UNIVERSAL REGULARITIES OF SYMMETRY AND FRACTALS IN GENOMIC DNA AND THEIR BINARY-GENOMIC NUMBERS, CONSCIOUSNESS AND GENETIC INTELLIGENCE

S.V. Petoukhov
Mechanical Engineering Research Institute of Russian Academy of Sciences
spetoukhov@gmail.com, http://petoukhov.com/

Abstract

The article is devoted to statistical analysis of nucleotide sequences of single-stranded DNAs of higher and lower organisms. Given molecular binary oppositions in the DNA-alphabet of nucleotides, the notion of binary-genomic numbers is introduced, which represent the genomic sequences of nucleotides in binary form. Data about universal rules of probabilities in genomic DNAs are presented, which are connected with a system of dichotomous fractal-like structures and which, as one can think, exist during billions of years of biological evolution as evolutionary invariants. The results obtained pay attention to the fact that the harmonious organization of living matter with its apparatus of genetic inheritance is based on algebraic harmony of the probabilities of a special stochastical-deterministical type. The possible applications of the results are discussed for the development of quantum bioinformatics, artificial intelligence of genomorphic kind, comparative analysis of genes and other bioinformation objects.

Keywords: genomic DNA, binary oppositions, binary-genomic numbers, statistical rules, stochastics and determinism, dichotomies, fractals of probabilities, evolutionary invariants.

1. Introduction

The idea of the harmonious organization of living organisms and the world as a whole has ancient roots in the history of the philosophical and religious teachings of India and other countries. This report is devoted to the author's results of identifying new forms of harmony in the genetic organization of living bodies. They are associated with genetic intelligence, under which we understand those part of the intellectual potentials of living organisms, which allows (based on genetic information in DNA and RNA molecules) building a whole body with its trillions of cells starting from one fertilized cell. At the same time, parental traits are genetically reproduced in living bodies by a multi-channel and noise-resistant manner, despite the strong noises and constant changes of food conditions and external influences. The author believes that consciousness can and should be considered as part or continuation of genetic intelligence (with using the achievements of genetics and algebraic biology).

The creators of quantum mechanics, P. Jordan and E. Schrödinger, pointed out the key difference between living bodies and inanimate ones: inanimate objects are controlled by the average random movement of their millions of particles and the movement of individual particles is not significant for the whole object; on the contrary, in a living organism, selected – genetic – molecules have a dictatorial influence on the entire organism due to quantum enhancement [1]. In accordance with this, to reveal the secrets and patents of living nature, it is necessary to study the universal rules of DNA informatics.

Genetics as a science began with the discovery by Mendel - in experiments on the crossing of organisms - of statistical rules of inheritance of traits and with his fundamental idea about the binary-oppositional basis of these rules as connected with dominant and recessive inheritance factors (alleles). This article proposes a significant expansion of the significance of statistical rules in genomic informatics. In biology, everything is associated with the stochastic interactions of molecules inside the cells, but, in parallel with it, there is an inheritance of traits from parents to descendants (dualism "Stochastics-Determinism"). Jordan argued that the laws of living organisms missed by science are probability laws [1]. In search of these missed laws of probabilities, the author turned to the study of statistical rules in information sequences of nucleotides in long genomic DNAs of higher and lower organisms from the GenBank. Fig. 1 shows an example of a fragment of nucleotide sequence in genomic DNA.

```
TCTGACCTGAGGAGAACTGTGCTCCGCCTTCAGAGTACCACCGAAATCTGTGCAGAGGACAACGCAGCTC
CGCCCTCGCGGTGCTCTCCGGGTCTGTGCTGAGGAGAACGCAACTCCGCCGTTGCAAAGGCGCGCCGCGC
CGGCGCAGGCGCAGAGAGGCGCGCCGCGCCGGCGCAGGCGCAGAGAGGCGCGCCGCGCCGGCGCAGGCGC
AGAGAGGCGCGCCGCGCCGGCGCAGGCGCAGAGAGGCGCGCCGCGCCGGCGCAGGCGCAGAGAGGCGCGC
```

Fig. 1: Example of fragments of a nucleotide sequence in genomic DNAs.

