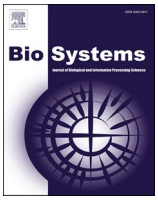
BioSystems 233 (2023) 105019

Contents lists available at ScienceDirect

BioSystems

journal homepage: www.elsevier.com/locate/biosystems



The principle "like begets like" in algebra-matrix genetics and code biology

Sergey V. Petoukhov

Mechanical Engineering Research Institute of Russian Academy of Sciences, M. Kharitonievskiy pereulok, 4, 101990, Moscow, Russia

ARTICLE INFO

Keywords:
Single-stranded DNA
Code biology
DNA alphabets
Binary opposition
Probabilities
Dichotomy
Fractal tree
Genomic DNA
Matrices
Hyperbolic numbers

ABSTRACT

The article is devoted to analysis of emergent properties of the system of binary oppositions in the genetic code ensemble. The epochal model of the double helix of DNA by Watson and Crick showed that the multiple reproduction of genetic information on DNA strands uses the ancient principle "like begets like" based on the simple complementarity in pairs of nucleobases. Each of these pairs is built on the binary opposition "purine-pyrimidine". But the system of DNA *n*-plet alphabets and genetic coding is much richer in types of binary oppositions, which also have some coding meanings related to this principle. The article contains the results of the application of the author's "method of hierarchy binary stochastics" (HBS-method) to the analysis of the quasi-stochastic organization of binary sequences of hydrogen bonds in genomic single-stranded DNAs. This analysis revealed hidden probability rules related to dichotomous fractal-like probability trees. The relationship between inherited bodily dichotomies in living organisms and the discovered probability dichotomies in information sequences of genomic DNAs is discussed. The encoding properties of molecular binary oppositions in the DNA nucleotide system allows the algorithmic construction of (2^n*2^n)-matrices of probabilities of *n*-plets in these binary sequences, which are matrix representations of 2^n-dimensional hyperbolic numbers. Connections of these multidimensional numbers with some inherited physiological phenomena and deep neural networks are noted. A unified algebra-numeric certification of the DNAs of genomes and genes - based on these multidimensional numerical systems - is proposed.

1. Introduction

The DNA double helix model by J.D. Watson and F. Crick gave a powerful impetus to the development of genetic research. It showed the world a recursive and binary oppositional algorithm for the complementary replication of DNA strands, which ensures the replication of the genetic information recorded on these strands. The seminal work by Watson and Crick was perceived as the discovery of a key secret of life, corresponding to the ancient notion that "like begets like" and the principle of binary oppositions. Scientists were struck by how simple and beautiful this explanation of the replication and preservation of genetic information based on the mechanism of complementarity turned out to be. It was emphasized that it is this complementarity that provides the most important properties of DNA as a carrier of hereditary information (see, for example, (Chapeville and Haenni, 1974)). Complementary replication of DNA strands occurs at an astonishing speed rate: for example, the well-known bacteria *E. coli* has a speed of replication of over 1000 bases per second (Bank, 2022).

This article presents original research concerning further revealing of importance of both the notion "like begets like" and the binary oppositional principle in the genetic code system on the basis of algebraic analysis of structural features of this system. Received results are discussed in connection with the key statement of Code Biology that biological evolution took place by two distinct mechanisms: by natural selection and by natural conventions providing a wide ensemble of biological codes (Barbieri, 2015).

Thinking about Code Biology and evolution by natural conventions, one should be aware of the "probability-determinism" dualism in living bodies. The deal is that in cells, individual molecules interact in a stochastic manner. In living bodies, everything is built on such stochastics. In particular, natural conventions that are studied in Code Biology are implemented against the background of stochastic processes. Even genetically identical cells of the same tissue have different levels of protein expression, sizes, shapes, etc. But against the background of this stochastics in the "small scale", bodily signs (or traits) in the "large scale" are inherited from parents as deterministic. Genetics as a science began with the discovery by Mendel of stochastic rules for the inheritance of traits when organisms were crossed. According to Mendel's law of independent inheritance of traits, information from the level of DNA molecules dictates the macrostructure of living bodies through many

E-mail address: spetoukhov@gmail.com.

https://doi.org/10.1016/j.biosystems.2023.105019
Received 8 July 2023; Received in revised form 1 September 2023; Accepted 1 September 2023
Available online 9 September 2023
0303-2647/© 2023 Elsevier B.V. All rights reserved.

independent channels, despite strong interference and noises. For example, the colors of hair, eyes and skin are inherited independently of each other. Accordingly, each organism is a multi-channel error-correcting coding machine based on the principles of stochastics.

The interrelation between stochastics and determinism in living nature has long troubled scientists. For example, P. Jordan, who was one of the founders of quantum informatics and the author of the first article on quantum biology, claimed that «*life's missing laws were the rules of chance and probability (the indeterminism) of the quantum world that were somehow scaled up inside living organisms*» (McFadden and Al-Khalili, 2018). According to Jordan's thoughts, the mechanisms of living organisms are associated with his 'amplifier theory', based on Bohr's notion of the 'irreversible act of amplification', required to bring the fuzzy quantum reality into sharp focus by 'observing' it. In addition, Jordan and some later E. Schrödinger noted the main difference between living and inanimate objects: inanimate objects are controlled by the average random movement of their millions of particles, whose individual influence is negligible, while in a living organism selected – genetic - molecules have a dictatorial influence on the whole living organism (McFadden and Al-Khalili, 2018). It is these hidden laws of chance and probability, postulated by Jordan, that we are looking for in the probabilistic characteristics of long DNA sequences of hydrogen bonds and nitrogenous bases (Petoukhov, 2019, 2020a,c). In continuation of these studies, new significant results of such author's researches are presented below in connection with Code Biology and new algebraic tools for modeling in it. Long DNA sequences are very important and convenient for such computerized research of possible biological rules of probabilities, which in Jordan's day was impossible. The main idea of all quantum mechanics is that everything in the world of appropriate objects is described only as probabilistic. Correspondingly we believe that the below described study of probabilities systems in structural organization of genomic DNAs is important and can be related to many works about a fundamental meaning of stochastics in Code Biology. For example, the idea about living bodies as biological quantum computers, where random connections between neurons can serve as codes for different tasks, were presented in (Liberman, 1990; Liberman and Minina, 1996).

Genomic DNAs sequences possess composition symmetries. According to the second rule by Chargaff, in long single-stranded DNAs percentages of cytosine C and guanine G are practically equal (%C ≈ %G); the same is true for percentages of adenine A and thymine T (%A ≈ %T). Besides this, percentage of any *n*-plet is approximately equal to percentage of its mirror-complementary *n*-plet: for example, percentages of the triplet ACG and its mirror-complementary triplet CGT (not TGC) is practically the same (Prabhu, 1993). Random sequences have no such properties. These and other facts indicate that genomic DNAs are complexly structured and not random at all, which is further substantiated in the article (Fimmel et al., 2019). This specific organization of genomic DNA needs further study in order to possibly reveal other hidden symmetries and patterns in them.

The article presents the discovery of general rules of stochastic-deterministic organization of information sequences of genomic DNAs of eukaryotes and prokaryotes. First of all, these rules concern binary sequences of hydrogen bonds in genomic DNAs. The described rules are connected with information dichotomies of probabilities and corresponding fractal-like trees of probabilities, which fundamentally differ from constructional dichotomies in biological bodies. The received phenomenological data and rules lead to new biological ideas.

In addition, the purpose of this article is to describe the author's results of an in-depth analysis of the system of binary-oppositional structures in these DNA alphabets and their algebraic-matrix representations. These results show that the molecular complementary replication of DNA strands is accompanied by the presence of an algebraic version of the principle "like begets like" in the named matrix representations of DNA alphabets. This algebraic version is based on binary-oppositional structures in the genetic molecular system, which can be represented by binary numbers and corresponding matrices of DNA alphabets. The received results allow thinking that the phenomenon "like begets like" (or a complementary replication in a wide sense) is systemic in the genetic organization and is connected with algebraic binary-oppositional features of biological organization. Correspondingly, the biological principle "like begets like" can be additionally modeled by algebraic-matrix methods and approaches. Such algebraic-matrix modeling of the genetic coding system gives new ways for studying and understanding a key role of the named principle in genetic and other inherited physiological complexes. It also gives - for further development of Code Biology - new phenomenological data and algebra-geometric tools effectively used in mathematical natural sciences and code information technologies. Below the author shows both new features of the genetic probabilistic laws and that described genetic matrices of hydrogen bonds probabilities are matrix representations of well-known hyperbolic numbers, with which structures of many inherited physiological phenomena are connected. Hyperbolic numbers, which are also related to hyperbolic geometry by Lobachevsky, are actively used around the world to create deep neural networks for artificial intelligence systems (see a survey (Peng et al., 2022)). The below described connection of the genetic code system with hyperbolic numbers and formalisms of algebraic geometry testifies in favor that the mentioned hyperbolic physiological phenomena did not appear out of nowhere, but are based on the genetic coding structures. These results enrich Code Biology by adding to its content the algebraic connection of the genetic coding system with a set of hyperbolic physiological phenomena, formalisms of algebraic geometry, and with the wide topic of hyperbolic deep neural networks having many practical applications. The main point of the article is a presentation of new results, which influence on further research and give new perspectives for development of Code Biology as the important scientific direction.

The genetic information in DNA molecules is represented in the form of sequences of four types of nucleobases: adenine A, guanine G, cytosine C, and thymine T. Along with this 4-letter DNA alphabet, other DNA alphabets exist: alphabets of 16 duplets, 64 triplets, 256 tetraplets, and other *n*-plets. In particular, the alphabet of 64 triplets is used in the genetic system to encode amino acids and termination signals of protein synthesis. Taking into account the existence of different alphabets of DNA *n*-plets turns out to be useful for revealing hidden regularities in the stochastic organization of genomic DNAs (Petoukhov, 2020a, 2021a, 2022a,b). This article presents new data on stochastic organization of information sequences of genomic DNAs.

2. Study of stochastic regularities in sequences of hydrogen bonds of genomic DNAs

In search of the missed - according to Jordan - laws of life, let us turn to the study of stochastic patterns in sequences of hydrogen bonds of genomic single-stranded DNAs. Hydrogen bonds are one of the most important components of life. It occurs in many biological structures (American Institute of Physics, 1999). In particular, hydrogen bonding plays the role of a promotional factor for intermolecular vibrational energy relaxation, and as a driving force for the occurrences of specific reaction channels in binary molecular complexes (Chatterjee et al., 2020). Hydrogen bonding is an important factor in the functioning of enzymes, which manage molecular processes in biological bodies (Shan and Herschlag, 1999; Trylska et al., 2004). The Nobel laureate L. Pauling emphasized: "*It has been recognized that hydrogen bonds restrain protein molecules to their native configurations, and I believe that as the methods of structural chemistry are further applied to physiological problems it will be found that the significance of the hydrogen bond for physiology is greater than that of any other single structural feature*" (Pauling, 1940). Hydrogen bonds determine properties of water and ice. Jellyfish are 98% water, which does not prevent them from being the most ancient species of multicellular animals with the richest evolutionary diversity and an abundance of life functions.

In the DNA double helix, complementary pairs of nucleotides A-T and C-G are connected by 2 and 3 hydrogen bonds, respectively. Accordingly, the entire sequence of nucleotides of a DNA strand can be represented as a binary sequence of digits 2 and 3 of the type 23323322 …. For specifics, let's focus on the analysis of the single-stranded DNA of the human chromosome N№1 presented in the GenBank (https://www.ncbi.nlm.nih.gov/nuccore/NC_000001.11). It contains a text of about 250 million nucleotides. As it was above noted, such chromosomal DNA sequence possesses internal symmetries and is not a random sequence at all. But perhaps it contains hidden symmetries unknown till now.

