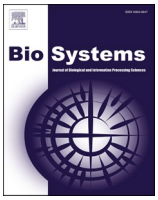
Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

BioSystems 198 (2020) 104273

Contents lists available at ScienceDirect

BioSystems

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



Hyperbolic rules of the cooperative organization of eukaryotic and prokaryotic genomes

Sergey V. Petoukhov

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

ARTICLE INFO

Keywords:
Hyperbolic rules
Harmonic progression
Quantum informatics
Tensor product
Oligomer sums method
Genomes
Genes
Viruses
Proteins

ABSTRACT.

The author's method of oligomer sums for analysis of oligomer compositions of eukaryotic and prokaryotic genomes is described. The use of this method revealed the existence of general rules for the cooperative oligomeric organization of a wide list of genomes. These rules are called hyperbolic because they are associated with hyperbolic sequences including the harmonic progression 1, 1/2, 1/3, .., 1/*n*. These rules are demonstrated by examples of quantitative analysis of many genomes from the human genome to the genomes of archaea and bacteria. The hyperbolic (harmonic) rules, speaking about the existence of algebraic invariants in full genomic sequences, are considered as candidates for the role of universal rules for the cooperative organization of genomes. The results concerns additionally the problem of the origin of life. The described phenomenological results were obtained as consequences of the previously published author's quantum-information model of long DNA sequences. The oligomer sums method was also applied to the analysis of long genes and viruses including the COVID-19 virus; this revealed, in characteristics of many of them, the phenomenon of such rhythmically repeating deviations from model hyperbolic sequences, which are associated with DNA triplets. In addition, an application of the oligomer sums method is shown to the analysis of amino acid sequences in long proteins like the protein Titin. The topics of the algebraic harmony in living bodies and of the quantum-information approach in biology are discussed.

1. Introduction

Living bodies are huge sets of various molecules, which have an amazing ability to inherit biological traits of organisms to the next generations. G. Mendel, in his experiments with plant hybrids, found that the transmission of traits under the crossing of organisms occurs by certain algebraic rules, despite the colossal heterogeneity and complexity of molecular structures of their bodies. This article represents new results of studying hidden algebraic rules in molecular genetic information structures.

One of the founders of quantum mechanics, who introduced also the term "quantum biology," P. Jordan 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). Taking into account the dictatorial influence of DNA and RNA molecules on the whole body, the author focused his research on a special analysis of numeric parameters of nucleotide sequences in single-stranded DNA of different genomes and their parts. As a result of this research, a new method of analysis of nucleotide sequences was created, which has led to discovering new numeric rules of cooperative oligomer organization of eukaryotic and prokaryotic genomes. These materials are described below. All initial data on nucleotide sequences for this analysis were taken from the GenBank.

It should be recalled that genomic nucleotide sequences are not random sequences. These sequences carry information transmitted in a noise-immune manner from generation to generation. They contain a great number of repeats and complementary palindromes. For example, in the human genome, about a third of DNA sequences are represented by complementary palindromes (Gusfield, 1997; McConkey, 1993). In evolutionary biology, the abundance of such complementary palindromes in genomes is seen as evidence of not random DNA sequences, that is, their irreducibility to a set of random mutations (see additional data in (Fimmel et al., 2019; Petoukhov, Tolokonnikov, 2020)).

For long nucleotide sequences of single-stranded DNA, the second Chargaff's rule is well known, which states that in such sequences the

E-mail address: petoukhovs@yandex.ru.

https://doi.org/10.1016/j.biosystems.2020.104273
Received 19 August 2020; Received in revised form 7 October 2020; Accepted 7 October 2020
Available online 13 October 2020

amount of guanine G is approximately equal to the amount of cytosine C and the amount of adenine A is approximately equal to the amount of thymine T. Many authors have devoted their works to the analysis and discussion of this rule (see, for example (Fimmel et al., 2019; Prabhu, 1993; Rapoport, Trifonov, 2012; Rosandic et al., 2016; Shporer et al., 2016; Yamagishi, 2017),). According to (Albrecht-Buehler, 2006), this rule applies to the eukaryotic chromosomes, the bacterial chromosomes, the double-stranded DNA viral genomes, and the archaeal chromosomes provided they are long enough. In connection with the hidden rules of long DNA sequences, Chargraff introduced the important term "a grammar of biology" (Chargaff, 1971), which is repeatedly used by his followers (see, for example (Yamagishi, 2017)).

Regarding the quantitative analysis of DNA sequences, researchers usually study quantities and percentages (or probability, or frequencies) of separate *n*-plets (that is separate oligomers, having their length *n*). For example, the second Chargaff's rule is based on such a study of the quantities of separate nucleotides A, T, C, and G. The work (Prahbu, 1993) studies quantities of separate *n*-plets. In contrast to such analytic approaches, the author suggests for analysis of long nucleotide sequences another method called the oligomer sums method. It allows studying the oligomer cooperative organization by the comparative analysis of total amounts of all *n*-plets, having fixed length *n*, from the certain equivalence classes of oligomers.

Below this analytic approach and the results of its application to many genomes and separate nucleotide sequences are represented. Besides, the article additionally shows that the oligomer sums method can be usefully applied to the analysis not only genomic sequences of nucleotides but also the analysis of amino acid sequences of long proteins. The presented study is a continuation of long term author's researches on biological symmetries.

2. The hyperbolic rule in the oligomer cooperative organization of all human nuclear chromosomes

The term "oligomer" refers to a molecular complex of chemical that consists of a few repeating units. Nucleobases - adenine A, thymine T, cytosine C, and guanine G - serve as such repeated units in DNA oligomers, which can have different lengths and which are also called *n*-plets, where *n* refers to the oligomer length. Each of nucleotide sequences in eukaryotic and prokaryotic genomes can be represented as a sequence of monomers (like as A-C-A-T-G-T- …), or a sequence of doublets (like as AC-AT-GT-GG- …), or a sequence of triplets (like as ACA-TGT-GGA- …), etc. In each of such fragmented representations, one can calculate total amounts of oligomers of various kinds in the analyzed sequence and then compare them.

The article describes the numerical analysis of sets of *n*-plets, which belong - in such fragmented representations of long DNA sequences - to the equivalence classes (or cooperative groupings) of A_1-oligomers, or T_1-oligomers, or C_1-oligomers, or G_1-oligomers correspondingly (their index 1 indicates that all oligomers of each class start with the same nucleotide A, or T, or C, or G). For example, the class of the A_1-oligomers contains the following *n*-plets: 4 doublets AA, AT, AC, and AG; 16 triplets AAA, AAT, AAC, AAG, ATA, …, AGG; etc. The total amount of different kinds of *n*-plets, which start with the same nucleotide, under fixed *n* is equal to 4^{n-1}.

To simplify a theoretical explanation, let us consider the example of an analysis of the oligomer cooperative organization of human chromosome №1 by the author's method of oligomer sums (abbreviation, the OS-method). The totality of data obtained by analyzing a nucleotide sequence by the OS-method is called its OS-representations. This method gives numeric sequences called oligomer sums sequences (or briefly, OS-sequences).

The application of the OS-method to the analysis of the human chromosome №1 includes the following steps, which are typical also for the analysis of other DNA and RNA sequences:

- Firstly, the DNA sequence of this chromosome is represented in the mentioned form of a set of its fragmented sequences of oligomers of certain lengths $n = 1, 2, 3, \ldots$;
- Secondly, phenomenological quantities S_A, S_T, S_C, and S_G of monomers A, T, C, and G correspondingly are calculated in the considered nucleotide sequence. In the human chromosome №1, the following quantities exist: $S_A = 67{,}070{,}277$, $S_T = 67{,}244{,}164$, $S_C = 48{,}055{,}043$, $S_G = 48{,}111{,}528$;
- Thirdly, in each of the fragmented representations of the DNA sequence under $n = 2, 3, 4, \ldots$, the corresponding total amounts $\Sigma_{A,n,1}$, $\Sigma_{T,n,1}$, $\Sigma_{C,n,1}$, and $\Sigma_{G,n,1}$ of *n*-plets are calculated in equivalence classes of A_1-oligomers, T_1-oligomers, C_1-oligomers, and G_1-oligomers (here, for example, the symbol $\Sigma_{A,3,1}$ refers to the total amount of triplets, which start with the nucleotide A). These total amounts regarding each of the classes are members of the appropriate OS-sequence of the class. For analysis of human chromosomes and various eukaryotic and prokaryotic genomes, the author usually takes $n = 1, 2, 3, \ldots, 19, 20$ or, in special cases, $n = 1, 2, 3, \ldots, 99, 100$.

One can remind here that genomic sequences in the GenBank sites usually contain some letters N, indicating that there can be any nucleotide in this place (https://www.ncbi.nlm.nih.gov/books/NBK21136/). By this reason, the total amount of all monomers A, T, C, G (that is the sum $S_A + S_T + S_C + S_G$), calculated for the sequence from the GenBank, is slightly less than the complete length of the DNA sequence, which is indicated in the GenBank. But practically this is not essential for the results of the application of the OS-method to analyze genomic sequences.

For human chromosome №1, phenomenological values of the total amounts of *n*-plets from the class of A_1-oligomers are shown in the graphical form for $n = 1, 2, 3, \ldots, 20$ in Fig. 2.1 at left (in blue). Here the abscissa axis represents the values of *n*, and the ordinate axis represents the values of the total amounts $\Sigma_{A,n,1}$ of *n*-plets, which start with the nucleotide A. The amazing result is that all 20 phenomenological points $[n, \Sigma_{A,n,1}]$ lie - with a high level of accuracy - along with the hyperbola $H_{A,1} = S_A/n = 67070277/n$ shown in red in Fig. 1, middle. Deviations of phenomenological quantities $\Sigma_{A,n,1}$ from model values S_A/n lie in the range $-0.030\% \div 0.024\%$, that is, they comprise only one-hundredths of a percent (Fig. 1, right). Initial data on this chromosome were taken in the GenBank: https://www.ncbi.nlm.nih.gov/nuccore/NC_000001.11.

This result is striking because it shows that knowing only the number of nucleotides A, that is, only one member of the number series shown in Fig. 1, at left, one can predict with the high accuracy all other 19 members, each of which is a sum of 4^{n-1} possible kinds of *n*-plets. The number of possible kinds of *n*-plets in these sums is growing rapidly, becoming astronomically huge: 4, 16, 64, 256, 1024, …, 4^{10}, …, 4^{19}. Of course, in the human chromosome №1, for example, not all possible 4^{19} kinds of the mentioned 20-plets exist but the total amount of all those kinds of 20-plets, which exist in this chromosome, is practically equal to $S_A/20$ with a high level of accuracy shown below.

Similar results were obtained when studying in this chromosome the total amounts of *n*-plets, which start with the nucleotide T (Fig. 2, at left), and with the nucleotide C (Fig. 2, at middle), and with the nucleotide G (Fig. 2, at right). The phenomenological values of the total amounts $\Sigma_{T,n,1}$, $\Sigma_{C,n,1}$, and $\Sigma_{G,n,1}$ of *n*-plets are also modeled effectively by appropriate hyperbolic progressions $H_{T,1}$, $H_{C,1}$, $H_{G,1}$ (2.1), which differ from each other only by their numerators S_T, S_C, and S_G:

$$H_{T,1} = S_T/n = 67244164/n,\ H_{C,1} = S_C/n = 48055043/n,\ H_{G,1} = S_G/n = 48111528/n \quad (2.1)$$

Fig. 3 shows phenomenological and model numeric values for the OS-representation of the classes of A_1-, T_1-, C_1-, and G_1-oligomers of the human chromosome №1 for $n = 1, 2, 3, \ldots, 20$. The model values of the

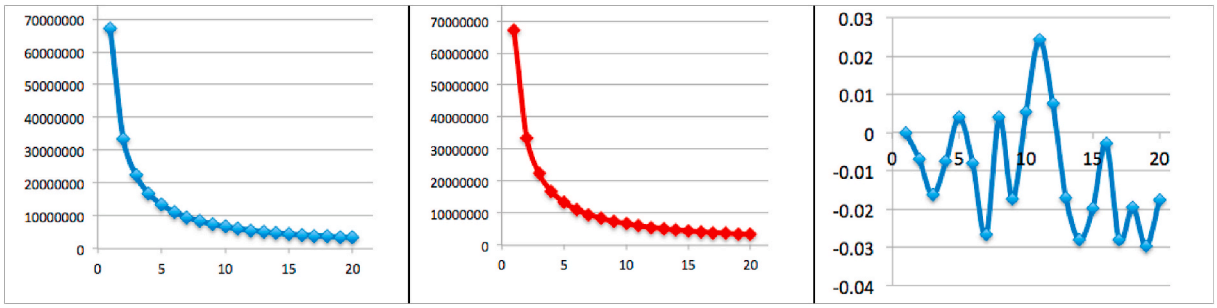


Fig. 1. The graphs of data for the case of the OS-sequences of *n*-plets from the class A_1-oligomers of the human chromosome №1. In these graphs, the abscissa axis represents the values $n = 1, 2, 3, \ldots, 20$. **Left**: the ordinate axis represents the set of phenomenological total amounts $\Sigma_{A,n,1}$ of *n*-plets beginning with the nucleotide A. **Middle**: the ordinate axis represents modeling values of the hyperbolic progression $S_A/n = 67070277/n$. The dots with coordinates $[n, S_A/n]$ belong to the shown hyperbola $H_{A,1} = S_A/n = 67070277/n$. **Right**: deviations of the real OS-sequence $\Sigma_{A,n,1}$ from the model hyperbolic progression S_A/n in percentages.

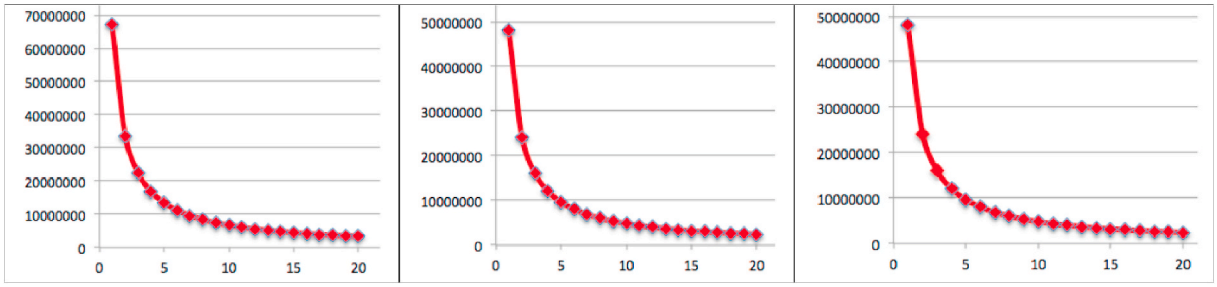


Fig. 2. Additional graph data to the OS-representation of the human chromosome №1. The abscissa axes represent the values $n = 1, 2, 3, \ldots, 20$. The ordinate axes show model values $H_{T,1}(n)$, $H_{C,1}(n)$, and $H_{G,1}(n)$ (in red) from (2.1), which practically coincide phenomenological values $\Sigma_{T,n,1}$, $\Sigma_{C,n,1}$, and $\Sigma_{G,n,1}$ of the total amount of *n*-plets, which start with the nucleotide T (at the left graph), the nucleotide C (at the middle graph), and the nucleotide G (at the right graph). The numerical data on this coincidence is shown below. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

total amounts of *n*-plets, which start with a certain nucleotide (A, T, C, or G), are calculated correspondingly as values of the hyperbolic progressions $H_{A,1} = S_A/n = 67070277/n$, $H_{T,1} = S_T/n = 67244164/n$, $H_{C,1} = S_C/n = 48055043/n$, and $H_{G,1} = S_G/n = 48111528/n$. Deviations of phenomenological values from model values are also shown in percent in accordance with the expression: 100 (1 – (real value)/(model value)). One can see that these deviations are much lesser than 0,2% in all cases.

The model hyperbolic progressions $H_{A,1} = S_A/n$, $H_{T,1} = S_T/n$, $H_{C,1} = S_C/n$, and $H_{G,1} = S_G/n$ serve as mathematical standards for the described phenomenological facts. These hyperbolic progressions differ from each other only in the magnitude of numerators in their expressions, and therefore they can be specified by the general expression (2.2):

$$H_{N,1}(n) = S_N/n, \tag{2.2}$$

where N refers to any of nucleotides A, T, C, or G; S_N refers to the number of corresponding monomers A, T, C, or G in the analyzed nucleotide sequence. If you know the total quantity S_N of the monomer N, you can predict - with a high level of accuracy - the total amounts of *n*-plets belonging to the class N_1-oligomers by using the general expression (2.2). These phenomenological facts testify in favor of the cooperative entity of the nucleotide sequence in the human chromosome №1.

