Morphogenesis Software based on Epigenetic Code Concept

N. Bessonov[2], O.Butuzova[6], A.Minarskiy[7], R. Penner[1,4], C. Soule[1], A. Tosenberger[5] and N. Morozova[1,3,6]

[1] IHES, France; [2]IPME, St-Petersburg, Russia, [3]UMR9198 CNRS; [4] Caltech, USA; [5] ULB, Belgium; [6]BIN, St-Petersburg, Russia; [7]Academic University, St-Petersburg, Russia.

1. Introduction

The process of morphogenesis is an evolution of a shape of an organism together with the differentiation of its parts. The discovery of differential gene expression (the spatial–temporal distribution of the gene expression pattern during morphogenesis), together with its key regulators (such as Hox genes), is one of the main recent achievements in developmental biology. Nevertheless, solely differential gene expression cannot explain the development of the precise geometry of an organism and its parts in space.

It looks plausible to suggest the existence of an additional biological code (epigenetic code) which bears information about geometrical pattern of an organism and thus coordinates the cascades of molecular events implementing a pattern formation (e.g. differential gene expression, directed protein traffic, growth of cytoskeleton). Though the concrete signal transduction pathways connecting the morphogenetic coding information and expression of a given sets of genes are not yet elucidated, we can suggest a set of postulates and possible approaches for discovering the correspondence between this code and its realization in a given geometry of an organism in space–time.

2. Epigenetic Code Concept

The set of theoretical conjectures on the geometry of morphogenesis built on the hypothesis of an existence of epigenetic cell surface code is a further development of a work published in Morozova, Shubin, 2012, Morozova, Penner, 2015, Minarsky et al., 2017.

First we suggest as a model that a cell fate and, correspondingly, a final pattern of a multicellular object, is coded by a biological code located on cell surfaces in such a way, that with each cell can be associated a corresponding matrix/coding array, reflecting a 3D pattern of distribution of a set of coding molecules on cell surfaces.

Next we suggest that a set of rules (*developmental laws*) for converting this coded information into instructive signals for cell events for a cell, as well as for transforming the coding arrays after each type of cell event, may be common to all living organisms.

In this case, development of an organism depends on a coding array of its initial cell, its zygote. Next after each cell division, daughter cells inherit a part of a *coding array* of the mother cell, thus providing a basis for differential developmental paths of cells containing the same DNA content.

We provide a set of arguments why coding molecules should be located on the cell surface, and suggest a set of experiments for the confirmation of this concept. Concerning the molecular character of coding molecules, our prevailing **assumption** is that such code may be provided by a pattern formed by a set of several types of oligosaccharide residues of glycoconjugates (glycoproteins and glycolipids), some specific features of which make them plausible candidates. However, in the full context of the general model, we can also consider any type of cell surface markers.

Next we give notions of *cell state* which is determined by its *coding matrix,* and *cell event,* which is a change of *cell state,* and formalize development as a graph (tree) of cell states connected by 5 types of cell events, corresponding to the processes of cell division, cell growth, death, cell movement and cell differentiation. We show that such a developmental tree with exact parameters of cell events has one-to-one correspondence with an embryo morphology at each time slice.

Next, we assume that there is a universal rule R_i for changes in the amount and composition of cell surface markers for each type i of *cell event.*

We suggest a mathematical formalism suitable to decipher these rules (*developmental laws)* for converting the coded morphogenetic information into instructive signals for cell events for a cell, and a corresponding software which gives a tool for determination of cell events based on the distribution of epigenetic code and the rules of epigenetic code change following cell events.

Hypotheses and conjectures.

1. Existence of morphogenetic code, determining a geometrical outcome of developing organism, located on cell surfaces. The arguments why coding molecules should be located on the cell surface are:

- The points of a cell surface correspond to the geometrical structure in 3D space thus giving an intrinsic metric which can be used for recording a spatial information
- A cell surface location gives a possibility of different distribution of this information within a set of dividing cells of an organism, hence providing a diversity of cell potencies for further differentiation
- This location gives a possibility of a feedback, i.e., the instruction to stop, when the task for a proper shape (for a cell, morphological domain, or a whole organism) is fulfilled
- This location gives a possibility to be involved in the signal transduction pathways. For example, received outer signals go to the nucleus or Golgi apparatus and influence the expression of specific set of genes or the process of protein traffic
- This location gives a possibility to influence a direct cell-cell communication
- The cell surface of an ovule is inherited as well as its DNA content
- A set of experimental data confirms the significance of cell surface information for pattern formation

2. Cell fate may be coded by a biological code located on cell surfaces in such a way that with each cell can be associated a corresponding matrix/array, containing this code

3. *Cell events of a cell* and its parameters implementing embryo development are functions of its coding matrix

4. There is a universal rule R_i for changes in the amount and composition of surface markers for each type i of *cell event* (cell division, cell growth, death, cell movement), which may be the same for all living organisms

5. After each *cell division* event daughter cells inherit one a part of a *coding matrix* of the mother cell, while the remaining part is created according to a rule R_d

6. The set of rules (*developmental laws*) for converting a coded information into instructive signal for cell event for a cell, as well as for transforming the coding arrays after each cell event, may be common for all living organisms.

We propose a model implementing these ideas and hope to find the general rules of a surface code coordination of the pattern formation by mathematical formalization. We do not pretend for now to suggest any molecular mechanisms underlying this coordination.

3. Morphogenesis Software

3.1. Mathematical formalization and modeling

For checking this hypothesis by numerical simulations, we create a Morphogenesis Software, allowing to create a first cell (zygote) with a given coding matrix, which will develop into an embryo according to this set of rules (https://cloud.mail.ru/public/2Gor/wWHkC4ndE). The program has an advanced interface that allows to observe the process of growth of an an embryo from a zygote simultaneously with a corresponding developmental graph (Figure 1), and to show the matrix of any selected cell during simulations. The program starts with a matrix of the first cell of an organism (zygote), which can be varied by the (computational) experimentalist and includes a set of rules for embryo development depending on cell matrix information. By applying these rules to all cells at each time step, the program presents a developmental process as a programmed consequence of *Cell events*, occurring at each time interval, and compute and

displays both geometrical structure of a developing embryo and a corresponding graph (tree) of *Cell events* (Figure 1).

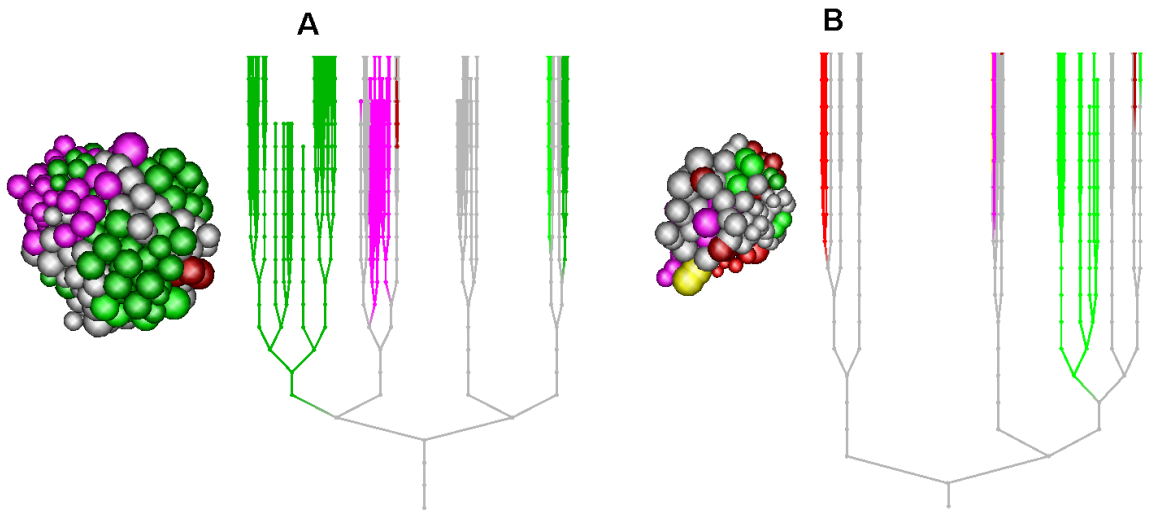


Figure 1. Developing embryo and corresponding Tree of Cell events. Different colors correspond to different types of differentiated cell.

This means that having a determined set of Rules, the main parameters of *Cell events* implementing embryo development are the functions of a code, located on a cell surface (a corresponding matrix, associated with each cell). We consider 5 types of *Cell events*, determined by the algorithm: cell division, cell growth (including changing a cell shape), movement, death and differentiation. Also the algorithm includes the rules for filling in new elements of a coding matrix after each type of cell event (one rule for each type of cell event).

We consider that a cell surface code can be written and transmitted in a form of a matrix $a_{¿}$, $i=1,\ldots,I$, $n=1,\ldots,N$ which has the following structure: N columns of the matrix $a_{¿}$ corresponds to N sectors on a cell surface, while each row i corresponds to one type of coding molecules I. An element of a matrix $a_{¿}$shows an amount of a given type of coding molecules i in a sector n of a cell surface, presented by an integer number.

Though it does not influence the calculations in a frame of a model for now, we would like to provide a suggestion for possible candidates for coding molecules. We think that it can be different types of oligosaccharides - short sugar residues of cell-surface glycoconjugates, i.e.,

proteins or lipids with sugar (glycol-) part. Oligosaccharides can be monosaccharides (mannose, glucose, galactose, rhamnose, fucose, xylose, etc.), or di- or tri-saccharides, combined from 2 or 3 monosaccharides.

In a simplest model we consider 8 sections of cell surface and 8 coding molecules (e.g. monosaccharide residues only), as shown on the software panel (Figure 2A). For considering a spatial orientation of the sectors they are numbered as a matrix $2\times2\times2$ with elements $A_{xyz},(x=0,1,y=0,1,z=0,1)$, (Figure 2B). In this case we can consider a vector M with I coding elements $M_i,(i=1,\ldots,I)$ corresponding to each sector, and we consider $I=8$ (marked as A, G, F, K, P, D, X, Y on the software panel).

A.

	1(AVR)	2(PVR)	3(ADR)	4(PDR)	5(AVL)	6(PVL)	7(ADL)	8(PDL)
A =	3	78	7	91	75	21	41	2
G =	97	95	73	29	18	37	87	66
F =	45	7	35	36	36	3	31	8
K =	96	11	84	15	7	4	49	5
P =	20	99	28	68	37	8	82	95
D =	10	61	6	79	90	70	22	28
X =	43	92	44	18	81	52	60	20
Y =	8	16	93	58	38	46	94	24

B .