To discover such possible hidden symmetries in quasi-stochastic (probabilistic) organization of the chromosomal DNA, we analyze it according to the author's "method of hierarchy binary stochastics" (or, briefly, the HBS-method), which resembles traditional Russian dolls "Matryoshka" based on the hierarchical nesting of objects of different sizes into each other (Fig. 1). In accordance with this method, consider first the studied binary sequence of the form 23322322 … as a sequence of single digits 2-3-3-2-2- …, in which we calculate the percentage of the digits 2 and 3, i.e., %2 and %3. Then we consider the same sequence as a sequence of digital duplets 23-32-23-22 … and calculate the percentage of each of the 4 possible types of duplets: %22, %23, %32, %33. Then, similarly, we consider the same sequence as a sequence of digital triplets, tetraplets, pentaplets, …, each time calculating the percentage of each of the types of triplets, tetraplets, etc. In other words, we consider a given binary DNA text as a set or matryoshka of many parallel texts (or *n*-plet layers), each written in its own *n*-plet alphabet of hydrogen bonds (for briefly, we denote hydrogen bond *n*-plets as H-*n*-plets).

Since genomic DNA's sequences have internal symmetries and are not random, there is no reason to initially assert that there are universal symmetrical interrelations between the *n*-plet percentages in different *n*-plet representations (or layers) of the genomic DNA. It is necessary to determine experimentally the *n*-plet percentages for each *n*-plet layer. Only after obtaining such numerical data for each *n*-plet layer, one can begin to search for possible hidden relationships between the percentages of different *n*-plets from different *n*-plet layers.

In the case of the DNA of human chromosome No. 1, we obtain - by the mentioned method - Table 1 of percentages of H-*n*-plets (percentages are given in fractions of a unit).

The data in Table 1 show that in some cases the percentage of a given *n*-plet in its *n*-plet layer of the DNA is well approximated by the product of the percentages of the monoplets of which it is composed. For example, %23 = 0,254635 and %2*%3 = 0,243151 are roughly equal. But this correspondence is not at all general and cannot be used as a general rule, since this Table contains examples of non-compliance with such a correspondence. For example, %2323 ≈ 0,0741 is almost twice as different from %3322 ≈ 0,0462, although these tetraplets consist of identical monoplets and their percentages should equal the same value %2*%2*%3*%3.

But all these different sets of percentages (or probabilities) of H-*n*-plets are unexpectedly closely and algorithmically interconnected each with other as any reader can check by using the phenomenological data in Table 1. Below these revealed interconnections are presented in detail.

Fig. 1. The example of the Russian dolls "Matryoshka" (from https://www.en.wikipedia.org/wiki/Matryoshka_doll#/media/File:Russian-Matroshka.jpg; permission is granted by the GNU Free Documentation License).

2.1. *Suffix dichotomies of percentages of n-plets of hydrogen bonds in DNA of human chromosome № 1*

Analysis of the phenomenological data in Table 1 reveals the following: the percentage of any H-*n*-plet is practically equal to the sum of percentages of those two H-(*n*+1)-plets, which are generated from this H-*n*-plet by addition to it the suffixes 2 and 3 as Fig. 2 illustrates. We call this type of dichotomy as a suffix dichotomy.

In phenomenological equalities of this type (Fig. 2), based on the suffix dichotomy of percentages of H-*n*-plets, their high accuracy is impressive - up to several decimal places. Fig. 2 represents only a few examples of high-precision equality of the percentages. Using numeric data from Table 1, any reader can check himself that such equalities of percentage dichotomies also hold with similar high precisions for all other possible variants of the noted percentages dichotomy, for example, %223 ≈ %223**2** + %223**3**, %232 ≈ %232**2** + %232**3**, %3332 ≈ %3332**2** + %3332**3**, etc. The author has systematically checked these percentage dichotomic equalities in many genomic DNAs, listed below, only for $n = 1, 2, 3, 4, 5$. He also selectively checked the fulfillment of such equalities for higher values of *n* getting positive results in every case, but he did not check at what values of *n* these dichotomic percentage equalities cease to hold.

The rule of the suffix dichotomy of H-*n*-plets percentages leads to fractal dichotomous trees of H-*n*-plets percentages, which is shown in Fig. 3 for the case of DNA of human chromosome N№ 1. H-*n*-plets starting with even digit 2 are conditionally called even H-*n*-plets (their percentages are marked by red in Fig. 3 and further), and H-*n*-plets starting with odd digit 3 are called odd H-*n*-plets (their percentage are marked by blue in Fig. 3 and further). Each of the levels of these fractal trees with percentage values of even and odd H-*n*-plets corresponds to a certain magnitude of *n*.

Levels of the trees of percentages in Fig. 3 contain different quantity of percentage summands, but their sums remain the same with high precision at all the levels: in the tree of even H-*n*-plets, the sums of percentages of even H-*n*-plets are equal to 0.5827÷0.5828, and in the tree of odd H-*n*-plets, the sums of percentages of odd H-*n*-plets are equal to 0,4172÷0,4173. These summary values on the different levels are equal to the percentages of H-monoplets: %2 and %3.

One can see in Fig. 3 (at the bottom) that, for example, the level with n = 5 contains a set of very different percentages of 32 H-pentaplets. The considered binary sequence of the single-stranded DNA has a quasi-stochastic character, but in this seeming stochastically set, the regular phenomenon of percentage dichotomies exists related to the known principle of even-odd numbers. More precisely, the corresponding **rule of percentage dichotomies** in the interrelation between the considered levels of H-*n*-plets is the following:

- the sum of the percentages of those two *n*-plets, whose binary numberings are almost identical and differ from one another only by suffixes 2 or 3, is practically equal to the percentage of that (*n*-1)-plet whose binary numbering is obtained from the named numberings by deleting these suffixes. For example, %3232 + %3233 ≈ %323.

Table 1
Phenomenological percent values of each of the H-*n*-plets in corresponding H-*n*-plets representations of the single-stranded DNA of human chromosome № 1 taken from the GenBank (https://www.ncbi.nlm.nih.gov/nuccore/NC_000001.11). In numeric representation, its binary sequence of hydrogen bonds contains about 250 million digits 2 and 3. H-*n*-plets starting with even number 2 are in red, and H-*n*-plets starting with odd number 3 are in blue (here $n = 1, 2, 3, 4, 5$).

Alphabet of **H-monoplets** having 2 members: 2, 3	
%2	%3
0,582757	0,417243

Alphabet of **H-duplets** having 4 members: 22, 23, 32, 33			
%22	%23	%32	%33
0,328129	0,254635	0,254622	0,162614

Alphabet of **H-triplets** having 8 members: 222, 223, 232, 233, 322, 323, 332, 333							
%222	%223	%232	%233	%322	%323	%332	%333
0,200289	0,127765	0,155746	0,098982	0,127812	0,126809	0,098968	0,063630

Alphabet of **H-tetraplets** having 16 members: 2222, 2223, 2232, 2233, 2322, 2323, 2332, 2333, 3222, 3223, 3232, 3233, 3322, 3323, 3332, 3333							
%2222	%2223	%2232	%2233	%2322	%2323	%2332	%2333
0,127550	0,072785	0,081762	0,046020	0,081593	0,074083	0,059814	0,039135
%3222	%3223	%3232	%3233	%3322	%3323	%3332	%3333
0,072757	0,055044	0,073902	0,052937	0,046241	0,052734	0,039127	0,024518

Alphabet of **H-pentaplets** having 32 members: 22222, 22223, 22232, 22233, 22322, 22323, 22332, 22333, 23222, 23223, 23232, 23233, 23322, 23323, 23332, 23333, 32222, 32223, 32232, 32233, 32322, 32323, 32332, 32333, 33222, 33223, 33232, 33233, 33322, 33323, 33332, 33333							
%22222	%22223	%22232	%22233	%22322	%22323	%22332	%22333
0,084172	0,043336	0,047984	0,024786	0,045577	0,036137	0,029506	0,016538
%23222	%23223	%23232	%23233	%23322	%23323	%23332	%23333
0,047895	0,033731	0,045148	0,028941	0,029592	0,030252	0,024935	0,014141
%32222	%32223	%32232	%32233	%32322	%32323	%32332	%32333
0,043374	0,029467	0,033744	0,021229	0,036013	0,037946	0,030323	0,022550
%33222	%33223	%33232	%33233	%33322	%33323	%33332	%33333
0,024897	0,021326	0,028765	0,023982	0,016667	0,022502	0,014179	0,010365

0,58276=%2 ≈ %22+%23=**0,58276**	**0,41724**=%3 ≈ %32+%33=**0,41724**
0,3281=%22 ≈ %222+%223=**0,3281**	**0,25462**=%32 ≈ %322+%323=**0,25462**
0,2546=%23 ≈ %232+%233=**0,2547**	**0,1626**=%33 ≈ %332+%333=**0,1626**
0,2003=%222 ≈ %2222+%2223=**0,2003**	**0,06363**=%333 ≈ %3332+%3333=**0,06364**

Fig. 2. At the top: the illustration of the general algorithm of the suffix dichotomy in interrelations between percentages of H-*n*-plets and H-(*n*+1)-plets in a genomic binary sequence of hydrogen bonds, which is analyzed by the described method. Here the symbol % H_n denotes a percentage of any H-*n*-plet under studying (values $n = 1, 2, 3, 4, 5$ were studied). **At the bottom**: for the case of DNA of the human chromosome №1, a few numeric examples are shown of high-precision equalities between a percentage of any H-*n*-plet and a sum of the percentages of two H-(*n*+1)-plets, which are generated from it by the addition of suffixes 2 and 3. Added suffixes 2 and 3 are highlighted by green. Rounded values of percentages are taken from Table 1.

This rule of percentage dichotomies is universal and holds for all genomic DNAs studied by the author and listed below. For instance, in DNA of human chromosome № 1, we have from the data of Table 1 the following example of interrelation between percentages of the representatives of alphabets of 8 H-triplets and 16 H-tetraplets (1):

$$\%323 = \mathbf{0{,}12681} \approx \%3232 + \%3233 = 0{,}07390 + 0{,}05294 = \mathbf{0.12684} \quad (1)$$

∑ %	Percentage of even H-*n*-plets starting with 2	Percentage of odd H-*n*-plets starting with 3	∑ %
0,58275	%2	%3	0,4172
0,58276	%23+%22	%33+%32	0,4172
0,58278	%233+%232+%223+%222	%333+%332+%323+%322	0,4172
0,58274	%2333+%2332+%2323+%2322 +%2233+%2232+%2223+%2222	%3333+%3332+%3323+%3322 +%3233+ %3232+%3223+%3222	0,4173
0,58267	%23333+%23332+%23323+%23322 +%23233+%23232+%23223+%23222 +%22333+%23332+%22323+%22322 +%22233+%22232+%22223+%22222	%33333+%33332+%33323+%33322 +%33233+%33232+%33223+%33222 +%32333+%32332+%32323+%32322 +%32233+%32232+%32223+%32222	0,4173

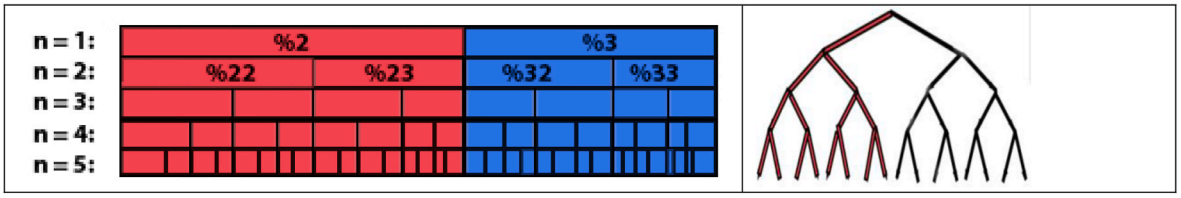


Fig. 3. At the top: the beginning of the fractal dichotomic percentage trees of H-*n*-plets, which is based on the algorithm of the suffix dichotomy of percentages of H-*n*-plets when passing from the H-*n*-plet representation of the H-sequence to its H-(*n*+1)-representation. The case of DNA of human chromosome № 1 is presented. Rounded numeric data of percentages are taken from Table 1. The left and the right columns present a practical invariance of sums of percentages of even and odd H-*n*-plets at all the levels. **At the bottom**: the diagram of comparative percent values at different levels of these dichotomic trees. At each level "*n*", lengths of intervals are proportional to percent values of corresponding H-*n*-plets from Table 1. The *n*th level contains percent values of 2^n H-*n*-plets. The total length of the strip of each layer corresponds to the summary value 1,0 of the percentages of its H-*n*-plets. The relations of dichotomies between the lengths of the corresponding intervals at neighboring levels are visible. The traditional scheme of dichotomous trees is also shown at the right.