As a result, on a wide set of genomes, the author established the existence of general statistical rules that are candidates for the role of universal rules (laws) of the statistical organization of genomic DNAs [2-4]. For specifics, for example, we will turn to the statistical patterns in the sequence of nucleotides of a single-stranded DNA of the human chromosome №1 containing about 250 million nucleotides of four species that make up the nucleotide alphabet of DNA: A- adenine, C - cytosine, G - guanine, T - thymine.

In this study, under the analysis of statistical features of nucleotide sequences of genomic DNAs of higher and lower organisms, the fact is used that the alphabet of DNA nucleotides is a carrier of the system of three pairs of binary-oppositional traits (or indicators) and contains three types of binary sub-alphabets:

1. two of these nucleotides are purines (A and G), having 2 rings in their molecules, and the other two nucleotides (C and T) are pyrimidines with 1 ring in the molecule, which gives the representation C=T=0, A=G=1. From the point of view of this binary trait, any sequence of DNA nucleotides is represented by a corresponding binary sequence. For example, the sequence GCATGAAGT is represented as 101011110;
2. two of these nucleotides are keto-molecules (T and G), and the other two (C and A) are amino-molecules, which gives the representation C=A=0, T=G=1. In terms of this trait, the same sequence GCATGAAGT is represented differently as 100110011;
3. two of these nucleotides (A and T) are linked in a complementary pair by 2 hydrogen bonds, and the other two nucleotides (C and G) are linked in a complementary pair by 3 hydrogen bonds. This gives the representation C=G=0, A=T=1. In terms of this feature, the same sequence GCATGAAGT is read as 001101101.

In other words, each DNA sequence of nucleotides is a carrier of three parallel messages in three different binary languages. At the same time, these three types of binary representations form a common logical system based on the logical trinity of DNA sub-alphabets, which is revealed by the logical operation of modulo-2 addition, denoted by the symbol $\oplus$, applied to them: modulo-2 addition of any two such binary representations of a DNA sequence yields a binary sequence identical to the third type of binary representation of the same DNA sequence: for example, using the above binary representations of the GCATGAAGT sequence, we have $101011110 \oplus 100110011 = 001101101$. Let us recall the rules of bitwise modulo-2 addition: $0 \oplus 0 = 0$; $0 \oplus 1 = 1$; $1 \oplus 0 = 1$; $1 \oplus 1 = 0$. Logical modulo-2 addition is actively used in quantum informatics [5].

Under representing the nucleotide sequence of any single-stranded genomic DNA as binary sequences based on the three specified binary DNA sub-alphabets, we obtain three types of binary sequences called binary-genomic numbers [6, 7]. The results of statistical analysis of these three types of binary-genomic numbers (or BG-numbers for short) are presented below.

The author's method (called the method of hierarchy of binary statistics) of analysis of statistical regularities of any of three types of binary-genomic numbers of any genomic DNA used in this analysis is as follows. The binary sequence of the BG-number is first considered as a row population of single symbols 0 and 1 with computer calculation of the percentage (probability) of each of these two symbols %0 and %1. Then this BG-number is considered as a row population of binary duplets 00, 01, 10, 11 with calculation of the probability of each of these 4 types of binary duplets %00, %01, %10, %11. Then similarly the same binary-genomic number is represented as a row population of triplets, tetraplets, etc. with each time calculating the percentages of each of the 64 types of binary triplets (%000, %001,..), each of the 256 types of binary tetraplets (%0000, %0001,..), etc. Thus, each BG-number is represented as a multilayer system of populations of binary n-plets, each population of n-plets of which, forming one of the layers of the system, is written by members of the corresponding dyadic group of n-bit numbers (or the binary alphabet of the n-bit Gray code). As a result, for the BG-number under consideration, we obtain a general set of probabilities of each type of n-plets in the genomic system of populations of binary n-plets under consideration. Since the BG-number is considered as a multilayer system of populations of binary n-plets, this system can be called a "population of populations" or a hyper-population of binary n-plets.

Below, the universal rules of binary-genomic numbers in the genomic DNAs of higher and lower organisms revealed by this method are explained using the example of the nucleotide sequence of single-stranded DNA of the human chromosome № 1, containing about 250 million nucleotides (https://www.ncbi.nlm.nih.gov/nuccore/NC_000001.11). Its binary (for example, based on the “purine-pyrimidine” sub-alphabet) representation is a super-huge binary-genomic number with about 250 million bits (the decimal analogue of such a number reaches 2^250000000). These mutually related binary-genomic numbers (based on the three specified types of DNA sub-alphabets) supposedly reflect the laws of quantum biophysics and quantum bioinformatics, genetic memory, algorithms of genetic biomechanics, etc. Mathematical natural science has not previously worked with systems

of such super-huge numbers (as far as the author knows). The author does not know of any analogs of BG-numbers in inanimate bodies. Let us proceed to the presentation of the results of the statistical analysis conducted by the author.