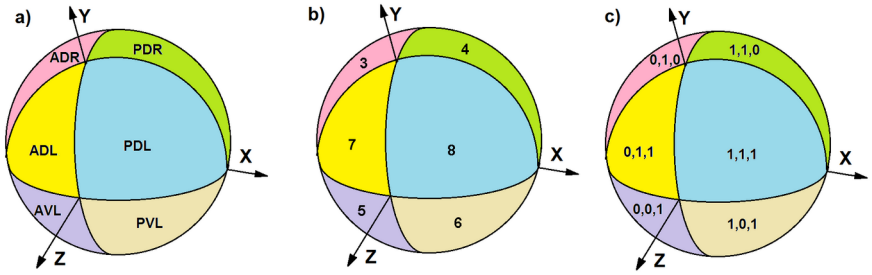


Figure 2. A coding matrix, associated with a cell. **A** - A coding matrix, shown in the panel of Morphogenesis Software. For each of the eight sectors of a cell the amount of 8 coding molecules A, G, F, K, P, D, X, Y in this sector are given as integer numbers. **B** - The 3D fitting of 8 columns of the matrix according to their position in a coordinate system, associated with zygote: $xyz, x=0,1, y=0,1, z=0,1$, where xyz corresponds to embryo axis AP, DV, LR. The corresponding spatial 3D coordinates are marked below each colon.

The algorithm includes two types of interactions between cells, namely, the adhesion and signaling, both determined by coding matrixes, associated with contacting cells. Later we are

planning to add also a long-distance signaling, determined by secreted factor(s), produced by some of the cells and influencing cell events of other cells in the area of their effective concentrations.

Though we are currently working out an appropriate mathematical formalism which will allow deciphering of possible rules determining *cell events* as operators acting on a corresponding matrix of a *cell state*, for the time being we include in the algorithm of Morphogenetic Software a set of simple but biologically relevant rules for the proof-of-concept of the cell surface code hypothesis.
The detailed description of the Morphogenesis Software see in Supplemented Materials.

3.2. Set of Rules for cell development.

Rule 1 - Choice of cell event depending on a cell coding matrix.

We will consider a constant time period corresponding to duration of each cell event. The pipe-line for the choice of a cell event corresponding to a coding matrix of a new appearing cell will be as following.
First, for each new cell we check if **t**he conditions for cell event *Apoptosis* (see below) are applicable. If yes, cell undergoes cell event *Apoptosis*. If no, we are checking the possibility of cell event *Division*. For that, we consider each set of 2 halves of cell surface which can be obtained during cell division in 3 possible division planes (x,y,z), namely, the “left” and “right” halves X_L and X_R for the division plane x, Y_L and Y_R halves for the division plane y, Z_L and Z_R halves for the division plane z.
Next we calculate a sum of all components (A,G,F,K,P,D,X,Y) for each half and determine Moment M of a cell for each axis:

$$\begin{aligned} Mx &= \left|X_R - X_L\right| \\ My &= \left|Y_R - Y_L\right| \\ Mz &= \left|Z_R - Z_L\right| \end{aligned} \quad ¿ \quad \begin{aligned} M_{min} &= min\left(M_x, M_y, M_z\right) \\ M_{max} &= max\left(M_x, M_y, M_z\right) \end{aligned}$$

We will consider Conditions 1 and 2:

Condition 1:

$$M_{max} - M_{min} > P \cdot m_h \cdot \lambda, \quad (1)$$

where P is a percentage ([0,1]) of the heterogeneity in zygotic matrix, showing the maximal dispersion of its elements; m_h is the highest possible amount of coding molecules in zygotic matrix (both are introduced as a parameters). Parameter λ is a parameter of inheritance (see below section 2).

Condition 2:

Both standard deviations $\sigma_{¿¿}$ and $\sigma_{¿}$of all elements a in both left and right halves of the surface related to the M_{max} direction satisfy $\sigma > e$, where $e = 0{,}1 \cdot P \cdot m_h$.

If both conditions 1 and 2 are fulfilled, a cell will undergo cell event *Division,* The division plane is determined by $M_{max} = max(M_x, M_y, M_z)$.

If the condition 1 is fulfilled, and the condition 2 - not, a cell undergoes *Internal cell event* of type *S*. If the condition 1 is not fulfilled, we determine a possibility to have cell event *Growth.* For that we check the condition:

If at least for 1 type of coding molecules $i(i = 1, \ldots 8)$ in a cell matrix its deviation (in a row) $\sigma_r < e$, there will be No cell event for this cell. Otherwise, cell undergoes cell event *Growth.*

We do not consider cell event *Movement* in this simplified version of the program, this is a work in progress.

3.3. Filling in new elements of cell matrices during *Cell events*

Rule 2. Filling in elements in daughter cells during Cell event Division (i.e. inheriting the spectrum of the daughter cell from the mother's one).

We will assume that one half of N sectors of each daughter cell (columns in a daughter cell matrix) will stay equal to the corresponding half of sectors of the mother cell (the ones related to outer sectors of a daughter cell), while each element in new (inner) sectors will be created *de novo* according to a rule R_2 (as schematically illustrated on Figure 3, colons in red color).

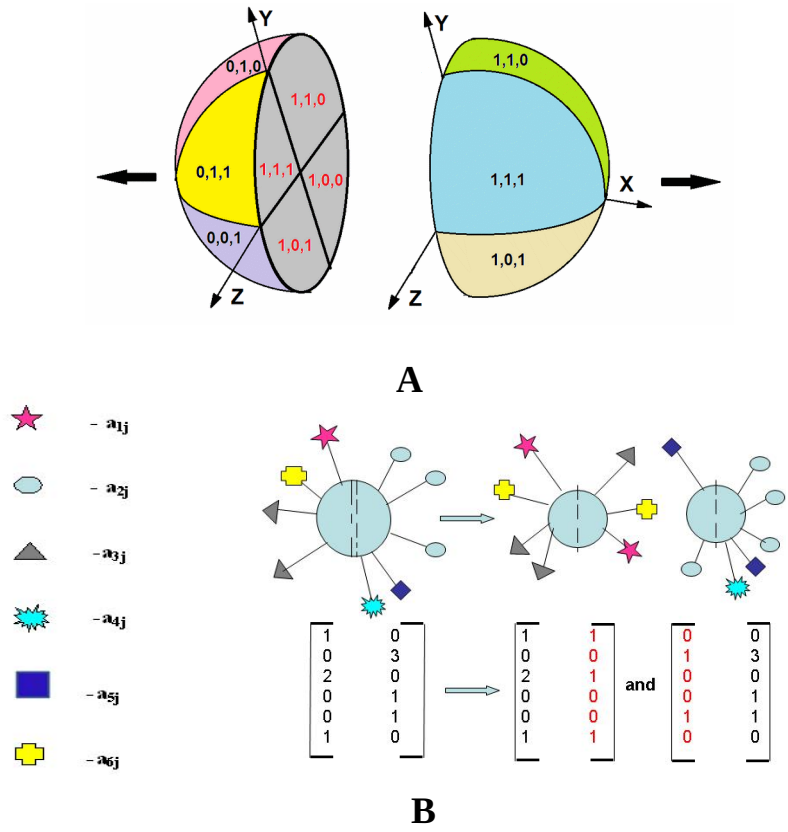


Figure 3. Illustration of a Rule for filling in the matrices, corresponding to daughter cells after division. **A**.- Spatial organization of the sectors on the two haves of mother cell before the division. **B**.-The simplified 2-sectors example of the Rule for filling in the matrices, corresponding to daughter cells after division. Each cell matrix containing 6 coding elements on a cell surface with 2 sectors.

According to this rule, each novel element i in a matrix of a daughter cell is a function the mother's cell matrix.

For mathematical formalization of the rule R_2 let us consider the coding matrix a_{ij} together with the spatial orientation of its n colons (as sectors of a cell surface) in a zygotic/embryonic coordinate system (axes AP, DV, LR) in R^3 which we will replace by xyz one. Then at any generation t a cell spectrum will be described by the matrix:

$$\hat{a}_t = \left[a^t_{ixyz} \right],$$

where a^t_{ixyz}- amount of molecules of i-th element in a sector with spatial coordinates x, y, z.

Having a division in L-direction $(L=x,y,z)$, we will notate the daughter cells as $(L,\varepsilon),\varepsilon=0,1$. The corresponding picture is*:* (L,0) -cell → (L,1)-cell , where → is a direction of the axe L. This means, that the daughter cell, emerging in the direction of axe L gets index $\varepsilon=1$, while one emerge in the opposite direction gets $\varepsilon=0$.

We suggest as a **Rule 2** that each new element *i* in a sector *n* of a daughter cell is a function of matrix of a mother cell. To discover this function as an operator, acting on a matrix (spectrum) of a mother cell returning two new matrices of daughter cells, is a direction of our current theoretical and experimental work (Morozova & Penner, 2015; Minarsky et al., 2017). For a time being, for a simplified version of a program, we suggest this function to be a linear combination of the same element *i* in all sectors of the mother's cell.

In this case the **Rule 2** of the filling in the elements of a spectrum of a (L,ε) -daughter cell during Cell event *Division* will be:

$$a_{i\ xyz}^{t+1}=a_{ixyz}^{t}\ if\ x=\varepsilon \tag{2}$$

$$a_{i\ xyz}^{t+1}=c_1 a_{ixyz}^{t}+c_2 a_{i\underline{x}\,yz}^{t}+c_3 a_{ix\underline{y}\,z}^{t}+c_4 a_{i\underline{xy}\,z}^{t}+c_5 a_{ixy\underline{z}}^{t}+c_6 a_{i\underline{x}\,y\underline{z}}^{t}+c_7 a_{ix\underline{yz}}^{t}+c_8 a_{i\underline{xyz}}^{t},\quad if\ x=¿\ \varepsilon \tag{3}$$

where: $\underline{x}=1-x\,;\ \underline{y}=1-y\,;\ \underline{z}=1-z\,;\ \underline{\varepsilon}=1-\varepsilon;$ c_i are chosen parameters.

Here (2) means conservation of a half of information of a mother cell in a half of the sectors of a daughter one, while in another half of daughter cell sectors each new element will be calculated as a linear combination of its amount in all sectors of the mother cell (3).

The law is written for the axis *x* be the axis of a division plane. For all other cases before and after the application of this rule the program will perform L - permutation, which is a permutation of sectors indexes which makes the index L to be the first, while the others moves in circle.

Note, that for each new element $a_{i\ xyz}^{t+1}$ in a daughter cell all odd coefficients c_i in formula (3) correspond to the elements of "mother" part of a coding matrix belonging to this cell, while the even coefficients c_i - to the "mother" part of a coding matrix, which will move to another daughter cell after division.

