

Application of a morphological similarity measure to the analysis of shell morphogenesis in Foraminifera

Maciej Komosinski¹, Agnieszka Mensfelt¹,
Paweł Topa^{2,3}, and Jarosław Tyszka²

¹ Institute of Computing Science, Poznan University of Technology
Piotrowo 2, 60-965 Poznan, Poland,
maciej.komosinski@cs.put.poznan.pl

² Institute of Geological Sciences, Polish Academy of Sciences
Research Centre in Cracow, Senacka 1, 31-002 Krakow, Poland

³ AGH University of Science and Technology, Department of Computer Science
Al. Mickiewicza 30, 30-962, Krakow, Poland

Abstract. This work evaluates the genotype-to-phenotype mapping defined by one of the models of growth of foraminifera. Foraminifera are simple unicellular organisms with very diverse morphologies. To analyze the mapping, a morphological similarity measure is needed that compares 3D structures. One of the key components of the similarity estimation algorithm is Singular Value Decomposition (SVD). Since this algorithm is heavily used and its performance is important, four SVD implementations have been compared in this work. Distance matrices of the phenotypes obtained for equally distant genotypes were computed using the similarity measure. For the visualization of the phenotype space, multidimensional scaling techniques were used. Visual comparison of the genotype and the phenotype spaces revealed characteristics and potential weaknesses of the analyzed model of foraminifera growth, and demonstrated usefulness of the proposed approach.

Keywords: Similarity, morphology, genotype, phenotype, foraminifera

1 Introduction

Applications of similarity measures for the analysis of three-dimensional constructs range from evolutionary design to artificial life and theoretical biology. Such measures are very useful when a large population of structures needs to be automatically compared. Estimating similarity allows one to classify morphologies, construct hierarchies of morphologies, discover clusters and investigate the correlation between phenotypes and fitness of individuals [8, 7, 9].

Modeling of organism morphogenesis benefits from the use of such a similarity measure as well. Although foraminifera are unicellular organisms whose shells (also called “tests”) are usually smaller than 1 mm, they have very diverse morphology. Taxonomy of those organisms include over 10,000 living and

This paper appeared in *Advances in Intelligent Systems and Computing*, vol. 391, pages 215–224, Springer, 2016.

http://dx.doi.org/10.1007/978-3-319-23437-3_18

fossil species, which are of great interest to biologists and micropalaeontologists. Foraminifera build shells consisting of one or more chambers. In the simplest case, a single chamber contains an opening called the aperture. Openings of subsequent chambers compose the communication line which is called the local communication path. There are two basic chamber morphologies in polythalamous (multichambered) foraminifera: globular and tubular. The models considered in this work are focused on foraminiferal shells composed of globular chambers. This new class of foraminifera is called Globothalamea and is based on results of molecular and morphogenetic modeling [14].

1.1 The Model of Foraminifera Morphogenesis

Models of growth of foraminifera are actively developed [18, 20, 17]. The model used in this work describes foraminiferal morphology with 7 parameters [19, 10] that determine the location and size of subsequent chambers [12]. The growth of a foraminifer starts from a single chamber for which the location of its center, O_0 , is arbitrarily defined. The location of the aperture U_0 is calculated according to the local communication path length minimization principle: it is a point on the surface of the shell which is nearest to the center, O_0 .

In the two-dimensional case, calculation of the new (i -th) chamber center, O_i , starts with the determination of the reference axis, which is the line passing through both the previous and the current aperture. In the first step, O_i is located exactly on the current aperture U_{i-1} . It is then moved along the reference axis according to the translation factor parameter TF . Next, it is deviated from the reference axis according to the deflection angle $\Delta\phi$. The size of the new chamber is determined by the scaling factors K_x and K_y . In the three-dimensional case the procedure is similar, yet there are two additional parameters: the rotation angle $\Delta\beta$ and the scaling factor K_z .