2. Statistical Analysis of Three Types of Binary-Genomic Numbers that Form a Logical Trinity

The presentation of the obtained results of statistical analysis of three types of binary-genomic numbers based on three types of binary DNA sub-alphabets ("purine-pyrimidine", "amino-keto", "strong and weak hydrogen bonds") will begin with BG-numbers based on the purine-pyrimidine sub-alphabet. As indicated above, in this sub-alphabet purines are designated by the binary blue symbol 0, and pyrimidines by the binary blue symbol 1 (**C=T=0, A=G=1**). Fig. 2 shows the diagram of probabilities of the corresponding purine-pyrimidine n-plets for the case of the nucleotide sequence of DNA of the human chromosome № 1, represented as a set of populations or layers of n-plets (or a set of texts, each written in its own alphabet of n-plets, forming a dyadic n-bit group of binary numbers or a group of code words of the n-bit cyclic Gray code [7, 8]).

When passing from the layer of n-plets to the next layer of (n+1)-plets in genomic DNA, the number of members in the alphabet of n-plets doubles: for example, in the alphabet of monoplets it is 2 (that is, 0 and 1), in the alphabet of duplets it is 4 (that is, 00, 01, 10, 11), etc. At the same time, in genomic DNA between adjacent layers of n-plets and (n+1)-plets of purines-pyrimidines, nature realizes regular dichotomous relationships of probabilities, expressed by the following rule, called the rule of the dichotomous fractal of probabilities in the texts of n-plets of the purine-pyrimidine type in genomic DNA (that is, in the binary-genomic numbers of the purine-pyrimidine type). This rule is illustrated in Fig. 2, which shows the values of the probabilities of n-plets of this type for the case of DNA of the human chromosome № 1 (the author systematically studied the implementation of this rule in various genomic DNAs for values of n = 1, 2, 3, 4).

The rule of dichotomic fractals of genomic probabilities of the n-plets based on purines and pyrimidines: For genomic DNA, presented in the form of a set of texts of n-plets of purine-pyrimidine type, the probability of each such n-plet is almost equal to the summary of the probabilities of two (n+1)-plets. For example, in the DNA of the human chromosome № 1, the duplet **01** in the text of the duplets has a probability of **0.2190**, and two triplets **010** and **011** in the text of the triplets have almost the same total amount of their probabilities **0.2191** (= 0.1160+0.1031).