∑%	Percentages of H-*n*-plets starting with 23
0,2546	%23
0,2547	%232+%233
0,2546	%2322+%2323+%2332+%2333
0,2546	%23222+%23223+%23232+%23233 +%23322+%23323+%23332+%23333

Fig. 4. The beginning of the dichotomic fractal tree having %23 at its top. The left column contains practically the same values of percentage summaries for each level.

The dichotomic percentage trees in Fig. 3 are called fractal (or fractal-like) since each of their numeric member (that is, a percent value of any H-*n*-plet) is a top of its own dichotomic tree of percent values of H-*n*-plets from the genomic DNA. Fig. 4 demonstrates this for the particular example, where %23 is the top of its own dichotomic fractal tree.

2.2. The rule of percentage equalities in the set of genomic n-plets of hydrogen bonds

The numeric data in Table 1 confirm else the following phenomenological rule:

- those two H-*n*-plets, which are read as mirror (or reversed) copies of each another (such as 223 and 322), always have almost the same percentages in the genomic H-*n*-plet sequence.

This rule of practical equalities of percentages of such *n*-plets further reduces the degree of stochasticity in the organization of H-*n*-plet sequences in genomic DNAs. To demonstrate examples of this rule, we write out all such pairs of reversed H-*n*-plets from Table 1 (percentages are rounded):

%23 = %32 = 0,2546; %223 = %322 = 0,1278; %233 = %332 = 0,0990; %2223 = %3222 = 0,0728; %2232 = %2322 = 0,082; %2233 = %3322 = 0,046; %2333 = %3332 = 0,0391; %2323 = %3232 = 0,074; %3233 = %3323 = 0,053; %22223 = %32222 = 0,043; %22232 = %23222 = 0,048; %22233 = %33222 = 0,025; %22323 = %32322 = 0,036; %22332 = %23322 = 0,0296; %22333 = %33322 = 0,017; %23223 = %32232 = 0,0337; %23233 = %33232 = 0,029; %23323 = %32332 = 0,0303; %23333 = %33332 = 0,014; %32233 = %33223 = 0,0213; %32333 = %33323 = 0,023 (2)

Even and odd (or red and blue) fractal trees in Fig. 3 are asymmetric each to another: at all the levels, sums of percentages of even H-*n*-plets are not equal to sums of percentages of odd H-*n*-plets. But is it possible to algorithmically build fractal trees with symmetrical left and right halves based on the phenomenological data on the percentage of H-*n*-plets from Table 1? Yes, you can. To construct such symmetric percentage trees, it suffices to take as their vertices the percentages of those two H-*n*-plets that are read as mirror (or reversed) copies of each other (for example, % 23 and %32, or %223 and %322); then the algorithm of the suffix dichotomies is applied for these vertices to construct sets of H-*n*-plets at other levels of the trees. Fig. 5 shows an example of so constructing symmetric fractal trees where percentage sums of even and odd H-*n*-plets are practically identical at all their levels.

Similar rules hold for the case of the prefix dichotomies of percentages of H-*n*-plets in genomic DNAs, where other fractal-like trees of percentages arise (see detail and numeric data in the preprint (Petoukhov, 2023)). It should be noted that, in a general case, a percentage of each H-*n*-plet simultaneously belongs to different fractal trees of percentages of H-*n*-plets. For example, %232 belongs to all above-described fractal trees. The stochastic organization of hydrogen bond sequences in genomic DNAs deals with interrelated nets of fractal dichotomic trees of percentages. To demonstrate the performance of similar rules on other genomic DNAs, the preprint (Petoukhov, 2023) presents detail numeric data about H-*n*-plet percentages in the plant *Arabidopsis thaliana*, and bacteria *Bradyrhizobium japonicum*.

The following relation of complementarities exists between even and odd fractal dichotomous trees of percentages in Figs. 3 and 5: the complementarity operation 2↔3 transforms each of the H-*n*-plets of the

∑%	Percentages of H-*n*-plets starting with 23	Percentages of H-*n*-plets starting with 32	∑%
0,2546	%23	%32	0.2546
0,2547	%232+%233	%322+%323	0.2546
0,2546	%2322+%2323+%2332+%2333	%3222+%3223+%3232+%3233	0.2546
0,2546	%23222+%23223+%23232+ %23233+%23322+%23323+ %23332+%23333	%32222+%32223+%32232+ %32233+%32322+%32323+ %32332+%32333	0.2546

Fig. 5. An example of symmetric family of even and odd fractal trees in DNA human chromosome № 1, which start from percentages of two mirror H-*n*-plets %23 and %32 and are based on the suffix dichotomy. Here percentage sums **∑%** of even and odd H-*n*-plets are practically identical at all the levels.

even fractal tree into the corresponding H-*n*-plet of the odd fractal tree and vice versa. Two such corresponding H-*n*-plets we call complementary H-*n*-plets. In the relation to the complementarity operation, these even and odd (or left and right, or Yin and Yang) fractal dichotomous trees of percentages are complementary each to another. It is additional evidence in favor of the key role of the principle of complementarity and dichotomies in the genetic system related to the ancient principle "like begets like".

Similar results and rules have been obtained by the author on a significant number of DNAs of eukaryotic and prokaryotic genomes, including the following:

- all 24 human chromosomes;
- all chromosomes of drosophila, mouse, worm, many plants;
- 19 genomes of bacteria and archaea;
- many extremophiles, living in extreme conditions, for example, radiation with a level 1000 times higher than fatal for humans.

These genomic DNAs were early analyzed by the author concerning another theme related to the hyperbolic rules of amounts of nucleotide *n*-plets in genomic DNAs (Petoukhov, 2020c). Presented results give pieces of evidence of the existence of a quantum-mechanical long-range bond in biological structures considered, for example, in the popular Frohlich's theory (Fröhlich, 1970, 1978, 1980, 1988).

The received results on quasi-stochastics of genomic DNAs testify in favor of the existence of universal rules and nontrivial algebraic invariants of a **globally** genomic nature, which remain unchanged over billions of years of biological evolution, during which millions of species of organisms die off and new ones appear (although **locally** genomic sequences are modified by mutations, natural selection mechanisms, etc.). At this stage of research, the described rules are a candidacy for the role of universal genetic rules. Additional studies of the widest set of genomic DNAs are required to test their universality.

One can added that all the described universal rules of percentage dichotomies in stochastic organization of genomic DNAs hold not only for binary sequences of hydrogen bonds but also for binary sequences of purines and pyrimidines, and for binary sequences of keto- and amino molecules in genomic DNAs (see detail and examples of numeric percentage data for genomic DNAs of some eukaryotic and prokaryotic organisms in the preprint (Petoukhov, 2023)).

3. Difference of dichotomies in biological bodies and in stochastics of genomic DNAs

Dichotomies in genetically inherited biological bodies and their transformations are well known. Their examples give, for example, the functioning of double-stranded DNA; branching in plants; the bronchial tree of the human lungs with its 23 levels of dichotomies (there are approximately $2^{23} = 8{,}388{,}608$ alveoli at the end of the bronchial tree branches (Medvedev, 2020); mitosis of somatic cells; dichotomous branching of neuronal axons and dendrites (there are no branching for three, four, five, and so on (Tsang, 2016, p. 235)). The inherited system of biorhythmic frequencies is also related to the dichotomous (or octave) ratio 2:1, which made it possible to create technical devices for resonant physiotherapy (Khazina, 2015).

Presented in Fig. 3, asymmetric pairs from the left and right dichotomous trees of information percentages of *n*-plets, which are connected with even and odd numbers 2 and 3, can be associated with two genetically inherited asymmetric cerebral hemispheres. The Yin-Yang theme (left and right, even and odd, feminine and masculine), that is cross-cutting for biology and culture, is systematized in a well-known book published in different languages with the characteristic title "Even and Odd. Asymmetry of the brain and sign systems" (Ivanov, 1978). This feature of the information dichotomy of percentages of hydrogen bonds 2 and 3 in genomic DNAs allows us to recall some historical facts that, for example, in Ancient China, the numbers 2 and 3 were considered the numbers of Yin and Yang (or female and male numbers) and also as numbers of the Earth and the Sky; these numbers served as the basis of ancient Chinese arithmetic.

But in genomic DNAs, in contrast to bodily biostructures, we encounter a fundamentally different type of dichotomy: the dichotomies of informational probabilistic characteristics in DNA information sequences. The vast dichotomous fractal networks of genomic DNAs probabilities can be considered as the information soil from which material living bodies and genetic intelligence grow. Material structures of living bodies do not arise from scratch, but have structural prototypes in a regular information-coding biosystem of probabilities in a variety of genetic languages with their families of 2^n-parameter alphabets of *n*-plets of DNAs and RNAs. In the beginning of this article, the importance of the "probability-versus-determinism" dualism for living bodies and their inheritance phenomena was presented.

One of the interesting areas of future research is the analysis of the relationship between the traditionally studied dichotomous fractals in the configurations of biological bodies and the above-described dichotomous percentage fractals in the quasi-stochastic (probabilistic) organization of information sequences of genomic DNAs. Fractals (or fractal-like structures) in the configurations of biological bodies are studied long ago, in particular, in connection with cancer, and is presented, for example, in the works (Pellionisz, 1989; Pellionisz et al., 2013; Werner, 2010). Fractal packing of chromosomes in the cell nucleus is known. The book (Tsang, 2016), which has the characteristic title "Fractal Brain Theory", contains rich material on biological fractals. Revealing the universal rules of stochastic dichotomies and their percentage fractals in genomic DNAs gives a new direction of thoughts to understand the basis of phenomena of material dichotomies in the inherited structures of biological bodies.

Life on Earth exists for at least 3.5 billion years, and all this time the genetics of organisms and their genetically inherited bodily configurations are built on dichotomies. For example, over the course of billions of years of life on Earth, it is typical for bacteria and prokaryotes in general to reproduce themselves by dichotomous division of the body into two halves having similar information-coding properties connected with dichotomies. What are the structural foundations for this "eternal" dichotomous phenomenon of bacterial reproduction, which is accompanied by the most complex processes of dichotomous separation of all dichotomously organized genetic information, together with the

accompanying multi-species protein and nucleic assemblies inside a bacterial cell?

