

Article

Spectral Decomposition of Mappings of Molecular Genetic Information in the System Basis of Single Nucleotide Functions

Ivan Stepanyan ^{1,2,*}  and Michail Lednev ² ¹ Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya Street, 117198 Moscow, Russia² Mechanical Engineering Research Institute of the Russian Academy of Sciences (IMASH RAN), M. Kharitonyevskiy Pereulok, 101990 Moscow, Russia; miklesus@mail.ru

* Correspondence: stepanyan_iv@pfur.ru or neurocomp.pro@gmail.com

Abstract: This paper presents and visualizes examples of large amounts of genetic information using a new class of cognitive computer graphics algorithms. These algorithms are related to the semiotics of perception and allow the interpretation of those properties of nucleotide sequences that are difficult to perceive by simple reading or by standard means of statistical analysis. This article summarizes previously presented algorithms for visualizing long nucleic acids based on the primary Hadamard–Walsh function system. The described methods allow us to produce one-dimensional mappings of nucleic acids by levels corresponding to their scale-integral physicochemical parameters and construct a spectral decomposition of the nucleotide composition. An example of the spectral decomposition of parametric representations of molecular genetic structures is given. In addition, a multiscale composition of genetic functional mappings visualizing the structural features of nucleic acids is discussed.

Keywords: big data; DNA; spectral analysis; decomposition; visualization; chromosomes; nucleotides; geometry



Citation: Stepanyan, I.; Lednev, M. Spectral Decomposition of Mappings of Molecular Genetic Information in the System Basis of Single Nucleotide Functions. *Symmetry* **2022**, *14*, 844. <https://doi.org/10.3390/sym14050844>

Academic Editor: Miroslav Miletin

Received: 18 February 2022

Accepted: 15 April 2022

Published: 19 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

DNA and RNA sequences are fundamental to the coding and processing of genetic information. Genetic information can be interpreted not only from character sequences but also from hidden signals within individual sequences [1]. At the same time, biosemiotics is concerned with the study of sign systems in living nature, in the context of which semiotics of text generation and semiotics of text perception are considered. The working hypothesis is that any character sequences can be transformed into numerical ones according to certain rules and then visualized in spaces of binary-orthogonal functions to reveal hidden regularities.

One of the problems of modern bioinformatics is the task of automatic annotation and the visualization of genetic information for the systematization, classification, and in-depth analysis of molecular genetic data. At the same time, the use of numerical methods is of great importance for bioinformatics. Existing methods of genetic sequence analysis are methods mainly related to statistical analysis. Therefore, among software products for the analysis of genetic nucleotide sequences (we mean data of the AGGCT . . . type, obtained from DNA of living organisms and stored in files or databases), algorithms based on statistical analysis are mainly presented.

Many publications are devoted to the problem of applying the latest mathematical tools to the analysis of genetic information. Eigenvectors derived from principal component analysis (PCA) are commonly used to estimate genealogy. In [2], the authors used connections between multidimensional scaling and spectral graph theory to propose an alternative to PCA. PCA cannot provide a more relevant characterization of ancestors than a spectral embedding approach created from a normalized Laplacian graph. On the other hand, the authors of [3] developed a method that groups genes and conditions by finding

distinctive patterns in gene expression data matrices. In addition, eigenvectors can be easily identified using linear algebra approaches, in particular, singular value decomposition. The authors then applied spectral biclustering to a sample of data sets.

In [4], a Fourier spectrogram based on a binary representation of the nucleotide sequence of the virus RNA genome was analyzed. The authors found that at low frequencies, the power spectrum deviates slightly from the expected behavior of a statistically independent sequence. Except for a small low-frequency region, the spectrum is dominated by random fluctuations around fixed values, and only one additional peak is associated with the use of triplet codons. The authors reached an important conclusion that opens an intriguing new scaling rule for the coronavirus genome: the genome's structure scales linearly with the power-law exponent, which characterizes the low-frequency region of the spectral density.

Numerous new works are devoted to the detection and visualization of the genome in DNA. This refers to both genomic targeting technologies and technologies for building graphical visual models. The relationship between the coding regions of the plant genome and the distribution of energy during the photochemical process of photosynthesis is analyzed in [5]. To identify resistant strains using the genomes of *Mycobacterium tuberculosis* (MTB), the authors applied a technique for genomic similarity analysis that links different levels of genome decomposition (discrete non-decimated wavelet transform) and the Hurst exponent [6]. To analyze data on the same samples from multiple sources, a new model of related components was created in [7], which directly includes partially shared structures. The method demonstrates excellent performance in signal evaluation and component selection. In [8], algorithms were developed to build three-dimensional (3D) genome models related to the structure of chromatin, considering the limitations of computational methods to reveal the building blocks of genome architecture. In [9], the authors presented a tool by which target genes can be efficiently and conditionally knocked out by editing the genome at any stage of development.

The article [10] presents an intuitive version of the method for the visual 3D representation and analysis of the adjacency of chromosome territory. The authors of the article proposed a cascade neural network architecture for processing background noise in chromosome images based on the noise reduction method [11]. This study does not rely on a sub-alphabetic system.

Researchers from [12] developed a method to extract three-dimensional information from two-dimensional records over time and created a 3D model of the X-chromosome. In [13], the authors presented a multiscale polymeric model of chromatin covering phase separation at the megabase and physics at the nucleosome level. Plotsr [14] is a tool for visualizing structural similarities and rearrangements across various genomes. It can be used to compare genomes on a chromosome-by-chromosome basis.

2. Materials and Methods. Algorithms for Visualization of Nucleic Acids in One-Dimensional Spaces of Physicochemical Parameters

As is well known, DNA is a nucleotide sequence, a system, which is a double helix that has many mathematical properties and symmetrical patterns [15,16]. The chemical formulas of purine and pyrimidine bases are given to demonstrate the Boolean properties of their physicochemical parameters: each nitrogenous base of the genetic code has three variants of its Boolean representation. S.V. Petukhov called these representation variants "binary sub-alphabets" [15]. These representations differ in the types of Boolean properties in the set of nitrogenous bases:

- G = C "3 hydrogen bonds"/A = T "2 hydrogen bonds";
- C = T "pyrimidines"/A = G "purines";
- A = C "amino"/G = T "keto".

This organization of the binary properties of chemical compounds constitutes a system of Walsh–Hadamard basic orthogonal functions, which is embedded in any DNA or RNA molecule by the listed molecular attributes of its constituent nucleotides. De-

tailed information about symmetries and Hadamard matrices in genetic coding, as well as genetic algebras, is detailed in the works of biomathematician S. V. Petukhov (see, for example, [16]). In view of the close connection between algebra and geometry (and hence the presence of a connection between genetic algebras and genetic geometries), the task of developing a method for nucleic acid visualization is set and solved. The following research is based on the hypothesis that visualization should reflect the symmetries of the nucleotide composition in some discrete space.