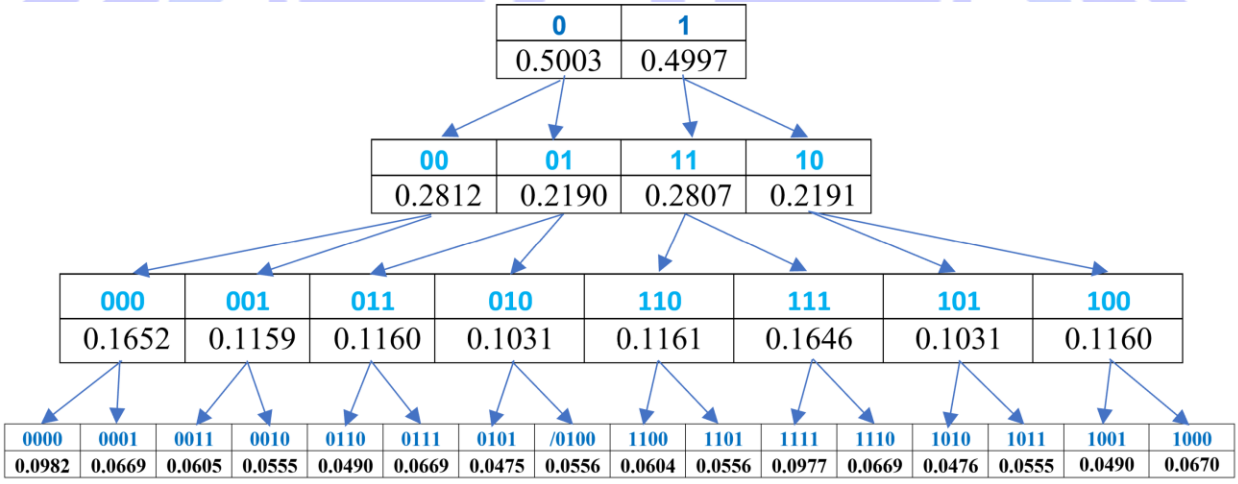


Fig. 2: Graphic illustration of the dichotomous relationship between the probabilities of n-plets of the purine-pyrimidine type in the corresponding n-plet representations of the nucleotide sequences of genomic DNAs. The percentage (probability) values are given as fractions of a unit for the case of DNA of human chromosome № 1 (rounded to the fourth decimal place). The symbol **0** denotes purines (C and T), and the symbol **1** denotes pyrimidines (A and G)

A similar probability rule holds for binary-genomic numbers determined by the amino-keto molecular opposition. Fig. 3 illustrates this rule with indication of the probabilities of such n-plets for the case of human first chromosome DNA.

The rule of dichotomous fractals of genomic probabilities of n-plets of the amino-keto type: For genomic DNA, represented as a set of texts from n-plets of the amino-keto type (that is, as corresponding binary-genomic

number), the probability of each such n-plet in an n-plet text is practically equal to the sum of the probabilities of two (n+1)-plets in an (n+1)-plet text.

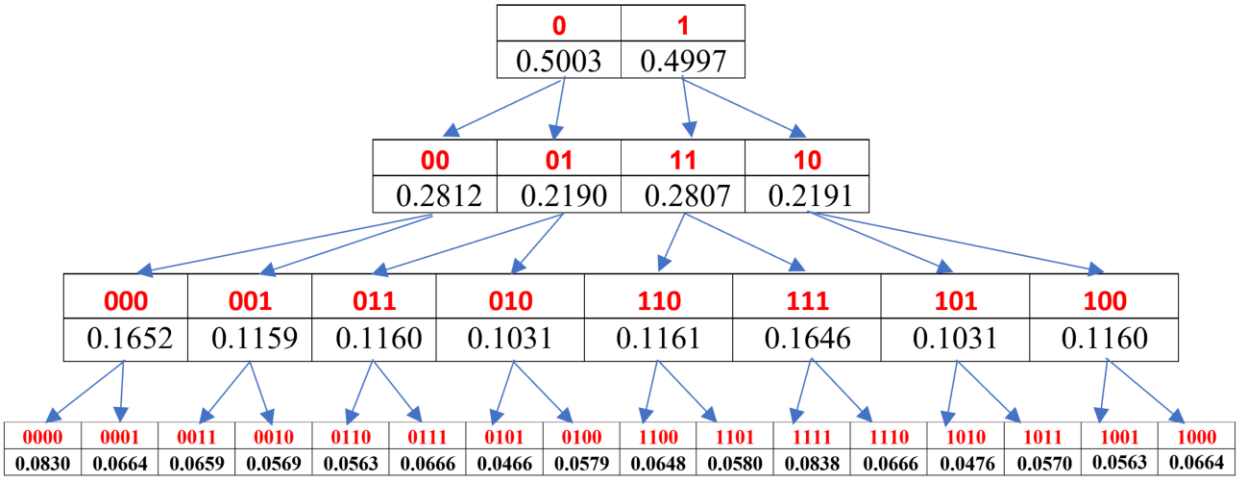


Fig. 3: Graphic illustration of the dichotomous relationship between the probabilities of amino-keto n-plets in the corresponding n-plet representations of nucleotide sequences of genomic DNAs. The percentage (probability) values are given as fractions of a unit for the case of human chromosome 1 DNA (rounded to the fourth decimal place). The symbol **0** denotes amino-molecules (C and A), and the symbol **1** denotes keto-molecules (T and G)

A similar rule of the dichotomous fractal of the probabilities of n-plet hydrogen bonds holds for binary-genomic numbers representing nucleotide sequences of genomic DNA based on binary oppositions of strong and weak hydrogen bonds (see data in preprint [9]). As stated above, in binary-genomic numbers of hydrogen bonds, complementary nucleotides C and G are designated by the symbol **0**, and complementary nucleotides A and T by the symbol **1**.