The author suggests the following possible answer to such questions. There is a world of families of probabilities, hidden from direct perception and structured on the basis of binary oppositions (like Yin-Yang), reminiscent, in particular, of binary oppositions in physics: positive and negative electric charges, the north and south poles of magnets, the forces of attraction and repulsion, etc. It is in the image and likeness of the binary organized families of probabilities of this multilayer world that genetically inherited biological bodies are built. Figuratively speaking, our bodies are clothes put on these binary structured families of information probabilities, which act as prototypes of biological structures and are endowed with their own forms of energy. To a certain extent, this is similar to the situation with the invisible man from the novel "The invisible man" by H.G. Wells, whose invisible figure becomes definable only when he is wearing clothes. It is also reminiscent of the ancient concepts of the manifested and the non-manifested worlds, and famous Plato's allegory about the world of ideas and the shadows on the cave wall, by which people living in a cave can judge the true hidden world of ideas. By studying the universal rules of genomic DNAs, we indirectly research the rules of this hidden world of binary families of probabilities, which is the progenitor of biological structures with their amazing properties.

The author believes that one of the promising approaches to the study of this hidden world with its binary-opposition features can be based on the mentioned Frohlich theory, which is supported by R. Penrose and many other authors (see the review in Petoukhov (2021d)). This theory testifies in favor of the fact that living bodies are a kind of analog of Bose-Einstein condensate, which plays a prominent role in the theory of the physical vacuum and is based on bosons (particles with integer spin, for example, photons). A quantum system consisting of an arbitrary number of bosons and an even number of fermions (particles with half integer spin) is itself a composite boson. Currently, the Bose-Einstein condensate is being studied by physicists in different countries for new technologies and theoretical concepts based on it, as well as the possibility of obtaining this condensate at room temperature (see, for example, (Henderson, 2018; Plumhof et al., 2014)). An important feature of the physical vacuum is its Yin-Yang property: the entities generated from it appear in the form of a mutually opposite pair, for example, particles - antiparticles, waves - antiwaves, etc. Under looking at biological structures from the standpoint of Frohlich's theory as special analogs of Bose-Einstein condensate that do not require ultra-low temperatures, the following thought suggests itself: this vacuum-like quantum state of organisms is responsible for many binary-oppositional (Yin-Yang) features of genetic and other inherited biological structures. In Marshall (1989), the idea of living organisms as an analog of the Bose-Einstein condensate was proposed as the basis for understanding human consciousness; Marshall's hypothesis and related materials are discussed in the book (Scaruffi, 2014).

The idea of the regular world of stochastic energy processes as the basis of inherited biological structures is important, among other things, for understanding genetic intelligence. By genetic intelligence, we mean that part of the intellectual potency of living organisms that allows - with using of biological codes - to build, for example, from one fertilized cell an organism with trillions of cells that parental characteristics are reproduced in it by a multichannel noise-resistant manner, despite strong noises and constantly changing conditions of food and external influences in the course of life (Petoukhov and He, 2023). In this case, we are talking about a systematic growth - in the course of ontogenesis - of the number of body parameters and a corresponding increase in the dimension of its configuration space of states (Petoukhov, 2022c). This approach echoes the following opinion about the activity of the brain by R. Kurzweil, a well-known artificial intelligence specialist: the brain is a probabilistic recursive fractal that seems extremely complex, but in fact it may turn out to be much simpler than it seems (Tsang, 2016, p. 50).

With the development of physics and the emergence of quantum mechanics, people drew attention to the fundamental importance of the world of probabilities for understanding and modeling the objects and processes of the environment that they see. The described algebra-genetic studies complement the theme of the importance of the world of probabilities. The presented studies are consonant with the works of V. Nalimov (https://en.wikipedia.org/wiki/Vasily_Nalimov), who considered it his goal to build a probabilistic model of the language, and then of consciousness in general (Nalimov, 2015). In intersection with these statements, the book (Hofstadter et al., 1980, p. 115) emphasizes: "*In other words, artificial intelligence could be brought closer to mathematical thinking, if it were possible to realize the metric properties of the human thinking space … We are ready to go further and say that consciousness itself is geometrically structured: existentially, a person is geometric … In our minds, when constructing texts through which we perceive the World, something very similar to what happens in morphogenesis happens. We are ready to see in the depths of consciousness the same geometric images that are revealed in morphogenesis*".

4. The HBS-analysis of the DNA sequence of hydrogen bonds in the gene of the Titin

Sequences of genomic DNAs consist of genes and fragments (introns) that do not code for proteins. For example, it is currently believed that there are 4234 genes on human chromosome № 1, which are associated with 890 genetic diseases, including Alzheimer's disease, Parkinson's disease, glaucoma, breast and prostate cancer, etc. The total length of the sequences of these genes is only a small part of this chromosomal DNA, the rest of it is occupied by introns.

Since, as shown above, the application of the HBS-method to the analysis of genomic DNAs yields important results, it is logical to apply it to the analysis of the constituent parts of genomic DNAs, that is, to the analysis of genes and introns. In this section, we provide only one illustrative example of the application of this perspective method to gene analysis for revealing hidden rules and symmetries in stochastic organization of genetic sequences. For brevity, this analysis is called HBS-analysis.

Let us consider the human *TTN* gene, having 81940 bp and encoding the protein Titin, which is the largest protein in humans. Titin, also known as connectin, is important in the contraction of striated muscle tissues. The length of the DNA sequence of this longest gene is about 3000 times shorter than the length of the DNA sequence of the human chromosome № 1 analyzed above (Table 1). But surprisingly, the sets of percentages of *n*-plets in this highly truncated DNA obey practically the same (except for the case $n = 3$) rules for the dichotomous interconnections of percentages of *n*-plets with different values *n* in the binary representations of the DNA.

Representing each nucleotide in the nucleotide sequence of this gene by its number of hydrogen bonds (A = T = 2, C=P=3), we obtain a binary sequence of digits 2 and 3. Table 2 shows the results of the analysis of this binary H-sequence of hydrogen bonds by the HBS-method.

First of all, analysis of the phenomenological data in Table 2 reveals the following suffix dichotomous interconnections between presented probabilities of H-*n*-plets in the case of the gene *TTN*. The percentage of any H-monoplet is practically equal to the sum of percentages of those H-duplets, which are generated from this H-monoplet by addition of the suffixes 2 and 3 to it, for example, %2 ≈ %22 + %23. The percentage of any H-tetraplet is practically equal to the sum of those H-pentaplets, which are generated from it by addition of the suffixes 2 and 3 to it, for example, %2222 ≈ %22222 + %2223. Fig. 6 shows these percentage dichotomous interconnections in detail with rounded percent values.

But for the case n = 3, that is, for the shown percentage set of H-triplets, its dichotomous interconnections with neighboring percentage sequences of H-duplets and H-tetraplets are significantly disturbed in the considered *TTN* gene. One can recall here that nucleotide triplets encode amino acids of proteins.

Table 2
Phenomenological percentages of the H-*n*-plets in the corresponding H-sequence of the DNA of the gene *TTN* ($n = 1, 2, 3, 4, 5$). Its H-sequence contains about 81940 digits 2 and 3. Initial data on this chromosome were taken from the GenBank: https://www.ncbi.nlm.nih.gov/nuccore/X90568.1. Even H-*n*-plets starting with an even number 2 are in red, and odd H-*n*-plets starting with an odd number 3 are in blue.

%2	%3
0,560679	0,439321

%22	%23	%32	%33
0,284672	0,276983	0,275031	0,163315

%222	%223	%232	%233	%322	%323	%332	%333
0,156665	0,106469	0,134954	0,072054	0,185626	0,130377	0,155494	0,05836

%2222	%2223	%2232	%2233	%2322	%2323	%2332	%2333
0,160672	0,090685	0,100824	0,050154	0,100737	0,056957	0,056027	0,025264
%3222	%3223	%3232	%3233	%3322	%3323	%3332	%3333
0,090644	0,060129	0,056975	0,031122	0,050243	0,030993	0,025347	0,013226

%22222	%2223	%22232	%22233	%22322	%22323	%22332	%22333
0,038626	0,03887	0,044179	0,023798	0,04332	0,041555	0,036063	0,017818
%23222	%23223	%23232	%23233	%23322	%23323	%23332	%23333
0,044179	0,044667	0,051501	0,03112	0,03417	0,036795	0,023005	0,013608
%32222	%32223	%32232	%32233	%32322	%32323	%32332	%32333
0,038199	0,031731	0,04009	0,027825	0,04406	0,039114	0,0371	0,018672
%33222	%33223	%33232	%33233	%33322	%33323	%33332	%33333
0,024652	0,024286	0,030815	0,025812	0,01733	0,018977	0,010923	0,007139

Suffix dichotomies between percentages of H-monoplets and H-duplets:

%2=**0,561** ≈ %22+%23=**0,562**; %3=**0,439** ≈ %32+%33=**0,438**

Suffix dichotomies between percentages of H-tetraplets and H-pentaplets:

%2222=**0,077** ≈ %22222+%22223=**0,077**;
%2223=**0,068** ≈ %22232+%22233=**0,068**;
%2232=**0,088** ≈ %22322+%22323=**0,085**;
%2233=**0,054** ≈ %22332+%22333=**0,054**;
%2322=**0,087** ≈ %23222+%23223=**0,089**;
%2323=**0,081** ≈ %23232+%23233=**0,083**;
%2332=**0,072** ≈ %23322+%23323=**0,071**;
%2333=**0,035** ≈ %23332+%23333=**0,037**;
%3222=**0,068** ≈ %32222+%32223=**0,070**;
%3223=**0,070** ≈ %32232+%32233=**0,068**;
%3232=**0,083** ≈ %32322+%32323=**0,083**;
%3233=**0,054** ≈ %32332+%32333=**0,056**;
%3322=**0,052** ≈ %33222+%33223=**0,049**;
%3323=**0,059** ≈ %33232+%33233=**0,057**;
%3332=**0,033** ≈ %33322+%33323=**0,036**;
%3333=**0,020** ≈ %33332+%33333=**0,018**

Fig. 6. Dichotomous interconnections between percent values of H-*n*-plets with n = 1 and n = 2, and also with n = 4 and n = 5 for the gene TTN of the human longest protein Titin in the case of the suffix dichotomies. Percentage values are taken from Table 2 and rounded to the third decimal place. In each of the 18 shown equations, percent values, in the left side and in the right side of the equation are practically coincide each other. These values are marked by bold numbers.

∑%	Percentage of H-*n*-plets starting with 2	Percentage of H-*n*-plets starting with 3	∑%
0,561	%2	%3	0,439
0,562	%23+%22	%33+%32	0,438
0,470	%233+%232+%223+%222	%333+%332+%323+%322	0,530
0,562	%2333+%2332+%2323+ %2322+%2233+%2232+ %2223+%2222	%3333+%3332+%3323+ %3322+%3233+ %3232+ %3223+%3222	0,438
0,563	%23333+%23332+%23323+ %23322+%23233+%23232+ %23223+%23222+%22333+ %23332+%22323+%22322+ %22233+%22232+%22223+ %22222	%33333+%33332+%33323+ %33322+%33233+%33232+ %33223+%33222+%32333+ %32332 +%32323+%32322+ %32233+%32232+%32223+ %32222	0,437

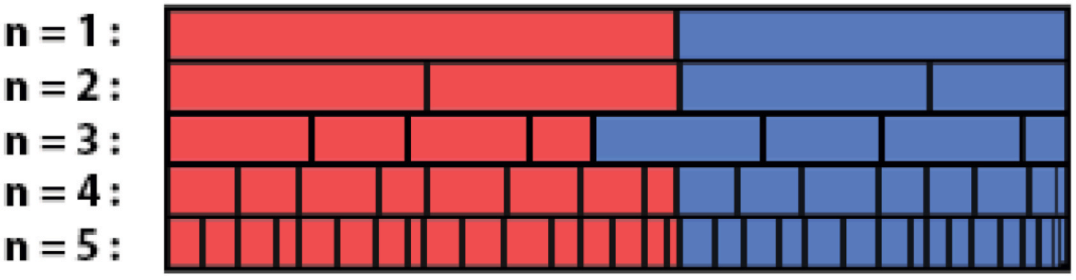


Fig. 7. At the top: numeric demonstration of the partial preservation of the dichotomy rule in the interconnections of the percentage sets of H-*n*-plets in the DNA of the gene *TTN* at n = 1, 2, 4, 5 (the left and the right columns present a practical invariance of sums of percentages of even and odd H-*n*-plets at these levels). In the case n = 3, the sequence of H-triplets gives a violation of this rule (marked in yellow). Rounded numeric data of percentages are taken from Table 2. **At the bottom**: the diagram of comparative percentages at different levels of these trees. At each level "n", lengths of intervals are proportional to percentages of corresponding H-*n*-plets from Table 2.