Another important remark is that the law is written in a way, providing for both L,0 - cell and L,1 - cell conservation of an outer half of information of a mother cell, while the part which will be

filled in with new information will correspond to the novel sectors on a cell division septum (for plants) or on the stretched out parts of the cell membrane (for animals).

In general, to specify the Rule of filling in the elements (inheritance), the 8 parameters $c_1 \dots c_8$ should be required. But we will consider biologically meaningful conditions/restrictions for c_i in (3) which allow minimizing the number of chosen parameters.

The main hypothesis about the conditions on filling in the elements is the requirement of *«Complementarity»* of nascent neighbor sectors in two daughter cells. According to our model, each of four sectors of a daughter cell, which were filled in a course of division, has a corresponding adjacent sector in another daughter cell, also been filled during the same cell event Division. We will call each pair of these new-formed adjacent sectors the «twins».

The suggested Rule of *«Complementarity»* proposes that the sum of each element in two «twins» will be the same for all pairs of twins. This means that the sum

$$a_i^{t+1}{}_{1\,yz(L,0-cell)} + a_i^{t+1}{}_{0\,yz(L,1-cell)}$$

is independent from y, z. This is possible if:

$$c_1 + c_2 = c_3 + c_4 = c_5 + c_6 = c_7 + c_8 = C\ . \qquad (4)$$

Let us consider a total sum for a given element in the cell:

$$A_i^t = \sum_{x,y,z} a_{ixyz}^t \qquad (5)$$

Then from (2), (3), (4) and (5) we will get:

$$a_i^{t+1}{}_{1\,yz(L,0-cell)} + a_i^{t+1}{}_{0\,yz(L,1-cell)} = C\,A_i^t \qquad (6)$$

for any y, z and any element i. Let us notate parameter C as a *parameter of Complementarity*.

Furthermore, we can assume existence of additional restrictions on the coefficients in formula (3) depending on the process of twins sectors formation. For example, we can consider symmetry or spirality pattern of twins sectors formation. This will give the following relations for the coefficients:

- in the case of the Symmetrical process:

$$c_3 = c_5;\ c_4 = c_6 \qquad (7)$$

- in the case of the *Spiral process:*

$$c_4 = c_5 = 0 \quad \text{for the left one} \tag{8}$$

$$c_3 = c_6 = 0 \quad \text{for the right one.} \tag{9}$$

Generalization of the requirements for the set of parameters

From the formulas (2), (3) we can obtain for each element i:

$$A_i^{t+1}{}_{(L,0-cell)} + A_i^{t+1}{}_{(L,1-cell)} = 2\lambda A_i^t, \tag{10}$$

where λ is parameter of inheritance, calculated from coefficients c_i, and characterizing an average change of a sum of each element in two daughter cells after division.

We will admit that $\lambda < 1$ which means that the total sum of all elements in a spectrum of a cell decreases with generations. This came from the fact that during a set of divisions and in a course of differentiation a potency of a cell, generally speaking, decreases, and thus, the amount of "status-significant" substances on its surface should decrease too.

Next we can re-write a moment M of a cell in a direction $L, (L = x, y, z)$ from formula (1) as:

$$M_L = \sum_i M_{i \cdot L}, \tag{11}$$

where $M_{i \cdot L}$ is a Moment of an element in each direction:

$$M_{i \cdot x} = \sum_{y,z} (a_{i1yz} - a_{i0yz})$$
$$M_{i \cdot y} = \sum_{x,z} (a_{ix1z} - a_{ix0z}).$$
$$M_{i \cdot z} = \sum_{x,y} (a_{ixy1} - a_{ixy0})$$

Then, taking into consideration formulas (2), (3), we will see that for each element *i* will be true:

$$M_i^{t+1}{}_{L(L,0-cell)} + M_i^{t+1}{}_{L(L,1-cell)} = 2\mu M_{iL}^t, \tag{12}$$

where L is the direction of cell division, μ is a coefficient which gives an average of a change of a longitudinal moment in two daughter cells after division. The parameter μ influences the dispersion of daughter cells by the sum of their elements, and thus on their possible diverse differentiation. Also, it can be seen that for $\mu > 1$, the average longitudinal moment of the

daughter cells increases, thus increasing the chance of continuation of the division in the same direction as in the mother cell.

Parameters λ and μ are *the chosen* parameters, directly connected with coefficients c_i:

$$c_1+c_3+c_5+c_7=\lambda+\mu-1\ , \tag{13}$$

$$c_2+c_4+c_6+c_8=\lambda-\mu\ . \tag{14}$$

Thus the parameter of **complementarity** C from formula (4) may be written as:

$$C=\frac{2\lambda-1}{4}=(c¿¿1+c_2+c_3+c_4+c_5+c_6+c_7+c_8)/4¿\ . \tag{15}$$

Next we can obtain the general formulas for the total sum of each element in daughter cells:

$$\begin{aligned} A_i^{t+1}{}_{(L,0-cell)}&=\lambda A_i^t-(1-\mu)M_{iL}^t \\ A_i^{t+1}{}_{(L,1-cell)}&=\lambda A_i^t+(1-\mu)M_{iL}^t \end{aligned} \tag{16}$$

which is important for detecting a cell differentiation; and for the longitudinal moment:

$$\begin{aligned} M_i^{t+1}{}_{(L,0-cell)}&=-(1-\lambda)A_i^t+\mu M_{iL}^t \\ M_i^{t+1}{}_{(L,1-cell)}&=(1-\lambda)A_i^t+\mu M_{iL}^t \end{aligned} \tag{17}$$

which is important for determining the further direction of division of daughter cells. Also, both formulas (16) and (17) are important for calculation concerning signaling (see section 3.4).

Importantly, the introducing of the parameters λ and μ decreases, according to (13), (14), the number of required parameters in the model. The specification of complementarity according to formulas (4), (15) decreases the additional (to λ and μ) parameters to 3 ones (for example, the coefficients c_1, c_3, c_5). The condition of symmetry of inheritance requests the setting of only 2 additional coefficients (c_1 and c_3), while the condition of spirality– only coefficient c_1.

Rule 3. Changes of surface markers during cell event Growth between its birth and disappearance (when it divides).

We consider two halves of the surface in the M_{max} direction.

In each colon of a half for which a sum of all components (A,G,F,K,P,D,X,Y) is smaller,

we apply: $a_{min}^{xyz}-e$

In each colon of a half for which a sum of all components (A,G,F,K,P,D,X,Y) is bigger,

we apply: $a_{max}^{xyz}+e$

Rule 4. During Internal cell event of type S a code of a cell is changed without any external cell event for a cell. The changing the code in this case is suggested as follows:

We consider Left and Right halves of the surface in the M_{max} direction;

1. If standard deviation σ of all elements a in any one half of the surface in the M_{max} direction satisfies the condition

   $\sigma > e$,

   while the other does not, then it will be the following changes in that half which does NOT satisfy the condition:

   $a_{max}+e$

   $a_{min}-e$

2. If BOTH deviations $\sigma_{¿¿}$ and $\sigma_{¿}$ of all elements a in Left and Right halves of the surface in the M_{max} direction satisfy

   $\sigma \le e$,

   then in each half the changes will be:

   $a_{max}-e$

$a_{min}+e$

Rule 5 for determination of apoptosis.

If all elements a_i in at least 1 sector of a matrix satisfy an in equation:

$\left|a_{ixyz}^{t}\right| < e$, then a cell will undergo apoptosis.

The cells which undergo apoptosis are marked in black color.

Rule 6 for determination of differentiation.

We will consider the process of differentiation as an *Internal cell event* of type D, which does not acquire a unit time period for itself, and thus can coincide with all other types of cell events. The differentiation status of a cell depends on the content of its matrix and is detected by the program in the moment of a cell appearance. If according to the suggested rules, a cell should undergo differentiation corresponding to its matrix information, then a cell will be marked by one of 8 different colors, reflecting its differentiation status.

The Rules for differentiation are determined by a proportion d_i of an element i in a total sum of all elements, $0<d_i<1$.

To determine if a cell undergoes a cell event *Differentiation* , the program performs a following check for each cell at the end of each cell event :

1. Calculate a sum of all coding elements in the matrix of a cell $A^t=\sum_i A_i^t$,

   and calculate a proportion of each element i in A^t cell: $d_i=¿A_i^t/A^t\vee ¿$.

2. If d_i is the maximal proportion among all elements *k*:

$$d_i=\left|\frac{A_i^t}{A^t}\right|=max_k\left|\frac{A_k^t}{A^t}\right| \quad and \quad d_i>d,$$

   where d is a chosen parameter, then a cell will get differentiation status of i-type.

3. When cell gets a differentiated status, a novel mode of inheritance is switched on for this cell. This means that for i-differentiated cells (cells with differentiation status of i-type) λ will be replaced by $\lambda=1$ in all formulas for calculating the parameters of rules.

It can be noticed that the choice $\mu=1$ provides the same differentiation status of the daughter cells as the mother's one, independently on λ (see from (16)).

Rule 7 for cell adhesion.

Adhesion is described as force F between two intercommunicating cells, which can have three possible states: strong, medium and zero, depending on a content of coding matrixes of the two contacting cells. The medium adhesion occurs when cells can change their mutual position, but do not come off from each other, the strong one - when cells can not change their relative position. The state “zero” corresponds to a case when a particular cell has no adhesion to other cells, thus enabling this cell to move in the body passing by other cells (which is a case, for example, for stem cells in Planaria). For such “zero adhesion” cases the adhesion depends only on a content of a matrix of one cell having “zero adhesion”, independently of the matrix of its neighbors. Current program considers only medium adhesion for all cells. Other cases are the work in progress.

3.4 Signaling

According to our theoretical assumptions (Minarsky et al., 2017), we consider that a signal for a cell is based on the information of the matrices of its neighboring cells.

We will assume that in the process of normal development without programmed cell event Movement the distance between daughter cells supposes to be me rather small, i.g., not exceeding 3R, where R is a cell radius.

As it was mentioned before (page 10), we call each pair of new-formed adjacent sectors of two daughter cells as the «twins»; thus, each cell division produces 4 pairs of «twins», and each cell in an embryo can have maximal amount of 8 «twins» relations with its neighbors.

We will suggest the following rules for signaling between cells:

1. If all distances T between «twins» sectors of a cell and its «twins» neighbors do not exceed R, a cell receives no signal.
2. If at least one distance T between «twins» sectors of a cell and its neighbors are greater than R but do not exceed 3R, a cell receives a small signal.
3. If in a cell which has N «twins» relations, the distance T between L «twins» sectors exceeds 3R, where $0 < L \leq N/2$, then a cell receives a middle signal.
4. If in a cell which has N «twins» relations, the distance T between L «twins» sectors exceeds 3R, where $L > N/2$, then a cell receives a strong signal.