By the corresponding compression of the ordinate axis in these Cartesian coordinate systems (that is by appropriate scaling of numerators S_A, S_T, S_C, and S_G), each of these four hyperbolic sequences $H_{A,1} = S_A/n$, $H_{T,1} = S_T/n$, $H_{C,1} = S_C/n$, and $H_{G,1} = S_G/n$ reduces to the hyperbolic sequence (2.3):

$$y = 1/n, \tag{2.3}$$

which we call the canonical (or reference) hyperbolic sequence of OS-representations (or the canonical OS-sequence) of nucleotide sequences. In mathematics, the sequence (2.4)

$$1/1, 1/2, 1/3, 1/4, 1/5, \ldots, 1/n \tag{2.4}$$

is known long ago as the harmonic progression (or the harmonic sequence) where each term is the harmonic mean of the neighboring terms. For this reason, the revealed hyperbolic sequences in genomes can be also called genomic harmonic progressions, and, in this mathematical sense, one can talk about the harmonic rules and the harmonious organization of genomes described below. The historically famous name “the harmonic progression” comes from the connection (2.4) with the series of harmonics in music. The sums of the first members of the harmonic progression (2.4) are called harmonic numbers. The cross-ratio (or the double ratio), which is the basic invariant of projective geometry, is equal to 4/3 for any four adjacent terms of the harmonic progression (the harmonic progression is projectively equivalent to arithmetic progressions, in which the cross-ratio of any four adjacent terms is also equal to 4/3). This connection of the harmonic progression with the basic invariant of the projective geometry is interesting with respect to a wide theme of inherited non-Euclidean symmetries in biological objects (Petoukhov, 1989).

The rich centuries-old history of the study of harmonic progressions and harmonic series is associated with the names of Pythagoras, Orem (d’Oresme), Leibniz, Newton, Euler, Fourier, Dirichlet, Riemann, and other researchers. The generalization of the harmonic series is known as the Riemann zeta function. Using musical terminology, where the term “timbre” refers to the totality of the set of sound frequencies in a prolonged sound, one can conditionally say that the oligomer sums method represents the analyzed nucleotide sequence as some “oligomer timbre”. The series of harmonic numbers serves as the discrete analog of the continuous function of natural logarithm ln(*n*) (Graham et al., 1994, p. 276); this, in particular, connects the harmonic progression (2.4) with Weber-Fechner logarithmic law, which is the main psychophysical law and dictates informatic peculiarities for all inherited sensory channels - vision, hearing, smell, etc, whose organs (eyes, ears, nose, etc.) very

N	1	2	3	4	5	6	7	8	9	10
A										
Real	67070277	33537501	22360413	16768845	13413532	11179286	9584038	8383461	7453552	6706672
Model	67070277	33535139	22356759	16767569	13414055	11178380	9581468	8383785	7452253	6707028
Δ%A	0.000	-0.007	-0.016	-0.008	0.004	-0.008	-0.027	0.004	-0.017	0.005
T										
Real	67244164	33620498	22412993	16808862	13445360	11207274	9606748	8405040	7470145	6724359
Model	67244164	33622082	22414721	16811041	13448833	11207361	9606309	8405521	7471574	6724416
Δ%T	0.000	0.005	0.008	0.013	0.026	0.001	-0.005	0.006	0.019	0.001
C										
Real	48055043	24024903	16012711	12013624	9612227	8005708	6865944	6008215	5336968	4803919
Model	48055043	24027522	16018348	12013761	9611009	8009174	6865006	6006880	5339449	4805504
Δ%C	0.000	0.011	0.035	0.001	-0.013	0.043	-0.014	-0.022	0.046	0.033
G										
Real	48111528	24057606	16040889	12028924	9625086	8021235	6869132	6013412	5348337	4813156
Model	48111528	24055764	16037176	12027882	9622306	8018588	6873075	6013941	5345725	4811153
Δ%G	0.000	-0.008	-0.023	-0.009	-0.029	-0.033	0.057	0.009	-0.049	-0.042

n	11	12	13	14	15	16	17	18	19	20
A										
Real	6095821	5588773	5160139	4792078	4472245	4192017	3946422	3726860	3531067	3354107
Model	6097298	5589190	5159252	4790734	4471352	4191892	3945310	3726127	3530015	3353514
Δ%A	0.024	0.007	-0.017	-0.028	-0.020	-0.003	-0.028	-0.020	-0.030	-0.018
T										
Real	6111970	5601854	5173904	4801395	4479492	4202773	3954021	3735327	3535288	3360459
Model	6113106	5603680	5172628	4803155	4482944	4202760	3955539	3735787	3539167	3362208
Δ%T	0.019	0.033	-0.025	0.037	0.077	0.000	0.038	0.012	0.110	0.052
C										
Real	4370502	4002753	3694018	3433636	3202830	3003511	2826568	2668499	2531448	2402186
Model	4368640	4004587	3696542	3432503	3203670	3003440	2826767	2669725	2529213	2402752
Δ%C	-0.043	0.046	0.068	-0.033	0.026	-0.002	0.007	0.046	-0.088	0.024
G										
Real	4374518	4013372	3701250	3435824	3210839	3006763	2830698	2673815	2532772	2407301
Model	4373775	4009294	3700887	3436538	3207435	3006971	2830090	2672863	2532186	2405576
Δ%G	-0.017	-0.102	-0.010	0.021	-0.106	0.007	-0.021	-0.036	-0.023	-0.072

Fig. 3. Phenomenological and model values to the OS-representations of the classes of A_1-, T_1-, C_1-, and G_1-oligomers in human chromosome №1 are shown for n = 1, 2, …, 20. The real total amounts of n-plets, which start with a certain nucleotide (A, T, C, or G), are indicated (in blue) jointly with their model values $H_{A,1}(n)$, $H_{T,1}(n)$, $H_{C,1}(n)$, and $H_{G,1}(n)$ from (2.1) (in red). The symbol Δ% refers to deviations of real values from model values in percent (the model values are taken as 100%). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

differ each other in appearance. This mathematical connection is that the natural logarithm is defined as the area under the hyperbola on the Cartesian plane (Conway, Guy, 1995, p. 250). It testifies that genetic and different psychophysical levels of inherited biological informatics are structurally interrelated on the algebra-harmonical basis (Petoukhov, 2016, 2020b).

Given the relationship of the harmonic progression (2.4) with the four OS-sequences for the four types of nucleotides A, T, C, and G, genomic sequences can be called tetra-harmonic sequences. Fig. 3 shows that the OS-sequences of the total amounts of n-plets from the classes of A_1-oligomers and T_1-oligomers differ little from each other. The same is true for the OS-sequences of the total amounts of n-plets from the classes of C_1- and G_1-oligomers. This fact is described by the expressions (2.5):

$$\Sigma_{A,n,1} \approx \Sigma_{T,n,1}, \ \Sigma_{C,n,1} \approx \Sigma_{G,n,1} \quad (2.5)$$

In the particular case at $n = 1$, expressions (2.5) demonstrate the second Chargaff's rule on the approximate equality between the amounts of nucleotides A and T, as well as C and G in long DNA sequences. Correspondingly the phenomenological fact, described by expressions (2.5), is a certain generalization of the second Chargaff's rule.

The results presented indicate, at least for the human chromosome №1, that there exist two general hyperbolic (or harmonic) rules regarding the total amounts of n-plets, which start with a certain nucleotide A, T, C, or G.

The first hyperbolic rule (about interrelations of oligomers in individual chromosomes):

- For any of classes of A_1-, T_1-, C_1-, or G_1-oligomers in individual chromosomes, the total amounts $\Sigma_{N,n,1}(n)$ of their n-plets, corresponding different n, are interrelated each other through the general expression $\Sigma_{N,n,1} \approx S_N/n$ with a high level of accuracy (here N refers to any of nucleotides A, T, C, or G; S_N refers to the number of monomers N; $n = 1, 2, 3, 4, \ldots$is not too large compared to the full length of the nucleotide sequence). The phenomenological points with coordinates $[n, \Sigma_{N,n,1}]$ practically lie on the hyperbola having points $H_{N,1} = S_N/n$.

The second hyperbolic rule (about the similarity in the pairs of OS-sequences):

- In individual chromosomes, two numeric OS-sequences expressing the total amounts of n-plets, which start with the nucleotide A and with the nucleotide T, are approximately identical. The same is true for two numeric OS-sequences expressing the total amounts of n-plets, which start with the nucleotide C and with the nucleotide G (in accordance with the expressions (2.5)). Here $n = 1, 2, 3, 4, \ldots$ is not too large compared to the full length of the nucleotide sequence.

The obtained results of the hyperbolic (or harmonic) interrelationship of the amounts of n-plets, belonging to the indicated classes of oligomers, are not trivial. Theoretical counter-examples of artificial

nucleotide sequences, which have not such interrelation, can be indicated. For example, for the case of the class of A_1-oligomers, one can mentally construct a long nucleotide sequence that contains many nucleotides A but does not have two adjacent nucleotides A, that is, does not contain a single AA doublet. Such a sequence does not have the hyperbolic interrelationship between the amounts of the nucleotide A and the total amounts of *n*-plets starting with A. It can be noted else that, in the same human chromosome №1, the comparison of amounts of different *n*-plets, consisting of only one type of nucleotides, for example, of the nucleotide A, shows the absence of the hyperbolic relationship between them. Really, in this case the amount of the nucleotide A is equal to 67,070,277, the amount of the doublets AA - 10,952,057, the amount of the triplets AAA – 2,837,038, the amount of the tetraplets AAAA – 856,207, and so on without their hyperbolic interrelation.

Let us continue the description of obtained results of the analysis of the human genome, which contains 22 autosomes and 2 sex chromosomes X and Y. These chromosomes are very different from each other in length, molecular weight, gene content, etc. What can be said about the other 23 human chromosomes? Are there hyperbolic rules similar to formulated rules for the human chromosome №1? Yes, the author has got a positive answer to this question. For each of 24 human chromosomes, knowing its quantity S_N of the monomer N (that is A, T, C, or G) allows you to calculate the total amounts of *n*-plets, which start with the oligomer N, with a high level of accuracy by using the general expression (2.2). Here $n = 1, 2, 3, \ldots$ but not very large in comparison with the length of the DNA sequence. Fig. 4 shows general confirmational results of studying all 24 human chromosomes by the OS-method under $n = 1, 2, 3, \ldots, 20$.

These results demonstrate that both hyperbolic (or harmonic) rules №1 and №2 hold true for each of the human chromosomes with a high level of accuracy.

One can show that the obtained phenomenological data also leads to the third hyperbolic rule related to normalized versions of the OS-sequences S_A/n, S_T/n, S_C/n, and S_G/n. Scaling the numerators S_A, S_T, S_C, and S_G by dividing by their total amount $S = S_A + S_T + S_C + S_G$, we obtain the corresponding scaling of all these OS-sequences, which are termed as "normalized OS-sequences" (2.6):

$$S_A/(nS),\ S_T/(nS),\ S_C/(nS),\ S_G/(nS) \tag{2.6}$$

It turns out that the normalized OS-sequences of all human chromosomes are similar to each other with a high level of accuracy as Fig. 5 shows regarding the first main members S_A/S, S_T/S, S_C/S, and S_G/S of these hyperbolic sequences.

The same results on the similarity of normalized OS-sequences S_A/nS, S_T/nS, S_C/nS, and S_G/nS in all chromosomes of a particular genome were obtained by the author when studying the genomes of a number of eukaryotes (until now, without a single exception in analyzed cases). Below appropriate results for some eukaryotic genomes are described. These results allow proposing the third hyperbolic (or harmonic) rule on

№	S_A	Range %	S_T	Range %	S_C	Range %	S_G	Range %
1	67070277	-0.030 +0.024	67244164	-0.025 +0.110	48055043	-0.088 +0.068	48111528	-0.106 +0.057
2	71791213	-0.079 +0.087	71987932	-0.075 +0.095	48318180	-0.097 +0.072	48450903	-0.105 +0.141
3	59689091	-0.021 +0.045	59833302	-0.097 +0.098	39233483	-0.130 +0.081	39344259	-0.034 +0.088
4	58561236	-0.065 +0.044	58623430	-0.036 +0.128	36236976	-0.039 +0.127	36331025	-0.117 +0.075
5	54699094	-0.052 +0.040	54955010	-0.071 +0.078	35731600	-0.012 +0.132	35879674	-0.103 +0.085
6	51160489	-0.039 +0.057	51151754	-0.049 +0.022	33520786	-0.092 +0.061	33516767	-0.029 +0.069
7	47058248	-0.104 +0.040	47215040	-0.061 +0.030	32317984	-0.086 +0.091	32378859	-0.076 +0.069
8	42641072	-0.061 +0.068	42581941	-0.111 +0.071	28600559	-0.110 +0.069	28600963	-0.068 +0.050
9	31752642	-0.134 +0.090	31733822	-0.083 +0.065	22487631	-0.099 +0.141	22470915	-0.079 +0.143
10	38875926	-0.081 +0.052	39027555	-0.067 +0.099	27639505	-0.058 +0.085	27719976	-0.118 +0.085
11	39286730	-0.032 +0.084	39361954	-0.062 +0.042	27903257	-0.139 +0.056	27981801	-0.086 +0.112
12	39370109	-0.096 +0.056	39492225	-0.097 +0.094	27092804	-0.076 +0.078	27182678	-0.073 +0.105
13	29224840	-0.067 +0.077	29320872	-0.107 +0.069	18341128	-0.107 +0.141	18346620	-0.130 +0.065
14	25606393	-0.109 +0.100	25819249	-0.040 +0.086	17733667	-0.137 +0.077	17782016	-0.056 +0.142
15	24508669	-0.085 +0.179	24553812	-0.127 +0.088	17752941	-0.090 +0.162	17825903	-0.067 +0.113
16	22558319	-0.122 +0.080	22774906	-0.143 +0.104	18172742	-0.146 +0.074	18299976	-0.146 +0.173
17	22639499	-0.141 +0.105	22705261	-0.146 +0.070	18723944	-0.134 +0.072	18851500	-0.144 +0.105
18	22087028	-0.160 +0.071	22109347	-0.169 +0.121	14574701	-0.090 +0.134	14594335	-0.160 +0.210
19	15142293	-0.160 +0.024	15282753	-0.062 +0.062	13954580	-0.103 +0.097	14061132	-0.057 +0.226
20	16455618	-0.106 +0.129	16643030	-0.099 +0.089	13037092	-0.062 +0.116	13098788	-0.092 +0.155
21	9943435	-0.161 +0.083	9882679	-0.206 +0.173	6864570	-0.134 +0.277	6852178	-0.373 +0.219
22	10382214	-0.175 +0.084	10370725	-0.036 +0.209	9160652	-0.258 +0.155	9246186	-0.143 +0.235
X	46754807	-0.078 +0.084	46916701	-0.102 +0.055	30523780	-0.116 +0.179	30697741	-0.135 +0.067
Y	7886192	-0.244 +0.097	7956168	-0.063 +0.185	5285789	-0.181 +0.407	5286894	-0.247 +0.142

Fig. 4. Some results of the analysis of all 24 human nuclear chromosomes by the oligomer sums method are represented. For each of the chromosomes, quantities S_A, S_T, S_C, and S_G of monomers A, T, C, and G are shown to define the model hyperbolic progressions (2.2). The columns «Range %" show ranges of deviations of real OS-series of corresponding *n*-plets ($n = 1, 2, \ldots, 20$) from their appropriate model values S_A/n, S_T/n, S_C/n, and S_G/n in percentages (in each case, an appropriate model value is taken as 100%). The left column shows chromosome numbers.

the total amounts of *n*-plets, which start with a certain nucleotide A, T, C, or G.

The third hyperbolic rule (about the similarity of chromosomes):

- All chromosomes of any individual eukaryotic genome have approximately the same normalized OS-sequences S_A/nS, S_T/nS, S_C/nS, and S_G/nS representing classes of A_1-, T_1-, C_1-, and G_1-oligomers ($n = 1, 2, 3, 4, \ldots$is not too large compared to the full length of the nucleotide sequence).

$$S_A/n + S_T/n + S_C/n + S_G/n = S/n \text{ or } S_A/nS + S_T/nS + S_C/nS + S_G/nS = 1/n \quad (2.8)$$

The author suggests that these hyperbolic rules are universal genetic rules. But at this stage of the study, they are only candidates for the role of universal rules, since the analysis of the widest variety of genomes is required to verify their universality.

Let us return to the harmonic progression (2.4) and recall its relation with the well-known concept of the harmonic mean. The harmonic mean *H* of the positive real numbers *x1, x2, …, xn,* is defined to be

$$H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}} \quad (2.7)$$

Knowing two neighboring members of the harmonic progression, one can calculate its next member by means of the harmonic mean relation. Here we can briefly mention that the harmonic mean is associated with the Pythagorean teaching on the musical harmony and the aesthetics of proportions, presented in the famous numerical triangle published 2000 years ago by Nichomachus of Gerasa in his book "Introduction into arithmetic". In accordance with this triangle, the Parthenon (Kappraff, 2006) and other great architectural objects were created because architecture was interpreted as the non-movement music, and the music was interpreted as the dynamic architecture (see more details in (Kappraff, 2000, 2002; Petoukhov, 2008; Petoukhov, He, 2010, Section 2, Chapter 4)). Since the harmonic mean is related to the harmonic progression, the author indicates magnitudes of the harmonic mean in some figures of the article for the comparison analysis of OS-sequences in different nucleotide sequences (Fig. 5 and many others).

Each genomic DNA sequence with its total amount S of all nucleotides A, T, C, and G also contains total amounts S/*n* of *n*-plets (that is, S/2 doublets, S/3 triplets, etc.). These total amounts are members of the hyperbolic progression S, S/2, S/3, …, S/*n*. Each member of this sequence S/*n* is the sum of the four OS-sequences S_A/n, S_T/n, S_C/n, and S_G/n (2.8):

These linear superpositions are valid for a wide variety of genomes that differ only in individual coefficients S_A, S_T, S_C, and S_G.