By analogy with the dichotomy trees of the percentages of H-*n*-plets in genomic DNAs shown in Fig. 3, it is possible to construct a tree of percentages of *n*-plets of hydrogen bonds 2 and 3 for the DNA of the analyzed gene. Fig. 7 shows that at $n = 3$ there is a significant violation of the dichotomy interconnections rule with the percent values of *n*-plets of neighboring levels (where $n = 2$ and $n = 4$).

By analogy with the above-presented case of the genomic DNA (2), the following rule of percentage equalities holds for the considered gene

n = 2	%23=**0,277** ≈ %32=**0,275**
n = 4	%2223=**0,0678** ≈ %3222=**0,0677** %2232=**0,088** ≈ %2322=**0,087** %2333=**0,035** ≈ %3332=**0,033**
n = 5	%22223=**0,039** ≈ %32222=**0,038** %22232=**0,044179** ≈ %23222=**0,044179** %22233=**0,024** ≈ %33222=**0,025** %22323=**0,042** ≈ %32322=**0,044** %22332=**0,036** ≈ %23322=**0,034** %22333=**0,018** ≈ %33322=**0,017** %23333=**0,014** ≈ %33332=**0,011**

Fig. 8. The numeric confirmation of the rule of percentage equalities for pairs of reversed H-*n*-plets in the gene *TTN* for cases of $n = 2, 4, 5$. Rounded numeric data of percentages are taken from Table 2.

under $n = 2, 4, 5$.

In genomic DNAs, the rule of percentage equalities in the set of genomic H-*n*-plets holds (2): those two H-*n*-plets, which are read as mirror (or reversed) copies of each another (such as 223 and 322), always have almost the same percent values in the genomic H-*n*-plets sequence. This rule holds as well for studied cases $n = 2, 4, 5$ in the analyzed gene *TTN* though the length of its sequence is about 3000 times shorter than the length of the DNA sequence of the human chromosome No. 1. Fig. 8 shows numerical data confirming this.

But for the case of $n = 3$, that is for the percentage sequence of H-triplets in the gene, this rule of percentage equalities is disturbed since this sequence has expressed percentages inequalities: %223 = 0,106 ≠ %322 = 0,186 and %233 = 0,072 ≠ %332 = 0,155. The reasons for such very special status of H-triplets sequence in the gene should be studied in the future.

5. Probability matrices, binary principles, and multidimensional numbers

The four nucleobases of DNA are interrelated by their symmetrical peculiarities into the united molecular ensemble having pairs of binary-oppositional traits or indicators (Fimmel et al., 2013; Petoukhov, 2008; Petoukhov and He, 2010; Stambuk, 1999):

- 1) Two letters are purines (A and G), and the other two are pyrimidines (C and T). From the standpoint of these binary-oppositional traits one can denote C = T = 0, A = G = 1;

	0	1
0	C	A
1	T	G

	00	01	10	11
00	CC	CA	AC	AA
01	CT	CG	AT	AG
10	TC	TA	GC	GA
11	TT	TG	GT	GG

	000	001	010	011	100	101	110	111
000	CCC	CCA	CAC	CAA	ACC	ACA	AAC	AAA
001	CCT	CCG	CAT	CAG	ACT	ACG	AAT	AAG
010	CTC	CTA	CGC	CGA	ATC	ATA	AGC	AGA
011	CTT	CTG	CGT	CGG	ATT	ATG	AGT	AGG
100	TCC	TCA	TAC	TAA	GCC	GCA	GAC	GAA
101	TCT	TCG	TAT	TAG	GCT	GCG	GAT	GAG
110	TTC	TTA	TGC	TGA	GTC	GTA	GGC	GGA
111	TTT	TTG	TGT	TGG	GTT	GTG	GGT	GGG

Fig. 9. The square tables of DNA-alphabets of 4 nucleotides, 16 duplets and 64 triplets with a strict arrangement of all components. Each of the tables is constructed by the coding algorithm using the binary numberings of its columns and rows based on molecular binary oppositions of the nucleobases (see explanations in the text).

- 2) Two letters are amino-molecules (A and C) and the other two are keto-molecules (G andT). From the standpoint of these traits one can denote A = C = 0, G = T = 1.

Taking this into account, it is convenient to represent DNA-alphabets of 4 nucleotides, 16 duplets, 64 triplets, …, 4^n n-plets in a form of appropriate square tables (Fig. 9) by the following coding algorithm. Entries of each column are enumerated by binary indicators "pyrimidine or purine" (C = T = 0, A = G = 1); for example, the triplet CAG and all other triplets in the same column are the combinations "pyrimidine-

	0	1
0	3	2
1	2	3

	00	01	10	11
00	33	32	23	22
01	32	33	22	23
10	23	22	33	32
11	22	23	32	33

	000	001	010	011	100	101	110	111
000	333	332	323	322	233	232	223	222
001	332	333	322	323	232	233	222	223
010	323	322	333	332	223	222	233	232
011	322	323	332	333	222	223	232	233
100	233	232	223	222	333	332	323	322
101	232	233	222	223	332	333	322	323
110	223	222	233	232	323	322	333	332
111	222	223	232	233	322	323	332	333

Fig. 10. The beginning of the family of matrices of n-plets of hydrogen bonds 2 and 3 with strict arrangements of all shown H-n-plets in the result of the binary representation of nucleotide n-plets in genetic matrices in Fig. 9 as described in the text.

purine-purine" and so this column is correspondingly enumerated 011. By contrast, entries of each row are numerated by binary indicators "amino or keto" (C = A = 0, T = G = 1); for example, the same triplet CAG and all other triplets in the same row are the combination "amino-amino-keto" and so this row is correspondingly numerated 001. By this coding algorithm, each of 4 letters, 16 duplets, 64 triplets, … takes automatically its own individual place and all components of these alphabets are arranged in a strict order in such alphabetic tables (Fig. 9). This strict ordering of the relative positions of all members of the DNA alphabets proves useful in revealing hidden regularities and rules in the genetic coding system (Petoukhov, 2008; Petoukhov and He, 2010, 2023; Boulay, 2022, 2023). In other words, the noted pairs of binary-oppositional traits in the DNA nucleotide alphabet are instruments of coding of special arrangements or orderings of members of DNA *n*-plet alphabets and belong to sets of biological codes including the coding amino acids by triplets, etc. (Barbieri, 2015). Let us demonstrate that these genetic matrices with special encoded orders of *n*-plets lead to useful algebraic tools for study DNA of genomes and genes.

Each of the 4 nucleotides A, T, C, and G is characterized by its number of hydrogen bonds 2 or 3 in its complementary pairs A-T and C-G in double-stranded DNAs: A = T = 2 and C = G = 3. Correspondingly, each nucleotide in the genetic matrices in Fig. 9 can be represented by its number of these hydrogen bonds. For example, in this case, the symbolic triplet CAG is represented by the digital sequence 323. In the result of such representation of all *n*-plets of nucleotides, the matrices of nucleotide *n*-plets in Fig. 9 are represented by corresponding matrices of digital *n*-plets of hydrogen bonds (that is, of H-*n*-plets) in Fig. 10.

In any DNA sequence of hydrogen bonds 2 and 3, analyzed by the HBS-method, each H-*n*-plet has its individual percent value. Representing each of H-*n*-plets in matrices in Fig. 10 by its amplitude of probability (that is, the square root from its percent value by analogy with quantum informatics), you get matrices of probability amplitudes of H-*n*-plets in Fig. 11.

Correspondingly, emergent properties of any DNA sequence of hydrogen bonds - as a complex system of H-*n*-plets - are related to (2^n*2^n)-matrices, which can serve as matrix operators. One can show that the matrices of (2*2)-, (4*4)-, (8*8)-orders in Fig. 11 are the matrix representations of 2^n-dimensional hyperbolic numbers (hyperbolic numbers are also known as split-complex numbers, double numbers, perplex number, or hyperbolic matrions) [https://www.en.wikipedia.org/wiki/Split-complex_number; Kantor and Solodovnikov, 1989; Petoukhov, 2008; Petoukhov and He, 2010]. Let us explain this with using so-called dyadic-shift decompositions of the (2^n*2^n)-matrices (Fig. 11), which are well-known in the theory of digital signals processing (Ahmed and Rao, 1975).

The (2*2)-matrix in Fig. 7 is decomposed into the sum of two sparse matrices e_0 and e_1 with appropriate coefficients as Fig. 8 shows. The set of these two matrices e_0 and e_1 is closed relative to multiplication and define the multiplication table of the known algebra of 2-dimensional hyperbolic numbers also showing in Fig. 8. These 2-dimensional hyperbolic numbers are usually written with the linear expression: $h = a + bj$, where *a* and *b* are real numbers, and j is the imaginary unit of hyperbolic numbers satisfying the condition $j^2 = +1$ (don't confuse with the imaginary unit i of complex numbers satisfying the condition $i^2 = -1$). In Fig. 12, the sparse matrix e_0 represents the real unit since $e_0^2 = e_0$, and the sparse matrix e_1 represents the imaginary unit j since $e_1^2 = e_0$. Each of these matrices is orthogonal, that is, the real specialization of a unitary matrix.

Fig. 13 shows the similar decomposition of the (4*4)-matrix from Fig. 11 into the sum of 4 sparse matrices s_0, s_1, s_2, and s_3 with coefficients, which are probability amplitudes of H-duplets. The set of these 4 matrices is also closed relative to multiplication and define the multiplication table of the known algebra of 4-dimensional hyperbolic numbers given in Fig. 13. Each of these sparse matrices is orthogonal. The sparse matrix s_0 represents the real unit, and the matrices s_1, s_2, s_3 represent imaginary unites of these 4-dimensional hyperbolic numbers, whose linear form is $as_0 + bs_1 + cs_2 + ds_3$ where a, b, c, d are real numbers.

Analogically by dyadic-shift decompositions of the (2^n*2^n)-matrices of probability amplitudes of *n*-plets under $n = 3, 4, 5, \ldots$, one can show

$$\begin{vmatrix} \sqrt{}\%3, \sqrt{}\%2 \\ \sqrt{}\%2, \sqrt{}\%3 \end{vmatrix} = \sqrt{}\%3 * \begin{vmatrix} 1, 0 \\ 0, 1 \end{vmatrix} + \sqrt{}\%2 * \begin{vmatrix} 0, 1 \\ 1, 1 \end{vmatrix} = \sqrt{}\%3 * e_0 + \sqrt{}\%2 * e_1$$

*	e_0	e_1
e_0	e_0	e_1
e_1	e_1	e_0

Fig. 12. The decomposition of the (2*2)-matrix of probability amplitudes of H-monoplets from Fig. 11 revealing that this matrix is the matrix representation of 2-dimensional hyperbolic number $\sqrt{}\%3 + j*\sqrt{}\%2$. The table of multiplication of basic elements of the algebra of 2-dimensional hyperbolic numbers is shown.