Thus, the statistical organization of all three considered types of binary-genomic numbers, representing genomic DNAs based on three types of binary oppositions in the nucleotide alphabet of DNA, turns out to be structurally interrelated and connected in a triune manner with the dichotomous fractals of probabilities, graphically shown in Fig. 4.

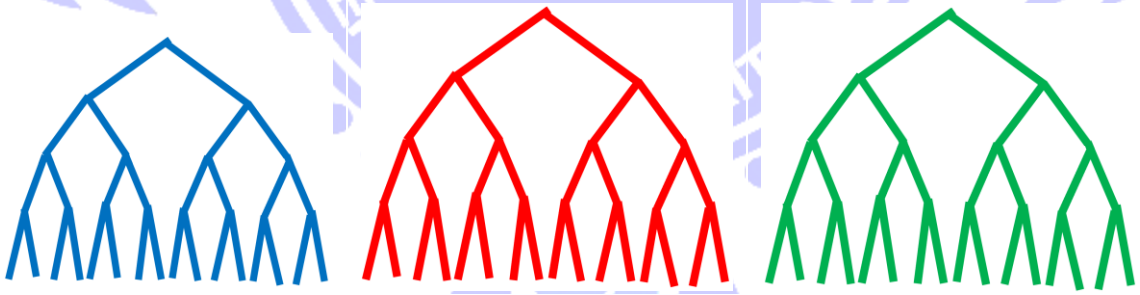


Fig. 4: Graphic illustration of the trinity of dichotomous probability fractals in the statistical organization of nucleotide sequences of genomic DNA when they are represented as binary-genomic numbers based on three types of binary oppositions in the DNA nucleotide alphabet: purines-pyrimidines (blue fractal), amino-keto (red fractal), strong-weak hydrogen bonds (green fractal)

These rules of dichotomous fractals in the statistical organization of genomic DNAs are valid for all genomic DNAs already studied in our laboratory:

1. all 24 human chromosomes;
2. all chromosomes of Drosophila, mice, worms, and many plants;
3. 19 genomes of bacteria and archaea;
4. many extremophiles living in extreme conditions, such as radiation with a level 1000 times higher than lethal for humans.

These rules are therefore candidates for the role of universal genomic rules. Studies of a significantly larger set of genomic DNAs are required to confirm the universality of these rules.

All genomic informatics turns out to be connected to a huge multilayer network of dichotomous fractals of probabilities. Thus, Jordan's prediction about the existence of laws of life missed by science, which are the laws of probability of the quantum world, is confirmed [1]. The discovered universal rules of genomic DNA indicate the existence of non-trivial algebraic invariants of global genomic nature, which remain unchanged over billions of years of evolution, during which millions of species of organisms die out and new ones arise (although locally genomic sequences change as a result of mutations, pressure of natural selection, etc.).

The discovered rules indicate that the information sequences of genomic DNAs are stochastic-deterministic, i.e. such in which stochastics is strongly limited by the presence of regular interrelations between the total probabilities of certain groups of n-plets from different n-plet layers. The phenomenological systems of dichotomous fractals of probabilities in genomic DNAs identified by the author can be modeled on the basis of algebraic formalisms of quantum informatics in the form of multilevel poly-qubit states in which the coefficients of computational basis states of different levels are numerically interconnected, which determines the unity of the system.

Let us briefly mention that the revealed universal rules of dichotomies of probabilities in genomic DNAs resemble the known phenomena of crystal twinning. They can also be considered from the point of view of crystallography, especially since DNA and RNA molecules, cytoskeletons, cell membranes, and many other biostructures are liquid crystal polymers, i.e. liquid crystals are involved in genetic informatics (this area of research can be conventionally called "crystal-genetics"). It is possible that the hidden world of probabilities (mentioned below) is structured on principles similar to crystallographic ones.

3. Biological Dichotomies

Dichotomies in genetically inherited bodies are well known: two hemispheres of the brain, dichotomous branching in plants, vessels and neurons; the bronchial tree of the human lungs, endowed with 23 levels of dichotomous branching; mitosis of somatic cells, etc. But in genomic DNAs, unlike in bodily structures, we encounter a fundamentally different type of dichotomy: dichotomies of probabilistic characteristics in DNA information sequences. Extensive dichotomous fractal networks of genomic DNAs probabilities are the soil from which living bodies and genetic intelligence grow. The material structures of living bodies do not arise out of nowhere, but have structural prototypes in the lawful system of genomic probabilities.