The motivation for this work is the further development of the previously published methods in the field of discrete geometry and DNA algorithms. As it is known, geometry is built on axiomatics; in this regard, any geometry can be used to represent DNA parameters. Riemannian geometry is of particular interest because it is reflected in cosmology and biomorphogenesis.

A class of algorithms for multiscale visualization of genetic information, which we developed earlier in [17], is implemented using the described system of binary orthogonal Walsh–Hadamard functions encoding physical and chemical features of DNA. This algorithm can display DNA in different parametric spaces.

Let us recall the main ideas of the algorithm. A more detailed version is published in [17]. DNA is divided into equal fragments of N nucleotides in length (N -plets). Each of the N -plets has its own binary representation over all its sub-alphabets. These representations define the coordinates of points in different parametric spaces. The parameters of the representation can be frequencies, decimal values of codes, or the number of some elements in N -plets. The author proposes to call such representations “genometric”. Mappings are specified by coordinates in parameter spaces (visualization spaces or parametric spaces). These algorithms and their generalizations can be called molecular-genetic or DNA algorithms to emphasize their difference from the well known genetic algorithms of the Holland type.

The Boolean properties of the set of nitrogenous bases organize the system of Walsh basis functions shown in Figure 1. Each of the functions corresponds to its own coordinate X , Y , and Z , which will be used in further visualizations.

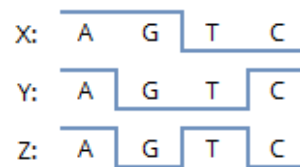


Figure 1. Walsh’s system of nucleotide base functions.

It should be noted that to standardize studies, it is recommended to agree on a unified coding method so that the obtained visualizations coincide with each other. Otherwise, rotations and reflections may be observed in the obtained visualizations. At the same time, the binary sub-alphabets are interconnected by the operation of addition modulo two and define a space with properties in which the coordinates of each point are interrelated.

The use of one-dimensional coordinate axes $\{X, Y, Z\}$ using sub-alphabets yields three different mappings. The depth of recorded changes is determined by a scalable parameter N , which sets the partitioning of the nucleotide sequence into fragments of equal length, N -plets. Figure 2 show the application of DNA algorithms previously described by us in [17]. Thus, the basic algorithms were used to construct Figure 2.

In Figure 2a, one axis is used to display the encoding of the physicochemical parameters of one of the sub-alphabets (parametric ordinate axis). The abscissa axis encodes the serial number of the N -plet in the sequence, while the ordinate axis encodes the decimal values (parameters) of the binary representation of each N -plet, ordered in ascending order. Thus, Figure 2b show an integral one-dimensional visualization of the total number of units in the codes of N -plets in the genetic sequence.

Figure 2c,e are two-dimensional mappings in which each of the axes corresponds to the decimal values of N -plets in the X - Y and X - Z sub-alphabets, respectively. Figure 2d,f

use identical parameters, but instead of decimal values, the numbers of units in N-plets are displayed. In Figure 2 and below, the darker areas correspond to the concentration of N-plets in the corresponding regions.

The resulting mappings make it possible to evaluate changes in the nucleotide composition when a given fragment of a molecule is fully read. As can be seen, the regions of the chromosome with different physicochemical parameters have an individual character (which can be traced in various types of visualization). Thus, the peculiarities of differences in the nucleotide composition become visible.

It should be noted that any error in the coding of binary-oppositional features entails the construction of visualizations in which noise artifacts appear, such as Sierpinski triangles or other structures. Thus, the coding option based on the correct interconnected set of orthogonal functions given by the parameters of four nucleotides is biomathematically justified.

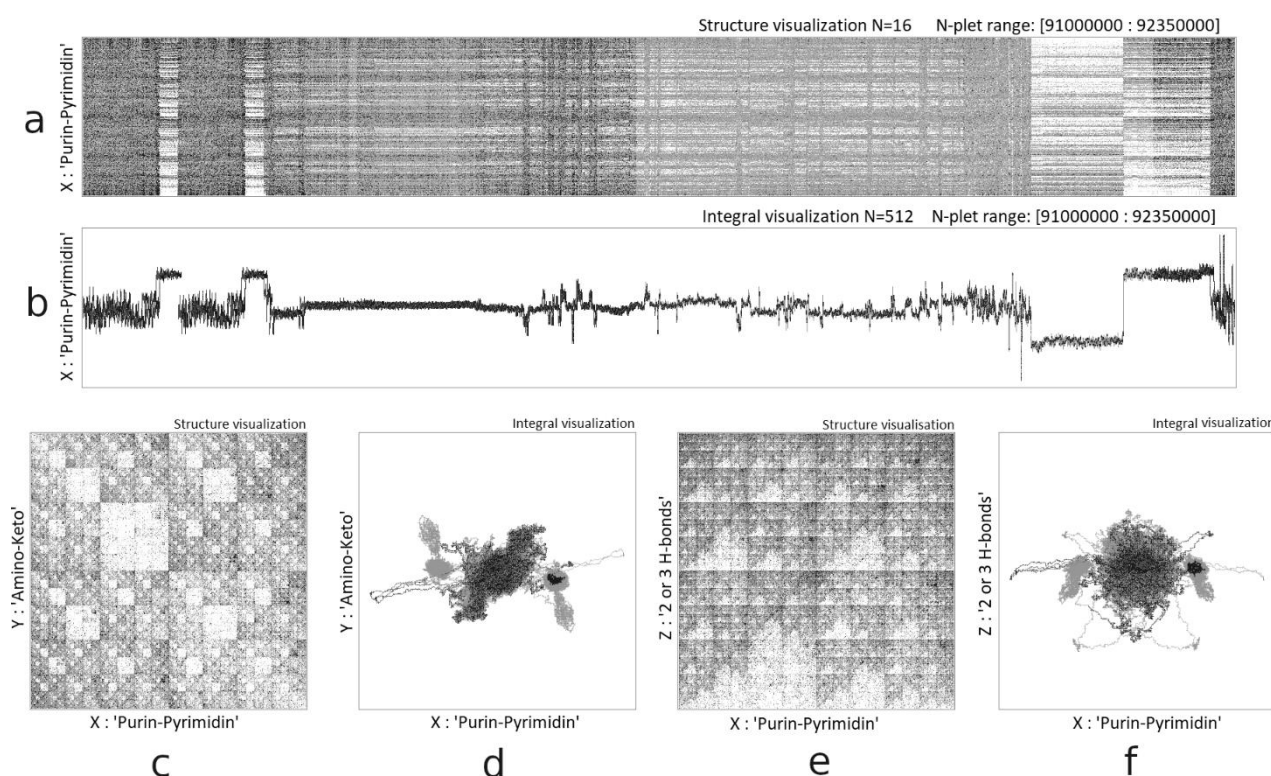


Figure 2. Various examples of visual mappings (projections) constructed using the basic algorithm of a long DNA fragment of a living organism: (a) one-dimensional structural representation in purine-pyrimidine sub-alphabet; (b) one-dimensional integral representation over purine-pyrimidine sub-alphabet; (c), two-dimensional structural representation over sub-alphabets keto-amino (X) and purine-pyrimidine (Y); (d), two-dimensional integral representation of keto-amino (X) and purine-pyrimidine (Y) sub-alphabets; (e), two-dimensional structural representation over hydrogen bond number (X) and purine-pyrimidine (Y) sub-alphabets; (f), two-dimensional integral representation over hydrogen bond number (X) and purine-pyrimidine (Y) sub-alphabets.