After computing the location and the size of the new chamber, the new aperture can be found – again, it is a point on the surface of the new chamber with the shortest distance to the previous aperture. The new aperture cannot be located inside any previous chamber. The 7 parameters of this model of growth are enumerated in Table 1, and sample foraminiferal morphologies are shown in Fig. 1.



Fig. 1: A sample growth sequence of a foraminifer. The number of chambers increases from 1 to 10. The dark spot is the aperture. Values of growth parameters are shown in Table 1.

Table 1: The parameters comprising foraminifera genotype. Angles are expressed in radians. The TF range ensures that subsequent chambers are joined together.

Symbol	Name	Range from	Range to	Example in Fig. 1
N	Number of chambers	1	15	1,2,4,6,8,9,10
K_x	Scale in x	1.00	1.10	1.00
K_y	Scale in y	1.00	1.10	1.00
K_z	Scale in z	1.00	1.10	1.00
TF	Translation factor	-1.00	1.00	-0.02
$\Delta\phi$	Deflection angle	-3.14	3.14	0.64
$\Delta\beta$	Rotation angle	-3.14	3.14	0.72

The 7 parameters can be considered high-level genes, and the resulting 3D shell morphology corresponds to a phenotype. The process of morphogenesis (growing a phenotype based on a genotype) can be considered a mapping between a genotype and a phenotype [11]. While this model of growth has been introduced to mimic and simulate biological processes, it is important to be able to analyze the genotype-to-phenotype mapping formally – in terms of the relationship between the space of genotypes and the space of phenotypes. This is where similarity measures are required.

Measuring similarity of genotypes is straightforward in case of numbers, as long as we assume that the traditional meaning of similarity of values is reasonable (the lower the difference between values, the higher their similarity). To measure similarities between morphologies, we need a procedure that compares three-dimensional structures. The concept that describes the correspondence between differences in genotypes and phenotypes is called *locality* [16]. High locality is obtained when neighboring genotypes are expressed as neighboring phenotypes. The locality has impact on the relative topology of both spaces, and in consequence, on optimization techniques and evolutionary processes.

2 Morphological Similarity Measure

2.1 The Algorithm

The similarity measure used to estimate differences in foraminiferal morphologies considers 3D structures as undirected graphs [8, 7, 9]. In case of foraminifera, each chamber constitutes one vertex, and all vertices are connected with edges producing a linear graph structure (Fig. 2). The similarity estimation algorithm consists of three main steps: alignment of the two structures that are compared, construction of the matching function, and calculation of the dissimilarity components. Two components are taken into account when computing distance between vertices: difference in vertex degree ($iDeg$) and geometrical distance ($iGeo$). The importance of each of the components can be adjusted using weights.



Fig. 2: A linear graph (10 vertices and 9 edges) representing the 10-chamber foraminiferal morphology from Fig. 1.

The approach taken for the alignment is based on the distribution of points in a three-dimensional space. To position the morphologies according to the distributions of vertices, the SVD transform [13] is used. It is applied separately to both structures. After the transform is computed, the center of the structure is located in the origin of the coordinate system. The axis with the highest variance becomes the first axis of the structure, and the axis with the second highest variance becomes the second axis of the structure. This method was proved to provide good alignment for the geometry of the structures [9].

The matching algorithm is a heuristic. Vertices in both structures are sorted by vertex degree in a descending order. The vertices with the same degree are grouped together. The procedure starts with groups of vertices having the highest degree in each structure. The algorithm tries to find a match (the least distant vertex from the other structure) for the vertices which are unmatched yet in both structures, starting from the vertices with the lowest indexes. When all of the vertices from a group are matched, the next group is taken into account. Once the matching function is constructed, the overall dissimilarity between the two compared structures can be determined.

It is desired for the dissimilarity measure to be a metric. For this purpose, it must satisfy non-negativity, identity of indiscernibles, symmetry and triangle inequality. The similarity measure outlined above always satisfies the first three conditions. The last condition – the triangle inequality – can be extremely rarely violated when the *iGeo* component is considered [9].