We provide the following rules for the response to the signaling:

In the case of a **small signal,** the cell event will *depend on* the matrix â of a cell, as it supposes to be for the optimal (coded) development.

In the case of a **middle signal,** the cell event will be *independent of* the matrix of a cell, and can be one of the 3 scenarios:

-cell movement

-de-differentiation (proliferation)

-stagnation (no cell event).

For a time being, the choice of the exact scenario should be chosen by experimentalist (see description in Supplemented materials**)**, and will be the same for all cells receiving a middle signal.

In order to provide a response to a **strong signal**, the program calculates for a cell a "*determining matrix*" $\hat{a}_d$, which is a "complementary to adjacent" matrix for the matrix $\hat{a}$, i.e. $\hat{a}_d=(\hat{a}_a)^{compl}$.
We will determine an "*adjacent matrix*" $\hat{a}_a$ for a cell with matrix $\hat{a}$ as a matrix with 8 colons of elements corresponding to 8 adjacent sectors of neighboring cells (for simplification we consider that 1 sector of a cell is adjacent to 1 sector of one of neighboring cells).
A *complementary* matrix to any given matrix is constructed from eight complementary sectors to eight sectors of a matrix of a cell. According to (4), (6), (13), (16), and (17), the formula for calculation of a complementary sector to a given sector of a cell will be:

$$(a^t_{i\,xyz})^{compl}=\mu\frac{C}{S_0}A^t_i-a^t_{i\,xyz} \tag{18}$$

where S_0 is a sum of odd coefficients c_i (see (13)). Actually, this is a simplified isotropic version of the formula, which derives from the calculations and includes a dependence of the complementary sector on the moment M^t_{iL} of the preceding division:

$$(a^t_{i\,xyz})^{compl}=\mu\frac{C}{S_0}A^t_i\pm\frac{1-\mu}{S_0}M^t_{iL}-a^t_{i\,xyz}, \tag{19}$$

where index L=0,1 indicates different daughter cells.
Indeed, from (4), (6) we have:

$$(a_i^{t+1}{}_{xyz})^{compl}+a_i^{t+1}{}_{xyz}=C\,A^t_i\,, \tag{20}$$

and the transition from A^t_i to A^{t+1}_iaccording to (13), (16), (17) results in two formulas for different daughter cells:

$$A^t_i=\frac{\mu}{S_0}A_i^{t+1}{}_{L(L,0-cell)}+\frac{1-\mu}{S_0}M_i^{t+1}{}_{L(L,0cell)} \tag{21a}$$

$$A^t_i=\frac{\mu}{S_0}A_i^{t+1}{}_{L(L,1-cell)}-\frac{1-\mu}{S_0}M_i^{t+1}{}_{L(L,1\,cell)} \tag{21b}$$

Taking A^t_i as an average of (21a) and (21b), we get (18), allowing direct calculation of a complementary sector to a given one without additional (moment) information.

Next using (18) (or, if possible, 21a,b) the eight sectors of a matrix $\hat{a}_d$ are calculated, which provides a "*complementary to adjacent"* matrix for a matrix $\hat{a}$.

We propose the following rule (Rule 8) for the response to the strong signal:

If a *total sum* A of all elements in a matrix â of a cell and a *total sum* A_d of all elements in a matrix $\hat{a}_d$ satisfy an inequation $A<A_d$, a cell will *undergo cell death* (apoptosis). If $A>A_d$, the cell A will *convert its matrix from* $\hat{a}$ *to* $\hat{a}_d$. This rule means, that in the case of strong signal (strong difference between its own spectrum and the "adjacent spectrum" of neighbors) the cell "older" than its environment is dying, while the behavior of the "younger" cell is totally determined by this environment., each one complementary to one sector of the matrix $\hat{a}_a$, $\hat{a}_d=(\hat{a}_a)^{compl}$.

Next cell event in this situation is performed according to a program (programmed development) with the matrix $\hat{a}_d$.

4.Results and Discussion

4.1. Computational experiments-description and general results

The computational experiments using Morphogenesis Software start with the generation of an initial cell of an organism (zygote) with an assigned matrix. Next a development of an embryo from a zygote is modeled following the set of suggested rules 1-8 up to the formation of early stages of emryogenesis (up to around 1500 cells). The influence of variations of

-different sets of parameters λ, μ,C,m_h,P,X,d

- the possible choices of the type of the initial matrix (random one or diagonal random one with different coefficient)

-possible choices of the mode of complementarity (symmetrical or spiral)

-the modes of signaling

on the modeling embryo were studied by changing them one by one with the same zygotic matrix. Next for the best chosen constant set of parameters and the list of other best choices the dependence of the development on the variations of zygotic matrixes were studied.

The computational experiments demonstrated the following results.

First, the computational experiments using Morphogenesis Software have shown that any change of one element in a zygotic matrix with keeping all other parameters being the same, causes well visible (and sometimes very essential) changes in the resulting shape of a computational embryo. This means an existence of one-to-one dependence of pattern formation (a shape of a developing embryo and differentiation of its cells) on a spectrum of coding molecules on an initial (zygotic) cell matrix.

Second, we have found that any developmental tree with pre-determined parameters of cell events has one-to-one correspondence with embryo morphology at each time slice.

Third, we have found that among millions of randomly generated zygotic matrices, next undergoing development following the same set of rules 1-8 with different sets of parameters λ, μ,C,m_h,P,X,d there exist several numerical embryos with the shapes well approximating the shapes of actual embryos, belonging to different plant and animal taxons at the equivalent (early) stages of development.

In order to prove this similarity by precise mathematical methodology, and also to illustrate the second statement, we have performed a comparison of the developmental trees of the actual organisms with the computational trees produced by Morphogenesis Software, followed by the comparison of corresponding shapes of the embryos.

4.2. Development of actual organisms, presented as a tree of cell events

We have studied the development of early stages of embryogenesis of actual organisms, with its next formalization as a graph (i.e., a rooted graph or tree) with vertexes corresponding to cells in particular cell states, and edges corresponding to cell events, representing cell fates. Namely, the anatomical sections of a developing embryos of three angiosperm species belonging to different classes and families: *Miryophyllum specatum* L. (Haloragaceae, dicot) [23], *Polygala major* Jacq. (Polygalaceae, dicot) and *Triglochin palustre* L. (Juncaginaceae, monocot) [24], were investigated. The analysis of these anatomical sections enabling to reveal cell events reflecting the fate of each cell, has resulted in construction of corresponding developmental graphs (trees) for these species, together with the reconstructions of a 3-dimensional shape of the embryos at progressive developmental stages (Fig. 4,5,6)

During the observed period of development only three types of cell events were observed: cell growth, cell division and cell differentiation, while no cell movement and no cell death events were noted. On the graphs a cell event *Growth* is manifested as an edge with a number coefficient showing cell enlargement in size. The planes of cell division are indicated by the marks on the two edges emerging from a corresponding vertex: the horizontal dash means transverse division (corresponding to x axes of zygote), the vertical dash means the longitude division (corresponding to y axes of zygote), and narrow ovals designate a division in a plane perpendicular to the xy one. The edge presented as a dotted line indicates the situation when during a particular step no cell event has happened for a cell.

Each horizontal layer of a graph corresponds to one step, which is a time period during which at least one cell event has occurred at least in one cell.

The first event of differentiation in an embryo is differentiation of an embryoderm, an outer layer of an embryo. In the developmental graph built for *Polygala major*, the vertexes which correspond to the differentiated cells of embryoderm are colored in black.

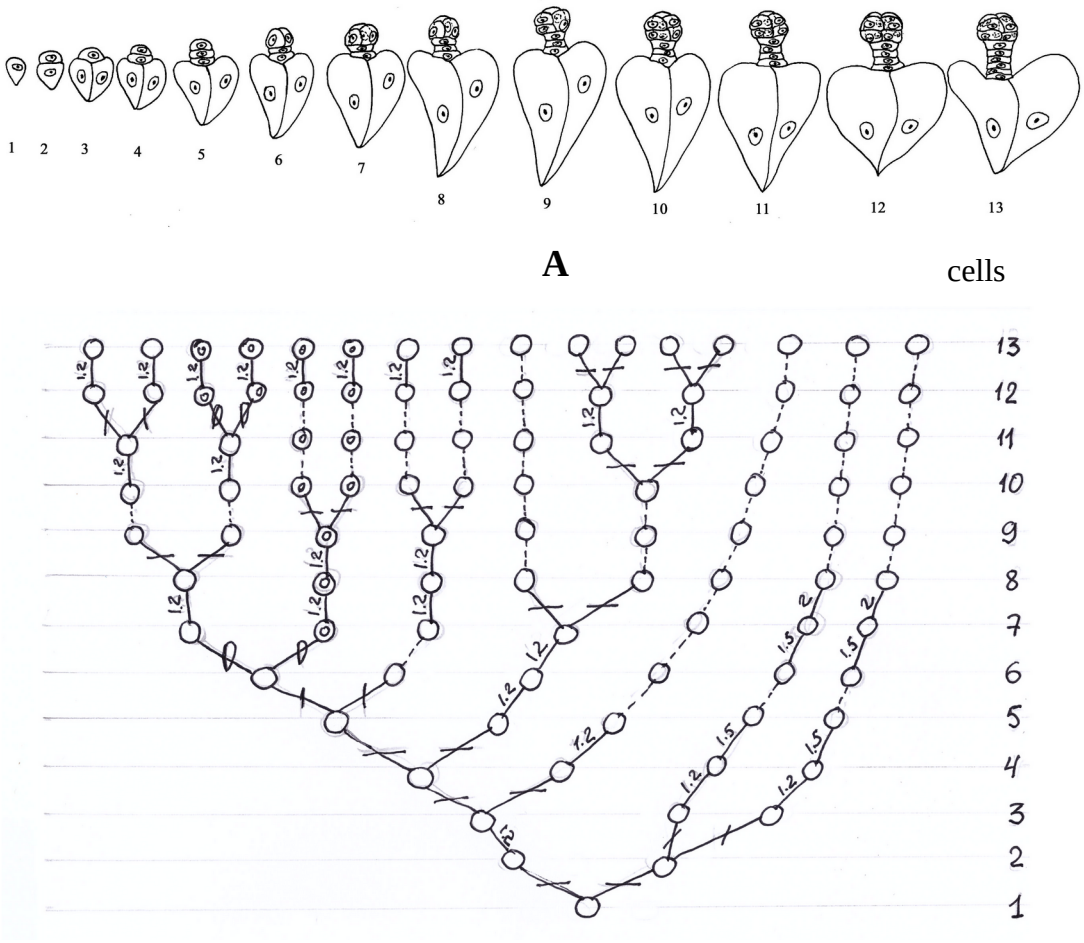
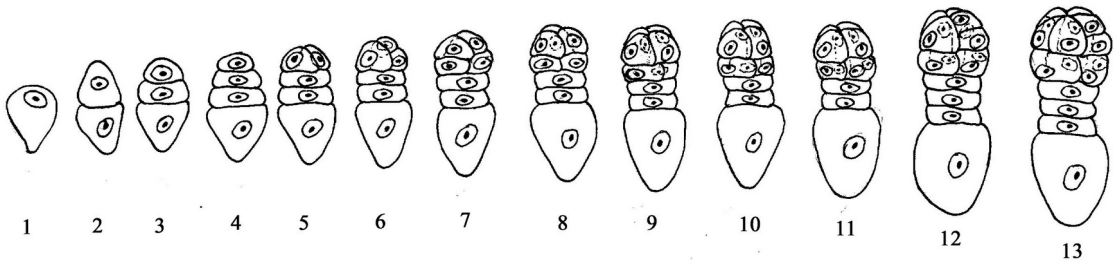
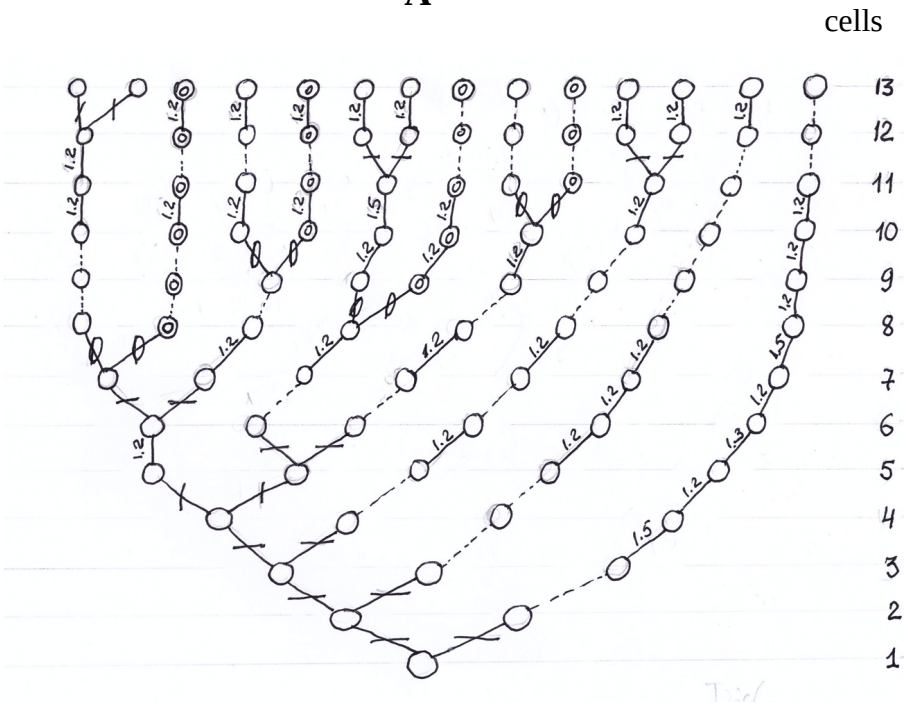