3. Spectral Decomposition and Multiscale Composition

Let us consider two modifications of the basic algorithm: spectral decomposition and multiscale composition of basic one-dimensional mappings. For the spectral decomposition of the mappings in Figure 2a,b, we will duplicate along the ordinate axis N times from value 1 to N. On each of the N levels, we will display only those parameters, the number of units in which is N. As a result of the mapping, the one-dimensional structural representation in the sub-alphabet will take the form shown in Figure 3.

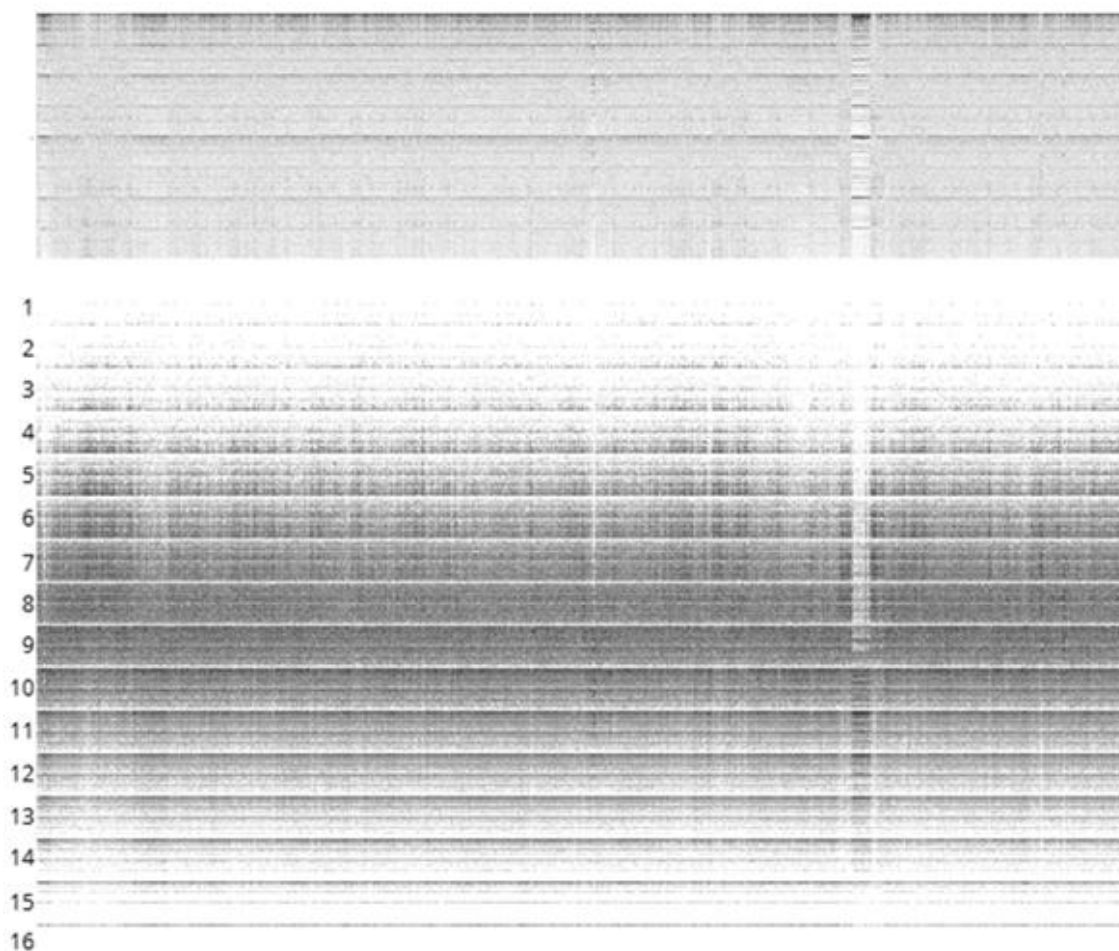


Figure 3. One-dimensional structural mapping of the *Drosophila melanogaster* X chromosome according to the hydrogen bond sub-alphabet (upper row) and spectral mapping (below). $N = 16$, length is 23,542,271 nucleotides.

Spectral decomposition makes one-dimensional visualizations more visible by highlighting regions with different nucleotide compositions within the chromosome. In this case, an analog of spectral decomposition by mass is implemented since the number of units counted in the display determines the number of purines, pyrimidines, or other physicochemical parameters by the sub-alphabetic coding system.

The presented version of one-dimensional parametric visualization allows us to display the composition of a molecule in its spatial arrangement, which is in a sense closer to the physical space than to the two-dimensional parametric space since the DNA molecule (more precisely, one of the two chains) in this case is represented by each of its sub-alphabets along the entire length. This display method retains a partial analogy with the physical space since the X-axis is used to represent the order number of the N-plet in the nucleotide sequence.

For multiscale composition and visualization in any sub-alphabet, one-dimensional mappings with different scales can be positioned relative to each other according to the following principle. The mapping width of N-plets at each scale is such that all nested N-plets are located strictly below the corresponding N-plets of the previous level, implementing the nesting principle. As a result of multiscale visualization, it is possible to trace the fractal ordering of genome structure on different scales (Figure 4).

It should be noted that one-dimensional visualization methods have certain advantages over two-dimensional ones since they make it possible to evaluate the features of changes with the possibility of binding to specific fragments of these molecules for a more detailed analysis of the individual fragments of the molecule. In this regard, one-

dimensional visualization methods seem very promising as a tool for further research in bioinformatics and comparative genomics.