Chrom	S_A/S	S_T/S	S_C/S	S_G/S	Harmonic mean
1	0.2910	0.2918	0.2085	0.2087	0.243
2	0.2984	0.2993	0.2009	0.2014	0.241
3	0.3013	0.3020	0.1980	0.1986	0.239
4	0.3086	0.3089	0.1910	0.1915	0.236
5	0.3018	0.3032	0.1971	0.1979	0.239
6	0.3021	0.302	0.1979	0.1970	0.239
7	0.2960	0.2970	0.2033	0.2037	0.241
8	0.2994	0.2990	0.2008	0.2008	0.240
9	0.2928	0.2926	0.2074	0.2072	0.243
10	0.2917	0.2929	0.2074	0.2080	0.243
11	0.2920	0.2926	0.2074	0.2080	0.243
12	0.2957	0.2966	0.2035	0.2042	0.242
13	0.3069	0.3079	0.1926	0.1926	0.237
14	0.2945	0.2970	0.2040	0.2045	0.242
15	0.2896	0.2901	0.2097	0.2106	0.244
16	0.2758	0.2784	0.2221	0.2237	0.247
17	0.2730	0.2738	0.2258	0.2273	0.248
18	0.3011	0.3014	0.1987	0.1989	0.240
19	0.2591	0.2615	0.2388	0.2406	0.250
20	0.2778	0.2810	0.2201	0.2211	0.247
21	0.2964	0.2946	0.2047	0.2043	0.242
22	0.2651	0.2648	0.2339	0.2361	0.249
X	0.3019	0.3029	0.1971	0.1982	0.239
Y	0.2985	0.3012	0.2001	0.2001	0.240

Fig. 5. Data for normalized OS-sequences $S_A/(nS)$, $S_T/(nS)$, $S_C/(nS)$, and $S_G/(nS)$ of all human chromosomes are shown for comparison. Here S_A, S_T, S_C, and S_G refer to phenomenological quantities of nucleotides A, T, C, and G; $S = S_A + S_T + S_C + S_G$. Harmonic means of the values S_A/S, S_T/S, S_C/S, and S_G/S in each chromosome are also indicated

3. The hyperbolic rules in all chromosomes of a fruit fly *Drosophila melanogaster* and some other model eukaryotic species

This Sections is devoted to the analysis - by the oligomer sums method (the OS-method) - of single-stranded DNA sequences of the complete sets of chromosomes of a few model eukaryotic organisms, which are used long ago in the study of genetics, development, and disease. Received results confirm that both hyperbolic (harmonic) rules regarding *n*-plets from the classes of A_1-, T_1-, C_1-, and G_1-oligomers hold for each of described chromosomes at $n = 1, 2, 3, 4, \ldots, 19, 20$ (although these rules are also satisfied for larger values of *n*, at least up to $n = 100$, but the data tables for such large *n* are too cumbersome to include in the presentations).

Let us start with a fruit fly *Drosophila melanogaster*, which is studied in biology labs for over eighty years. All initial data about its chromosomes were taken from the GenBank - https://www.ncbi.nlm.nih.gov/genome/?term=drosophila+melanogaster. Resulting data in Fig. 6 confirm that - for all the chromosomes - the model hyperbolic progressions $H_{A,1}(n) = S_A/n$, $H_{T,1}(n) = S_T/n$, $H_{C,1}(n) = S_C/n$, and $H_{G,1}(n) = S_N/n$ from the expression (2.2) practically coincide with the real sequences of total amounts of *n*-plets from the classes A_1-, T_1-, C_1-, and G_1-oligomers at $n = 1, 2, 3, \ldots, 20$. In all shown cases, the deviations of real sequences from model hyperbolic progressions are less than 1% as data in the tabular columns « Range %» indicates. This means that the formulated hyperbolic (harmonic) rules are fulfilled in the considered genome.

№	S_A	Range %	S_T	Range %	S_C	Range %	S_G	Range %
X	6732793	-0.196 ÷0.057	6774766	-0.125 ÷0.090	4975870	-0.198 ÷0.139	4992722	-0.148 ÷0.213
2L	6853032	-0.217 ÷0.178	6836080	-0.219 ÷0.090	4912017	-0.239 ÷0.313	4912383	-0.251 ÷0.350
2R	7272860	-0.259 ÷0.128	7235562	-0.144 ÷0.304	5395216	-0.195 ÷0.222	5376598	-0.222 ÷0.323
3L	8143548	-0.142 ÷0.196	8198331	-0.126 ÷0.206	5825673	-0.211 ÷0.108	5824515	-0.262 ÷0.169
3R	9205526	-0.143 ÷0.152	9197619	-0.145 ÷0.132	6833716	-0.170 ÷0.169	6817898	-0.231 ÷0.192
4	425241	-1.759 ÷0.488	436669	-0.423 ÷0.744	232566	-1.463 ÷1.299	236655	-0.855 ÷1.369
Y	1056780	-0.494 ÷0.314	1008635	-0.125 ÷0.431	682725	-0.268 ÷0.659	661579	-0.512 ÷0.386

Fig. 6. The results of the analysis of all chromosomes of *Drosophila melanogaster* by the OS-method. The left column shows symbols of chromosomes. S_A, S_T, S_C, and S_G refer to the quantities of nucleotides A, T, C, and G in appropriate chromosomes. The columns "Range %" show deviations of real sequences from the model hyperbolic progressions $H_{A,1}(n) = S_A/n$, $H_{T,1}(n) = S_T/n$, $H_{C,1}(n) = S_C/n$, and $H_{G,1}(n) = S_G/n$ at $n = 1, 2, 3, \ldots, 20$ (the model values are taken as 100%).

Chrom	S_A/S	S_T/S	S_C/S	S_G/S	Harmonic mean
X	0.2868	0.2886	0.2120	0.2127	0.244
2L	0.2915	0.2907	0.2089	0.2089	0.243
2R	0.2877	0.2862	0.2134	0.2127	0.245
3L	0.2909	0.2929	0.2081	0.2081	0.243
3R	0.2872	0.2869	0.2132	0.2127	0.245
4	0.3195	0.3280	0.1747	0.1778	0.228
Y	0.3099	0.2958	0.2002	0.1940	0.239

Fig. 7. Data of normalized OS-sequences S_A/nS, S_T/nS, S_C/nS, and S_G/nS of all chromosomes of *Drosophila melanogaster* are shown for comparison. Here $S = S_A + S_T + S_C + S_G$. Harmonic means of the values S_A/S, S_T/S, S_C/S, and S_G/S in each chromosome are also indicated.

Fig. 7 shows data of normalized OS-sequences for all chromosomes of *Drosophila melanogaster*.

Similar results, which confirm the hyperbolic rules in eukaryotic genomes, are received by the oligomer sums method for all the analyzed eukaryots including the following:

- The free-living soil nematode *Caenorhabditis elegans* by the OS-method. This nematode is widely used as a model organism in genetics for a long time. The *Caenorhabditis elegans* nuclear genome is approximately 100 Mb, distributed among 6 chromosomes;
- the laboratory mouse *Mus musculus*, which has 21 chromosomes and is a major model organism for basic mammalian biology, human disease, and genome evolution;
- a plant *Arabidopsis thaliana.* This small flowering plant has 5 chromosomes and is used for over fifty years to study plant mutations and for classical genetic analysis. It became the first plant genome to be fully sequenced; it has a small genome of ~120 Mb.

All initial data on these genomes were taken from the GenBank. Detailed numerical results of the analysis are presented in the preprint (Petoukhov, 2020e).

4. Analysis of long genes by the oligomer sums method

Before proceeding to the analysis of prokaryotic genomes, it is useful to show the applicability of the oligomer sums method to the analysis of genes whose sequences are much shorter than DNA sequences in chromosomes. The application of the method unexpectedly reveals the phenomenon of regular rhythmic deviations of the sequences of real total sums of *n*-plets in the described genes from the corresponding model hyperbolic progressions.

Let us first consider the human *TTN* gene encoding the largest known protein Titin. Titin, also known as connectin, is important in the contraction of striated muscle tissues. Figs. 8–9 show some results of the analysis - by the oligomer sums method - of the nucleotide sequence of

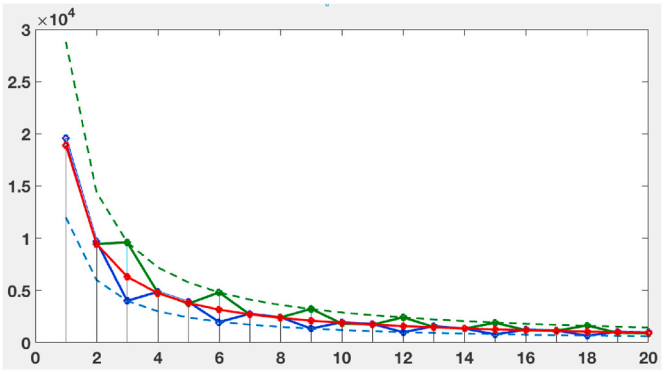


Fig. 9. The graph, uniting two graphs from Fig. 8 for the *TNT* gene, is shown. The blue dot line and the green dot lines correspond to those additional hyperbolic progressions $11979/n$ and $28788/n$, which model real total amounts of $3m$-plets. Other parts of this united graph are the same as in Fig. 8. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

the *TTN* gene (numeric results will be represented below). Initial data on its nucleotide sequence are taken in the GenBank https://www.ncbi.nlm.nih.gov/nuccore/X90568.1. This gene contains 26,373 nucleotides A, 19569 nucleotides T, 17097 nucleotides C, and 18,901 nucleotides G, that is $S_A = 26{,}373$, $S_T = 19{,}569$, $S_C = 17{,}097$, and $S_G = 18{,}901$ for the model hyperbolic progressions (2.2). It can be especially noted that, in this gene, the amounts of nucleotides A and T are significantly different (26,373 and 19,569), that is, the second Chargaff's rule on their approximate equality in long sequences is not satisfied here since this nucleotide sequence is not enough long for the Chargaff's rule.

Fig. 8 shows the sequences of the highly regular significant deviations of the real total amounts of *n*-plets, which start with the nucleotide T and the nucleotide G, from model hyperbolic progressions $S_T/n = 19569/n$ and $S_G/n = 18901/n$. One should note that all these significant deviations happen only at $n = 3, 6, 9, \ldots, 3m$, that is only for cases of $3m$-plets (here $m = 1, 2, 3, \ldots$). Correspondingly these significant deviations can be called « triplet-deviations».

Fig. 9 shows the graph, which unites both graphs from Fig. 8 and demonstrates a few interesting features of the highly regular series of these triplet-deviations.

Firstly, one can see in Fig. 9 that, in classes of T_1-oligomers and G_1-oligomers, the triplet-deviations happen in opposite directions (or, figuratively speaking, in antiphase):

- in the class of T_1-oligomers, they decrease real values compared with model values of the hyperbolic progression $19569/n$;

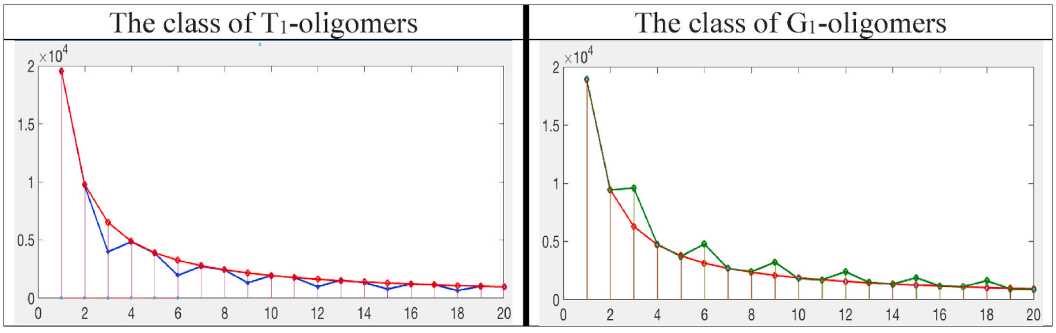


Fig. 8. Graphical representations of the results of the analysis - by the oligomer sums method – of the human *TTN* gene. The OS-sequences of its total amounts of *n*-plets, which start with the nucleotide T (left) and the nucleotide G (right), are shown. The red lines refer to model hyperbolic progressions S_T/n and S_G/n correspondingly, where $S_T = 19{,}569$ and $S_G = 18{,}901$ are quantities of nucleotides T and G in the gene; $n = 1, 2, 3, \ldots, 20$ as shown at the abscissa axes. The blue line (left) and the green line (right) with dots on them refer to the real OS-sequences of the total amounts of such *n*-plets. The ordinate axes indicate the total amounts of *n*-plets. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

- in the class of G_1-oligomers, they increase real values in comparison with model values of the hyperbolic progression 18901/*n*.

Secondly, under triplet-deviations, real total amounts of 3*m*-plets from the classes of T_1-oligomers and G_1-oligomers belong correspondingly to other hyperbolic progressions 11979/*n* and 28788/*n*. These hyperbolic progressions are indicated by the blue dot line and the green dot line in Fig. 9. Where did these numerators of model hyperbolas come from? Each of these numerators is associated with the total amount of triplets ($n = 3$) in an appropriate class of oligomers in this gene: the total amount of triplets starting with nucleotide T is equal to 3993, and the total amount of triplets starting with nucleotide G is equal to 9596. To calculate the first values of the model hyperbolas, each of these amounts of triplets must be tripled, giving the shown numerators 11,979 and 28,788.

Similar triplet-deviations exist in the OS-representations not only of the *TTN* gene but also of other long genes, prokaryotic genomes, and viruses in different degrees as the author has discovered in the analysis of a limited set of nucleotide series by the OS-method. In the genetic code system, triplets have an important meaning, which differs from other *n*-plets: they encode amino acids and punctuations of protein synthesis. One can believe that the phenomenon of the triplet-deviations is related to this special meaning of triplets. For this reason, the deeper analysis of triplet-deviations in different species can be useful to study the secrets of the genetic system and biological evolution.

Fig. 9 demonstrated the highly regular rhythmic triplet-deviations for $n = 1, 2, 3, \ldots, 20$, but similar rhythmic triplet-deviations exist in a much wider range of values *n*.

Fig. 10 shows in graphical forms percentage values of the highly regular rhythmic deviations of the real total amounts of *n*-plets, which start with the nucleotide T and with the nucleotide G in the *TTN* gene, from the appropriate model values 19569/*n* and 18901/*n*. Two cases of the range of values *n* are represented there: $n = 1, 2, 3, \ldots, 20$, and $n = 1, 2, 3, .., 100$.

The nucleobases T and G are keto-nucleobases. Figs. 9 and 10 draw attention to the phenomenon of long-range correlations in the *TTN* gene between sequences of the triplet-deviations in classes of T_1-, and G_1-oligomers: the triplet-deviations in these sequences happen in opposite directions as above mentioned. Such binary oppositions, which meet in different long genes, prokaryotic genomes, and viruses regarding the classes of different N_1-oligomers (here N refers to A, T, C, or G), should be specially studied in future since they bear important information and are associated with other binary-opposition features of molecular genetic systems.

The following Fig. 11 shows the OS-sequences of the total amounts of *n*-plets, which start with two other nucleotides A and C in the *TTN* gene. This gene contains 26,373 nucleotides A and 17,097 nucleotides C; correspondingly $S_A = 26{,}373$ and $S_C = 17{,}097$ for the model hyperbolic progressions (2.2).

One can see in Fig. 11 that the class of C_1-oligomers has regular sequences of the significant triplet-deviations at 3*m*-plets shown by the blue line. The class of A_1-oligomers has not such regular sequences of significant deviations; besides, its deviations are essentially less than deviations in the class of C_1-oligomers. In the class of A_1-oligomers, the real and model values differ little from each other, and therefore, in Fig. 11, the red line of model values covers the line of real values.

Fig. 12 shows the numeric results of the analysis of the *TTN* gene by the oligomer sums method.

The preprint (Petoukhov, 2020e) shows similar results of the analysis of some other genes, including one of the short genes of human histones, by the oligomer sums method. The author notes else that not all long genes have regular sequences of the pronounced triplet-deviations in their OS-representations. The comparison analysis of the OS-representations of different genes is a new research field. Certain triplet-deviations between real and model values under 3*m*-plets are also found in the OS-representations of entire chromosomes of humans and other organisms, but in a much less pronounced form than in cases of individual genes.