2.2 Comparison of Performance of Four SVD Implementations

Since the SVD transform is frequently computed during estimation of morphological similarity, its performance is important. For this reason, we have compared four C++ implementations of SVD: two open-source libraries, a routine from Numerical Recipes, and the MATLAB® Math Library. Table 2 lists these libraries.

Fig. 3 compares the performance of the four SVD implementations on various 3D structures. The performance was similar for two sets containing morphologies with the low average number of vertices (Figs. 3a and 3b); MATLAB and Eigen libraries were slightly slower. For the set with high average number of vertices, MATLAB library outperformed other libraries and the Eigen library

Table 2: Comparison of libraries/routines for SVD.

Name	Type	Description
OpenCV	BSD license	Open-source library for computer vision [3]
Eigen	MPL2 license	Template C++ library [1]
NR	Commercial	Numerical Recipes [15, 4]
MATLAB	Commercial	C++ interface for a computing environment [2]

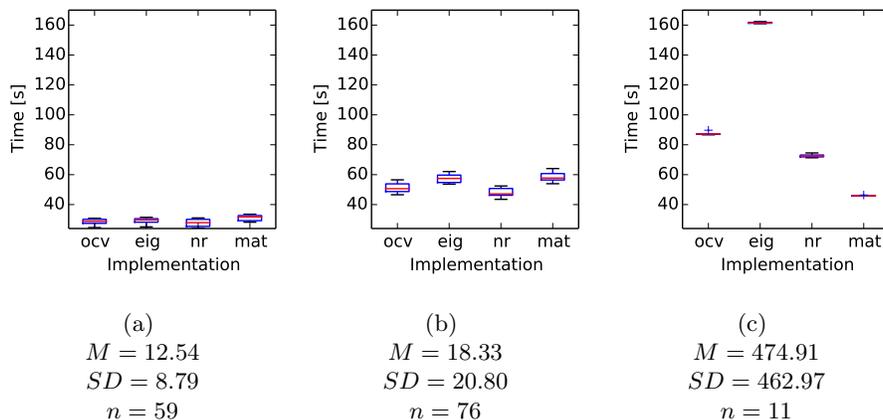


Fig. 3: Time of calculating the full distance matrix ($n \times n$) for three sets of morphologies of increasing complexity. M and SD are the mean and the standard deviation of the number of vertices in a set containing n morphologies. All tests were performed on Intel Core i7-4770 with 8 GB of RAM running Windows 7. The one-thread program was compiled with Visual Studio 2014.

was significantly slower than other libraries (Fig. 3c). Since foraminifera have simple morphologies that usually have no more than 20 vertices, both NR and OpenCV were considered best choices in terms of speed and ease of use.

3 Application of Similarity Measure to the Analysis of the Genotype–Phenotype Mapping in Foraminifera

To evaluate the characteristics of the genotype-to-phenotype mapping in foraminifera morphogenesis, the relationship between genotype and phenotype spaces was analyzed. 5 out of 7 genes were kept constant, while the remaining two were varied from the minimal to the maximal value (Table 3). For the purpose of the analysis of each pair of parameter values, $32 \times 32 = 1024$ genotypes and the same number of corresponding phenotypes were generated (we used 32 evenly spaced values for each of the two varied parameters).

For the visualization of the phenotype space, multidimensional scaling was employed [6]. The distance matrix that resulted from estimating similarity of all

Table 3: Parameter values used in the analysis. Constant values are indicated by a single number. Ranges of the varied parameter values are shown in square brackets.