Figure 4. Embryogenesis of Myriophyllym specatum. **A**. The scheme of the first stages up to 16-cellular proembryo. **B**. The corresponding tree of Cell events. The number of stage is indicated as 1-13.

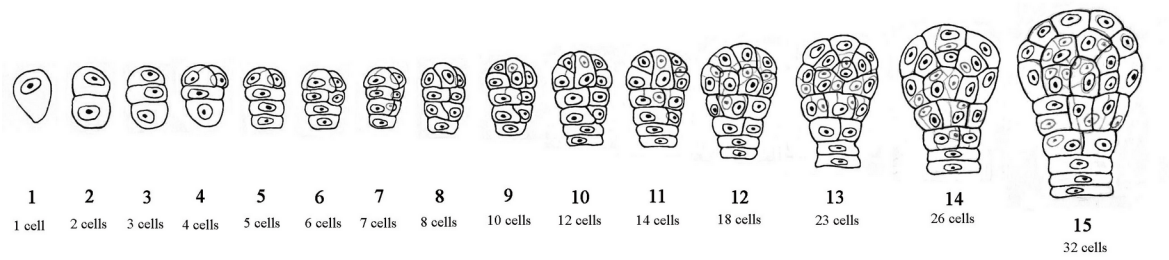


A

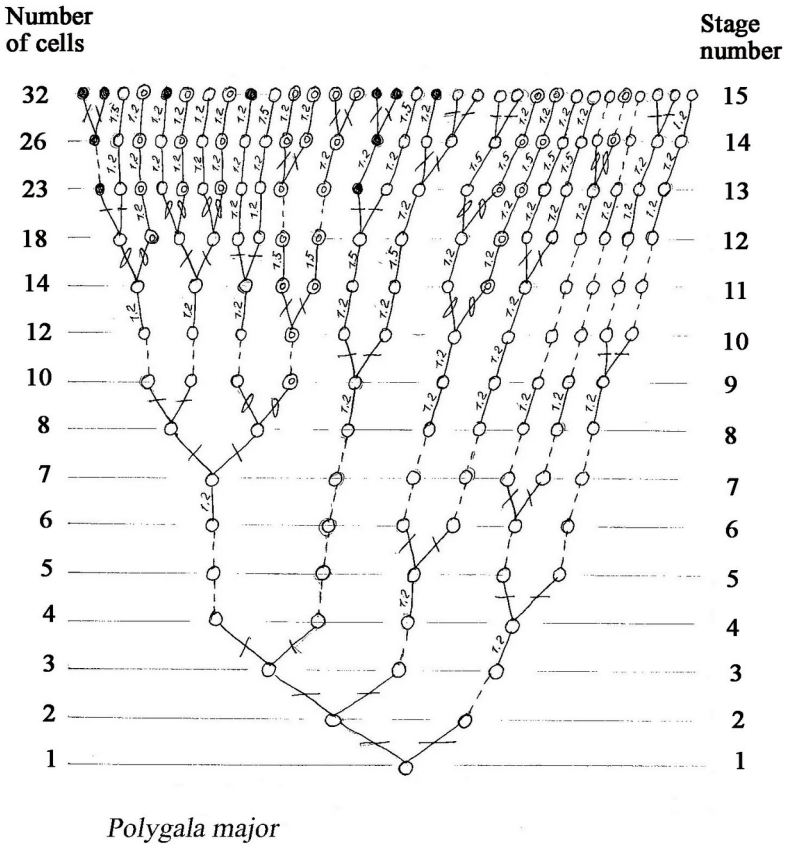


B

Figure 5. Embryogenesis of Triglochin palustre. **A**. The scheme of the first stages up to 14-cellular proembryo. **B**. The corresponding tree of Cell events. The number of stage is indicated as 1-13.



A



B

Figure 6. Embryogenesis of Polygala major. **A**. The scheme of the first stages up to proembryo. **B**. The corresponding tree of Cell events. The number of stage is indicated as 1-15.

B

4.3. Comparison of actual and computational trees and shapes

The trees, obtained from ten millions randomly generated zygotic matrices, developing into the embryos by Morphogenesis Software, where compared with each of three developmental trees of the actual organisms, presented on the Figs 4,5,6. The sourse code for matrices generation, embryo growth and trees comparison can be found at:

https://cloud.mail.ru/public/PiiN/31Y2ZmYGu

and the algorithm for the developmental graphs (trees) comparison is presented in Supplemental Materials. The embryos were developed with the "Random" choice of the initial matrix, the constant set of all parameters (λ=0,345; μ=0,655;C=$\frac{(2\lambda-1)}{4}$;c_1=c_3=$\frac{2}{3}$C; m_h=100;P=0,75,X=0,1;d=0,2),the symmetry pattern of matrixes complementarity and with the request to follow the rules for normal signaling. The best zygotic matrices found for the developmental trees, most close to the actual ones, and the corresponding trees, are presented in Supplemental Materials. As the computational trees includes edges for both "Internal cell event" (marked with 1 on the corresponding edge) and "No cell event" (marked with 0 on the corresponding edge), while the information about the actual trees provides "No cell event" edges only (as currently we do not have an information about molecular markers, providing matrices, and thus, about their changing, i.e., "Internal cell event"), for the trees comparison algorithm we have considers both types of events to be equivalent.

It is important to note that this computations were done having the most simple case of initial matrix (8X8) which gives the amount of possible matrixes as 101^{64}, which provides much less than 0.1% of the maximal amount of possible trees (around $5^{(2¿¿n+1-1)¿}$, where 5 is the amount of possible cell events in the actual cases of plant species (cell division in three directions (x,y,z), growth and no cell event), n is the number of steps(levels) of the tree). And even for this very simplified case

(only 8 possible coding molecules, while in the theory we assume all possible combinations of mono- ,di- and three-saccharides, and only 8 sectors of the cell surface, which should be increased) we have got the putative zygotic matrices, providing the trees with 0,89% of similarity with the actual ones. This provide a good proof-of-concept and a possibility to search next for a

more complex matrices allowing a complete correspondence of the computational trees to the actual ones.S

Next we have checked that the computational trees, identical to the trees of three presented plant species, give the same embryo shapes as the actual species have. The results produced by Morphogenesis Software up to the step 13 are shown on Figures 7,8,9. These embryo shapes where compared to the corresponding embryo shapes of three plant species, presented on the Figs 4,5,6 using standard image comparison software (https://opencv.org). The results of comparison have shown high level of similarity for all three plant species. Namely, the distance between the original shapes and the computational ones (calculated as a simple Hausdorff distance measure between shapes defined by contours) were scored as 11, 13, 14 units correspondingly, while a threshold of good similarity of two images stated in the software is 40 units and the best matching is stated to be below 9 units.