5. The hyperbolic rules in bacterial genomes of different groups both from bacteria and archaea

Let us turn now to prokaryotic genomes. The Section represents results of the analysis of nucleotide sequences of all 19 bacterial genomes of different groups both from Bacteria and Archaea, which are listed in the article on the second Chargaff`s rule (Rapoport, Trifonov, 2012, p. 2): "*Nucleotide disparities for prokaryotic coding sequences were taken from bacterial genomes of different groups both from Bacteria and Archea. All*

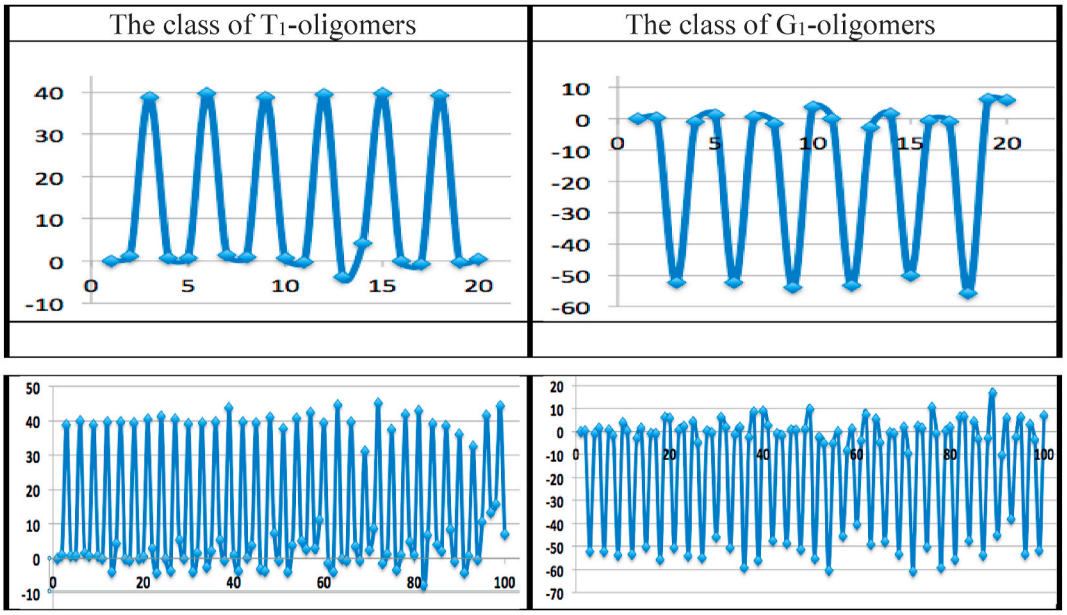


Fig. 10. Percentage representations of highly regular rhythmic sequences of the triplet-deviations of the real amounts of *n*-plets, which belong to classes of T_1-, and G_1-oligomers, from the appropriate model hyperbolic values 19569/*n* and 18901/*n* in the *TTN* gene. Here $n = 1, 2, 3, \ldots, 20$ (upper row) and $n = 1, 2, 3, \ldots, 100$ (bottom row) as shown at the abscissa axes. The ordinate axes show percentages of the deviations (the model values are taken as 100%).

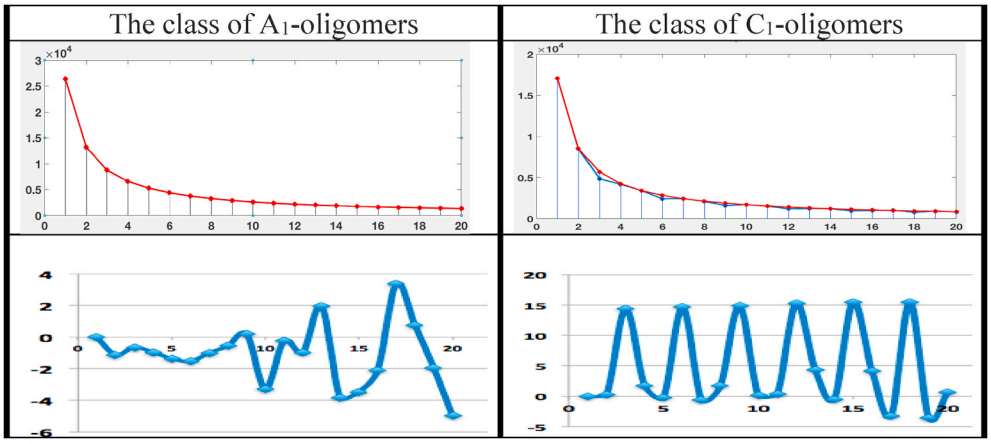


Fig. 11. Graphical representations of the results of the analysis - by the oligomer sums method – of the human *TTN* gene regarding the sequences of the total amounts of *n*-plets, which start with the nucleotide A (left) and the nucleotide C (right). Here $n = 1, 2, 3, \ldots, 20$ (at the absciss axes). **Upper row**: the red lines refer to model hyperbolic progressions $S_A/n = 26373/n$ and $S_C/n = 17097/n$ correspondingly. The ordinate axes show the total amounts of appropriate *n*-plets. The class of C_1-oligomers has regular sequences of the significant triplet-deviations at $3m$-plets shown by the blue line. **Bottom row**: percentage representations of the sequences of deviations of the real total amounts of *n*-plets of these classes from the appropriate model hyperbolic values $26373/n$ and $17097/n$ (the ordinate axes show these percentages). The model values are taken as 100%. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

n	1	2	3	4	5	6	7	8	9	10
A										
Real	26373	13334	8848	6656	5346	4463	3805	3315	2924	2724
Model	26373	13187	8791	6593	5275	4396	3768	3297	2930	2637
Δ%	0	-1.119	-0.648	-0.952	-1.354	-1.536	-0.993	-0.557	0.216	-3.287
T										
Real	19569	9677	3993	4857	3885	1964	2755	2426	1332	1943
Model	19569	9784.5	6523	4892	3914	3262	2796	2446	2174	1957
Δ%	0	1.099	38.786	0.721	0.736	39.782	1.451	0.823	38.740	0.710
C										
Real	17097	8522	4876	4199	3426	2431	2458	2101	1617	1707
Model	17097	8549	5699	4274	3419	2850	2442	2137	1900	1710
Δ%	0	0.310	14.441	1.761	-0.193	14.687	-0.638	1.690	14.880	0.158
G										
Real	18901	9437	9596	4773	3731	4798	2687	2400	3231	1820
Model	18901	9451	6300	4725	3780	3150	2700	2363	2100	1890
Δ%	0	0.143	-52.309	-1.011	1.302	-52.309	0.487	-1.582	-53.849	3.709

n	11	12	13	14	15	16	17	18	19	20
A										
Real	2403	2219	1989	1956	1819	1683	1499	1454	1415	1384
Model	2398	2198	2029	1884	1758	1648	1551	1465	1388	1319
Δ%	-0.228	-0.967	1.957	-3.833	-3.458	-2.104	3.375	0.762	-1.941	-4.956
T										
Real	1782	986	1563	1339	788	1224	1160	660	1032	974
Model	1779	1631	1505	1398	1305	1223	1151	1087	1030	978
Δ%	-0.169	39.537	-3.833	4.206	39.598	-0.077	-0.772	39.292	-0.199	0.455
C										
Real	1548	1207	1258	1227	963	1024	1038	803	932	849
Model	1554	1425	1315	1221	1140	1069	1006	950	900	855
Δ%	0.404	15.283	4.346	-0.474	15.511	4.170	-3.211	15.459	-3.574	0.684
G										
Real	1716	2416	1493	1330	1892	1190	1123	1635	933	890
Model	1718	1575	1454	1350	1260	1181	1112	1050	995	945
Δ%	0.132	-53.389	-2.688	1.487	-50.151	-0.735	-1.005	-55.706	6.211	5.825

Fig. 12. Real and model values to the OS-representations of the classes of A_1-, T_1-, C_1-, and G_1-oligomers in the human *TTN* gene are shown for $n = 1, 2, \ldots, 20$. The real total amounts of *n*-plets, which start with a certain nucleotide (A, T, C, or G), are indicated jointly with their model values $H_{A,1}(n) = 26373/n$, $H_{T,1}(n) = 19569/n$, $H_{C,1}(n) = 17097/n$, and $H_{G,1}(n) = 18901/n$ (in red). The symbol Δ% refers to deviations of real values from model values in percent (the model values are taken as 100%). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

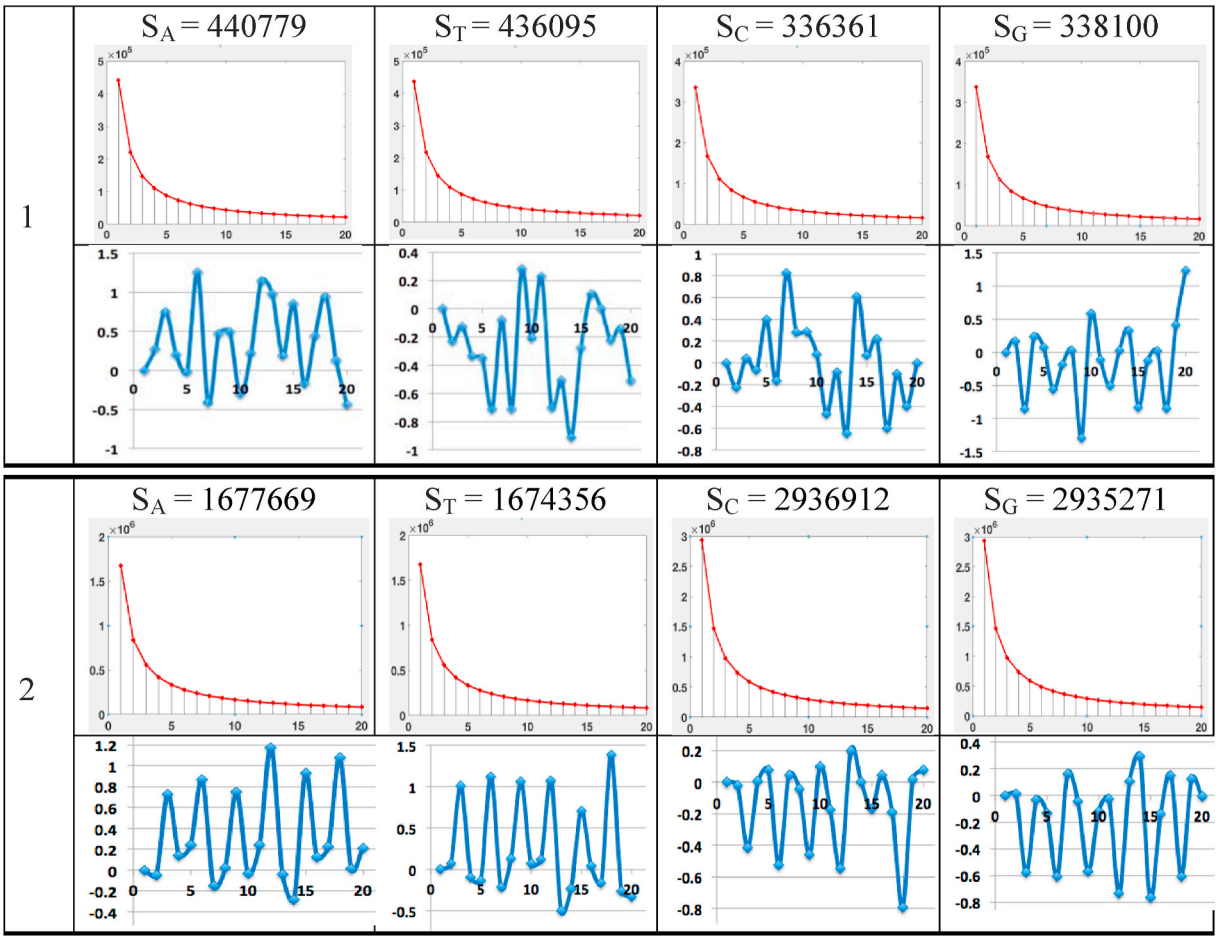


Fig. 13. Graphical representations of the results of the analysis - by the oligomer sums method – of the following bacterial genomes mentioned in (Rapoport, Trifonov, 2012, p. 2): 1) *Aquifex aeolicus;* 2) *Bradyrhizobium japonicum*. For each of genomes two rows of resulting data are shown at $n = 1, 2, \ldots, 20$ plotted along the abscissa axes: the top rows demonstrate that model hyperbolic progressions S_A/n, S_T/n, S_C/n, S_G/n (red lines) almost completely cover the OS-sequences of phenomenological values (the ordinate axes show appropriate values); the bottom blue lines show in percent slight alternating deviations of phenomenological values from model values. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

together 19 genomes were used: Aquifex aeolicus, Acidobacteria bacterium, Bradyrhizobium japonicum, Bacillus subtilis, Chlamydia trachomatis, Chromobacterium violaceum, Dehalococcoides ethenogenes, Escherichia coli, Flavobacterium psychrophilum, Gloeobacter violaceus, Helicobacter pilory, Methanosarcina acetivorans, Nanoarchaeum equitans, Syntrophus aciditrophicus, Streptomyces coelicolor, Sulfolobus solfataricus, Treponema denticola, Thermotoga maritima and Thermus thermophiles".

Fig. 13 shows the results of the analysis of the first three prokaryotic genomes form this list by the oligomer sums method. Similar results of the analysisi of all other genomes from the list are shown in the preprint (Petoukhov, 2020e). These results demonstrate that the hyperbolic rule No. 1 is fulfilled for all the listed genomes of prokaryotes: the model hyperbolic progressions $H_{A,1}(n) = S_A/n$, $H_{T,1}(n) = S_T/n$, $H_{C,1}(n) = S_C/n$, and $H_{G,1}(n) = S_G/n$ from the expression (2.2) practically coincide with the OS-sequences of real total amounts of n-plets from the classes A_1-, T_1-, C_1-, and G_1-oligomers at $n = 1, 2, 3, \ldots, 20$. Because of this coincidence, the model hyperbolic progressions, which are represented by red lines in the graphs of Fig. 13, almost completely cover the sequences of real values (the blue lines in the lower graphs show in percent slight alternating deviations of real values from model values).

The initial data on these prokaryotic genomes were taken from the GenBank:

1) Aquifex aeolicus VF5, complete genome, 1,551,335 bp, accession AE000657, version AE000657.1, HYPERLINK "https://www.ncbi.nlm.nih.gov/nuccore/AE000657.1?report = genbank" https://www.ncbi.nlm.nih.gov/nuccore/AE000657.1?report=genbank;
2) Bradyrhizobium japonicum strain E109, complete genome, 9,224,208 bp, accession CP010313, HYPERLINK "https://www.ncbi.nlm.nih.gov/nuccore/CP010313.1?report = genbank" https://www.ncbi.nlm.nih.gov/nuccore/CP010313.1?report=genbank.

Similar results, received in the analysis of other 17 genomes of bacteria and archaea by the OS-method, are presented in the preprint (Petoukhov, 2020e).

6. Analysis of genomes of microorganisms living in extreme environments

Of particular interest is the analysis of the genetic characteristics of microorganisms (extremophiles) living under extreme conditions of high and low temperatures, radiation, acidic and alkaline environments, drying, etc. Study of extremophiles is useful for many practical and theoretical problems. The https://en.wikipedia.org/wiki/Extremophile website contains a table of extremophiles. For the analysis of their genomes by the oligomer sums method, the author used 1–2 organisms from each category of the table. Thus, the genomic data of the following extremophiles were taken in the GenBank and analyzed: 1) *Pyrolobus fumarii 1A*, which lives in submarine hydrothermal vents; 2)

Synechococcus lividus PCC 6715, which lives in low temperature conditions; 3) *Chroococcidiopsis thermalis PCC 7203*, which lives in conditions of desiccation; 4) *Pyrococcus furiosus DSM 3638*, which lives in submarine hydrothermal vents; 5) *Psychrobacter alimentarius strain PAMC 27889*, which lives in soda lakes; 6) *Clostridium paradoxum* JW-YL-7 = DSM 7308 strain JW-YL-7 ctg1, which lives in volcanic springs, acid mine drainage; 7) *Deinococcus radiodurans R1*, which lives in conditions of cosmic rays, X-rays, radioactive decay; 8) *Halobacterium sp. NRC-1*, which lives in conditions of high salt concentration.

The results of the analysis of these genomic data shows the fulfillment of the hyperbolic rule №1 of oligomeric sums for all listed extremophiles. The extremal living conditions of these microorganisms do not affect the subordination of their genomes to the described hyperbolic (harmonic) rules of the algebraic invariance, which are true for the genomes of other prokaryotes and eukaryotes.

Fig. 14 shows the results of the analysis for the first two extremophiles from the list. Similar results for all other listed extremophiles are presented in the preprint (Petoukhov, 2020e).

7. Analysis of giant viruses by the oligomer sums method

This Section represents examples of studying genomes of different viruses by the oligomer sums method. The focus is on giant viruses (Fig. 15).

The results, presented in this Section, show the fulfillment of the hyperbolic (harmonic) rule № 1 for the viruses considered and provide material for comparative analysis of different OS-sequences in genetics.

8. Analysis of the COVID-19 virus by the oligomer sums method

Let us turn now to the analysis - by the oligomeric sums method - of the COVID-19 virus, which led to a pandemic. The initial data on its nucleotide sequence was taken by the author from the site https://www.ncbi.nlm.nih.gov/nuccore/MN908947.3, where the following is written about it: severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome, GenBank: MN908947.3, LOCUS MN908947, 29,903 bp ss-RNA linear VRL 18-MAR-2020.

Figs. 16–17 show some results of such an analysis of the virus.