	0	1
0	√%3	√%2
1	√%2	√%3

	00	01	10	11
00	√%33	√%32	√%23	√%22
01	√%32	√%33	√%22	√%23
10	√%23	√%22	√%33	√%32
11	√22	√%23	√%32	√%33

	000	001	010	011	100	101	110	111
000	√%333	√%332	√%323	√%322	√%233	√%232	√%223	√%222
001	√%332	√%333	√%322	√%323	√%232	√%233	√%222	√%223
010	√%323	√%322	√%333	√%332	√%223	√%222	√%233	√%232
011	√%322	√%323	√%332	√%333	√%222	√%223	√%232	√%233
100	√%233	√%232	√%223	√%222	√%333	√%332	√%323	√%322
101	√%232	√%233	√%222	√%223	√%332	√%333	√%322	√%323
110	√%223	√%222	√%233	√%232	√%323	√%322	√%333	√%332
111	√%222	√%223	√%232	√%233	√%322	√%323	√%332	√%333

Fig. 11. The beginning of the family of matrices of probability amplitudes of H-*n*-plets for a voluntary DNA, which is studied as a binary sequence of its hydrogen bonds 2 and 3 by the HBS-method. Probability amplitudes of the *n*-plets starting with 3 hydrogen bonds are marked by blue and probability amplitudes of *n*-plets starting with 2 hydrogen bonds are marked by red.

$$\begin{vmatrix} \sqrt{}\%33, \sqrt{}\%32, \sqrt{}\%23, \sqrt{}\%22 \\ \sqrt{}\%32, \sqrt{}\%33, \sqrt{}\%22, \sqrt{}\%23 \\ \sqrt{}\%23, \sqrt{}\%22, \sqrt{}\%33, \sqrt{}\%32 \\ \sqrt{}\%22, \sqrt{}\%23, \sqrt{}\%32, \sqrt{}\%33 \end{vmatrix} = \sqrt{}\%33 \begin{vmatrix} 1,0,0,0 \\ 0,1,0,0 \\ 0,0,1,0 \\ 0,0,0,1 \end{vmatrix} + \sqrt{}\%32 \begin{vmatrix} 0,1,0,0 \\ 1,0,0,0 \\ 0,0,0,1 \\ 0,0,1,1 \end{vmatrix} + \sqrt{}\%23 \begin{vmatrix} 0,0,1,0 \\ 0,0,0,1 \\ 1,0,0,0 \\ 0,1,0,0 \end{vmatrix} +$$

$$+ \sqrt{}\%22 \begin{vmatrix} 0,0,0,1 \\ 0,0,1,0 \\ 0,1,0,0 \\ 1,0,0,0 \end{vmatrix} = \sqrt{}\%33 * s_0 + \sqrt{}\%32 * s_1 + \sqrt{}\%23 * s_2 + \sqrt{}\%22 * s_3$$

*	s_0	s_1	s_2	s_3
s_0	s_0	s_1	s_2	s_3
s_1	s_1	s_0	s_3	s_2
s_2	s_2	s_3	s_0	s_1
s_3	s_3	s_2	s_1	s_0

Fig. 13. The dyadic-shift decomposition of the (4*4)-matrix of percent values of H-duplets from Fig. 7 revealing that this matrix is the matrix representation of the shown 4-dimensional hyperbolic number where s_0 represents the real unit, and s_1, s_2, and s_3 represent imaginary units of the algebra of 4-dimensional hyperbolic numbers. The table of multiplication of basic elements of this algebra is shown where the bold frame marks the subalgebra of 2-dimensional hyperbolic numbers.

that each of these matrices is a matrix representation of a corresponding 2^n-dimensional hyperbolic number (Petoukhov, 2023; Petoukhov and He, 2023). Each of such the 2^n-dimensional numbers determines a corresponding vector in an appropriate 2^n-dimensional vector space. More precisely, these vectors $\vec{V}_n$, where the index n denotes a belonging of the vector to 2^n-dimensional space, have the following sets of blue and red coordinates (3) in accordance with Figs. 11–13, etc.:

The set of vectors $\vec{V}_n$ (3) forms a hierarchical system where all vectors of percentages are interconnected by using the phenomenological rule of suffix dichotomies: the sum of percentages of two $(n+1)$-plets, which have the same root and differ only by their suffixes, is practically equal to the percentage of the n-plet with the same root. This gives a convolution of the coordinates of the vector $\vec{V}_{n+1}$ into the coordinates of the vector $\vec{V}_n$. Accordingly, knowing the coordinates of the vector $\vec{V}_n$ with a large value of n, we can go down to the coordinates of the vector $\vec{V}_2$. The author supposes that each of the coordinates of these vectors can have some code meaning, and then the entire hierarchy of vectors $\vec{V}_n$ can serve as a code percentage tree with a conserve memory of code meanings at its different levels under a grows of the value n. Each of the coordinates of vectors $\vec{V}_n$ can serve as an initial vector for its separate hierarchy of 2^n-dimensional vectors as it was explained by the example in Fig. 4.

$$\vec{V}_2 = [\sqrt{}\%3, \sqrt{}\%2]; \qquad \vec{V}_4 = [\sqrt{}\%33, \underline{\sqrt{}\%32}, \underline{\sqrt{}\%23}, \sqrt{}\%22];$$

$$\vec{V}_8 = [\sqrt{}\%333, \underline{\sqrt{}\%332}, \sqrt{}\%323, \underline{\sqrt{}\%322}, \underline{\sqrt{}\%233}, \sqrt{}\%232, \underline{\sqrt{}\%223}, \sqrt{}\%222];$$

$$\vec{V}_{16} = [\sqrt{}\%3333, \underline{\sqrt{}\%3332}, \underline{\sqrt{}\%3323}, \underline{\sqrt{}\%3322}, \underline{\sqrt{}\%3233}, \underline{\sqrt{}\%3232}, \sqrt{}\%3223, \underline{\sqrt{}\%3222},$$
$$\underline{\sqrt{}\%2333}, \sqrt{}\%2332, \underline{\sqrt{}\%2323}, \underline{\sqrt{}\%2322}, \underline{\sqrt{}\%2233}, \underline{\sqrt{}\%2232}, \underline{\sqrt{}\%2223}, \sqrt{}\%2222], \ldots \qquad (3)$$

In accordance with the described phenomenological rule of practical equalities of percentages of those H-n-plets, which are reversed copies of each another (such as 223 and 322), values of many coordinates in each of these vectors are repeated twice. All such coordinates are marked by bold and underlines in (3). It emphasizes the existence - in these vectors of genomic DNAs - of some dichotomous symmetries, associated with the concept of left and right (in the case of vectors of DNAs of genes, such symmetries are absent). The lengths $|\vec{V}_n|$ of all the vectors are equal to one both in the case of genomic DNAs and in the case of gene DNAs:

$$|\vec{V}_2| = \%3 + \%2 = 1; \quad |\vec{V}_4| = \%33 + \%32 + \%23 + \%22 = 1; \ \ldots. \qquad (4)$$

For example, in the case of the H-sequence of human chromosome № 1 (Table 1), the vectors (3) get the following numeric form (percentage values are rounded):

To transform into each other such 2^n-dimensional vectors of equal length, but belonging to spaces of different dimensions, the author introduced the algebraic operation of the tensor-unitary transformation (Petoukhov, 2022c,d). This operation is reminiscent of the unitary (or orthogonal) transformation operations, which preserve the lengths of vectors in a fixed-dimensional space and are important for quantum computer science. Unitary (or orthogonal) transformations play a key

$$\vec{V}_2 = [\sqrt{}0.417, \sqrt{}0.583]; \qquad \vec{V}_4 = [\sqrt{}0.162, \underline{\sqrt{}0.255}, \underline{\sqrt{}0.255}, \sqrt{}0.328];$$

$$\vec{V}_8 = [\sqrt{}0.063, \underline{\sqrt{}0.099}, \sqrt{}0.127, \underline{\sqrt{}0.128}, \underline{\sqrt{}0.099}, \sqrt{}0.156, \underline{\sqrt{}0.128}, \sqrt{}0.200];$$

$$\vec{V}_{16} = [\sqrt{}0.024, \underline{\sqrt{}0.039}, \underline{\sqrt{}\%0.053}, \underline{\sqrt{}0.046}, \underline{\sqrt{}0.053}, \underline{\sqrt{}0.074}, \sqrt{}0.055, \underline{\sqrt{}0.073},$$
$$\underline{\sqrt{}0.039}, \sqrt{}0.060, \underline{\sqrt{}0.074}, \underline{\sqrt{}0.082}, \underline{\sqrt{}0.046}, \underline{\sqrt{}0.082}, \underline{\sqrt{}0.073}, \sqrt{}0.127]; \ldots \qquad (5)$$

role in quantum computing, where all calculations are based on them, and any unitary transformation can serve as a logic gate. Tensor-unitary transformations are of interest for the development of quantum computing. These transformations can be considered as step-by-step memory expansion operations with conserving all memory at previous levels and also as multi-reproduction operations for structures.

For further development of evolutionary and personalized genetics, and also for biotechnology and pharmacology, it is useful to have tools of compare analysis of different DNAs using a fundamental algebraic basis, which relies on fundamental genomic rules and which can be applied to all DNAs. Taking into account the described results, the author proposes a program of binary stochastic certifications of different DNA sequences of genomes, genes, introns, etc. on the basis of representation of their families of percentages of H-*n*-plets by appropriate hierarchies of 2^n-dimensional hyperbolic numbers and their vectors. 2^n-dimensional hyperbolic numbers corresponding to the same value n (for example, $n = 3$) can be added, multiplied, divided with each other, each time obtaining a new 2^n-dimensional hyperbolic number (only you must take into account the existence of so-called zero divisors in this algebra (Kantor and Solodovnikov, 1989)).

In this approach, one can compare different single-stranded DNAs not as sequences of chemical elements but as families of 2^n-dimensional numbers or vectors from algebras of 2^n-dimensional hyperbolic numbers. It allows applying modern tools of metric-vector analysis, tensor-matrix theory, formalisms of quantum informatics, etc. to reveal hidden regularities and affinity relations in a huge set of DNA sequences. This universal algebraic approach seems to be useful, for example, to study the degree of closeness and relationship of separate species on the evolutionary tree of organisms. Binary sequences of hydrogen bonds of DNAs of genomes, genes, introns, etc. are all covered and embraced by a huge algebraic multi-layered network of 2^n-dimensional hyperbolic numbers with many dichotomous and other relationships between them. In the author's opinion, an international program is needed to systematically study this algebraic network and the relationships between its members in the interests of evolutionary biology, personal genetics, and biotechnologies. One should add that this program can include studying not only binary sequences of hydrogen bonds but also binary sequences of purines (A, G) and pyrimidines (C, T), and binary sequences of keto-molecules (G, T) and amino-molecules (A, C) in single-stranded DNAs, which have similar quasi-stochastic regularities as it was described in the preprint (Petoukhov, 2023).