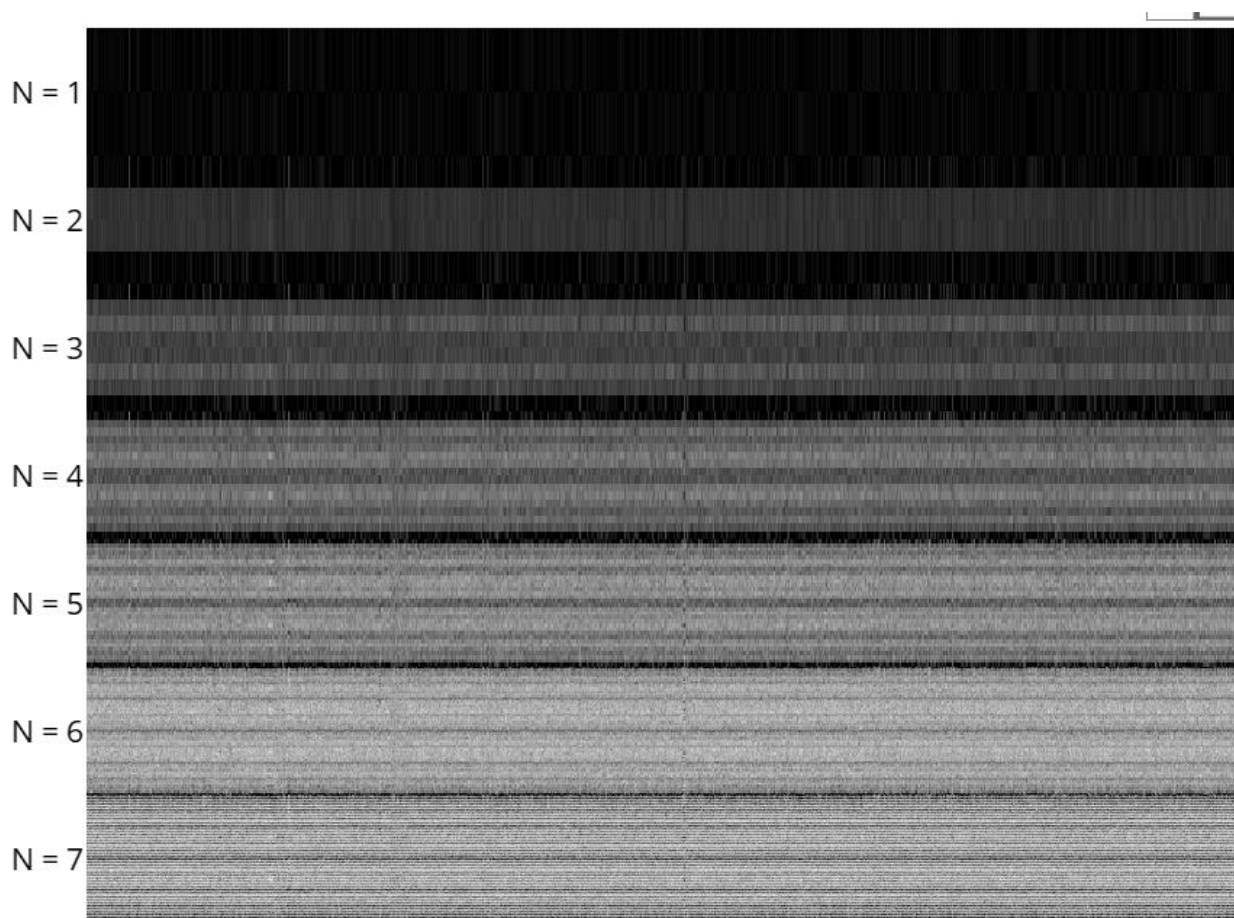


Figure 4. Multiscale composition of linear mappings of binary representations by the sub-alphabet of hydrogen bonds of the first chromosome of *Morella rubra* cultivar (first 21,342,335 nucleotides).

The interpretation of DNA/RNA properties based on the family of presented algorithms is possible because each display point encodes physicochemical properties such as purine/pyrimidine, 2/3 hydrogen bonds, and keto or amino groups. The coordinates of the points corresponding to DNA/RNA fragments of length N nucleotides are completely one-to-one specified by the functions described in the mapping algorithm. Therefore, the interpretation of mappings can only be unambiguous and related precisely to the physicochemical parameters of molecules. In this case, the internal ordering of the molecules, which often has a quasi-fractal character, manifests itself.

4. Cyclic Mappings

We will consider the following class of mappings related to cyclic operations and non-Cartesian coordinate systems. In this class, the mappings take into account the same parameters: subalphabetic coding and the number of units in N -plets. To obtain a new mapping type, let us assign to each N -plet a specific radius vector centered at the point $(0; 0)$ (a certain radius vector centered at $(0; 0)$). The angle of the radius vector lies in the interval $[0; 2\pi)$, and it divides the circle in steps of $2\pi/2^N$ into equal parts so that $2\pi = 2^N$. Thus, the circle makes it possible to display all variants of N -plets as radius vectors. We set the length of each radius vector equal to the number of units in the N -plet. Examples of organism DNA visualizations for different values of parameter N are shown in Figure 5. This figure shows color visualization, where instead of a grayscale corresponding to the frequency of occurrence (intensity) of each point, a color palette is used, in which the most

intense frequencies are indicated by red color, and the least intense by blue. It is possible to trace a specific display structure resembling a trace from the impact of an acoustic signal in the form of the so-called Chladni figures with asymmetric elements [18].

The described procedure, in which a particular radius vector is plotted from the point of origin, will be called the basic procedure. The basic procedure in Figure 6 was used only once for one sub-alphabet. Suppose that we choose the point obtained by applying another sub-alphabet for the same N-plet as the origin of coordinates for the first sub-alphabet. In this case, we will obtain the visualizations shown in Figure 6.

If in the basic algorithm, which is used only for one sub-alphabet, instead of the number of units, the value of the N-plet is used as the radius, then we are able to obtain the mapping, an example of which is shown in Figure 7a.