The discovery of universal dichotomous fractals of genomic probabilities (and other universal rules of statistical organization of genomic DNA) allows us to understand in a new way the phenomena of dichotomies in the hereditary structure of biological bodies. For example, why, over the course of billions of years of life on Earth, is it common for bacteria and prokaryotes to reproduce by dichotomous division of the body into two halves? On what structural and energetic foundations is this “eternal” dichotomous phenomenon of bacterial reproduction based, accompanied by the most complex process of dichotomous division of all dichotomously organized genetic information together with the accompanying complex protein and nucleic acid ensembles of the bacterial cell?

The following answer is possible: there is a world of families of probabilities hidden from direct perception, structured on the basis of binary oppositions (like Yin-Yang). It is in the image and likeness of the binary organized families of probabilities of this multilayered world that genetically inherited biobodies are built. Figuratively speaking, our bodies are like clothes put on these binary structured families of probabilities, which act as prototypes of biological structures. This is similar to the situation with the invisible man from the novel by Herbert Wells, whose invisible figure appears only when he is dressed. This also recalls the ideas of the manifested and unmanifested worlds and Plato's famous allegory about the world of ideas and shadows on the wall of a cave, by which people living in a cave can judge the true hidden world of ideas.

4. Probabilities, Brain, and Consciousness

As is known, our brain, containing 86 billion nerve cells and one quadrillion interneuronal contacts, operates on probabilistic principles, since the transmission of each impulse from neuron to neuron can be influenced by an immense number of factors [10]. But based on the stochastic flows of neural impulses in the brain, we obtain a deterministic idea of specific objects in the surrounding world and can successfully operate with them biomechanically. Uncovering the secrets of this relationship between stochastics and determinism in living things is a fundamental task of science. Since the brain is inherited through the genetic coding system, one can think that by learning the universal rules of the stochastic organization of genomic information, we indirectly

learn the probabilistic principles of the brain and the biomechanics of sensorimotor systems. Their disclosure is important for medicine, artificial intelligence systems, ergonomics, brain-computer interfaces, etc.

These questions are directly related to the long-known and still unresolved "biomechanical problem of N.A. Bernstein", a classic of biomechanics [11]. This problem is that the general target task of body movement is performed precisely regardless of the imprecision of its constituent motor subtasks. For example, when repeating a precise hammer blow on a nail, a person each time uses different trajectories, velocities and accelerations of body parts with changes in both joint flexion and the activity of many muscles of each joint with many motor neurons of each muscle. The degree of freedom problem in motor control states that there are many ways for a person or animal to perform a movement to achieve the same goal, using redundant neurophysiological degrees of freedom. The question of how the nervous system "selects" a subset of these almost infinite degrees of freedom is a fundamental problem in understanding motor control and motor learning. In other words, under normal circumstances there is no simple one-to-one correspondence between a motor task and its motor solution.

The results of our study of genomic probabilities can be used to understand and model the above-mentioned biomechanical phenomena, as well as to develop quantum bioinformatics and genomorphic artificial intelligence. Fractal-like principles in the organization of genomic informatics can be considered, among other things, in connection with fractal-like constructions known in ancient Indian and ancient Chinese philosophy [11, 12].

5. Some Concluding Remarks

Statistical analysis of the genomic DNAs nucleotide sequences of higher and lower organisms has revealed the existence in them of universal rules of probabilities persisting over billions of years of biological evolution. This confirms the long-standing thought of P. Jordan that the laws of living organisms missed by science are probability laws [1]. The obtained results draw attention to the fact that the harmonic organization of living matter with its genetic inheritance is based on an algebraic harmony of probabilities of a special stochastic-deterministic type. In particular, this harmony of probabilities manifests as a triunity of dichotomous probability fractals (Fig. 4).

Accounting for this harmony of genomic probabilities enables the development of: quantum bioinformatics of multilevel poly-qubit states; approaches to artificial intelligence of the genomorphic type and new genetic algorithms; new methods and tools for the analysis of genes and other bioinformatic objects. This algebraic harmony of genomic probabilities echoes a number of provisions about fractal-like structures figuring in ancient Indian and ancient Chinese philosophy. The author believes that many provisions of ancient Indian philosophy about the schemes of the universe did not arise from nowhere, but are a reflection of the structural regularities of the molecular-genetic system with its stochastic-deterministic and cyclical organization.