Visualization	N	K_x	K_y	K_z	TF	$\Delta\phi$	$\Delta\beta$
Fig. 4	5	1	1	1	-0.1	$[-3.14; 3.14]$	$[-3.14; 3.14]$
Fig. 5	5	1	1	1	$[-0.99; 0.99]$	$[-3.14; 3.14]$	0
Fig. 6	5	1	1	1	$[-0.99; 0.99]$	0	$[-3.14; 3.14]$

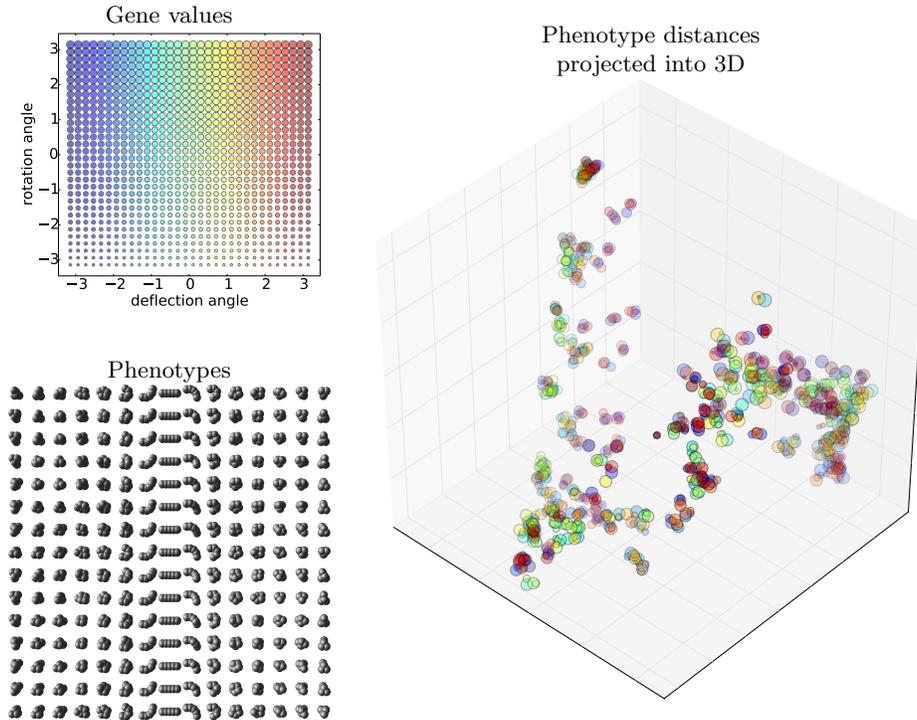


Fig. 4: The relationship between the genotype and the phenotype spaces for different values of deflection angle $\Delta\phi$ and rotation angle $\Delta\beta$. The number of phenotypes shown in the grid is reduced from 32×32 to 15×15 for legibility. The 3D projection of the phenotype distance matrix preserves 62% of total variance.

pairs of phenotypes was projected into three dimensions where the Euclidean distances best approximate the original distance matrix. These 3D coordinates were plotted using different colors and sizes that correspond to variable parameter values (genes). A small random jitter was added to 3D coordinates to avoid identical morphologies (with zero dissimilarity) to be plotted one over the other, and to expose their density.

Phenotype grid in Fig. 4 reveals symmetry of the deflection and rotation angles. It can also be seen in the 3D distances plot – morphologies corresponding

to the same absolute values of the angles are grouped together. For instance, on the right side of the distance plot, groups comprising of two small and two big circles are visible. Sizes of the circles in those groups correspond to extreme rotation angle values, and colors correspond to opposite deflection angle values.

Extreme values of the angles yield the deflection of about $+180$ and -180 degrees, both of which locate the center of the new chamber in a similar position. This indicates that for locality to be preserved and for successful search in the genotype space, the operation of modifying both angles should respect their cyclic nature, and for the two angles, the genotype grid should be considered a torus.