Morella rubra cultivar Y2012-145 chromosome 1, projection XY

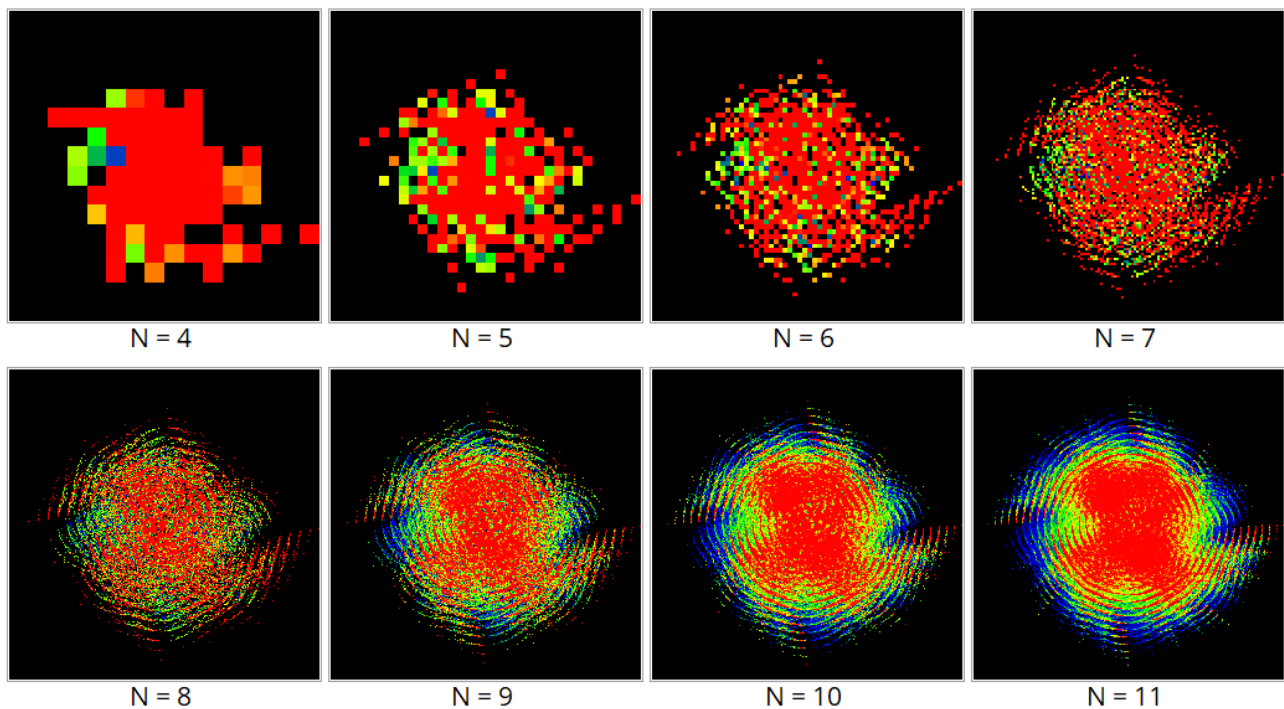


Figure 5. Chromosome of *Morella rubra* cultivar, length 30,276,728. The radius is equal to the number of units in the N-plet.

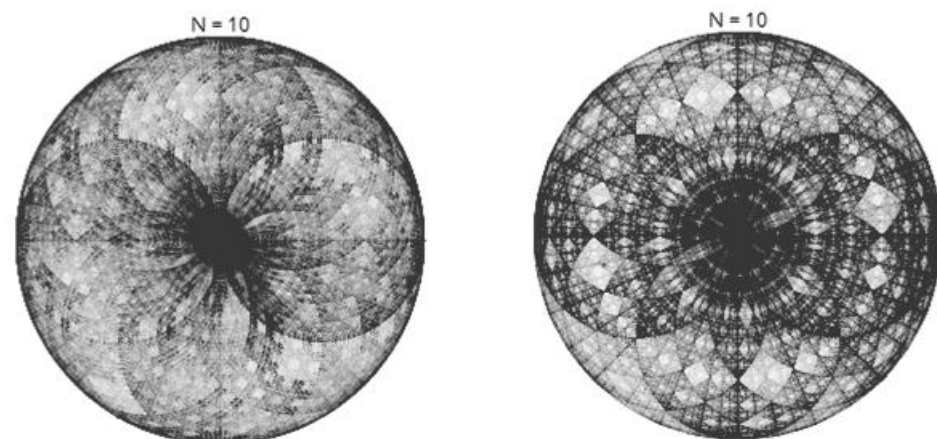


Figure 6. Ring projections of the first chromosome of the *Morella rubra* cultivar, length 30,276,728, AGtc-aGTc pair at N = 10.

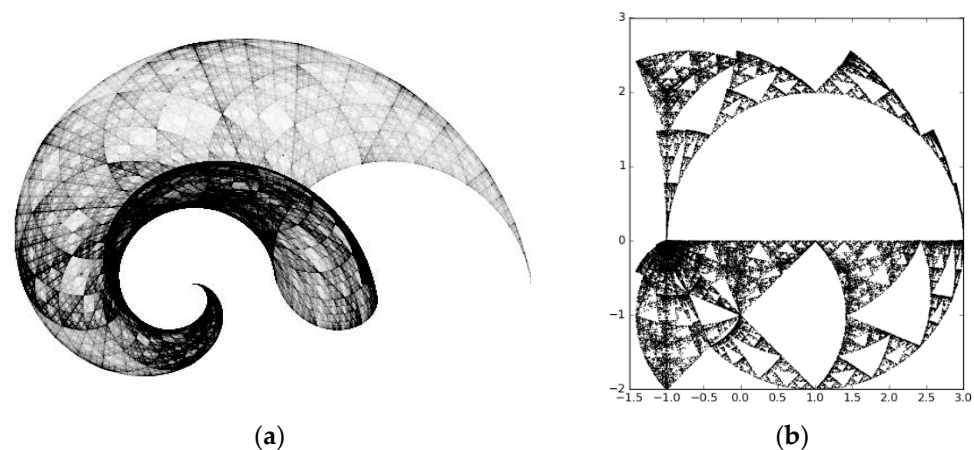


Figure 7. The radius is equal to the size of the N-plet (a). *Morella rubra* cultivar, length 30,276,728, pair AGTc-aGTc, $N = 16$ and Sierpinski Quasi-Triangle (visualization of the ring, (b)).

If the basic procedure is applied three times to all N-plets with a single radius (for all three sub-alphabets), we obtain the object shown in Figure 7b. However, if we apply the basic algorithm to the third sub-alphabet with a unit radius in the orthogonal plane, with the transition to the third dimension, we obtain the 3D visualizations that are shown with different scales in Figure 8 (for all three sub-alphabets).