In particular, Figs. 16 and 17 show that this virus in its OS-representations has under $n = 3, 6, 9, \ldots, 3m$ such deviations of

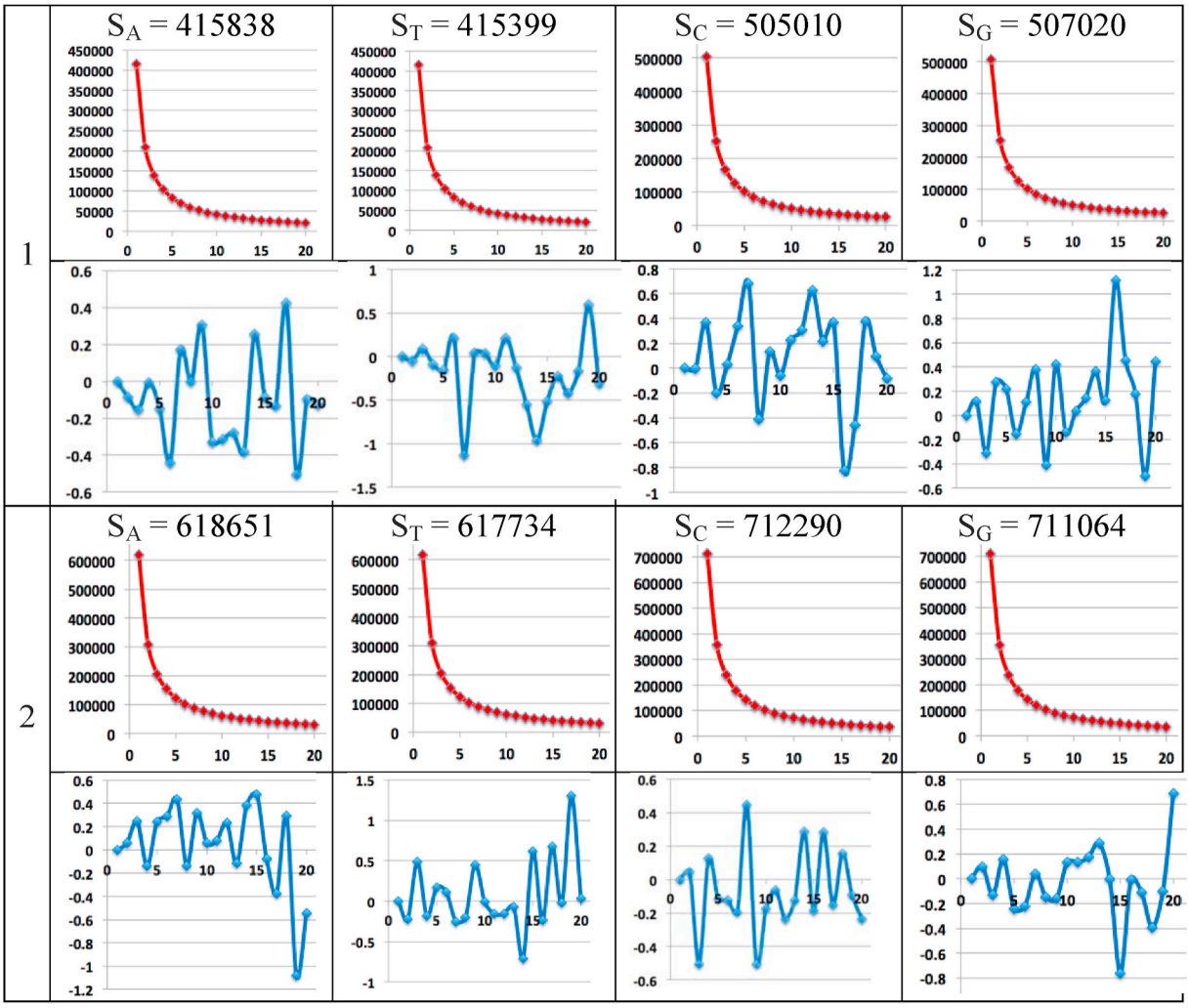


Fig. 14. The results of the analysis the following extremophiles: 1) *Pyrolobus fumarii 1A*, complete genome, 1843267 bp, https://www.ncbi.nlm.nih.gov/nuccore/NC_015931.1; 2) *Synechococcus lividus PCC 6715* chromosome, complete genome, 2,659,739 bp, https://www.ncbi.nlm.nih.gov/nuccore/NZ_CP018092.1 All abscissa axes show the values $n = 1, 2, \ldots, 20$. The red hyperbolic lines demonstrate model hyperbolic progressions S_A/n, S_T/n, S_C/n, S_G/n, which almost completely cover the OS-sequences of phenomenological values (the ordinate axes show appropriate values). Blue lines show in percent slight alternating deviations of phenomenological values of the OS-sequences from model values S_A/n, S_T/n, S_C/n, S_G/n (here S_A, S_T, S_C, and S_G denote number of nucleotides A, T, C, and G in the genomes). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

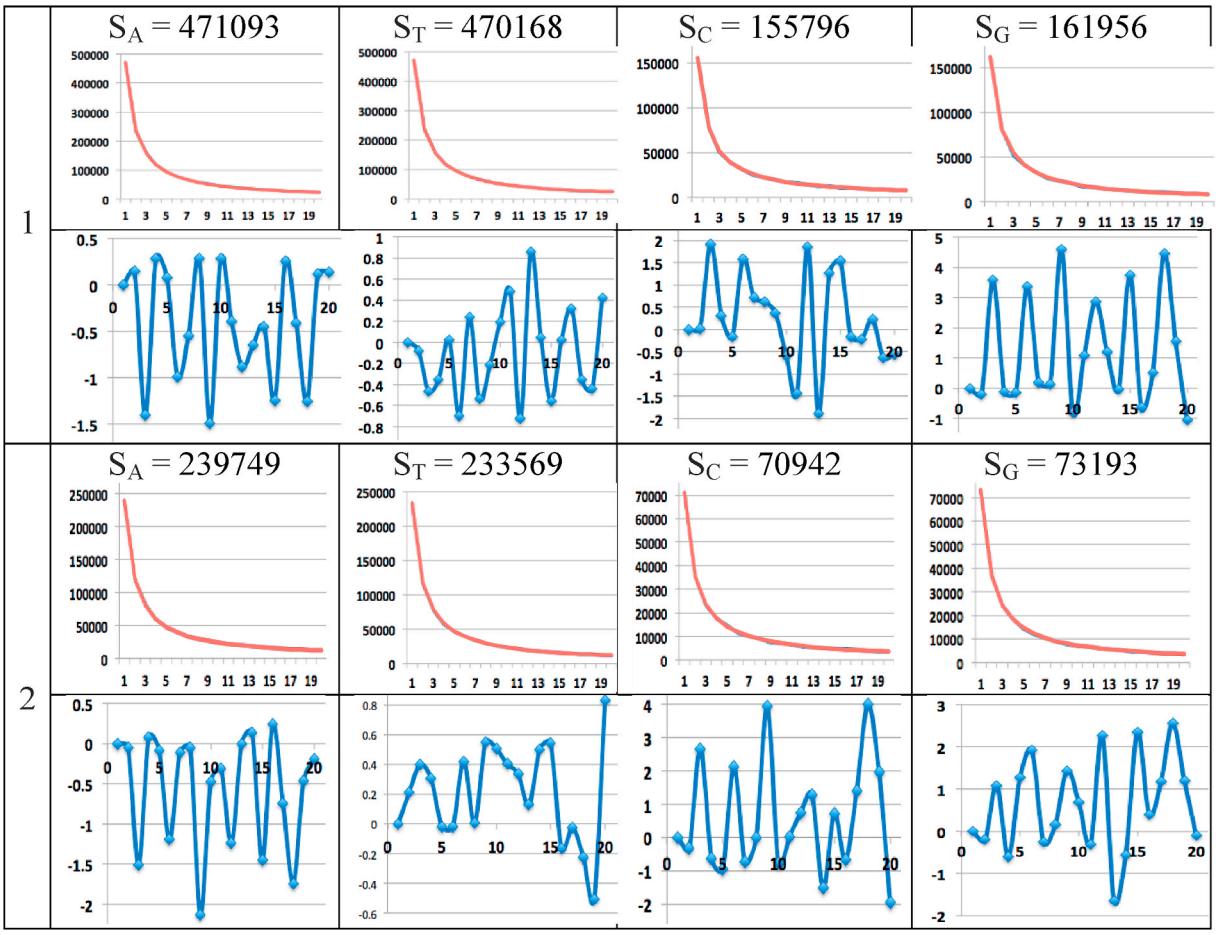


Fig. 15. The results of the analysis - by the oligomer sums method – the following giant viruses: 1) Megavirus chiliensis, complete genome, 1,259,197 bp, NCBI Reference Sequence: NC_016072.1, https://www.ncbi.nlm.nih.gov/nuccore/NC_016072.1; 2) Cafeteria roenbergensis virus BV-PW1, complete genome, 617,453 bp, NCBI Reference Sequence: NC_014637.1, https://www.ncbi.nlm.nih.gov/nuccore/NC_014637.1. All abscissa axes show the values n = 1, 2, …, 20. The red hyperbolic lines demonstrate that model hyperbolic progressions S_A/n, S_T/n, S_C/n, S_G/n (red lines) almost completely cover the OS-sequences of phenomenological values (the ordinate axes show appropriate values). Blue lines show in percent slight alternating deviations of real values of the OS-sequences from model values S_A/n, S_T/n, S_C/n, S_G/n (here S_A, S_T, S_C, and S_G denote number of nucleotides A, T, C, and G in these viruses). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

phenomenological values from model values, which resemble the triplet-deviations in human genes, which were described above in Fig. 4.1-4.5. Perhaps the harmfulness of this virus to humans is related to this similarity. It should also be noted that - in the classes of pyrimidines C_1- and T_1-oligomers - these deviations occur in opposite directions in a coordinated manner, which indicates a particular consistency in the structure of the nucleotide sequence of this virus concerning pyrimidines classes.

9. DNA epi-chains and the hyperbolic rules for oligomer sums

This Section presents some results of the study of special subsequences of long nucleotide sequences in single-stranded DNA by the oligomer sums method. These subsequences are termed «DNA epi-chains» (Petoukhov, 2019a). The author's initial results testify that the above described hyperbolic rules of oligomer sums for genomes are also fulfielld for these epi-chains; it gives new materials to a known theme of fractal-like structures in genetics.

By definition, in a nucleotide sequence N_1 of any DNA strand with sequentially numbered nucleotides 1, 2, 3, 4, … (Fig. 18a), epi-chains of different orders k are such subsequences that contain only those nucleotides, whose numeration differ from each other by natural number k = 1, 2, 3, …. For example, in any single-stranded DNA, epi-chains of the second order are two nucleotide subsequences $N_{2/1}$ and $N_{2/2}$ in which their nucleotide sequence numbers differ by k = 2: the epi-chain $N_{2/1}$ contains nucleotides with odd numerations 1, 3, 5, …(Fig. 18b), and the epi-chain $N_{2/2}$ contains nucleotides with even numerations 2, 4, 6, … (Fig. 18c). By analogy, epi-chains of the third order are those three nucleotide subsequences $N_{3/1}$, $N_{3/2}$, and $N_{3/3}$, each of which has sequence numbers that differ by k = 3: these epi-chains contain nucleotides with numerations 1, 4, 7, … or 2, 5, 8, … or 3, 6, 9, …, respectively (Fig. 18d-a). The epi-chain of the first order N_1 coincides with the nucleotide sequence of the DNA strand (Fig. 18a).

The term "epi-chain" was coined from the Ancient Greek prefix epi-, implying features that are "on top of" DNA strands. In any DNA strand, each nucleotide belongs to many epi-chains having different orders k. The symbol "N" in the designation of DNA epi-chains corresponds to the first letter in the word "nucleotides". In the designation "$N_{k/m}$" of single-stranded DNA epi-chains, the numerator "k" in the index indicates the order of the epi-chain, and the denominator "m" indicates the numeration of the initial nucleotide of this epi-chain along the DNA strand (Fig. 18a). For example, the symbol $N_{3/2}$ refers to the epi-chain of the third order with the initial nucleotide having the number 2 in the DNA strand: 2-5-8- … (Fig. 18e).

Each DNA epi-chain of k-th order (if k = 2, 3, 4, ….) contains k times fewer nucleotides than the DNA strand and has its own arrangements of nucleobases A, T, C, and G. But unexpectedly, despite on these differences, OS-sequences of the total amounts of those n-plets, which start

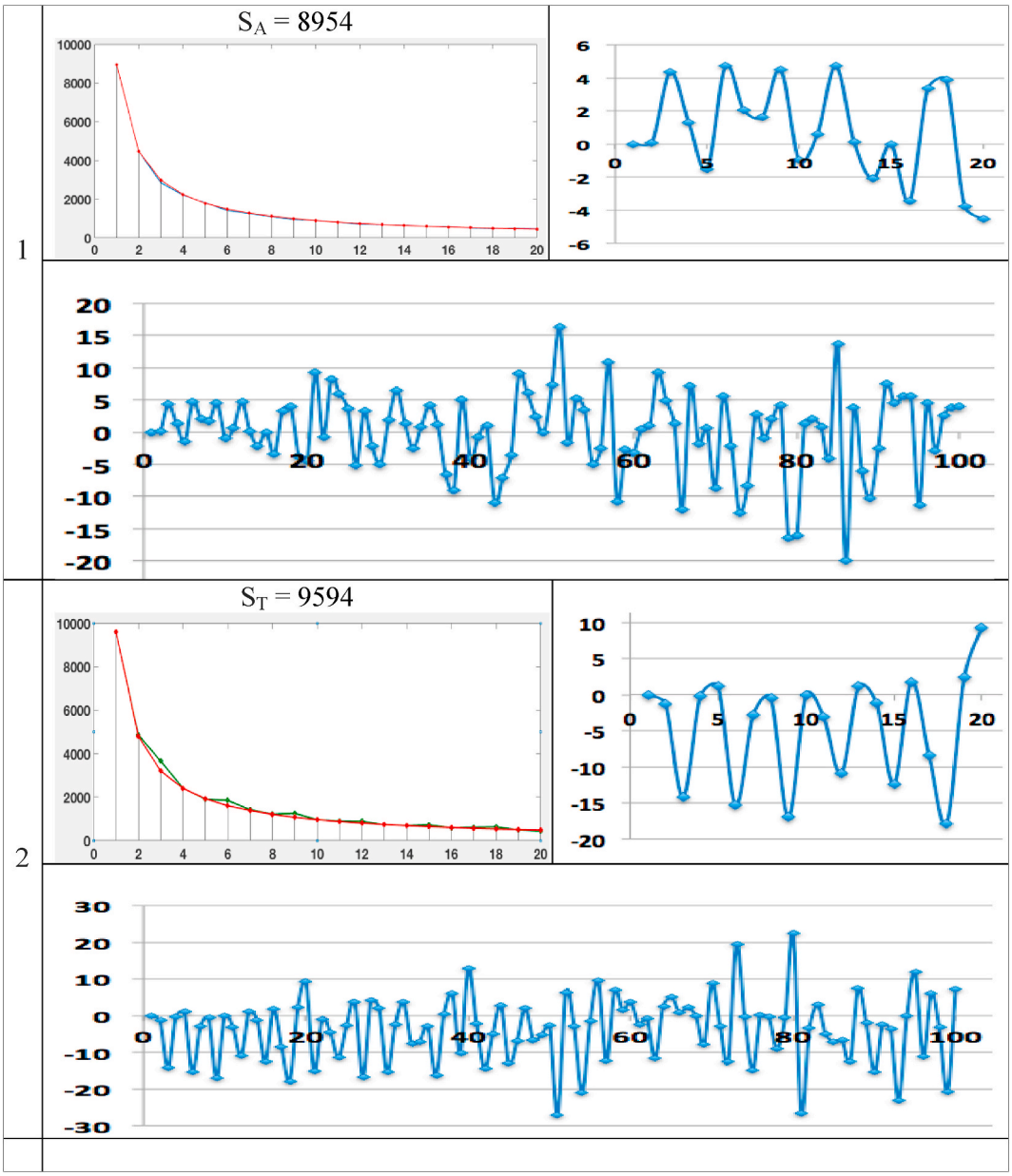


Fig. 16. The graphs for the cases of the OS-sequences of n-plets from the classes of A_1-oligomers (rows 1) and T_1-oligomers (rows 2) of the coronavirus 2 isolate Wuhan-Hu-1, complete genome, GenBank: MN908947.3, LOCUS MN908947, 29,903 bp. The abscissa axes represent the values $n = 1, 2, 3, \ldots, 20$ and $n = 1, 2, 3, \ldots, 100$. In the graphs with red lines the ordinate axes represent the set of phenomenological total amounts $\Sigma_{A,n,1}$ and $\Sigma_{T,n,1}$ of n-plets beginning with the nucleotides A and T. The graphs with blue lines show deviations of phenomenological OS-sequences $\Sigma_{A,n,1}$ and $\Sigma_{T,n,1}$ from the model hyperbolic progressions $S_A/n = 8954/n$ and $S_T/n = 9594/n$ in percentages. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

with a nucleotide A, or T, or C, or G, are modeled by very similar hyperbolic progressions as in the complete DNA strand and as in its epi-chains (at this stage of the research, the author studied OS-representations of epi-chains only in cases of epi-chains with relatively small orders k).

Figs. 19–23 explain these results in graphical forms by examples of the OS-representations of epi-chains $N_{2/1}$, $N_{3/1}$, $N_{4/1}$, $N_{10/1}$, and $N_{50/1}$ in the human chromosome №1 (the OS-representation of this complete chromosome was presented above in Figs. 1–3).

Figs. 19–23 show that in these epi-chains, which are sparse subsequences of the complete DNA sequence, the same hyperbolic rule No. 1 is fulfilled, which was formulated above for complete DNA sequences in eukaryotic and prokaryotic genomes. The rule is fulfilled in these epi-chains with the same high accuracy as in the complete DNA of the sequence.

Similar results were obtained by the author in study of epi-chains in the single-stranded DNA of other analyzed genomes (see some corresponding data in (Petoukhov, 2019a)). These results allow formulating the fourth hyperbolic (or harmonic) rule of eukaryotic and prokaryotic genomes, which is considered by the author as a candidate for the role of a universal genetic rule (it is necessary to further investigate the widest variety of genomes to verify a degree of its universality).

The fourth hyperbolic rule (about interrelations of oligomers in epi-chains of long DNA sequences):

- In any nuclear chromosome of eukaryotic genomes and in prokaryotic genomes, the hyperbolic rules №№ 1 and 2 are fulfielld not only for the complete nucleotide sequences but also for their epi-chains of the order k (where $k = 2, 3, 4, \ldots$is not too large compared to the full length of the nucleotide sequence).