Acknowledgments

The author thanks Yu. I. Manin, E.G. Rajan, V. I. Svirin, I. V. Stepanyan, G. K. Tolokonnikov, and S.Ya. Kotkovsky for useful thematic discussions.

About the Author

Sergey V. Petoukhov, Prof., Dr. Sci., is currently Head of Laboratory of biomechanical systems research in Mechanical Engineering Research Institute of the Russian Academy of Sciences; Chief researcher of the “Center of interdisciplinary researches of musical creativity” of the Moscow State Conservatory by P.I. Tchaikovsky; Editor-in-Chief of “International Journal of Mathematical Sciences and Computing” (Hong Kong). Laureate of the State prize of the USSR; Academician of the Academy of Quality Problems (Russia, from 2000); Full Professor (The European Academy of Informatization, Belgium, 2004); the Chinese government inserted his name in the «List of Outstanding Scientists in the World» in 2012 Chairman of Advisory Board of «International Symmetry Association», Budapest, Hungary, from 2003 till now; Honorary chairman of Board Directors of «International Society of Symmetry in Bioinformatics», USA, 2005; He is the author of approximately 250 scientific works including the following 8 books. S. V. Petoukhov is a specialist in theoretical and mathematical biology (in particular, in problems of genetic coding and bioinformatics), biomechanics, crystallography, theories of symmetry and self-organization.

References

[1] McFadden J., Al-Khalili J. The origins of quantum biology. *Proceedings of the Royal Society A*, Vol. 474, Issue 2220, pp. 1-13 (2018). (https://doi.org/10.1098/rspa.2018.0674).

[2] Petoukhov S.V. The Principle "Like Begets Like" in Molecular and Algebraic-Matrix Genetics. Preprints 2022, 2022110528 (2023a) (doi: 10.20944/preprints202211.0528.v3).

[3] Petoukhov S.V. The principle "like begets like" in algebra-matrix genetics and code biology. *Biosystems* 233 (2023b), 105019, https://doi.org/10.1016/j.biosystems.2023.105019

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

[5] Nielsen M.A., Chuang I.L. *Quantum Computation and Quantum Information*. - Cambridge Univ. Press, 2010, https://doi.org/10.1017/CBO9780511976667 .

[6] Petoukhov S.V. Binary-genomic numbers and symmetries of the universal rules of statistical organization of genomic DNAs. – Bioinformatics of Genome Regulation and Structure / Systems Biology. –Novosibirsk, ICG SB RAS, pp. 2163-2165, 2024. doi 10.18699/bgrs2024-12.3-09.

[7] Petoukhov S.V. Cyclic Gray codes in modeling inherited cyclic biostructures and analysis of statistical rules of genomic DNAs. - Preprints 2024, 2024020713, from 13 February 2024, https://doi.org/10.20944/preprints202402.0713.v1 .

[8] Petoukhov S.V. Genetic code, the problem of coding biological cycles, and cyclic Gray codes. – Biosystems, vol. 246, December 2024, 105349, December 2024, 105349, https://doi.org/10.1016/j.biosystems.2024.105349

