information Sciences*, pages 584–591. Springer.

Graner, F. and Glazier, J. A. (1992). Simulation of biological cell sorting using a two-dimensional extended potts model. *Physical review letters*, 69(13):2013.

Hogeweg, P. (2000). Evolving mechanisms of morphogenesis: on the interplay between differential adhesion and cell differentiation. *Journal of theoretical biology*, 203(4):317–333.

Joachimczak, M. and Wróbel, B. (2009). Evolution of the morphology and patterning of artificial embryos: scaling the tricolour problem to the third dimension. In *European Conference on Artificial Life*, pages 35–43. Springer.

Knabe, J., Schilstra, M., and Nehaniv, C. L. (2008). Evolution and morphogenesis of differentiated multicellular organisms: autonomously generated diffusion gradients for positional information. *Artificial Life XI*.

Kumar, S. and Bentley, P. J. (2003). Computational embryology: past, present and future. In *Advances in evolutionary computing*, pages 461–477. Springer.

Lacquaniti, F., Ivanenko, Y. P., d'Avella, A., Zelik, K., and Zago, M. (2013). Evolutionary and developmental modules. *Frontiers in Computational Neuroscience*, 7:61.

Levine, M. and Davidson, E. H. (2005). Gene regulatory networks for development. *Proceedings of the National Academy of Sciences*, 102(14):4936–4942.

Miller, J. F. (2004). Evolving a self-repairing, self-regulating, french flag organism. In *Genetic and Evolutionary Computation Conference*, pages 129–139. Springer.

Mouret, J.-B. and Clune, J. (2015). Illuminating search spaces by mapping elites. *arXiv preprint arXiv:1504.04909*.

Pollack, J. and Lowell, J. (2016). Developmental encodings promote the emergence of hierarchical modularity. In *ALIFE 2016, the Fifteenth International Conference on the Synthesis and Simulation of Living Systems*, pages 344–351. MIT Press.

Savill, N. J. and Hogeweg, P. (1997). Modelling morphogenesis: from single cells to crawling slugs. *Journal of theoretical biology*, 184(3):229–235.

Stanley, K. O. and Miikkulainen, R. (2003). A taxonomy for artificial embryogeny. *Artificial life*, 9(2):93–130.

Wolpert, L. (1969). Positional information and the spatial pattern of cellular differentiation. *Journal of theoretical biology*, 25(1):1–47.