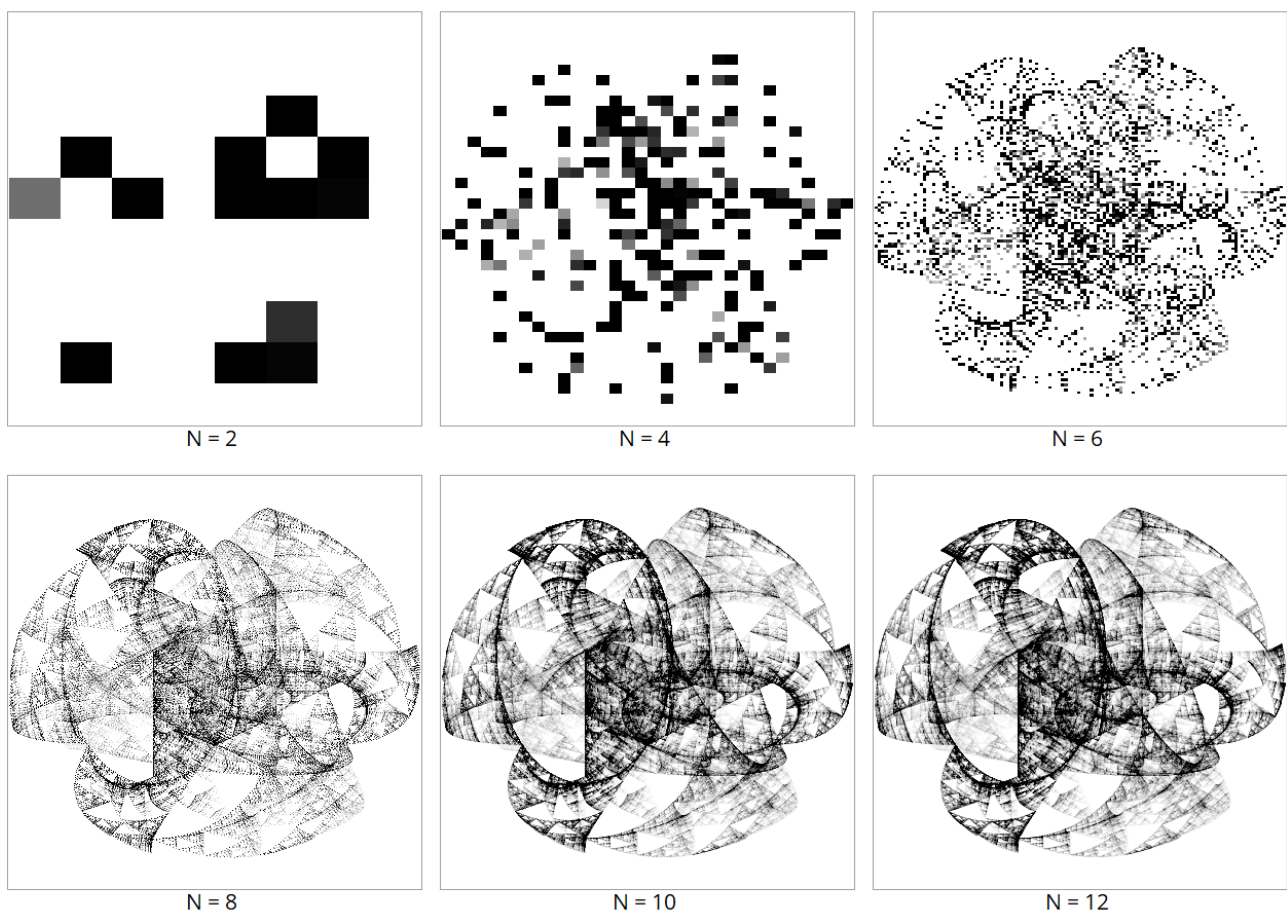


Figure 8. Ring 3D displays of DNA at different values of the scaling parameter N .

All the listed visualization variants differ from those published earlier [17]. Particularly compelling are the visualizations constructed using cyclic algorithms as they approach the objects of Riemann geometry. All living organisms, as is known from [15], are consistent

in their forms with the principles of Riemann geometry. However, the approaches visualization of DNA sequences outlined in this article requires additional consideration. At the same time, the approaches outlined are promising for comparative genomics, the visualization of individual DNA sections and complete chromosomes, as well as the application of color markers with mutations or differences from reference values to visualizations.

Note that in the described mappings, three sub-alphabets can be considered as a three-channel representation over these sub-alphabets. This fits well conceptually with the theory of color perception (RGB—red, green, blue), according to which the eye perceives three primary colors: red, green, and blue, and their combinations produce all other colors. This theory is considered by S. Petoukhov in connection with genetic matrices (for example, [16]). Therefore, each visualization channel can be mapped to one of three colors. The intensity of each color of each visualization point is different, so 2D and 3D representations allow you to take into account combinations of colors in proportion to the contribution from each of the three channels. For this, the hexadecimal color model in programming #RRGGBB seems to be convenient, where RR, GG, and BB are the amount of red, green, and blue. This makes it possible to enhance the color component in renderings and opens new possibilities for parametric rendering in accordance with the method described. The disadvantage of the method is a significant increase in computation time, including the need to recalculate the proportions of the contribution of each of the three colors in the color of the N-plet.

At the same time, it should be noted that genetic mechanisms determine the appearance (including coloration, morphogenesis, and structure) of various types of organisms and the structure of the visual analyzer in humans and animals (and related cognitive functions). A typical example is the Optix gene, which is involved in eye development in *Drosophila* and is simultaneously responsible for the coloration of butterfly wings [19].

Let us explain what information can be obtained using each of these methods. Figures 2–4 show long nucleotide sequences from one end of the chromosome to the other. Heterogeneity—genetic “inserts”—is clearly visible in Figure 2a, and the genetic “inserts” are also shown in Figure 2b. Figure 3 show the DNA in more detail than Figure 2a because it combines the methods used in Figure 2a,b. Figure 4 demonstrate the principle of fractal nesting of genetic information, and the corresponding method is more interesting from a theoretical point of view than from a practical one: the quasi-fractal ordering of DNA is visible at different levels of magnification. Figures 5–8 show exotic mappings, which are of interest more for theoretical mathematics than for applied biology. For practical applications in biomedicine, the relevant methods have yet to be elaborated.

Figure 9 show one-dimensional integral visualizations of chromosomes of various organisms. The same coloring method is used as in Figure 5, where red corresponds to the maximum concentration and blue to the minimum. For each chromosome, three sub-alphabets are shown in one-dimensional projections, and pairs of sub-alphabets are shown in three two-dimensional projections.

Each visualization point corresponds to a DNA fragment of length N nucleotides and encodes by its coordinates the number of elements in the corresponding sub-alphabet. In this case, one-dimensional visualization along the vertical axis contains the number of elements in a fragment of the nucleotide chain; on the horizontal axis, the sequence number of the elements is shown. Two-dimensional visualizations display only the number of elements for each pair of sub-alphabets. Significant differences in the chromosome structure of various organisms are clearly visible.