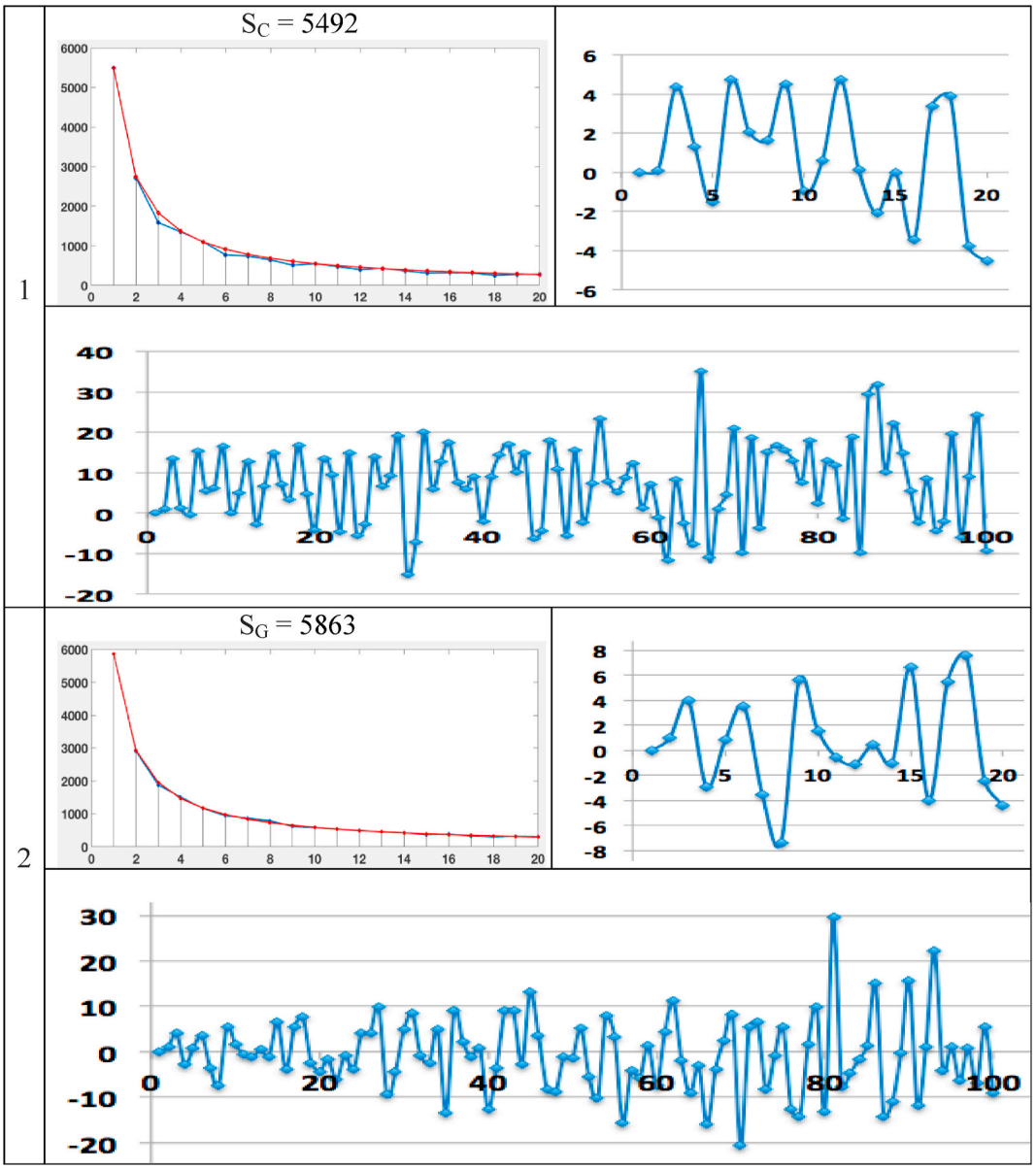


Fig. 17. The graphs for the cases of the OS-sequences of *n*-plets from the classes of C_1-oligomers (rows 1) and G_1-oligomers (rows 2) of the coronavirus 2 isolate Wuhan-Hu-1, complete genome, GenBank: MN908947.3, LOCUS MN908947, 29,903 bp. The abscissa axes represent the values $n = 1, 2, 3, \ldots, 20$ and $n = 1, 2, 3, \ldots, 100$. In the graphs with red lines the ordinate axes represent the set of phenomenological total amounts $\Sigma_{C,n,1}$ and $\Sigma_{G,n,1}$ of *n*-plets beginning with the nucleotides C and G correspondingly. The graphs with blue lines show deviations of phenomenological OS-sequences $\Sigma_{C,n,1}$ and $\Sigma_{G,n,1}$ from the model hyperbolic progressions $S_C/n = 5492/n$ and $S_G/n = 5863/n$ in percentages. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

The numeric data, which determine the graphs in Figs. 19–23, are represented in the preprint [Petoukhov, 2020e].

Fig. 24 shows that normalized values of amounts S_A, S_T, S_C, and S_G of each of nucleotides A, T, C, and G are practically identical in all considered epi-chains of the human chromosome №1, that is, they are independent of the epi-chain order.

10. The representation of the DNA alphabets by their binary-oppositional traits in matrix genetics

The described phenomenological rules in the genetic systems were discovered as a result of the development of a matrix-algebraic approach to modeling the genetic coding system. Some features of this author's model approach are presented below.

Science does not know why the DNA alphabet of nucleotides consists of only 4 relatively simple molecules A, T, C, and G. But science knows that this alphabet is endowed with a system of binary-opposition traits (or indicators):

- 1) in the double helix of DNA, there are two complementary pairs of nucleotides: the nucleotides C and G of the first pair are connected by three hydrogen bonds, and the nucleotides A and T of the second pair by two hydrogen bonds. Given these oppositional indicators, one can represent C = G = 1 and A = T = 0;
- 2) the two nucleotides are keto molecules (G and T), and the other two are amino molecules (A and C). Given these oppositional indicators, one can represent G = T = 1 and A = C = 0.

Taking this into account, it is convenient to represent DNA alphabets of 4 nucleotides, 16 doublets and 64 triplets in the form of square tables,

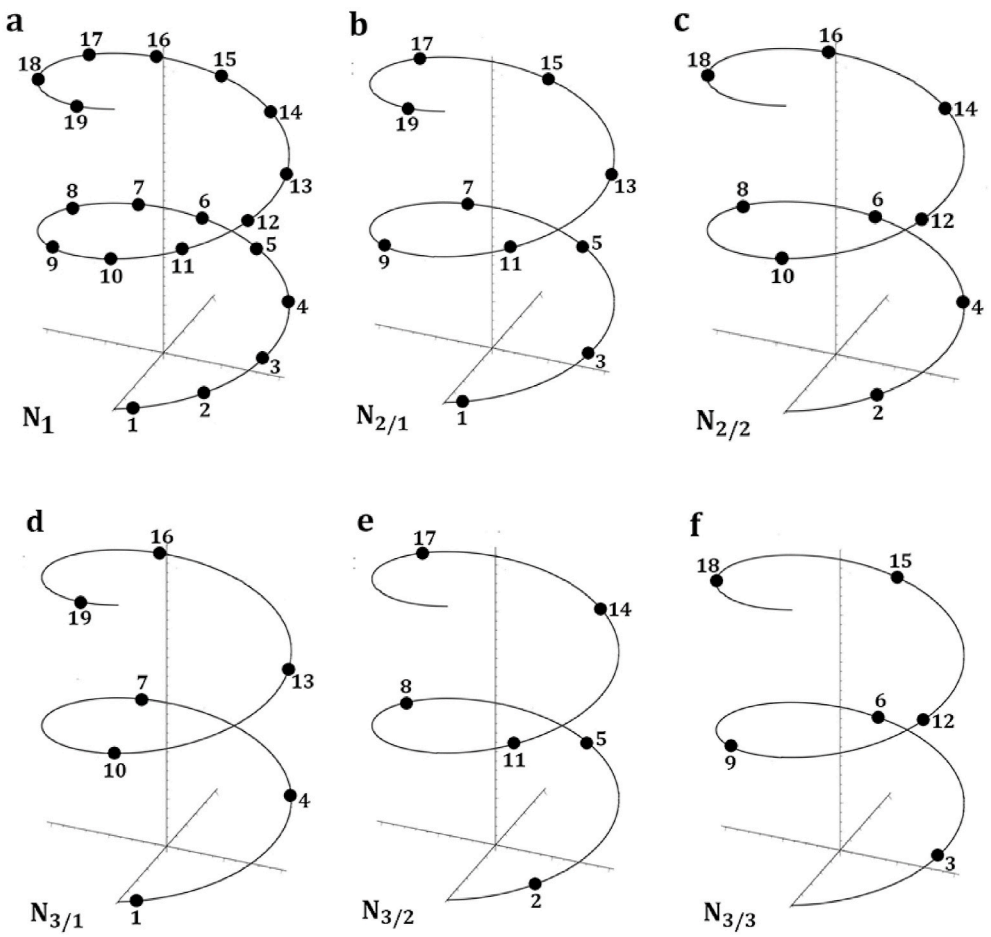


Fig. 18. A schematic representation of a single-stranded DNA and its initial epi-chains of nucleotides, denoted by black circles. **a**, a sequence N_1 of numerated nucleotides of the DNA strand. **b**, an epi-chain of the second order $N_{2/1}$ beginning with nucleotide number 1. **c**, an epi-chain of the second order $N_{2/2}$ beginning with nucleotide number 2. **d**, an epi-chain of the third order $N_{3/1}$ beginning with nucleotide number 1. **e**, an epi-chain of the third order $N_{3/2}$ beginning with nucleotide number 2. **f**, an epi-chain of the third order $N_{3/3}$ beginning with nucleotide number 3.

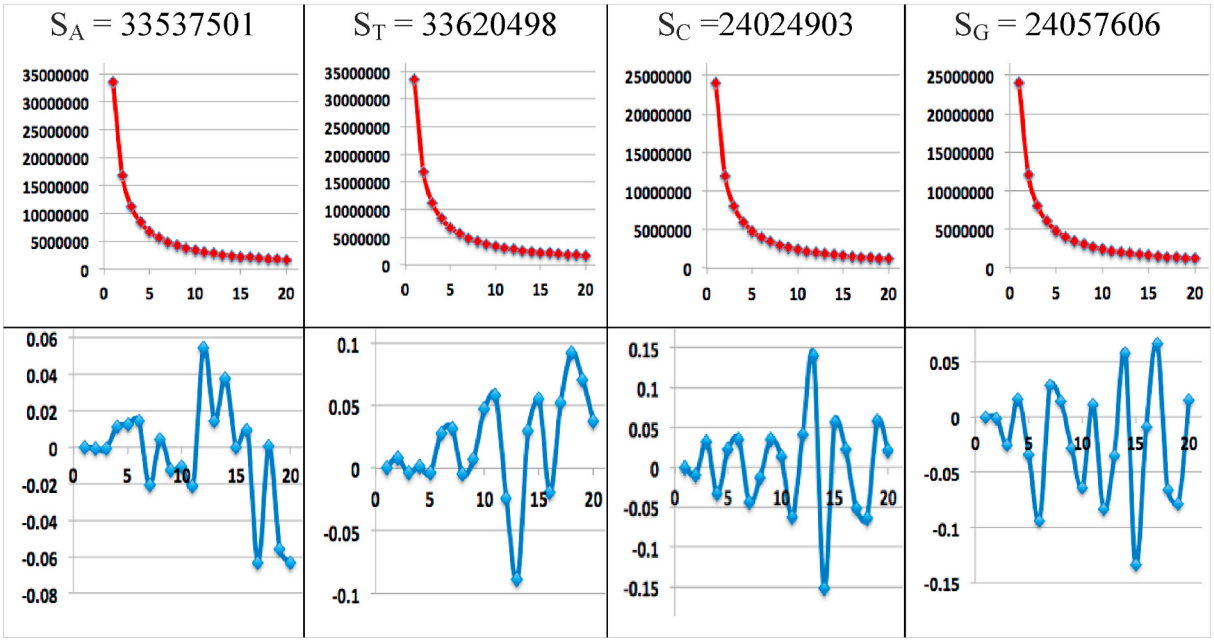


Fig. 19. The results of the analysis - by the oligomer sums method – the nucleotide sequence of the epi-chain of the second order $N_{2/1}$ (Fig. 18b), which consists of nucleotides with serial numerations 1-3-5-7-9- … in the DNA sequence of the human chromosome N° 1. All abscissa axes show the values $n = 1, 2, \ldots, 20$. The top row demonstrates that the model hyperbolic progressions S_A/n, S_T/n, S_C/n, S_G/n (red lines) almost completely cover the OS-sequences of real total amounts of those n-plets, which start with a nucleotide A, or T, or C, or G in this epi-chain correspondingly (the ordinate axes show appropriate amounts). The bottom row shows in percent slight alternating deviations of phenomenological values of the OS-sequences from model values. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

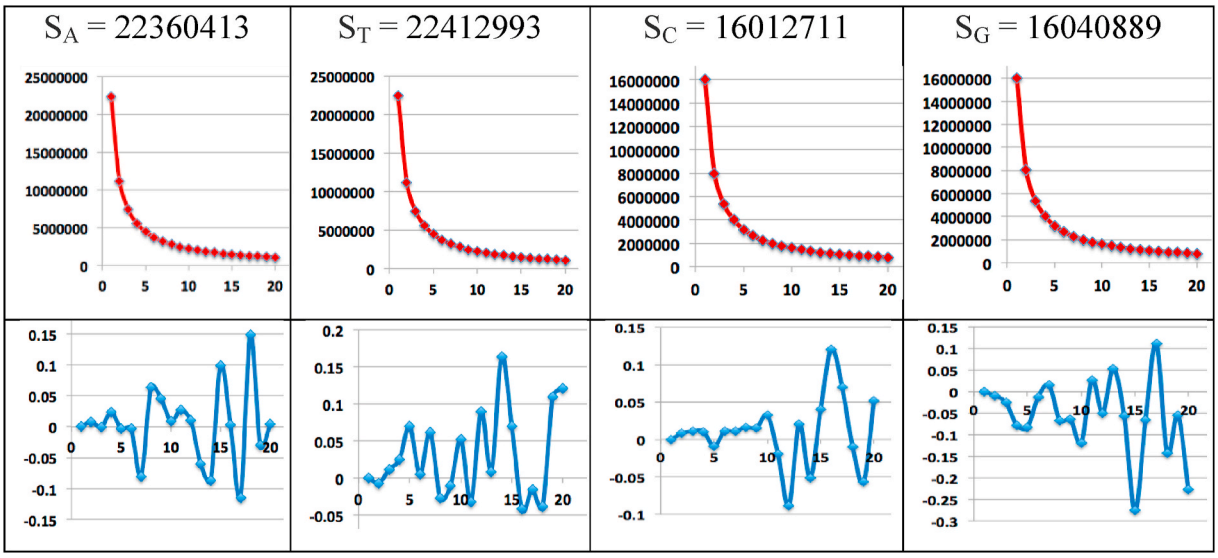


Fig. 20. The results of the analysis - by the oligomer sums method – the nucleotide sequence of the epi-chain of the third order $N_{3/1}$ (Fig. 18d), which consists of nucleotides with serial numerations 1-4-7-10-13- … in the DNA sequence of the human chromosome № 1. The top row demonstrates that the model hyperbolic progressions S_A/n, S_T/n, S_C/n, S_G/n (red lines) almost completely cover the OS-sequences of real total amounts of those *n*-plets, which start with a nucleotide A, or T, or C, or G in this epi-chain correspondingly. The bottom row shows in percent slight alternating deviations of real values of the OS-sequences from model values. All denotations are the same as in Fig. 19. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

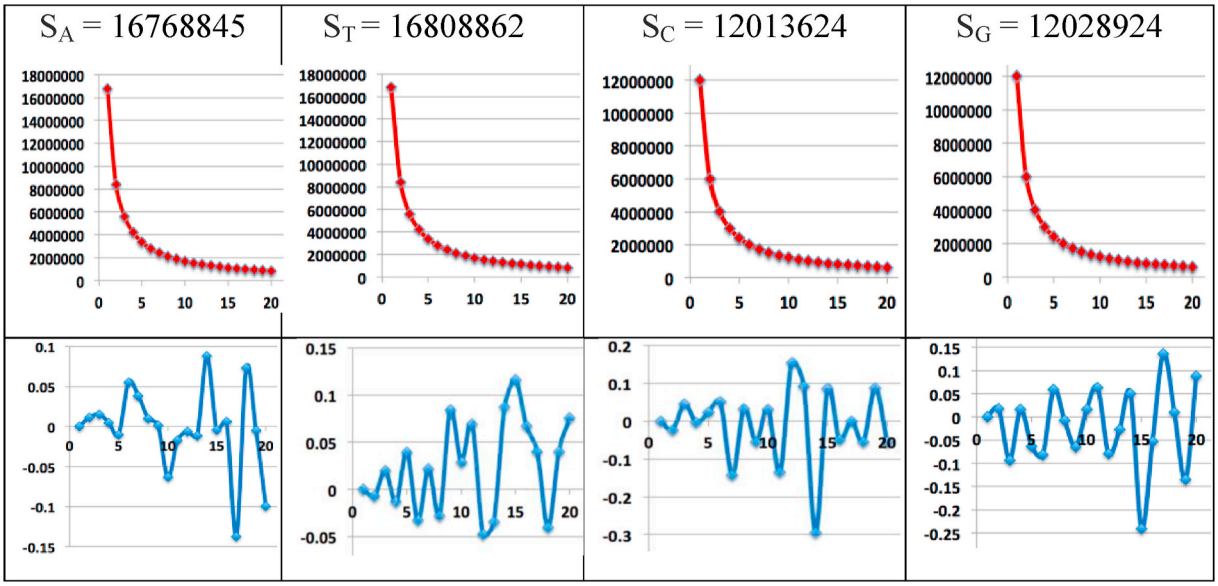


Fig. 21. The results of the analysis - by the oligomer sums method – the nucleotide sequence of the epi-chain of the 4th order $N_{4/1}$, which consists of nucleotides with serial numerations 1-5-9-13- … in the DNA sequence of the human chromosome № 1. The top row demonstrates that the model hyperbolic progressions S_A/n, S_T/n, S_C/n, S_G/n (red lines) almost completely cover the OS-sequences of real total amounts of those *n*-plets, which start with a nucleotide A, or T, or C, or G in this epi-chain correspondingly. The bottom row shows in percent slight alternating deviations of real values of the OS-sequences from model values. All denotations are the same as in Fig. 19. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

the columns of which are numbered in accordance with oppositional indicators "3 or 2 hydrogen bonds" (C = G = 1, A = T = 0), and the rows in accordance with oppositional indicators "amino or keto" (C = A = 0, G = T = 1). In such tables, all letters, doublets, and triplets automatically occupy their strictly individual places (Fig. 25).