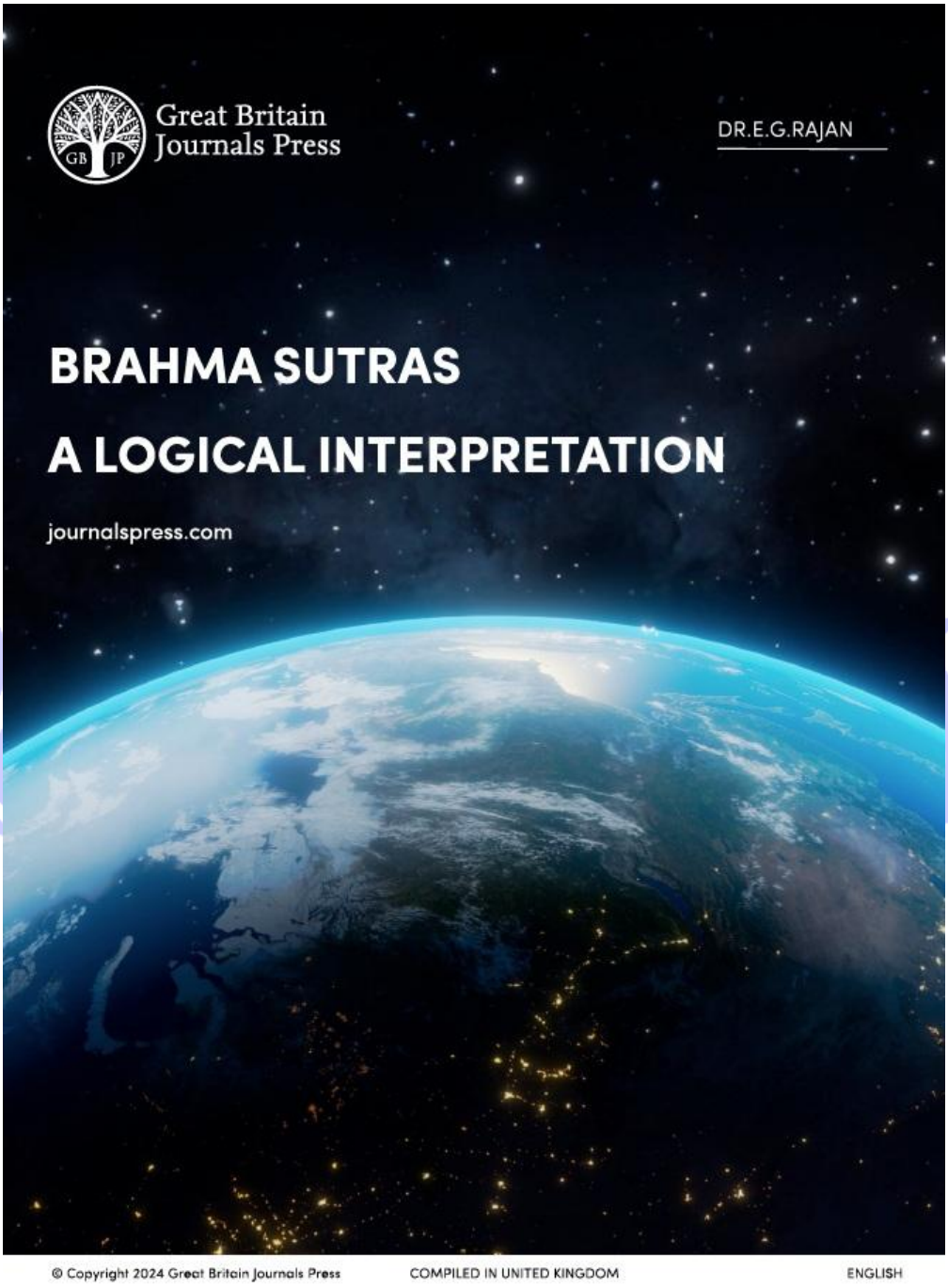
[9] Petoukhov S. Algebraic Gestalt-Archetypes of Probabilities in Genomic DNAs, Cyclic Gray Codes, Quantum Bioinformatics. *Preprints* **2025**, 2025010198, from 03.01.2025, https://doi.org/10.20944/preprints202501.0198.v1]).

[10] Siny Yu.E. On probabilistic principles of realization of genotype in phenotype and information processing in the brain. - International scientific journal "Innovative Science", 2016. - №10-3. – C. 30-40. - ISSN 2410-6070. (Siniy Yu. E. O veroyatnostnyx principax realizatsii genotipa v fenotipe and pererabotke informatsii v mozge. – Mezhdunarodny nauchny zhurnal "Innovatsionnaia nauka"). (in Russian).

[11] Petoukhov S.V., He M. Algebraic biology, matrix genetics, and genetic intelligence. – Singapore, World Scientific, 2023, 616 p., https://doi.org/10.1142/13468].

[12] Rajan E.G., Brahma sutras-A logical interpretation, Great Britain Journals Press, United Kingdom, 2024.

Prof. Petoukhov was felicitated as Distinguished Researcher of Russian Federation on April 12, 2025


Great Britain
Journals Press
DR.E.G.RAJAN
BRAHMA SUTRAS
A LOGICAL INTERPRETATION
journalspress.com
© Copyright 2024 Great Britain Journals Press
COMPILED IN UNITED KINGDOM
ENGLISH