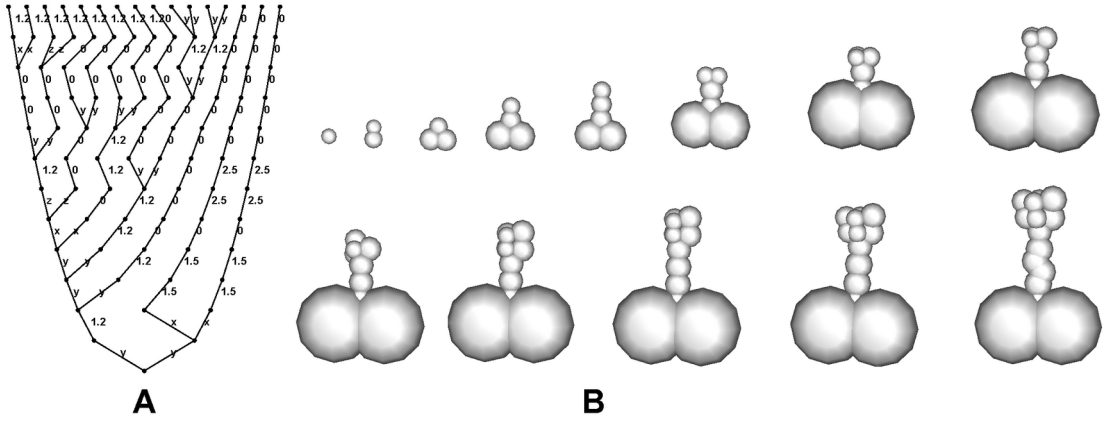


Figure 7. Embryogenesis of Myriophyllym specatum produced by Morphogenesis Software. A. Computational tree, similar to the actual tree presented on the Figure 4. B. Corresponding computational embryos up to the step 13.

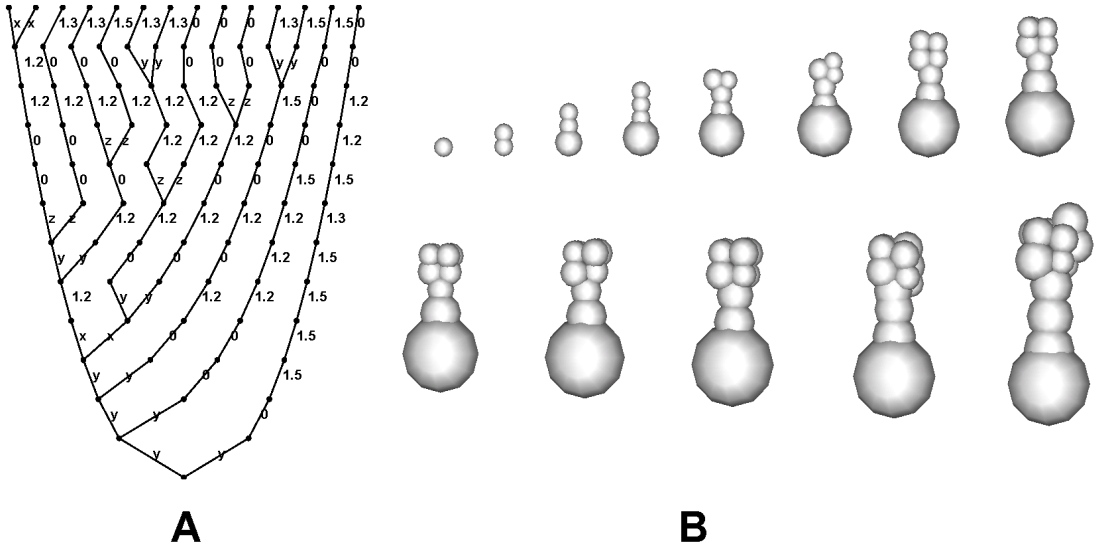


Figure 8. Embryogenesis of Triglochin palustre produced by Morphogenesis Software. A. Computational tree, similar to the actual tree presented on the Figure 6.B. Corresponding computational embryos up to the step 13.

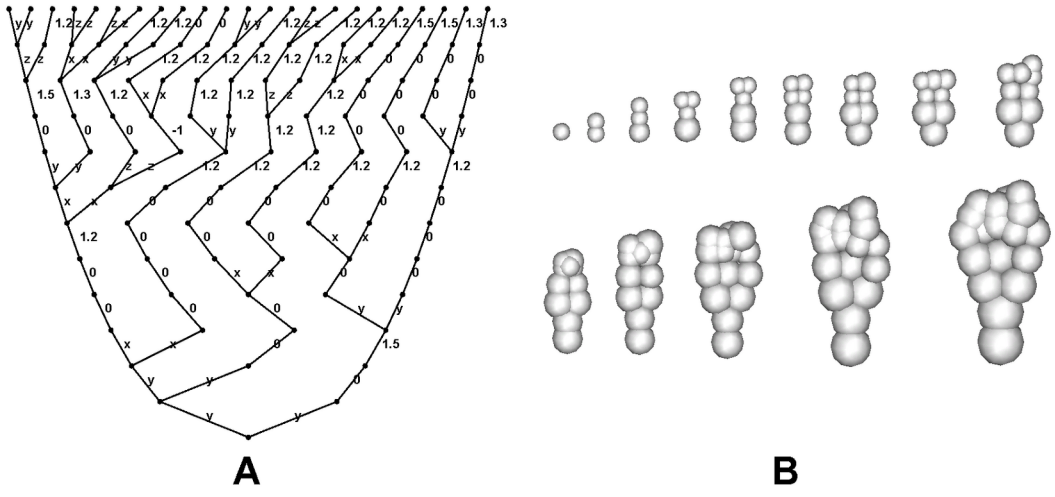


Figure 9. Embryogenesis of Polygala major produced by Morphogenesis Software.
A. Computational tree, similar to the actual tree presented on the Figure 6. B. Corresponding computational embryos up to the step 15.

This approach gives precise numerical results of comparison of the developmental processes, leading to the similar shapes of the objects, thus proving the fundamental nature of the suggested theory and of the acceptability of the chosen simplified rules.

Thus we provide a frame of mathematical formalism and corresponding software demonstrating a connection between epigenetic code on a cell surface, a consequent cell event for a cell and the geometry of an embryo. The experiments proving the main hypothesis about the existence of a cell surface morphogenetic code, and if so, providing a concrete data for building a mathematical model on a set of concrete data is a work in progress.

Materials and Methods

The Morphogenesis Software is constructed for exploring problems related to the processes of morphogenesis. The program consists of the executable file Morphogenesis.exe and is designed to work under the operating system Windows XP and above.

The development of an embryo from a zygote is modeled in the three-dimensional computational area, the size of which increases automatically during a course of embryo development. Cells are modeled as spheres, having a set of constant and variable parameters.

The computational time (CPU time) for a growth of organism consisting from several thousand cells is equal approximately to several seconds.

For each case of modeling of the development from the zygote one should specify in a program:

- parameters λ, μ, c_1, c_3(for symmetrical case),m_h,P,X,d
- the possible choices of the type of the initial matrix (random one or diagonal random one with different coefficient)
-possible choices of the mode of complementarity (symmetrical or spiral)
-the modes of signaling

The instruction for using the software and managing the parameters and the modes of the model is given in Supplemental Materials.

Main parameters of the model visualization

Parameters that determine the mutual spatial cell arrangement: m – mass of a cell, p - viscosity of the medium, h - the cell size.

At each step in time, between a pair of any neighboring cells, for example, for two cells 1 and 2 the forces acting between them will be $\vec{F_1}=\vec{e}f$, $\vec{F_2}=-\vec{e}f$ respectively, where

$$f=\begin{cases} k_f\left(1-\frac{h_1}{h}\right)\left(1-\frac{h_2}{h}\right), at\ h\le h_2, \\ 0, at\ h>h_2 \end{cases} \quad (22)$$

$\vec{r_2},\vec{r_1}$ are radius-vectors; k_f, h_1, h_2 are specified parameters$(h_1<h_2)$, $h=|\vec{r_2}-\vec{r_1}|$, $e=(\vec{r_2}-\vec{r_1})/h$.

After the forces between all cells are determined, the corrective movement of each cell is carried out according to the equation

$$m\frac{d\vec{v}}{dt}=-p\vec{v}+\vec{F},\vec{v}=\frac{d\vec{r}}{dt} \quad . \quad (23)$$

By choosing the parameters m , p, k_f in equations (22) and (23), different adhesion between cells can be modeled.

Acknowledgements

We thanks Z. Nikiticheva (BIN RAN) for providing the anatomical sections of the embryos, N. Pakudin and M.Gavrilovich for image comparison analysis. The work of N. Morozova and O.Butuzova was supported by grants 16-54-21014 and 15-54-45001 from the Russian Foundation for Basic Research and was carried out within the framework of the state assignment No. AAAA-A18-118030790063-6 to BIN RAS. The work of N. Bessonov, O.Butuzova, A.Minarskiy, A.Tosenberger was supported by IHES program on mathematical biology (Simons foundation).

References

1. Zablackis E, York WS, Pauly M, Hantus S, Reiter WD, Chapple CC, et al. Substitution of L-fucose by L-galactose in cell walls of Arabidopsis mur 1. Science. 1996;272(5269): 1808-1810.

2. Fry SC, Aldington S, Hetherington PR, Aitken J. Oligosaccharides as signals and substrates in the plant cell wall. Plant Physiol. 1993;103(1): 1-5.

3. McNeil M, Darvill AG, Fry SC, Albersheim P. Structure and function of the primary cell walls of plants. Ann Rev Biochem. 1984;53; 625-663.

4. Johnson SW, Alhadeff JA. Mammalian alpha-L-fucosidases. Comp Biochem Physiol B. 1991;99(3): 479-488.

5. Fry SC. Oligosaccharins as plant growth regulators. Biochem Soc Symp. 1994;60: 5-14.

6. Mohnen D, Hahn MG. Cell wall carbohydrates as signals in plants. Semin Cell Biol. 1993 Apr;4(2): 93-102.

7. Albersheim P, Darvill A, Augur C, Cheong J, Eberhard S, Hahn MG, et al. Oligosaccharins: oligosaccharide regulatory molecules. Acc Chem Res. 1992;25(2): 77-83.

8. Riou IF, Darribere T, Boucaut IC. Cell surface glycoproteins change during gastrulation in *Pieurodeles waltlii*. J Cell Sci. 1986;82: 23-40.

9. Taatjes DJ, Roth J. Glycosylation in intestinal epithelium. Int Rev Cytol. 1991;126: 135-193.

10. Tran Thanh Van K, Tonbart P, Cousson A, Darvill AG, Gollin DJ, Chelf P, et al. Manipulation of the morphogenetic pathways of tobaco by oligosaccharins. Nature. 1985;314(6012): 615-617.

11. de Vries SC, Booij H, Janssens R, Vogels R, Saris L, LoSchiava F, et al. Carrot somatic embryogenesis depends on phytohormone-controlled presence of correctly glycosylated extracellular proteins. Genes Dev. 1988;2: 462-476.

12. Bourrillon R, Aubery M. Cell surface glycoproteins in embryonic development. Inter Rev Cytol. 1989;116: 257-338.

13. Ito M, Takata K, Saito S, Aoyaki T, Hirano H. Lectin binding pattern in normal human gastric mucosa: A light and electron microscopy study. Histochemistry. 1985;83: 189-193.

14. Nemanic MK, Whitehead IS, Elias PM. Alterations in membrane sugars during epidermal differentiation: visualization with lectins and role of glucosidases. J Histochem Cytochem. 1983;86(4): 415-419.