Phenotype grid in Fig. 5 demonstrates that the positive translation factor and the extreme deflection angle values produce similar morphologies to the negative translation factor and the deflection angle values close to 0. In the 3D distances plot, this is represented by lines comprised of red and blue circles (extreme values of the deflection angle) and of light blue and green circles (deflection angle values near 0). Morphologies corresponding to the extreme deflection angle and the

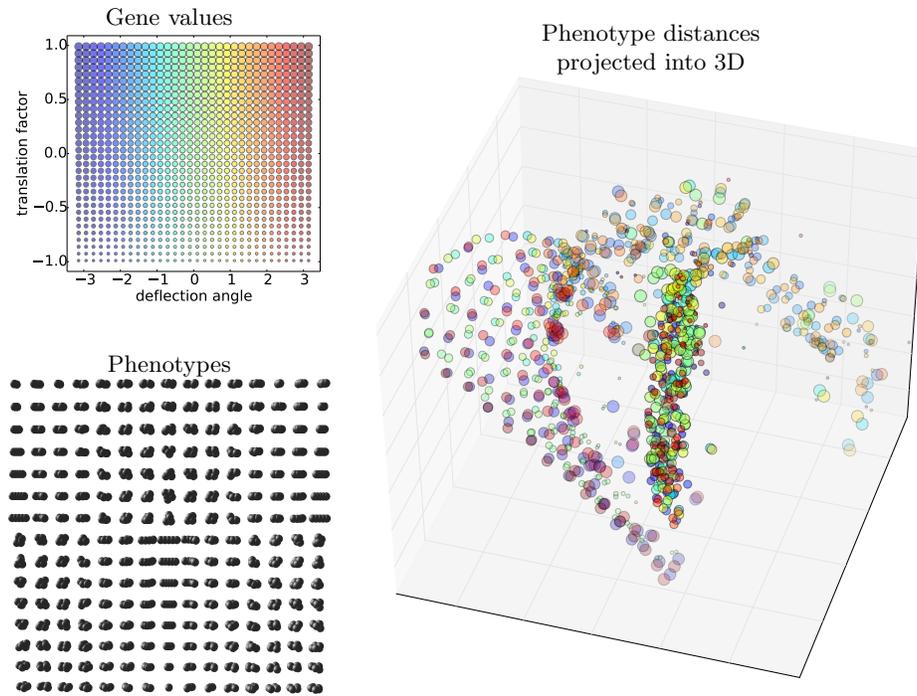


Fig. 5: The relationship between the genotype and the phenotype spaces for different values of translation factor TF and deflection angle $\Delta\phi$. The number of phenotypes shown in the grid is reduced from 32×32 to 15×15 for legibility. The 3D projection of the phenotype distance matrix preserves 47% of total variance.

negative translation factor, or the deflection angle value near 0 and the positive translation factor, are more similar to each other – in the 3D distance plot they form a dense, elongated cluster. However, phenotypes with small positive and negative values of translation factor that are very close in the genotype space are mapped into distant phenotypes (different morphologies) – this discontinuity is clearly visible in the phenotype grid.

Fig. 6 reveals two groups, one of which is dense and the other is sparse. Closer analysis of both groups indicates that for $TF \leq 0$, the value of the rotation angle has no influence on morphology. Such morphologies form the linear dense group, where their geometries are different only because of different values of translation factors. Morphologies in the sparse group ($TF > 0$) depend on both the translation factor and the rotation angle. The presence of such distinctive groups and the lack of smooth transition between them indicates that there is a discontinuity in the interpretation of gene values, and this would be disadvanta-