These three tables (Fig. 25) are not only simple tables but they are members of the tensor family of matrices: the second and the third tensor (Kronecker) powers of the matrix [G, T; C, A] generate similar arrangements of 16 doublets and 64 triplets inside matrices [G, T; C, A]$^{(2)}$ and [G, T; C, A]$^{(3)}$ as shown in Fig. 25. One can note here that the classes of G_1-, T_1-, C_1-, and A_1-oligomers, analyzed in the previous Section as related to the hyperbolic rules, are connected by a special manner with the tensor family of the matrices [G, T; C, A]$^{(n)}$ where the symbol (*n*) refers to an appropriate tensor power. More precisely, in Fig. 25, each of (2*2)-quadrants of the matrix [G, T; C, A]$^{(2)}$ contains a complete set of 4 doublets, which start with one of nucleotides G, T, C, and A; each of (2^2*2^2)-quadrants of the matrix [G, T; C, A]$^{(3)}$ contains a complete set of 16 triplets, which start with one of the nucleotides G, T, C, and A. In general, each of $(2^{n-1}*2^{n-1})$-quadrants of the matrix [G, T; C, A]$^{(n)}$ contains a complete set of 4^{n-1} *n*-plets, which start with one of the nucleotides G, T, C, and A.

The genetic code is called a "degenerate code" because 64 triplets

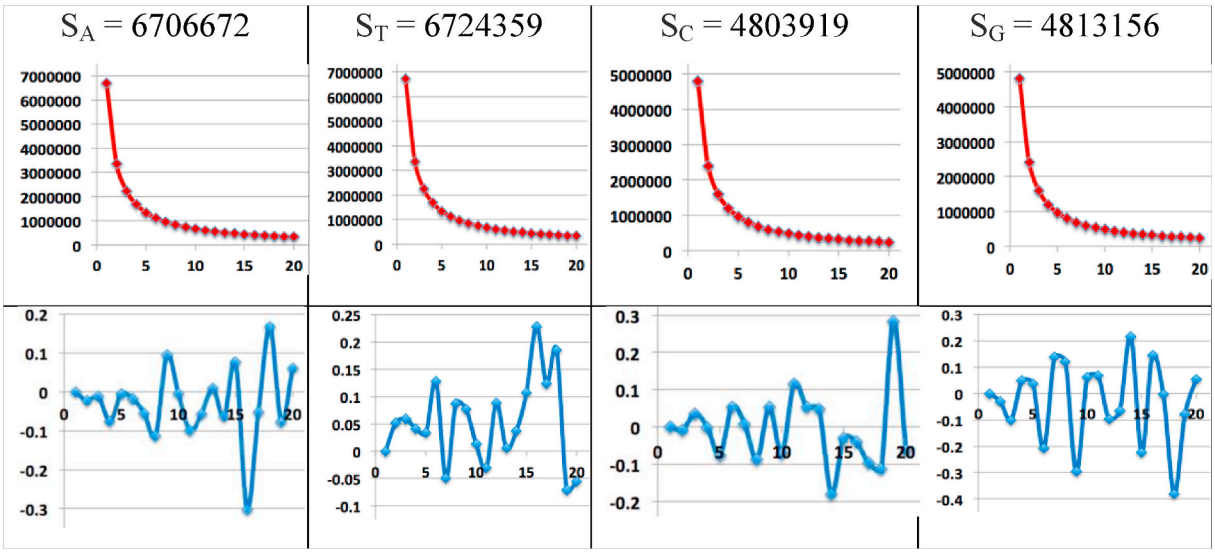


Fig. 22. The results of the analysis - by the oligomer sums method – the nucleotide sequence of the epi-chain of the 10th order $N_{10/1}$, which consists of nucleotides with serial numerations 1-11-21-31-41- … in the DNA sequence of the human chromosome №1. The top row demonstrates that the model hyperbolic progressions S_A/n, S_T/n, S_C/n, S_G/n (red lines) almost completely cover the OS-sequences of real total amounts of those *n*-plets, which start with a nucleotide A, or T, or C, or G in this epi-chain correspondingly. The bottom row shows in percent slight alternating deviations of real values of the OS-sequences from model values. All denotations are the same as in Fig. 19. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

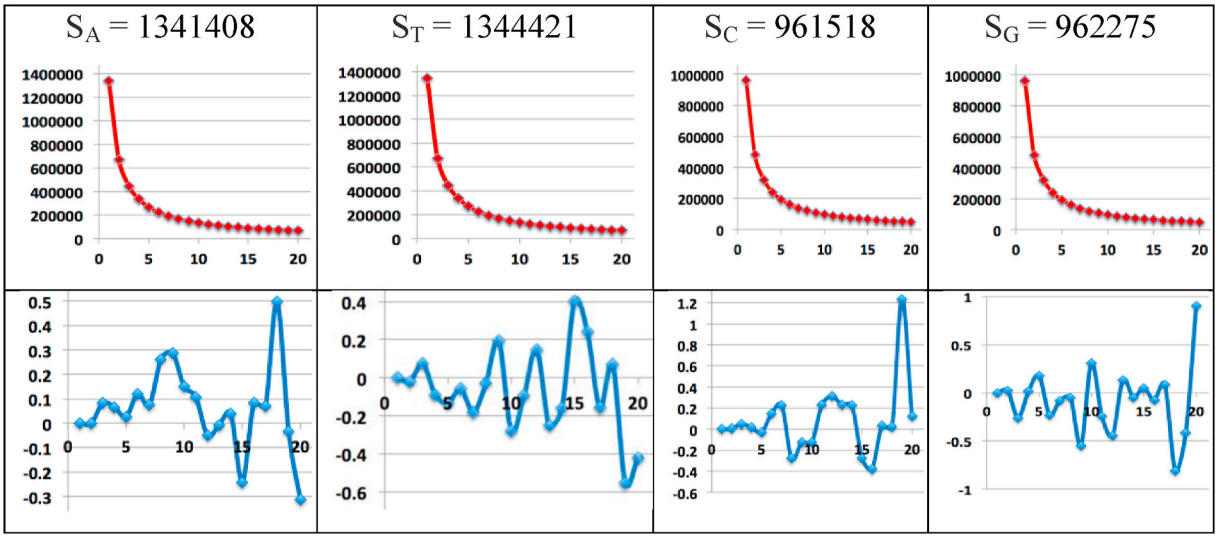


Fig. 23. The results of the analysis - by the oligomer sums method – the nucleotide sequence of the epi-chain of the 50th order $N_{50/1}$, which consists of nucleotides with serial numerations 1-51-101-151-201- … in the DNA sequence of the human chromosome №1. The top row demonstrates that the model hyperbolic progressions S_A/n, S_T/n, S_C/n, S_G/n (red lines) almost completely cover the OS-sequences of real total amounts of those *n*-plets, which start with a nucleotide A, or T, or C, or G in this epi-chain correspondingly. The bottom row shows in percent slight alternating deviations of real values of the OS-sequences from model values. All denotations are the same as in Fig. 19. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Epi-ch.	$S_A/(S_A+S_T+S_C+S_G)$	$S_T/(S_A+S_T+S_C+S_G)$	$S_C/(S_A+S_T+S_C+S_G)$	$S_G/(S_A+S_T+S_C+S_G)$
$N_{1/1}$	0.2910	0.2918	0.2085	0.2087
$N_{2/1}$	0.2910	0.2917	0.2085	0.2088
$N_{3/1}$	0.2910	0.2917	0.2084	0.2088
$N_{4/1}$	0.2910	0.2917	0.2085	0.2088
$N_{10/1}$	0.2910	0.2918	0.2084	0.2088
$N_{50/1}$	0.2910	0.291	0.2086	0.2088

Fig. 24. The normalized values $S_P/(S_A + S_T + S_C + S_G)$ of amounts S_A, S_T, S_C, and S_G of each nucleotide A, T, C, and G are practically identical in all considered epi-chains of different orders 1, 2, 3, 10, and 50 in the human chromosome №1, that is, they are independent of the epi-chain orders.

	1	0
1	G	T
0	C	A

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

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

Fig. 25. The square tables of DNA-alphabets of 4 nucleotides, 16 doublets, and 64 triplets with a strict arrangement of all components. Each of the tables is constructed in line with the principle of binary numeration of its column and rows on the basis of binary-oppositional indicators of nucleobases G, T, C, and A (see explanations in the text).

encode 20 amino acids and stop-codons so that several triplets can encode each amino acid at once, and each triplet necessarily encodes only a single amino acid or a stop-codon. The (8*8)-matrix of 64 triplets (Fig. 25) was built formally without any mention of amino acids and stop-codons. Nothing data preliminary exist on a possible correspondence between triplets and amino acids. How can these 20 amino acids and stop-codons be located in this matrix of 64 triplets? There are a huge number of possible options for the location and repetition of separate amino acids and stop-codons in 64 cells of this matrix. More precisely, the number of these options is much more than 10^{100} (for comparison, the entire time of the Universe existence is estimated in modern physics at 10^{17} s). But Nature uses - from this huge number of options - only a very specific repetition and arrangement of separate amino acids and stop-codons, the analysis of which is important for revealing the structural organization of the information foundations of living matter.

	111	110	101	100	011	010	001	000
111	**PRO** CCC	**PRO** CCA	**HIS** CAC	**GLN** CAA	**THR** ACC	**THR** ACA	**ASN** AAC	**LYS** AAA
110	**PRO** CCG	**PRO** CCT	**GLN** CAG	**HIS** CAT	**THR** ACG	**THR** ACT	**LYS** AAG	**ASN** AAT
101	**ARG** CGC	**ARG** CGA	**LEU** CTC	**LEU** CTA	**SER** AGC	**STOP** AGA	**ILE** ATC	**MET** ATA
100	**ARG** CGG	**ARG** CGT	**LEU** CTG	**LEU** CTT	**STOP** AGG	**SER** AGT	**MET** ATG	**ILE** ATT
011	**ALA** GCC	**ALA** GCA	**ASP** GAC	**GLU** GAA	**SER** TCC	**SER** TCA	**TYR** TAC	**STOP** TAA
010	**ALA** GCG	**ALA** GCT	**GLU** GAG	**ASP** GAT	**SER** TCG	**SER** TCT	**STOP** TAG	**TYR** TAT
001	**GLY** GGC	**GLY** GGA	**VAL** GTC	**VAL** GTA	**CYS** TGC	**TRP** TGA	**PHE** TTC	**LEU** TTA
000	**GLY** GGG	**GLY** GGT	**VAL** GTG	**VAL** GTT	**TRP** TGG	**CYS** TGT	**LEU** TTG	**PHE** TTT

Fig. 26. The location and repetition of 20 amino acids and 4 stop-codons (denoted by bold) in the matrix of 64 triplets [C, A; G, T][(3)] (Fig. 25) for the Vertebrate Mitochondrial Code. The symbol "stop" refers to stop-codons.

Fig. 26 shows the real repetition and location of amino acids and stop-codons in the case of the Vertebrate Mitochondrial Code, which is the most symmetrical among known dialects on the genetic code. This genetic code is called the most ancient and "ideal" in genetics (Frank--Kamenetskii, 1988) (other dialects of the genetic code have small differences from this one, which is considered in the theory of symmetries as the basis from the structural point of view).

The location and repetition of all amino acids and stop-codons in the matrix of 64 triplets have the following algebraic feature (Fig. 26):

- Each of sixteen (2*2)-sub-quadrants, forming this genetic matrix and denoted by bold frames, is bisymmetrical: each of its both diagonals contains an identical kind of amino acids or stop-codon.

Such bisymmetric (2*2)-matrices [a, b; b, a] are well known in algebra as matrix representations of two-dimensional hypercomplex numbers called hyperbolic numbers: $a+bj$ where "a" and "b" are real numbers, and the imaginary unit j satisfies $j^2 = +1$ (Kantor, Solodovnikov, 1989). Hyperbolic numbers are used in physics and mathematics and they have also synonymical names: "split-complex numbers", "double numbers" and "perplex numbers". The collection of all hyperbolic numbers forms algebra over the field of real numbers (Harkin, Harkin, 2004; Kantor, Solodovnikov, 1989). The algebra is not a division algebra or field since it contains zero divisors. Addition and multiplication of hyperbolic numbers are defined by the expressions (10.1):

$$\underline{a}*1+b*j \quad \Leftrightarrow \quad \begin{vmatrix} a, b \\ b, a \end{vmatrix} = a \begin{vmatrix} 1, 0 \\ 0, 1 \end{vmatrix} + b \begin{vmatrix} 0, 1 \\ 1, 0 \end{vmatrix}$$

Fig. 27. The decomposition of the bisymmetric matrix [a, b; b, a] into two sparse matrices representing real and imaginary units of hyperbolic numbers correspondingly.

$$(x + jy)+(u + jv)=(x + u)+j(y + v);\ (x + jy)\ (u + jv)=(xu + yv)+j(xv + yu) \quad (10.1)$$

This multiplication is commutative, associative, and distributes over addition.

Hyperbolic numbers have the matrix form of their representation in a form of bisymmetric matrix [*a*, *b*; *b*, *a*]. Fig. 27 shows the decomposition of such matrix into two sparse matrices, the first of which is the matrix representation of the real unit and the second one is the matrix representation of the imaginary unit j.

$$a*1 + b*j \Leftrightarrow \begin{vmatrix} a, & b \\ b & a \end{vmatrix} = a\begin{vmatrix} 1, & 0 \\ 0 & 1 \end{vmatrix} + b\begin{vmatrix} 0, & 1 \\ 1 & 0 \end{vmatrix}$$

Regarding the hyperbolas from the hyperbolic rules formulated above (Figs. 1 and 2, etc.), it can be noted the following:

- 1) the transformation of one point of the hyperbola to another point is determined by the transformation of the hyperbolic rotation, in which the hyperbole glides along with itself. Such a transformation is determined by a special bisymmetric matrix [a, b; b, a] representing a special form of hyperbolic numbers (the hyperbolic rotations are known in the special theory of relativity under the name of the Lorentz transformation);
- 2) in the hyperbolic sequence S/*n*, each its member can be considered as a point at the hyperbolic plane and interpreted as a corresponding hyperbolic number S/*n* + *n*j having its matrix representation [S/*n*, *n*; *n*, S/*n*].

If each amino acid and stop-codon is represented by some characteristic parameter (for example, the number of carbon atoms in these organic formations or numbers of protons in its molecular structure, etc.), then a numerical (8*8)-matrix arises (Fig. 28) with bisymmetric (2*2)-sub-quadrants representing hyperbolic numbers *a*+*b*j. In other words, this phenomenologic arrangement of amino acids and stop-codons in the matrix of 64 triplets is associated to the multiblock union of matrix presentations of 16 two-dimensional hyperbolic numbers.

An additional confirmation of the relationship of DNA alphabets with bisymmetric matrices, representing hyperbolic numbers, is given by returning to the tensor family of genetic matrices (Fig. 26), which were constructed very formally on the basis of binary-oppositional attributes of the four nucleotides A, G, C, and T. Taking into account that two nucleotides are purines (A and G) and the other two are pyrimidines (C and T), one can replace in these matrices the nucleotides A and G with the traditional symbol of purines R, and the nucleotides C and T with the traditional symbol of pyrimidines Y. In such representation of the considered genetic matrices, a tensor family of bisymmetric matrices arises, whose entries are combinations of purines R and pyrimidines Y (Fig. 29). These bisymmetric matrices represent 2-dimensional, 4-dimensional, and 8-dimensional hyperbolic numbers (10.2):

$$R + j_1Y;\ RR + j_1RY + j_2YR + j_3YY;\ RRR + j_1RRY + j_2RYR + j_3RYY + j_4YRR + j_5YRY + j_6YYR + j_7YYY \quad (10.2)$$

where all j_k are imaginary units of these hyperbolic numbers with the property $j_k^2 = +1$. The algebras of these 2-, 4-, and 8-dimensional hyperbolic numbers, which are called also as hyperbolic matrions, have corresponding multiplication tables for their basis units (Petoukhov, 2008; Petoukhov, He, 2010).