Speaking about hyperbolic numbers in the genetic code system, it should be noted their close connection with hyperbolic geometry by Lobachevsky, to which structures of many genetically inherited physiological phenomena are related (Ganea et al., 2018; Ghaninia et al., 2022; Zhang et al., 2023; Zhou and Sharpee, 2021; Zhou et al., 2018). The presented results about the connection of the genetic code system with hyperbolic numbers testify in favor that the mentioned hyperbolic physiological phenomena did not appear out of nowhere, but are based on the genetic coding structures. The genetic code system possesses remarkable noise-immunity properties and hyperbolic geometry is related to noise-immunity information coding (Preskill, 2016). In addition, hyperbolic numbers and hyperbolic geometry are actively used around the world to create deep neural networks for artificial intelligence systems (see a survey (Peng et al., 2022)). These results enrich Code Biology by adding to its content the algebraic connection of the genetic coding system with a set of hyperbolic physiological phenomena and with the wide topic of hyperbolic deep neural networks having many practical applications in fields of coding and proceedings of information. In our opinion, it gives new perspectives for development of Code Biology as the important scientific direction.

One should recall here that, in modern science, multidimensional numbers have played the role of the magic tool for development of theories and calculations in the field of problems of heat, light, sounds, fluctuations, elasticity, gravitation, magnetism, electricity, current of liquids, the quantum-mechanical phenomena. But in algebraic biology, multi-dimensional number systems do not practically used till recent time. Multidimensional numerical systems in algebraic biology can serve as a bridge for the convergence of physics and biology in order to their mutual enrichment.

It should be additionally noted one more binary-oppositional feature in molecular genetic system described in Rumer (1968) and Fimmel and Strüngmann (2016): the binary opposition of triplets with strong and weak roots separates the set of 64 triplets into the subset of 32 triplets with strong roots and 32 triplets with weak roots. This binary opposition has a coding meaning as well since leads to transformation of genetic percentage matrices (like matrices in Fig. 11) into matrices, which present 4-dimensional split-quaternions by Cockle and their algebraic extensions, which serve as one more multidimensional numeric tool to study binary sequences of single-stranded DNAs (Petoukhov, 2008; 2022e; Petoukhov and He, 2023). In this approach, new options of realization of the ancient principle "like begets like" in genetic matrices are additionally revealed (Petoukhov, 2023; Petoukhov and He, 2023).

One can else note the following usefulness of representing of ensembles of percent values of H-*n*-plets of different DNA sequences of hydrogen bonds in such form of families of genetic (2^n*2^n)-matrices, which are algorithmically constructed on binary-oppositional features of 4 nucleobases A, T, C, G (see Fig. 9). Each such numeric square matrix has its eigenvalues, eigenvectors, and characteristic polynomial, which can be used to study hidden emergent properties of the considered DNA sequence as a complex system. Such discovered connections of DNA H-sequences with characteristic polynomials bring the algebraic genetics closer to algebraic geometry, which is a branch of mathematics studies zeros of multivariate polynomials [https://en.wikipedia.org/wiki/Algebraic_geometry#:~:text=Algebraic%20geometry%20is%20a%20branch,about%20these%20sets%20of%20zeros]. Modern algebraic geometry is based on the use of abstract algebraic techniques, mainly from commutative algebra, for solving geometrical problems about these sets of zeros. Algebraic geometry occupies a central place in modern mathematics and has multiple conceptual connections with such diverse fields as coding theory, complex analysis, topology, and number theory. It is the applications of algebraic geometry in coding theory that make credit cards and the internet work. Now formalisms and achievements of algebraic geometry can be used in matrix genetics and algebraic biology to provide a progress of knowledge on secrets of coded organization of living bodies. In turn, the structural features of the genetic informatics system can be useful for the further development of algebraic geometry and the expansion of its applications.

Fig. 14 shows an example of the characteristic polynomial for the (4*4)-matrix of percent values of 4 duplets of hydrogen bonds (from Fig. 9), based on these percentages for human chromosome № 1 DNA from Table 1 (this polynomial is calculated using the online calculator https://mathforyou.net/online/matrices/charpoly/).

From the described point of view, stochastic organization of any DNA sequence of hydrogen bonds can be characterized by a corresponding set of characteristic polynomials of (2^n*2^n)-matrices of percentages of its H-*n*-plets. Hidden interrelations in such sets of characteristic polynomials for probabilistic (2^n*2^n)-matrices of H-*n*-plets, presenting DNAs of different genomes and genes, form interesting topic for future research.

W =

%33	%32	%23	%22
%32	%33	%22	%23
%23	%22	%33	%32
%22	%23	%32	%33

=

0,163	0,255	0,255	0,328
0,255	0,163	0,328	0,255
0,255	0,328	0,163	0,255
0,328	0,255	0,255	0,163

$$P(W, \lambda) = \lambda^4 - 0.652\lambda^3 - 0.315854\lambda^2 - 0.03301122\lambda - 0.000517792275$$

Fig. 14. At the top: the matrix W of percentages of duplets of hydrogen bonds in the case of DNA of human chromosome № 1 (Table 1). Percentages are rounded. **At the bottom**: the characteristic polynomial P(W, λ) of this matrix.

6. Some concluding remarks

This article shows connections of the ancient principle "like begets like" not only with double-stranded DNAs but also with holistic families of structural molecular ensembles of the genetic coding system in their algebraic-matrix presentations. The author believes that this principle plays a key role in genetics, and therefore, within the framework of Code Biology, many features of inherited biological structures can and should be studied precisely in connection with this principle. These include, for example, the following: morphogenetic symmetries; biological fractal-like patterns; the golden section, which since the Renaissance is regarded as a mathematical symbol of self-reproduction; universal rules of stochastic organization of information sequences of genomic DNAs connected with dichotomies of probabilities and corresponding fractal-like dichotomous trees of probabilities. They also include our visual perception with its optical system transmitting external images to the retina in complementary inverted and reduced forms, which are complementary replicated in the brain for decoding estimations. It concerns also the existence of mirror neurons in the brain of human and animals, which are considered by many as participants of the cognitive activity, an origin of languages, automatic imitation, and other things connecting with code biology (Ferrari and Rizzolatti, 2014; Heyes, 2010).

The described results present not only phenomenological genetic rules concerning new symmetries of genetic probabilities but also algebraic tools, which are related to these rules and have many applications in the fields of information coding and artificial intelligence: algebras of 2^n-dimensional hyperbolic numbers, connected with hyperbolic geometry by Lobachevsky and with deep neural networks of hyperbolic kinds; polynomials and matrices of algebraic geometry; theory of noise-immunity coding, fractals, and dichotomous graphs. They show that known data by different authors about hyperbolic structures in inherited physiological phenomena can be interpreted as based on hyperbolic structures in the genetic code system. These results support the author's hypothesis that the systems of mirror neurons and DNAs complementary replications are not isolated parts of the organism, but they are included in a holistic bio-algebraic complex realizing the inherited principle "like begets like".

On this way, the discovery of universal rules of dichotomies of probabilities in information sequences of genomic DNAs, which fundamentally differ from inherited constructional dichotomies in biological bodies, shows that widely knows dichotomies in biological bodies have prototypes in sets of dichotomies of probabilities in informatics of genomic DNAs. This allowed formulating the author's thought about existence of a hidden world of binary organized quasi-stochastic essences, which are hidden progenitors of biological structures and perhaps are connected with Bose-Einstein condensate and the Frohlich's theory of a quantum-mechanical long-range bond in biological structures.

The described results show that the noted principle is essential for studying and modeling algebraic peculiarities of molecular ensembles of the genetic code system with its set of binary-oppositional features participating in biological coding. New biological symmetries, connected with this principle, were revealed in the families of the genetic matrices and in stochastic organization of information sequences of genomic DNAs. Complementary replication (or interconnections) in a wide sense is a systemic phenomenon in the genetic organization, including dichotomous fractal-like trees of probabilities of genomic DNAs (Petoukhov, 2023; Petoukhov and He, 2023). The presented materials testify in favor that the multicomponent information system of DNA molecules is permeated with many code properties, that is, it is richer in code structures than it was known so far. Using these code properties allowed discover previously unknown emergent properties of genomic informatics systems and of their quasi-stochastic organization, associated, in particular, with multidimensional numerical systems described in this article.

What could be the bioinformatic meaning of such a multi-layered organization of genomic informatics? One of the possible directions of searching for answers to this question is associated with the development of a non-near analogy between this multi-layered organization and Packet Data Network (PDN) used in the Internet and in the technique of multichannel noise-immune communication [https://en.wikipedia.org/wiki/Public_data_network]. According to this technology, the transmitted information is cut into fragments of a specified length, which are transmitted along different routes. It is possible that data on the quasi-stochastic organization of each of the *n*-plet representations of genomic DNA sequences are also transmitted via different routes and addresses for different aims.

The newly received knowledge about the algebraic features of the genetic molecular systems opens new approaches to understanding interconnections of the genetic system with structural peculiarities of inherited physiological systems. All physiological systems should be coordinated with the genetic code to be genetically encoded for their transmission to the next generations. This determines the importance of studying the algebraic features of the molecular genetic system for understanding the origin and modeling of structures of inherited physiological complexes, and also for the development of evolutionary biology and genetic biomechanics. The results obtained are applicable for the further development of code biology (Barbieri, 2015). The ideology of code biology gets at its disposal the tools of algebraic-numerical analysis from the arsenal of modern mathematical science and computer science. These tools include the tensor-matrix apparatus, metric vector analysis in multidimensional configuration spaces, quantum informatics formalisms. On this path, code biology becomes an important part of modern mathematical natural science. In study of genetic and other biological sequences, the described author's method of hierarchical binary stochastics (the UBS-method) seems to be important and is recommended for further using.

It should be added that since the philosophical works of Martin Heidegger, there exists an idea that language is smarter than us. A rich language has an extensive history of development and effective application to reality. Using this or that language, we indirectly use all experience of its formation and applications to reality. As soon as the language is smarter than us, then a good language can guide us, suggesting new solutions and directions of search. The author encountered this feature of the well-developed and widely used scientific language of matrix-tensor analysis, which is one of the foundations of modern mathematical natural science: after receiving the first confirmations of this language adequacy for modeling a genetic coding system, the author began to exploit its rich and interconnected capabilities in the new scientific field, that is, in matrix genetics, guided by inner features of this algebraic language and obtaining new and valuable biological results. In other words, genetic structures are translated by the author into this algebraic language with final receiving completely new meanings and universal regularities to which this language has brought us.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

Acknowledgments

Some results of this paper have been possible due to long-term cooperation between Russian and Hungarian Academies of Sciences on the theme "Non-linear models and symmetrologic analysis in biomechanics, bioinformatics, and the theory of self-organizing systems", where the author was a scientific chief from the Russian Academy of Sciences. The author is grateful to G. Darvas, E. Fimmel, A.A. Koblyakov, S. Ya. Kotkovsky, M. He, Z.B. Hu, Yu.I. Manin, V. Rosenfeld, I.S. Soshinsky, I.V. Stepanyan, V.I. Svirin, and G.K. Tolokonnikov for their collaboration.

References

Ahmed, N., Rao, K., 1975. Orthogonal Transforms for Digital Signal Processing. Springer-Verlag Inc., New-York.

American Institute of Physics, 1999. The Secret Nature of Hydrogen Bonds. - ScienceDaily. www.sciencedaily.com/releases/1999/01/990121074852.htm.

Bank, E., 2022. How much time does it take for a DNA molecule to replicate? Sciencing. https://sciencing.com/much-time-dna-molecule-replicate-21660.html.

Barbieri, M., 2015. Code Biology. A New Science of Life. Springer-Verlag, ISBN 978-3-319-14534-1. https://doi.org/10.1007/978-3-319-14535-8.

Boulay, J.Y., 2022. Numbering of the Twenty Proteinogenic Amino-acids 3/2 Ratios Inside the Genetic Code. https://doi.org/10.13140/RG.2.2.13001.83046. Preprint.