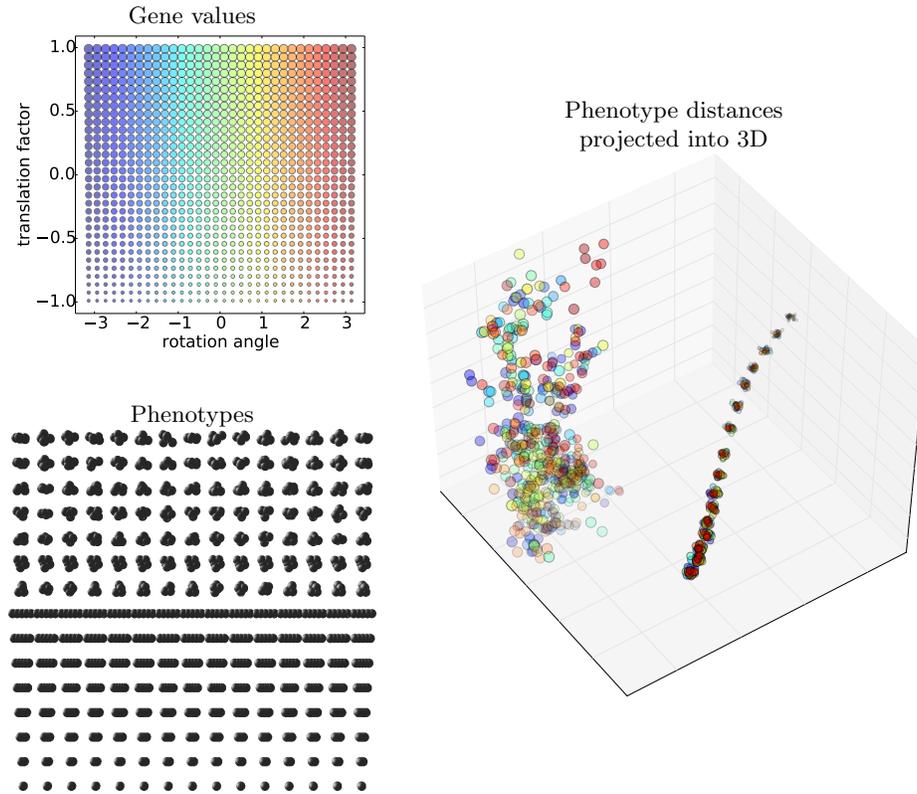


Fig. 6: The relationship between the genotype and the phenotype spaces for different values of translation factor TF and rotation angle $\Delta\beta$. The number of phenotypes shown in the grid is reduced from 32×32 to 15×15 for legibility. The 3D projection of the phenotype distance matrix preserves 53% of total variance.

geous from the search and optimization point of view. An interesting discovery is the fact that rotation angle has any influence at all for $TF > 0$. Since deflection angle is zero in this experiment, based on the theoretical model, rotation should not have any influence on morphology. However, the implementation uses a finite number of point samples on the chamber spheres to find the communication path with minimal length, and the analysis presented here may have discovered an artifact caused by this sampling.

4 Conclusions

Visual comparison of the genotype and the phenotype spaces performed for three pairs of parameters revealed characteristics and potential weaknesses of the foraminifera model of morphogenesis. Genotypes with extreme values of translation and rotation angles correspond to similar morphologies. Genotypes with equally distributed values of rotation angle and translation factor are mapped into two distinct groups of morphologies with no smooth transition possible. Although further investigation of the model is needed, preliminary results reported here suggest low locality of the mapping. Note however that the mapping was not devised with optimization in mind, it was rather expected to model biological reality in an extremely simple way, using just a few key parameters. Low locality may also be a property of biological genotype-to-phenotype mappings [5] and as such, it may be considered a feature that should be included in the model, not a disadvantage.

There is a great potential for application of this methodology to real organisms, although there are some challenges to tackle. The fundamental problem is that high-level genes are in reality represented by complex genetic and epigenetic processes responsible for morphogenesis [18]. A more realistic approach could take into account real molecular genetic data based on DNA, RNA, or protein sequences [14, 20]. Real morphologies of foraminiferal shells are also based on chamber arrangement patterns – however, chambers are not defined by their theoretical centers. The most promising method to test would be to follow apertures and foramina that form graph-like communication lines reconstructed based on high-resolution X-ray computed tomography.

A more detailed analysis of the model of foraminifera growth is needed, including a numerical evaluation of the genotype–phenotype mapping. The results of such evaluation increase the understanding of the relationships between genes and phenes, and facilitate the development of an improved model of foraminifera morphogenesis.

Acknowledgment

The research presented in the paper received partial support from Polish National Science Center (UMO-2013/09/B/ST10/01734).

References

- [1] Eigen library. <http://eigen.tuxfamily.org>.
- [2] Matlab. <http://www.mathworks.com/products/matlab/>.