15. Morozova N, Bragina E, Vasiljeva V. Dynamics and role of glycoconjugates of plant cell wall in embryogenesis. Phytomorphology. 2006;56(3,4): 1-7.

16. Ponder BA. Lectinhistochemistry. In: Polak S, van Noorden S, editors. Immunocytochemistry. Practical applications in pathology and biology. Bristol; 1983. pp. 129-142.

17. Zuzack IS, Tasca RI. Lectin-induced blocage of developing processes in preimplantation mouse embryos *in vitro*. Gamete Res. 1985;12(3): 275-290.

18. Suprasert A, Pongchairerk U, Pongket P, Nishida T. Lectin histochemical characterization of glycoconjugates present in abomasal epithelium of the goat. Kasetsart J Nat Sci. 1999;33: 234-242.

19. Grossin KL, Edelman GM. Specific binding of cytotactin to sulfated glycolipids. J NNeurosci Res. 1992 Dec;33(4): 631-638.

20. Friedlander DR, Hoffman S, Edelman GM. Functional mapping of cytotactin: proteolytic fragments active in cell-substrate adhesion. J Cell Biol. 1988;107: 2329-2340.

21. Tan SS, Crossin KL, Hoffman S, Edelman GM.Asymmetric expression in somites of cytotactin and its proteoglycan ligand is correlated with neural crest cell distribution. Proc Natl Acad Sci USA. 1987 Nov;84(22): 7977-7981.

22. Edelman GM. Topobiology: an introduction to molecular embryology. Basic Books, Inc.; 1988.

23. Haloragaceae. In: Yakovlev MC, editor. Comparative embryology of flowering plants. Brunelliaceae-Tremandraceae. L.:Nauka; 1985;3. pp.

24. Juncaginaceae. In: Yakovlev MC, editor. Comparative embryology of flowering plants. Butomaceae-Lemnaceae. L.:Nauka; 1990;5. pp.

Supplemented materials 1. Morphogenesis Software - Instruction for users

The program (https://cloud.mail.ru/public/2Gor/wWHkC4ndE) is launched by the button RUN on the top of the main window of **Morphogenesis Software,** which opens the dialog box "Begin" (Figure 1S).

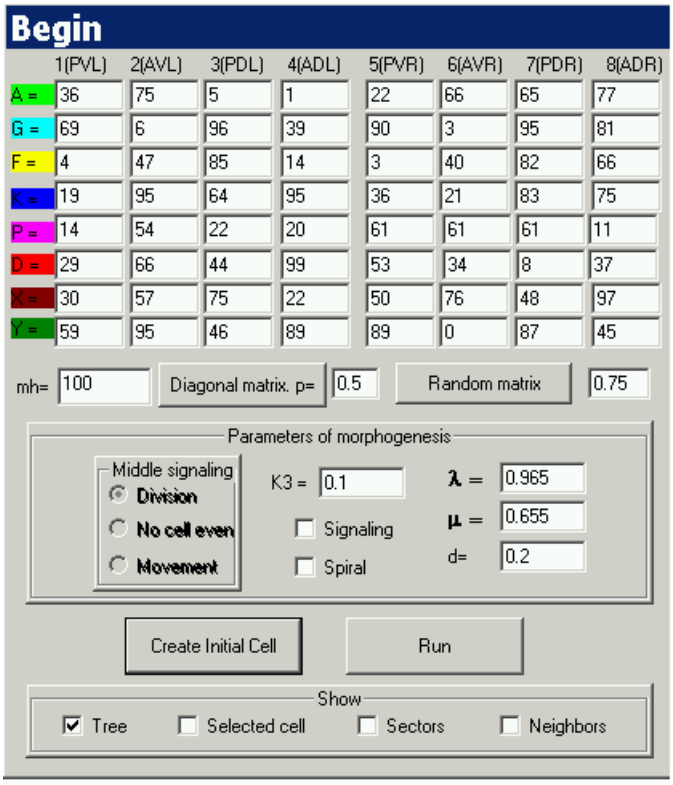


Figure 1S. Dialog panel “Begin”

It allows to set a matrix of a zygote (the initial cell). It means that for each of the eight sectors of the initial cell the amount of the 8 coding substances A, G, F, K, P, D, X, Y in this sector can be assigned , thus providing a coding matrix, associated with this cell. The matrix can be set in 3 following ways:

- each of the 64 coding numbers is manually entered into the fields of matrix elements indicated on the Begin dialog box.

- each coding number is set randomly in the chosen range from 0 to m_h by clicking the button “Random matrix”; the parameters m_h(the maximal amount of coding molecule) and P (a percentage ([0,1]) of heterogeneity in a zygotic matrix, showing the maximal dispersion of its elements) should be set in a corresponding dialog boxes on the panel.

- each coding number is set randomly in the chosen range from 0 to m_h, but with a diagonal predominance of the values located on the diagonal of the matrix - by clicking the button "Random matrix". The magnitude of the predominance is indicated by the parameter *p* and is determined by the following rule: if an element lies on the diagonal of the matrix then its value is chosen randomly in the range $[p m_h, m_h]$, if an element does not lie on the diagonal of the matrix, then its value is randomly set in the range $[0, p m_h]$.

The set range from 0 to m_h will be automatically kept for all cells of the embryo.

Next a set of parameters λ, μ, d, k_3 (=X) implementing the developmental Rules for Cell events should be specified in the corresponding windows of the panel "Begin".

After setting the matrix and specifying a set of parameters the program can be started by the button "Create Initial Cell", which will create a zygote with a given matrix. The development is started by the button "Run" on the panel "Begin", and the visualization of the development of the organism and of the corresponding graph(tree) occurs based on the given initial matrix and parameters (See Fig.1). On the graph, the edge corresponding to cell event "Growth" is marked with a coefficient of growth (which is the enlargement of cell volume during this cell event), the "Internal cell event" is marked by number "1" on the corresponding edge, and "No cell event" is marked with 0 on the corresponding edge). To stop the development one should press the computer keys "CTRL Q".

The check box "Signaling" at the bottom of the Begin panel allows to switch on the developmental program with signaling (see Section 3.4). Without checking this check box the behavior of cells is modeled as independent on the signaling. The check box "Sectors" allows marking the location of the numerated sectors on the surfaces of the cells, which is important for studding signaling. Also, on this panel there is a choice of a possible respond to the middle signaling: cell division, cell growing or no cell event.

The button "Advanced" on the top of the main window opens the dialog panel "Advanced" (Figure 2S), which displays a set of elements of software managing. The most important ones are:

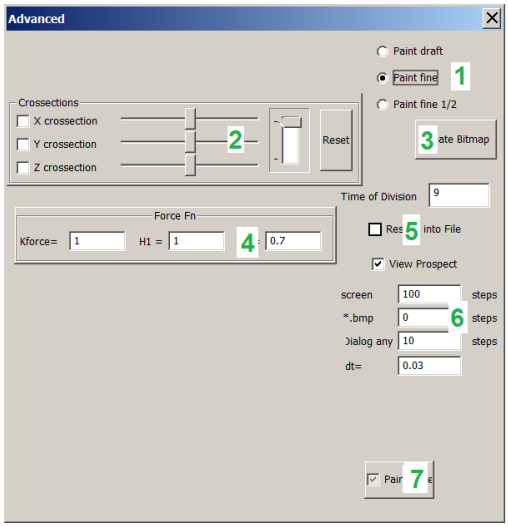


Figure 2S. Dialog Panel Advanced

-the element (shown on Figure 2S, point 2) which allows to make slices (sections) of the embryo perpendicular to one of the axes of coordinates x, y, z (Figure 3S a,b). The position of each slice on the chosen axe can be changed by the corresponding cursor, thus allowing observing all internal structure of the body.

-the element (shown on Figure 2S. point 3) which saves the current image of the main window (embryo, developmental tree, initial matrix) in the bmp file.

-the element (shown on Figure 2S. point 7) which allows displaying a tree of cell events alongside to a growing embryo (see Figure 1).

The remaining elements on this dialog panel (not mentioned here) are currently set by default and not required in the regular work with the program.

Any cell of an embryo visible on the screen can be marked by a mouse click, and simultaneously the chosen cell and its path from the zygote will be shown on the corresponding graph (tree) as a fat line (Figure 3S, c,d) For this, the check box “Selected cell” should be marked at the Begin panel. And vice versa, by the clicking on any of free vertexes of the graph detects and marks a corresponding cell in the embryo. For marking or detecting an internal cell it is necessary to make corresponding sections.

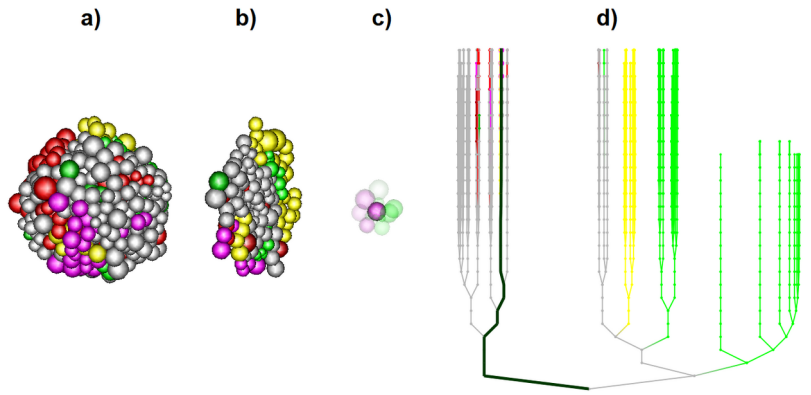


Figure 3S. Examples of a software managing. a) growing embryo; b) the cross-section of an embryo; c) a selected cell (bright colored) surrounded by its neighbors (faint colored); d) corresponding developmental tree.

The epigenetic matrix of this cell can be seen by clicking the button " Selected Cell " on the top of main window opening the dialog panel "Selected Cell" (Figure 4S). The matrix of a chosen cell is presented together with two additional matrixes $\hat{a}_a$ and $\hat{a}_d$, important for studying signaling (see Section 3.4). There is also an option to see a chosen cell with a set of its neighboring cells (Figure 3S,c), specifying a desirable neighborhood by the "area" buttons on the dialog panel "Selected Cell".