Boulay, J.Y., 2023. Numbering of the twenty proteinogenic amino acids and new alphanumerical nomenclature proposal to them. Symmetry: Cult. Sci. 34 (1), 61–86. https://doi.org/10.26830/symmetry_2023_1_061.

Chapeville, F., Haenni, A.-L., 1974. Biosynthese des proteins. Hermann Collection, Paris Methodes in French).

Chatterjee, P., Biswas, S., Chakraborty, T., 2020. Hydrogen bonding effects on vibrational dynamics and photochemistry in selected binary molecular complexes. J. Indian Inst. Sci. 100, 155–165. https://doi.org/10.1007/s41745-019-00158-0.

Ferrari, P.F., Rizzolatti, G., 2014. Mirror neuron research: the past and the future. Philos. Trans. R. Soc. Lond. B Biol. Sci. 369 (1644), 20130169 https://doi.org/10.1098/rstb.2013.0169.

Fimmel, E., Danielli, A., Strüngmann, L., 2013. On dichotomic classes and bijections of the genetic code. J. Theor. Biol. 336, 221–230. https://doi.org/10.1016/j.jtbi.2013.07.027.

Fimmel, E., Strüngmann, L., 2016. Yury Borisovich Rumer and his 'biological papers' on the genetic code. Phil. Trans. R. Soc. A 374, 20150228. https://doi.org/10.1098/rsta.2015.0228.

Fimmel, E., Gumbel, M., Karpuzoglu, A., Petoukhov, S., 2019. On comparing composition principles of long DNA sequences with those of random ones. Biosystems 180, 101–108.

Fröhlich, H., 1970. Long range coherence and the action of enzymes. Nature 228, 1093. https://doi.org/10.1038/2281093a0.

Fröhlich, H., 1978. Coherent electric vibrations in biological systems and cancer problem. IEEE Trans. Microw. Theor. Tech. 26, 613–617.

Fröhlich, H., 1980. The biological effects of microwaves and related questions. Adv. Electron. Electron Phys. 53, 85–152. New York, USA: Academic Press.

Fröhlich, H., 1988. In: Fröhlich, Herbert (Ed.), Introduction. Theoretical physics and biology, Biological Coherence and Response to External Stimuli. Springer, pp. 3–24. ISBN 978-0-387-18739-6.

Ganea, O., Bŕecigneul, G., Hofmann, T., 2018. Hyperbolic neural networks. Adv. Neural Inf. Process. Syst. 31, 5345–5355.

Ghaninia, M., Zhou, Y., Knauer, A.C., Schiestl, F.P., Sharpee, T.O., Smith, B.H., 2022. Hyperbolic odorant mixtures as a basis for more efficient signaling between flowering plants and bees, 17 PLoS One 13 (7), e0270358. https://doi.org/10.1371/journal.pone.0270358. PMID: 35830455; PMCID: PMC9278781.

Henderson, S., 2018. Bose-einstein condensate induced at room temperature. In: Novus Light Technologies Today. https://www.novuslight.com/bose-einstein-condensate-induced-at-room-temperature_N8055.html.

Heyes, C.M., 2010. Where do mirror neurons come from? Neurosci. Biobehav. Rev. 34 (4), 575–583. https://doi.org/10.1016/j.neubiorev.2009.11.007.

Hofstadter, D. R. Gödel, Escher, Bach, 1980. An Eternal Golden Braid. Vintage Books.

Ivanov Vyach. Vs. Even and odd, 1978. Asymmetry of the Brain and Sign Systems. Soviet radio, Moscow, p. 184 (in Russian; the book was translated into German in 1983; Hungarian - 1986; Japanese - 1988; Latvian - 1990).

Kantor, I.L., Solodovnikov, A.S., 1989. Hypercomplex Numbers. Springer-Verlag, Berlin, New York, ISBN 978-0-387-96980-0.

Khazina, L.V., 2015. Octaves, Fractals, Evolution and Neurogenetics. LAP Lambert Academic Publishing, ISBN 978-3-659-76893-4, 13, in Russian).

Liberman, E.A., 1990. How Does a Living Cell Work. Znanie, Moscow, 5-07-000739-4 (in Russian).

Liberman, E.A., Minina, S.V., 1996. Cell molecular computers and biological information as the foundation of nature's laws. Biosystems 38, 173–177.

Marshall, I.N., 1989. Consciousness and Bose-Einstein condensates. New Ideas Psychol. 7, 73–83.

McFadden, J., Al-Khalili, J., 2018. The origins of quantum biology. Proc. R. Soci. A 474 (2220), 1–13. https://doi.org/10.1098/rspa.2018.0674.

Medvedev, A.E., 2020. Method for constructing an asymmetric human bronchial tree in normal and pathological conditions. Math. Biol. Bioinf. 15 (2), 148–157. https://doi.org/10.17537/2020.15.148.

Nalimov, V.V., 2015. I Scatter My Thoughts. On the Road and at the Crossroads. Center for Humanitarian Initiatives, Moscow, 978-5-98712-521-2.

Pauling, L., 1940. The Nature of the Chemical Bond and the Structure of Molecules and Crystals: an Introduction to Modern Structural Chemistry, second ed. Oxford University Press.

Pellionisz, A., 1989. In: Cotterill (Ed.), Neural geometry: towards a fractal model of neurons, Models in Brain Function. Cambridge Univ. Press.

Pellionisz, A.J., Graham, R., Pellionisz, P.A., Perez, J.C., 2013. In: Manto, E., Gruol, D., Schmahmann, J., Koibuchi, N., Rossi, F. (Eds.), Recursive genome function of the Cerebellum: geometric unification of neuroscience and genomics, Recursive Genome Function of the Cerebellum: Geometric Unification of Neuroscience and Genomics. Springer. https://doi.org/10.1007/978-94-007-1333-8_61. Chapter: 61.

Peng, W., Varanka, T., Mostafa, A., Shi, H., Zhao, G., 2022. Hyperbolic deep neural networks: a survey. IEEE Trans. Pattern Anal. Machi. Intell. 44 (12), 10023–10044.

Petoukhov, S.V., 2008. Matrix Genetics, Algebras of the Genetic Code, Noise Immunity, 978-5-93972-643-6. RCD, Moscow, p. 316. in Russian).

Petoukhov, S.V., 2019. Nucleotide Epi-Chains and New Nucleotide Probability Rules in Long DNA Sequences, 201904001. Preprints 2019. https://www.preprints.org/manuscript/201904.0011/v2.

Petoukhov, S.V., 2020a. Hyperbolic Numbers in Modeling Genetic Phenomena, p. 181. https://doi.org/10.20944/preprints201908.0284.v4. Preprints 2019, 2019080284, version 4.

Petoukhov, S.V., 2020c. The Rules of Long DNA-Sequences and Tetra-Groups of Oligonucleotides arXiv:1709.04943v6 [q-bio.OT].

Petoukhov, S.V., 2021a. Modeling inherited physiological structures based on hyperbolic numbers. Biosystems 199, 104285. https://doi.org/10.1016/j.biosystems.2020.104285. ISSN 0303-2647.

Petoukhov, S.V., 2021d. Algebraic harmony and probabilities in genomes. Long-range coherence in quantum code biology. Biosystems 209, 104503.

Petoukhov, S.V., 2022a. Binary oppositions, algebraic holography, and stochastic rules in genetic informatics. Biosystems 221, 104760. https://doi.org/10.1016/j.biosystems.2022.104760.

Petoukhov, S.V., 2022b. The stochastic organization of genomes and the doctrine of energy-information evolution based on bio-antenna arrays. Biosystems, 104712. https://doi.org/10.1016/j.biosystems.2022.104712. ISSN 0303-2647.

Petoukhov, S.V., 2022c. Genetic intelligence, the doctrine of energy-informational evolution based on bio-antenna arrays, and quantum informatics. Biomachsystems 6 (4), 45–64 (in Russian).

Petoukhov, S.V., 2022d. Rules of Stochastics of Genomic DNAs, Biological Dualism "Stochastics-Determinism", and Tensor-Unitary Transformations, 2022, 2022080435. https://doi.org/10.20944/preprints202208.0435.v1. Preprints.

Petoukhov, S.V., 2022e. Binary oppositions, algebraic holography, and stochastic rules in genetic informatics. Biosystems 221, 104760. https://doi.org/10.1016/j.biosystems.2022.104760.

Petoukhov, S.V., 2023. The Principle "Like Begets Like" in Molecular and Algebraic-Matrix Genetics, 2022, 2022110528. https://doi.org/10.20944/preprints202211.0528.v3. Preprints.

Petoukhov, S.V., He, M., 2010. Symmetrical Analysis Techniques for Genetic Systems and Bioinformatics: Advanced Patterns and Applications. IGI Global, USA.

Petoukhov, S.V., He, M., 2023. Algebraic Biology, Matrix Genetics, and Genetic Intelligence. – Singapore. World Scientific, p. 616. https://doi.org/10.1142/13468.

Plumhof, J.D., Stoferle, Th, Mai, L., Scherf, U., Mahrt, R.F., 2014. Room-temperature Bose-Einstein condensation of cavity exciton-polaritons in a polymer. Nat. Mater. 13 (3), 247–252. https://doi.org/10.1038/nmat3825.

Prabhu, V.V., 1993. Symmetry observation in long nucleotide sequences. Nucleic Acids Res. 21, 2797–2800.

Preskill, J., 2016. Stability, Topology, Holography: the Many Facets of Quantum Error Correction. Presentation at American Physical Society. http://theory.caltech.edu/~preskill/talks/APS-March-2016-preskill.pdf.

Rumer, YuB., 1968. Codon systematization in the genetic code. Dokl. Akad. Nauk SSSR 183 (1), 225–226.

Scaruffi, P., 2014. Thinking about Thought. CreateSpace Independent Publishing Platform. ISBN-13: 978-1503361065.

Shan, S.O., Herschlag, D., 1999. Hydrogen bonding in enzymatic catalysis: analysis of energetic contributions. Methods Enzymol. 308, 246–276. https://doi.org/10.1016/s0076-6879(99)08013-1.

Stambuk, N., 1999. Circular coding properties of gene and protein sequences. Croat. Chem. Acta 72 (4), 999–1008.

Trylska, J., Grochowski, P., McCammom, J.A., 2004. The role of hydrogen bonding in the enzymatic reaction catalyzed by HIV-1 protease. Protein Sci. 13 (2), 513–528.

Tsang, W.H., 2016. Fractal Brain Theory. Lulu.com, ISBN 978-1-326-75322-1, p. 530.

Werner, G., 2010. Fractals in the nervous system: conceptual implications for theoretical neuroscience. Front. Physiol. 1 https://doi.org/10.3389/fphys.2010.00015.

Zhang, H., Rich, P.D., Lee, A.K., Sharpee, T.O., 2023. Hippocampal spatial representations exhibit a hyperbolic geometry that expands with experience. Nat. Neurosci. 26 (1), 131–139. https://doi.org/10.1038/s41593-022-01212-4. Epub 2022 Dec 29. PMID: 36581729; PMCID: PMC9829541.

Zhou, Y., Sharpee, T.O., 2021. Hyperbolic geometry of gene expression. iScience 24 (3), 102225. https://doi.org/10.1016/j.isci.2021.102225. PMID: 33748711; PMCID: PMC7970362.

Zhou, Y., Smith, B.H., Sharpee, T.O., 2018. Hyperbolic geometry of the olfactory space. Sci. Adv. 4 (8), eaaq1458 https://doi.org/10.1126/sciadv.aaq1458. PMID: 30167457; PMCID: PMC6114987.