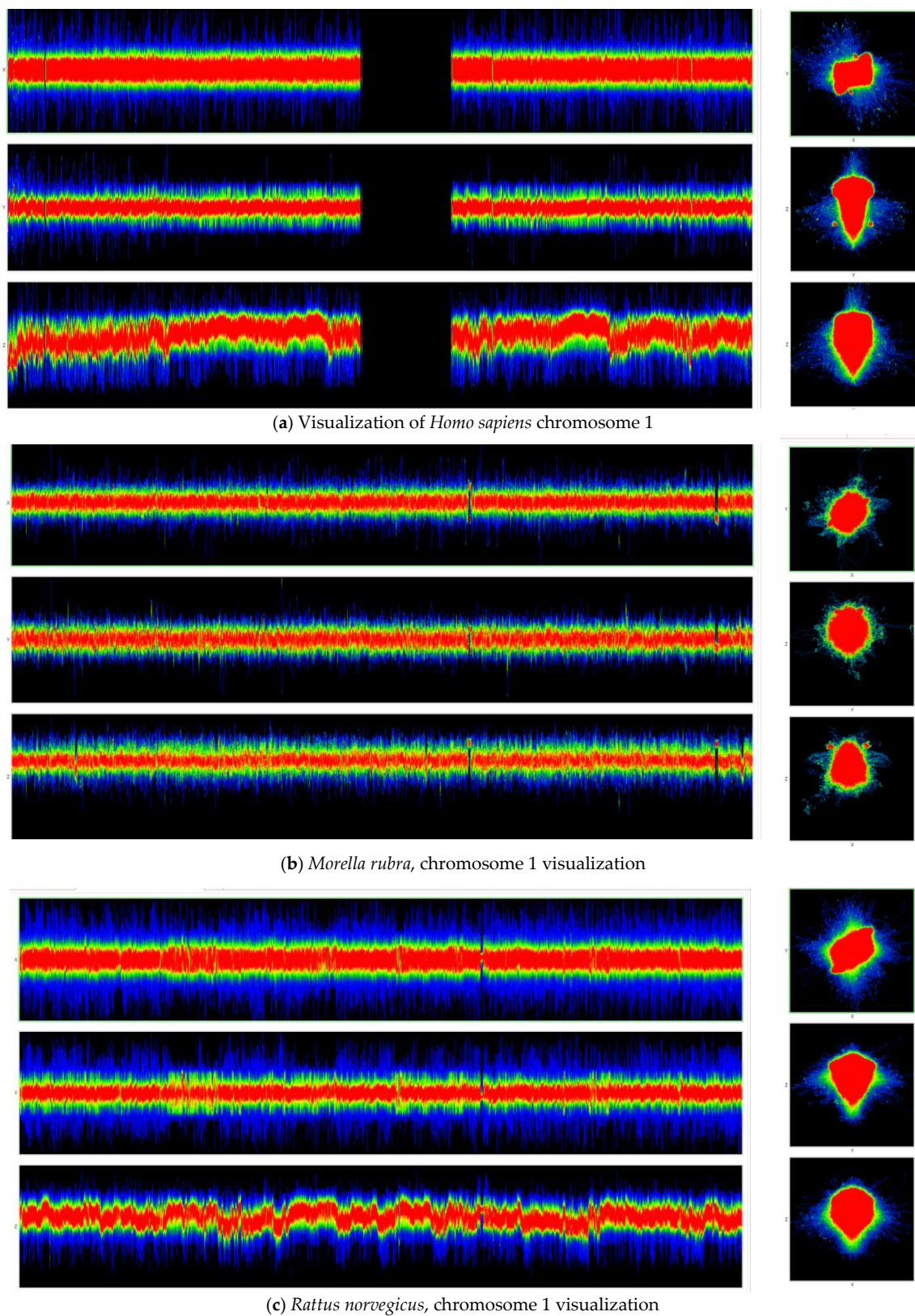


Figure 9. Color integral representations of the first chromosomes of *Homo sapiens* (a), *Morella rubra* (b), and *Rattus norvegicus* (c) with the scaling parameter $N = 300$. For one-dimensional representations (three horizontal rows on the left), the first row shows X, the second row shows Y, and the third row shows Z. For two-dimensional representations (square displays on the right), coordinates (top to bottom, abscissa/ordinate) are X/Y, Y/Z, X/Z.

5. Conclusions

In summary, the proposed set of methods for studying various traits allows one to process, analyze, compare, and generalize genetic texts. The main system properties of the DNA algorithms proposed in this paper are as follows:

1. Multiscaling. The possibility of clustering with various variants of the free parameter—the scaling factor N with the preservation of the internal structure of the display due to quasi-fractal properties of nucleotide composition of DNA (an example in Figure 2). The choice of coefficient N allows one to “adjust the sharpness” of visualization in different variants of the algorithm.
2. Three-dimensionality. The maximum number of mapping dimensions is three by the number of sub-alphabets.
3. Displaying information in parametrical spaces by means of nucleotide sub-alphabetic functions. This makes it possible to reveal the ordered structure of an information signal.
4. Ordering. The nature of the mappings can be related to the entropy level of the analyzed sequence (signal), based on the heuristic that the noise gives a chaotic visual pattern.
5. Quasi-fractality. Patterns are finite, and the fractal structure disappears at small scales. Quasi-fractality is especially evident in long genetic nucleotide sequences, where clusters may contain subclusters.
6. Symmetry. As a rule, mappings based on the described DNA algorithms are characterized by various types of symmetry.

This work belongs to the field of mathematical biology; a new computational perspective on the geometry and visualization of genetic nucleotide sequences is presented. This mathematical direction is of separate interest to biology. In particular, some visualization techniques can be used in comparative genetics to demonstrate differences between species or individuals of the same species by highlighting different areas in the visualizations. Visualizations, such as the one in Figure 2, make it possible to visualize the nucleic composition of the chromosome and large amounts of molecular genetic information.

The algorithms presented in this paper demonstrate different variants of visualizations with quasi-fractal and other characteristic patterns. Speaking about fractal and quasi-fractal structures, it should be noted that the present study deals with topological structures in spaces of fractional dimensionality. Fractional dimensionality arises from the properties of the system of “genetic” orthogonal Walsh functions that are closed by the operation of modulo two addition. Two dimensions are required to specify the coordinates of a three-dimensional point. Thus, the obtained space has ultrametric properties.

In this case, the research is based on a strict system of methods, which is not based on the arbitrary transformation of the genetic sequence but on a well-defined system of basic functions, which are mutually determined by the parameters of the four nucleotides. The presented new method of mapping large genetic data was prompted by the nature of DNA, namely the physicochemical properties of nucleotides that represent the system of orthogonal Walsh functions. All coding options using these functions lead to different reflections and rotations of the final visualizations, i.e., to solutions that are invariant with respect to symmetry transformations.

The article shows how this method allows one to construct mappings of rather long nucleotide sequences in various parametric spaces. Of note, these mappings are often quasi-fatal in nature. Moreover, the more the nucleotide composition differs, the more different the final mappings will be. This can be seen particularly well in the example of bacteria, whose DNA atlas was published in our monograph [20]. In some cases, the developed method significantly simplifies the comparative analysis of living organisms’ DNA due to new means of visualization.