- [3] OpenCV library. <http://opencv.org/>.
- [4] SVD implementation. <http://www.nr.com/webnotes/nr3web2.pdf>.
- [5] Jacobo Aguirre, Javier M. Buldú, Michael Stich, and Susanna C. Manrubia. Topological structure of the space of phenotypes: the case of RNA neutral networks. *PloS one*, 6(10):e26324, 2011. doi:10.1371/journal.pone.0026324.
- [6] Trevor F. Cox and M.A.A. Cox. *Multidimensional Scaling, Second Edition*. Chapman and Hall/CRC, 2000.
- [7] Maciej Komosinski, Grzegorz Koczyk, and Marek Kubiak. On estimating similarity of artificial and real organisms. *Theory in Biosciences*, 120(3-4):271–286, December 2001. doi:10.1007/s12064-001-0023-y.
- [8] Maciej Komosinski and Marek Kubiak. Taxonomy in Alife. Measures of similarity for complex artificial organisms. In Jozef Kelemen and Petr Sosík, editors, *Advances in Artificial Life*, volume 2159 of *LNCS*, pages 685–694. Springer, Berlin/Heidelberg, 2001. doi:10.1007/3-540-44811-X_79.
- [9] Maciej Komosinski and Marek Kubiak. Quantitative measure of structural and geometric similarity of 3D morphologies. *Complexity*, 16(6):40–52, 2011. doi:10.1002/cplx.20367.
- [10] Maciej Komosinski, Agnieszka Mensfelt, Paweł Topa, Jarosław Tyszka, and Szymon Ulatowski. Foraminifera: genetics, morphology, simulation, evolution. <http://www.framsticks.com/foraminifera>, 2014.
- [11] Maciej Komosinski and Szymon Ulatowski. Genetic mappings in artificial genomes. *Theory in Biosciences*, 123(2):125–137, September 2004. doi:10.1016/j.thbio.2004.04.002.
- [12] Paweł Labaj, Paweł Topa, Jarosław Tyszka, and Witold Alda. 2D and 3D numerical models of the growth of foraminiferal shells. In *Proceedings of the 1st International Conference on Computational Science: Part I*. Springer-Verlag, 2003.
- [13] Carl Meyer. *Matrix Analysis and Applied Linear Algebra*. SIAM: Society for Industrial and Applied Mathematics, February 2001.
- [14] Jan Pawłowski, Maria Holzmann, and Jarosław Tyszka. New supraordinal classification of foraminifera: Molecules meet morphology. *Marine Micropaleontology*, 100:1–10, 2013.
- [15] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, New York, NY, USA, 3rd edition, 2007.
- [16] Franz Rothlauf. Chapter 3.3: Locality. In *Representations for genetic and evolutionary algorithms*, pages 73–95. Springer-Verlag, 2006.
- [17] Paweł Topa, Maciej Komosinski, Maciej Bassara, and Jarosław Tyszka. eVolutus: a configurable platform designed for ecological and evolutionary experiments tested on Foraminifera. In Aleksandra Gruca et al., editors, *Proceedings of International Conference on Man-Machine Interactions ICMMI 2015*, volume 391 of *Advances in Intelligent Systems and Computing*. Springer, 2015. doi:10.1007/978-3-319-23437-3_23.
- [18] Paweł Topa, Jarosław Tyszka, Samuel S. Bowser, and Jeffrey L. Travis. DPD model of foraminiferal chamber formation: Simulation of actin meshwork–plasma membrane interactions. In *Parallel Processing and Applied Mathematics, 9th International Conference, Poland, LNCS vol. 7204*, pages 588–597. Springer, 2012.
- [19] J. Tyszka and P. Topa. A new approach to modeling of foraminiferal shells. *Paleobiology*, 31(30):526–541, 2005.
- [20] Jarosław Tyszka et al. Morphogenetic bridge between molecules and morphology. In Tatiana Hromic Margarita Marchant, editor, *International symposium on Foraminifera FORAMS 2014, Chile*. The Grzybowski Foundation, 2014.