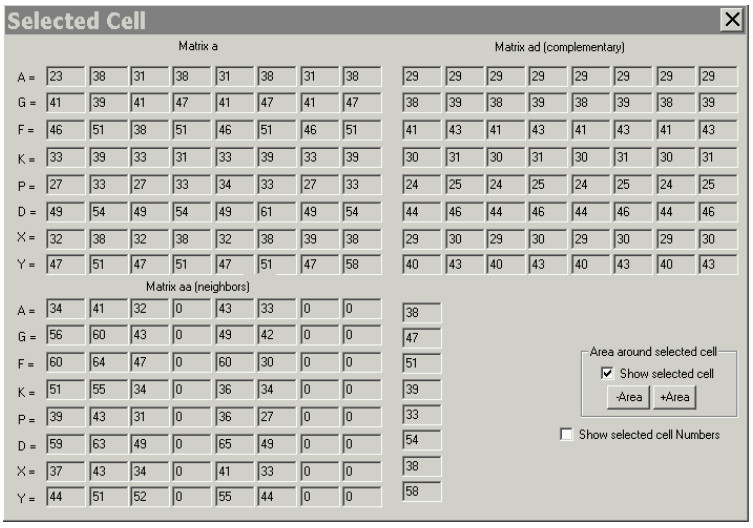


Figure 4S. Dialog panel "Selected Cell". Left upper table is a matrix (a) of a selected cell, left bottom table is an "adjacent matrix" ($\hat{a}_a$¿, which is combined from the corresponding adjacent

sectors of the neighbor cells; the upper right table is a “complementary to adjacent” matrix ($\hat{a}_d$). Zero in the “adjacent matrix” means the absence of a neighbor cell adjacent to a given sector of a cell.

Supplemented materials 2. Algorithm for trees comparison

Let us consider a set of cell events $C=\{\varnothing, I, G, X, Y, Z, A\}$, with elements: null element ($\varnothing$), identity (no change), growth, divisions in x, y and z directions respectively, and apoptosis.

Def: **Tree of cell events** $T=(V,E,e)$, where V is a set of vertices (cell events), $E \subseteq V \times V$ is a set of directed edges (cells states - epigenetic code…), and $e: V \to C$ a mapping from vertices of the tree T to the set of cell events C, such that (V,E) is a binary tree.

The **level** $l(v)$ of a vertex $v \in V$ is the length of the path from the root of the tree to the vertex.

The maximal level of the tree we denote with l_{max}.

Remark: *The choice of cell events being represented by vertices is not the best one, but at this stage it is easier to define cell division as a single element of a graph (a vertex) instead of having two elements (two edges). Hence, for simplicity, in this description cell events are represented by vertices, and later this can be changed.*

Def: If $T=(V,E,e)$ is a tree of cell events, we define the **parent function** which gives the n-th ancestor of a node as $p: V \times N_0 \to V$ such that $\forall a \in V$ and $\forall n \in \{0,\ldots,l(v)\}$ there $\exists b_0, b_1, \ldots, b_n \in V$ such that $(b_{i-1}, b_i) \in E \ \ \forall i \in \{1,\ldots,n\}$. In case when $n>l(v)$, we define $p(v,n)=\varnothing$. Also we define $p(v,0)=v$.

Def: If $T=(V,E,e)$ is a tree of cell events for any two vertices $v,w \in V$ we define a **path** $\langle v,w \rangle$ between them as the ordered n-tuple $(v=v_0,v_1,\ldots,v_n=w)$ such that $\forall\, i \in \{1,\ldots,n\}$ is $(v_{i-1},v_i) \in E$, and that $\forall\, i,j \in \{1,\ldots,n\}, i \neq j$ is $v_i \neq v_j$. n is then the path length.

NB: In any tree of cell events $T(V,E,e)$ for every two vertices there exists a path between them and that path is unique.

Def: If $T=(V,E,e)$ is a tree of cell events, for every cell event $c \in C$ we define its corresponding **cell event δ-function** $\delta_c : V \to C$ as

$$\delta_c(v)=\begin{cases}1, e(v)=c\\0, e(v)\neq c\end{cases}, v \in V. \quad (1)$$

Hence we have 6 different cell event δ-functions (δ_X, δ_Y, etc.).

Def: If $T=(V,E,e)$ is a tree of cell events, for every cell event $c \in C$ we define its corresponding **cell event measure** $m_c : P \times V \to N_0$

$$m_c(v,w)=\sum_{i=0}^{n} \delta_c(v_i) \quad (2)$$

where $\langle v,w \rangle=(v=v_0,v_1,\ldots,v_n=w) \in P_T$ is the path between the two vertices, and P_T is the set of all paths of the tree T. Hence we have 6 different cell event measures (m_X, etc.).

Def: For all leaves $v_1,\ldots,v_k$ of level n of the tree $T=(V,E,e)$ we define:

- The **sum** m_c^n of measures of all paths from the root of the tree to the nodes $v_1,\ldots,v_k$ as:

  $$m_c^n=\sum_{i=1}^{k} m_c(v_i). \quad (3)$$

- The **mean** $\overline{m}_c^n$ of measures of all paths from the root of the tree to the nodes $v_1,\ldots,v_k$ as:

$$\overline{m}_c^n=\frac{1}{k}\sum_{i=1}^{k} m_c(v_i). \quad (4)$$

- The **standard deviation** σ_c^n of measures of all paths from the root of the tree to the nodes $v_1,\ldots,v_k$ as:

$$\sigma_c^n=\sqrt{\frac{\sum_{i=1}^{k}\left(m_c(v_i)-\overline{m}_c^n\right)^2}{k}}. \quad (5)$$

- The **normalized standard deviation** s_c^n of measures of all paths from the root of the tree to the nodes $v_1,\ldots,v_k$ as:

$$s_c^n=\frac{\sigma_c^n}{\overline{m}_c^n\sqrt{k-1}}. \quad (6)$$

NB: If $\overline{m}_c^n=0$, then we define s_c^n as 1.

Def: For two trees $T_1=(V_1,E_1,e_1)$ and $T_2=(V_2,E_2,e_2)$ we define the **difference between trees** as:

$$\Delta(T_1,T_2)=1-\frac{1}{n+1}\sum_{i=0}^{n}\Delta^n, \quad (7)$$

where Δ^n is given by

$$\Delta^n(T_1,T_2)=\prod_{c\in C}\left(1-\frac{\left|m_c^n(T_1)-m_c^n(T_2)\right|}{m_c^n(T_1)+m_c^n(T_2)}\right)\left(1-\left(s_c^n(T_1)-s_c^n(T_2)\right)\right). \quad (8)$$

NB: If for $m_c^n(T_1)+m_c^n(T_2)=0$, then we take the part of the product for $c \in C$ in equation (8) as 1.

The similarity between trees is then 1-$\Delta(T_1,T_2)$.

3 examples of pairs of tree structures to compare (see figures below).

P1:

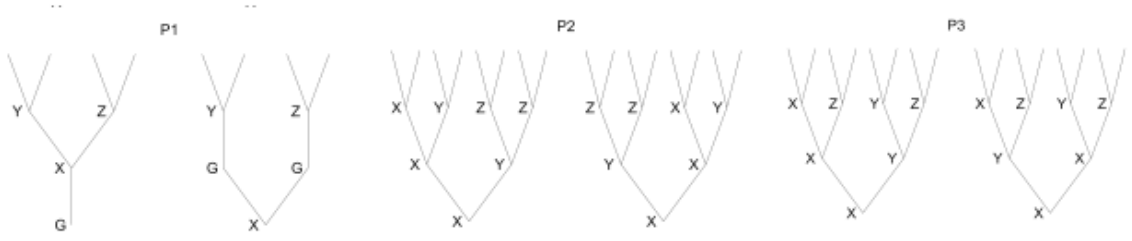
- Uneven number of branches during development leading to the similar final outcome.
- Cell events inverted in time (at level 1 and level 2).
- Similarity score by levels: L0 - 0%, L1 - 67%, L2 - 100%
- Total similarity score of development: 56%

P2:

- Inversion of branches (left and right) at level 2.
- Left and right side inverted.
- Similarity score by levels: L0 - 100%, L1 - 100%, L2 - 100%
- Total similarity score of development: 100%

P3:

- Inversion at level 2, but level 3 is not inverted (does not follow the inversion made at level 2).
- Similarity score by levels: L0 - 100%, L1 - 100%, L2 - 93%
- Total similarity score of development: 98%



Supplemented materials 3. The best zygotic matrices found for the developmental trees, most close to the actual ones, and the corresponding trees.

The presented integers should be read row by row (8 integers in each) to provide a matrix.

Myriophyllum spicatum

Score:12;1.00;1.00;1.00;0.95;0.95;0.94;0.93;0.92;0.90;0.89;0.88;0.87;

InitMatrix:69;7;66;58;43;36;31;67;46;55;33;31;90;25;91;81;20;80;6;99;81;3;35;70;44;95;41;81;62;96;24;35;28;81;82;57;46;93;88;43;96;14;26;48;90;6;25;78;51;14;25;99;72;4;8;29;44;68;98;82;81;8;11;46

Polygala major

Score:12;1.00;0.78;0.80;0.79;0.79;0.80;0.79;0.79;0.79;0.80;0.81;0.81;

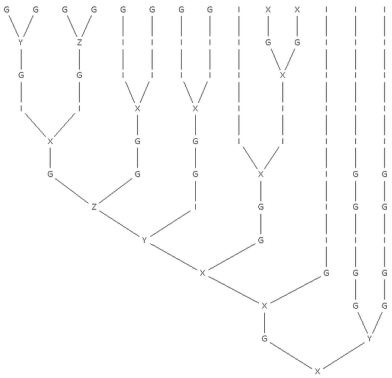
InitMatrix:45;23;6;65;99;4;71;96;85;80;29;17;29;61;97;13;89;9;36;45;93;3;77;63;87;74;11;66;27;53;26;37;51;42;25;40;43;19;46;27;80;44;37;50;83;32;91;98;45;88;20;27;83;16;15;96;48;6;79;28;52;25;98;42

Triglochin palustris

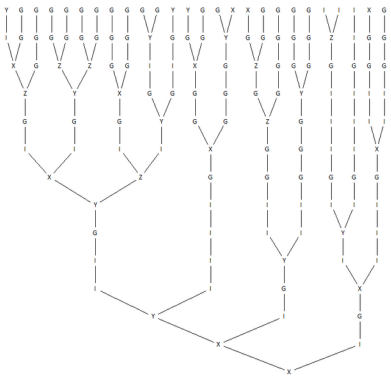
Score:12;1.00;1.00;0.98;0.92;0.88;0.87;0.84;0.84;0.83;0.83;0.82;0.82;

InitMatrix :72;29;97;46;44;37;51;62;55;90;94;19;86;25;93;10;17;30;7;90;94;80;92;74;20;42;45;62;34;2;96;61;32;5;8;4;69;2;41;10;54;47;97;56;44;54;96;67;41;72;25;5;52;37;83;17;23;32;56;2;24;11;43;2

Myriophyllum spicatum



Polygala major



Triglochin palustris