The method developed in this study employs new visualization algorithms, which follow the predictive properties of the molecular structure of nucleotides and makes the perception of large genetic data much easier. However, if the DNA structure of different organisms is identical, it is difficult to notice the difference in the mappings without additional focus on differing fragments.

The method seems promising for pre-processing information for machine learning methods since it is possible to identify the difference in various DNA structures by highlighting the different regions and then feed the resulting patterns as input parameters to recognition algorithms.

Thus, this research not only presents a new way to perceive the phenomenon of genetic coding by optimizing mental work but also provides a new method of representing the information in artificial intelligence systems. In this regard, the study can be useful as a tool for DNA visualization and for the construction of characteristic patterns of arbitrary information encoded in the tetra code.

On the whole, the outlined direction seems very promising. It allows us to take a new look at the phenomenon of genetic coding, using such mathematical tools as spectral representation and cyclic structures. The authors believe that the presented algorithms are fundamental because all mappings based on the “quartet” of nucleotide functions encode sub-alphabets.

Author Contributions: All authors contributed equally to the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This paper has been supported by the RUDN University Strategic Academic Leadership Program.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: This paper was supported by the RUDN University Strategic Academic Leadership Program.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Mendizabal-Ruiz, G.; Román-Godínez, I.; Torres-Ramos, S.; Salido-Ruiz, R.A.; Vélez-Pérez, H.; Morales, J.A. Genomic signal processing for DNA sequence clustering. *PeerJ* **2018**, *6*, e4264. [[CrossRef](#)] [[PubMed](#)]
2. Lee, A.B.; Luca, D.; Klei, L.; Devlin, B.; Roeder, K. Discovering genetic ancestry using spectral graph theory. *Genet. Epidemiol.* **2009**, *34*, 51–59. [[CrossRef](#)]
3. Kluger, Y.; Basri, R.; Chang, J.T.; Gerstein, M. Spectral Biclustering of Microarray Data: Coclustering Genes and Conditions. *Genome Res.* **2003**, *13*, 703–716. [[CrossRef](#)]
4. Tan, H.S. Fourier spectral density of the coronavirus genome. *bioRxiv* **2020**. [[CrossRef](#)]
5. Quero, G.; Bonnacarrère, V.; Simondi, S.; Santos, J.; Fernández, S.; Gutierrez, L.; Garaycochea, S.; Borsani, O. Genetic architecture of photosynthesis energy partitioning as revealed by a genome-wide association approach. *Photosynth. Res.* **2020**, *150*, 97–115. [[CrossRef](#)] [[PubMed](#)]
6. Ferreira, L.M.; Sáfadi, T.; Ferreira, J.L. Evaluation of genome similarities using a wavelet-domain approach. *Rev. Soc. Bras. Med. Trop.* **2020**, *53*, e20190470. [[CrossRef](#)] [[PubMed](#)]
7. Gaynanova, I.; Li, G. Structural learning and integrative decomposition of multi-view data. *Biometrics* **2019**, *75*, 1121–1132. [[CrossRef](#)] [[PubMed](#)]
8. Lin, D.; Bonora, G.; Yardımcı, G.G.; Noble, W.S. Computational methods for analyzing and modeling genome structure and organization. *WIREs Syst. Biol. Med.* **2019**, *11*, e1435. [[CrossRef](#)] [[PubMed](#)]
9. Wang, X.; Ye, L.; Lyu, M.; Ursache, R.; Löytynoja, A.; Mähönen, A.P. An inducible genome editing system for plants. *Nat. Plants* **2020**, *6*, 766–772. [[CrossRef](#)] [[PubMed](#)]
10. Tkacz, M.A.; Chromiński, K.; Idziak-Helmcke, D.; Robaszkiewicz, E. Novel visual analytics approach for chromosome territory analysis. *PeerJ* **2021**, *9*, e12661. [[CrossRef](#)] [[PubMed](#)]
11. Altinsoy, E.; Yang, J.; Tu, E. An improved denoising of G-banding chromosome images using cascaded CNN and binary classification network. *Vis. Comput.* **2021**, 1–14. [[CrossRef](#)]
12. Lappala, A.; Wang, C.-Y.; Kriz, A.; Michalk, H.; Tan, K.; Lee, J.T.; Sanbonmatsu, K.Y. Four-dimensional chromosome reconstruction elucidates the spatiotemporal reorganization of the mammalian X chromosome. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2107092118. [[CrossRef](#)] [[PubMed](#)]
13. Lappala, A.; Lee, J.; Tan, K.; Sanbonmatsu, K. 4D Chromosome Organization: Combining Polymer Physics, Knot Theory and High Performance Computing. In Proceedings of the APS March Meeting 2022, Chicago, IL, USA, 14–18 March 2022.
14. Goel, M.; Schneeberger, K. Plotsr: Visualising structural similarities and rearrangements between multiple genomes. *bioRxiv* **2022**. [[CrossRef](#)] [[PubMed](#)]

15. Petoukhov, S.V. Quantum biology, universal cooperative rules in genomes, and algebraic harmony in living bodies. In Hyojeong Academic Foundation, Korea. Available online: <http://petoukhov.com/PETOUKHOV%20ARTICLE%20FOR%20KOREA.pdf> (accessed on 1 April 2022).
16. Petoukhov, S.V. Hyperbolic rules of the cooperative organization of eukaryotic and prokaryotic genomes. *Biosystems* **2020**, *198*, 104273. [[CrossRef](#)] [[PubMed](#)]
17. Stepanyan, I.V.; Petoukhov, S.V. The Matrix Method of Representation, Analysis and Classification of Long Genetic Sequences. *Information* **2017**, *8*, 12. [[CrossRef](#)]
18. Mobarakeh, P.S.; Grinchenko, V.T.; Soltannia, B. Bending vibrations of bimorph piezoceramic plates of non-canonical shape. *Int. Appl. Mech.* **2019**, *55*, 321–331. [[CrossRef](#)]
19. Schember, I.; Halfon, M.S. Common Themes and Future Challenges in Understanding Gene Regulatory Network Evolution. *Cells* **2022**, *11*, 510. [[CrossRef](#)] [[PubMed](#)]
20. Stepanyan, I.V.; Lednev, M.Y. *Algorithms for Visualization of Molecular Genetic Sequences in Spaces of Binary-Orthogonal Walsh Functions: Monograph*; (Electronic Edition of Network Distribution); KDU: Moscow, Russia, 2020; 193p, ISBN 978-5-7913-1159-7. Available online: <https://bookonline.ru/node/5373> (accessed on 1 April 2022).