It should be noted two following aspects. Firstly, the multiplication table of the algebra of 2^n-dimensional hyperbolic numbers has a fractal-like character since it contains multiplication tables of the algebras of 2^{n-1}-, 2^{n-2}-, …, 2-dimensional hyperbolic numbers (Fig. 29 shows an example of this). 2^n-dimensional hyperbolic numbers can be generated by the tensor power (*n*) of the 2-dimensional bisymmetric matrix [R, Y; Y, R]$^{(n)}$. Secondly, coordinates of hyperbolic numbers in (25) have some relation to analysis of genomes by the oligomer sums method and by its modifications. For example, in 2-dimensional hyperbolic number R + j_1Y, the coordinate R is equal to total number of purines A and G, and the coordinate Y is equal to total number of pyrimidines C and T in the considered DNA sequence. The coordinate RY in the expression (10.2) is equal to the total amounts of doublets, which start with purines A and G and end with pyrimidines C and T, and so on.

The presented algebraic features of the genetic coding system supplement the following statement in a number of author's publications (Petoukhov, 2008, 2016, 2018a; Petoukhov, He, 2010, etc.). The genetic code is not just a mapping of one set of elements to other sets of elements by type, for example, of a phone book in which phone numbers encode names of people. But the genetic code is inherently an algebraic code, akin to a certain degree to those algebraic codes that are used in modern communication theory for noise-immune transmission of information. Algebraic features of the genetic code are related to the noise-immune properties of this code and the whole genetic system.

One can explain the meaning and possibilities of algebraic codes by the example of transmitting a photograph of the Martian surface from Mars to Earth using electromagnetic signals. On the way to the Earth, these signals travel millions of kilometers of interference and arrive at

5	5	6	5	4	4	4	6
5	5	5	6	4	4	6	4
6	6	6	6	3	0	6	5
6	6	6	6	0	3	5	6
3	3	4	5	3	3	9	0
3	3	5	4	3	3	0	9
2	2	5	5	3	11	9	6
2	2	5	5	11	3	6	9

Fig. 28. The numeric analog of the symbolic (8*8)-matrix of amino acids and stop-codons from Fig. 26 for the case of representing each of amino acids by numbers of its carbon atoms (stop-codons are conditionally represented by zero).

R	Y
Y	R

RR	RY	YR	YY
RY	RR	YY	YR
YR	YY	RR	RY
YY	YR	RY	RR

RRR	RRY	RYR	RYY	YRR	YRY	YYR	YYY
RRY	RRR	RYY	RYR	YRY	YRR	YYY	YYR
RYR	RYY	RRR	RRY	YYR	YYY	YRR	YRY
RYY	RYR	RRY	RRR	YYY	YYR	YRY	YRR
YRR	YRY	YYR	YYY	RRR	RRY	RYR	RYY
YRY	YRR	YYY	YYR	RRY	RRR	RYY	RYR
YYR	YYY	YRR	YRY	RYR	RYY	RRR	RRY
YYY	YYR	YRY	YRR	RYY	RYR	RRY	RRR

	1	j_1	j_2	j_3	j_4	j_5	j_6	j_7
1	1	j_1	j_2	j_3	j_4	j_5	j_6	j_7
j_1	j_1	1	j_3	j_2	j_5	j_4	j_7	j_6
j_2	j_2	j_3	1	j_1	j_6	j_7	j_4	j_5
j_3	j_3	j_2	j_1	1	j_7	j_6	j_5	j_4
j_4	j_4	j_5	j_6	j_7	1	j_1	j_2	j_3
j_5	j_5	j_4	j_7	j_6	j_1	1	j_3	j_2
j_6	j_6	j_7	j_4	j_5	j_2	j_3	1	j_1
j_7	j_7	j_6	j_5	j_4	j_3	j_2	j_1	1

Fig. 29. Top: the tensor family of bisymmetric genetic matrices [R, Y; Y, R]$^{(n)}$ received from the tensor family of matrices in Fig. 26 by the replacement there of purines A and G by the symbol R, and of pyrimidines C and T by the symbol Y. **Bottom**: the multiplication table of basis units 1 and j_k of the algebra of 8-dimensional hyperbolic numbers, which contains multiplication tables, marked by bold lines, of the algebras of 2- and 4-dimensional hyperbolic numbers (Petoukhov, 2008; Petoukhov, He, 2010).

the Earth in a very weakened and distorted form. But, magically, based on these mutilated signals on Earth, a high-quality photograph of the surface of Mars is recreated. The secret of this magic lies in the fact that from Mars not the information signals about this photo are sent, but algebraically encoded versions of these signals that are quite other. At receivers on Earth, these algebraically encoded signals are algebraically decoded into signals, which recreate the original photographic image of the surface of Mars. It should be emphasized that algebraic coding of information in the theory of noise-immune communication actively uses the mathematical apparatus of matrices, which is also used in quantum informatics and quantum mechanics as matrix operators. The author's works are aimed at studying algebraic properties of the genetic coding system for revealing hidden information rules algebraically encoded in the molecular genetic system. This article is part of a set of long-term author's studies of the genetic system by the methods of matrix analysis and modeling combined under the general name "matrix genetics" (Petoukhov, 2008, 2011, 2016, 2017, 2019b,c; Petoukhov, He, 2010; Petoukhov, Petukhova, 2017a,b).

Let's continue the presentation of confirmational data on the existence of hyperbolic (or harmonic) rules in the cooperative oligomeric organization of the eukaryotic and prokaryotic genomes.

11. The quantum-information model of the oligomer cooperative organization in genomes and its confirmed predictions

The Section is devoted to the connections of the described phenomenological hyperbolic (harmonic) rules in genomes with the concepts and mathematical formalisms of quantum informatics.

One of the creators of quantum mechanics P.Jordan in his work on quantum biology claimed that life's missing laws were the rules of chance and probability of the quantum world (Jordan, 1932; McFadden, Al-Khalili, 2018). From the standpoint of Jordan's statement, the study of probabilities or frequencies of *n*-plets (monoplets, doublets, triplets, etc.) in long DNA sequences is important for discovering hidden biological laws and for developing quantum biology. The phenomenological hyperbolic rules about the total amounts of certain oligomers in the genomes described above allow us to study their connection with the probability rules of these groups of oligomers in the genomes. Let us explain this.

Till now we considered the total amounts $\Sigma_{N,n,1}$ of certain *n*-plets, which start with the first nucleotide N (A, T, C, or G), and we discovered that, in different genomes, these amounts correspond to hyperbolic OS-sequences S_N/n with a high accuracy, where S_N refers to the total number of the nucleotide N. The whole sequence of all nucleotides in a long single-stranded DNA can be considered as a sequence of oligomers of a certain length *n*, whose amount is equal to S/n. Each such oligomer starts with one of four nucleotides A, T, C, or G. Therefore the total amount S/n of consecutive oligomers of length *n* in the analyzed DNA sequence is the sum of all oligomers of length *n* starting with A, or T, or C, or G:

$$S/n = \Sigma_{A,n,1} + \Sigma_{T,n,1} + \Sigma_{C,n,1} + \Sigma_{G,n,1} \quad (11.1)$$

The collective probability (percentage, or frequency) $P_n(N_1)$ of all $\Sigma_{N,n,1}$ *n*-plets starting with the nucleotide N, relative to the amount S/n (11.1), is determined by the expression (11.2):

$$P_n(N_1) = \Sigma_{N,n,1}/(S/n) \approx (S_N/n)/(S/n) = S_N/S = P(N) \quad (11.2)$$

The expression (11.2) shows that the collective probability $P_n(N_1)$ is independent of *n* and is approximately equal to the probability (frequency) $P(N) = S_N/S$ of the nucleotide N in the genomic sequence having S nucleotides.

For example, the human chromosome №1, which was considered above (Figs. 1–3), has the total amount of nucleotides $S = S_A + S_T + S_C + S_G$ = 67,070,277 + 67,244,164 + 48,055,043 + 48,111,528 = 230,481,012. The probability P(A) of the nucleotide A is equal to S_A/S = 67,070,277/230,481,012 ≈ 0.2910. From the data in Fig. 3, one can verify that, in this chromosome, the collective probabilities $P_n(A_1)$ of total amounts of *n*-plets ($n = 2, 3, \ldots, 20$) starting with the nucleotide A are also equal to this value P(A) = 0.2910 with a high level of accuracy independently of *n*. A similar situation holds with respect to the nucleotides T, C, and G.

It is also useful to note the opposite: if, for a genome, the phenomenological probabilities of *n*-plets $P_n(N_1)$ (where $n = 1, 2, 3, \ldots$) are initially known, and their compliance with the rule - of type $P(N) \approx P_n(N_1)$ - of approximate equality of collective probability of *n*-plets is also known, then connection (11.2) allows us to construct a hyperbolic OS-sequence of the sums $\Sigma_{N,n,1}$ of *n*-plets (11.3):

$$\Sigma_{N,n,1} = P_n(N_1)*S/n \quad (11.3a)$$

This is noted here because the author previously discovered and published (Petoukhov, 2018b) the rules of the approximate equality of the collective probabilities of *n*-plets for $n = 1, 2, 3, \ldots$ Given the expressions (11.2) and (11.3), the hyperbolic rules of the OS-sequences and these rules for the approximate equality of the collective probabilities of *n*-plets are equivalent. Both of them reflect in different languages the oligomer cooperative organization of genomes. This is useful to note because the author has published an effective mathematical model for the rules of collective probability, which is obviously applicable also to the above formulated hyperbolic rule № 1 (Petoukhov, 2018b; Petoukhov et al., 2019).

One should emphasize the following important aspect of the OS-representations of genomic sequences. Each nucleotide of a DNA sequence is a participant of those sets of its different *n*-plets (doublets, triplets, etc.), whose total amounts are members of OS-sequences of this DNA; in other words, each DNA nucleotide makes its small contribution immediately to many members of the OS-sequences. Figuratively speaking, each DNA nucleotide is "smeared" (or distributed) over many members of the DNA OS-sequence (this "smearing" over many members of the OS-sequence is also true for each DNA doublet, triplet, etc.). Correspondingly, OS-sequences reflect a sort of an interrelation over all *n*-plets in DNA sequences. Or, in other words, the oligomer sums method represents any long nucleotide sequence as a multi-partite (or many-body) system having a cooperative state regarding many its interrelated oligomers of different lengths $n = 1, 2, 3, \ldots$

This has some analogies with the well-known problem of multi-partite entanglement in quantum informatics described, for example, in (Walter et al., 2017; Horodecki et al., 2009; Gühne, Tóth, 2009; Amico et al., 2008).

Quantum entanglement is the physical phenomenon that occurs when a pair or group of particles is generated, interact, or share spatial proximity in a way such that the quantum state of each particle of the pair or group cannot be described independently of the state of the others. In quantum informatics, entangled states play very important roles. The study and use of entangled states are one of the main problems of quantum computing: "… *entanglement is a key element in effects such as quantum teleportation, fast quantum algorithms, and quantum error-correction. It is, in short, a resource of great utility in quantum computation and quantum information. … entangled states play a crucial role in quantum computation and quantum information*" (Nielsen, Chuang, 2010, p. XXIII and p. 96).

Quantum systems with many degrees of freedom are ubiquitous in nature, particularly in the context of condensed matter theory. "*It is hence not surprising that important classes of states, such as ground states of local Hamiltonians, are multi-partite entangled states. … Recent years have seen an enormous increase in interest at the intersection of quantum information and condensed matter theory that stems from the insight that notions of entanglement are crucial in the understanding of quantum phases of matter …. Another family of quantum many-body states that can be efficiently described is the classes of bosonic and fermionic Gaussian states. They both*

arise naturally in the context of quantum many-body models in condensed matter physics, but their bosonic variant is also highly useful in quantum optics when it comes to describing systems constituted of several quantum modes of light …. Relatedly, multi-partite entangled states serve as resources to a number of important protocols in quantum information theory in which more than two parties come together. A prominent example of such a multi-party quantum protocol is quantum secret sharing, in which a message is distributed to several parties in such a way that no subset is able to read the message, but the entire collection of parties is. …. Multi-partite entanglement does not only facilitate processing or transmission of information but also allow for applications in metrology" (Walter et al., 2017, pp. 15, 18, 20, 23). The entanglement refers to the nonlocal properties of quantum states that cannot be explained classically.

Distinguish entanglement of distinguishable and indistinguishable (identical) particles. The state of system K of distinguishable particles in a pure state is determined by the state vector $|\psi>$ in the Hilbert space **H**, which is the tensor product of the subspaces corresponding to each particle:

$$H = H_1 \otimes H_2 \otimes \ldots \otimes H_K \qquad (11.3b)$$

If the particles are not entangled, then the state of the system is defined as the tensor product of the state vectors $|\psi^{(i)}>$ of the subsystems:

$$|\psi> \; = \; |\psi^{(1)}> \otimes \psi^{(2)}> \otimes \ldots \otimes |\psi^{(K)}> \qquad (11.4)$$

If the vector cannot be expressed in this form (11.4), then they say that the particles are quantum entangled.

The tensor product gives a way of putting separate vector spaces together to form larger vector spaces and it is one of the basis instruments in quantum informatics. The following quotation speaks about the meaning of the tensor product: "*This construction is crucial to understanding the quantum mechanics of multiparticle systems*" (Nielsen, Chuang, 2010, p. 71) But above Section 3 described that the DNA alphabets of 4 nucleotides, 16 doublets, 64 triplets, …, 4^n *n*-plets, which have binary-oppositional systems of molecular traits, are interrelated by the tensor product of matrices representing them: these genetic matrices of DNA alphabets are members of a single tensor family $[G, T; C, A]^{(n)}$ (Fig. 25). This fact is one of the arguments in favor of the adequacy of the quantum-information approach to the study of genetic informatics and living bodies as informational entities.

The author believes that in eukaryotic and prokaryotic genomes we have some special case of multi-partite entangled states in genomic systems of many oligomers (in some analogy with the case of groups of many particles). This can be termed as "the genomic entanglement" or as "the genomic tetra-entanglement" since genomic sequences contain 4 kinds of nucleotides A, T, C, and G. It should be emphasized that the author doesn't declare an existence of ordinary physical quantum entanglement in the genomes, but only that the mathematical apparatus of the theory of quantum informatics is suitable for a modeling the considered genetic sequences. Any long DNA sequence of nucleotides can be analyzed as a multicomponent quantum system, whose quantum state is determined by the tensor product of the quantum states of its subsystems, represented by sets of oligomers of different fixed length *n* where $n = 1, 2, 3, \ldots$

Let us turn to the above-mentioned author's model of properties of genomic sequences expressed by the expressions (11.2) and (11.3) (Petoukhov, 2018b; Petoukhov et al., 2019). This model is based on the tensor products and some other formalisms of quantum informatics and concerns, first of all, the hyperbolic rule №1 of the oligomer cooperative organization of genomes. The model introduced the notion "genetic qubits" based on different pairs of binary-oppositional molecular traits of adenine A, guanine G, cytosine C, and thymine T. Appropriate 2*n*-qubit systems in separable pure states were constructed, where nucleotides A, T, C, and G (and also DNA doublets and other *n*-plets) were represented by appropriate computational basis states in Hilbert spaces of corresponding dimensionalities. For example, cytosine C was represented as the computational basis state $|00>$ of the 2-qubit system in the 4-dimensional Hilbert space, thymine T - as the computational basis state $|01>$, guanine G – as the computational basis state $|10>$, and adenine A - as the computational basis state $|11>$ of the same 2-qubit system. Correspondingly, 16 doublets were represented as 16 computational basis states of the 4-qubit system in the 16-dimensional Hilbert space: for example, the doublet CC was represented as the computational basis state $|0000>$, the doublet CT – as $|0001>$, …, etc. This model can be used for a deeper understanding of the genomic entanglement.

An effective model should not only explain known phenomenological data but also predict unknown data to search them in natural systems. Let us show now that the proposed quantum-informational model has predictive power, allowing us to open previously unknown properties of genomic DNA sequences. Really, the noted model allowed a prediction not only the hyperbolic rule №1 described above but also many other non-trivial interrelations in genomic structures. In a limited volume of this article, the author can show only a few following brief examples.

11.1. About additional confirmations of the model predictions

For example, the model predicts the following. Till now we considered OS-sequences, whose members are total amounts of *n*-plets, which start with a certain « attributive » nucleotide, for example, with the nucleotide A. In this case, we calculate the total amounts of oligomers in the following sets: 4 doublets AT, AC, AG, AA; 16 triplets ATT, ATC, ATG, ACC, …; and so on. But what results arise if one calculates, in the same genome, the total amounts in quite other sets of *n*-plets having the same attributive nucleotide A at their second positions, that is the following sets: 4 doublets TA, CA, GA, AA; 16 triplets TAT, TAC, TAG, CAC, …; and so on for $n = 2, 3, 4, \ldots$? And what results arise if one calculates, in the same genome, total amounts in the sets of *n*-plets, which have the same nucleotide at their third positions, that is the following sets: 16 triplets TTA, TCA, TGA, CCA, …; 64 tetraplets TTTA, TCTA, TGCA, …; and so on for $n = 3, 4, 5, \ldots$? The quantum-information model predicts that in all such cases the resulting OS-sequences will be practically identical to the hyperbolic-like OS-sequence of the total amounts of *n*-plets with the same attributive nucleotide at their first position. These model predictions also apply to cases of sets of *n*-plets, which have the same attributive nucleotide at their 4th, 5th, 6th, …, *k*th positions for $n = k, k+1, k+2, \ldots$(here *k* is not too large compared to the full length of the genomic sequence).

These model predictions are confirmed by direct calculations of total amounts of corresponding sets of *n*-plets in different genomes. Figs. 30 and 31 show examples of such confirmations by the comparisons of different OS-sequences calculated for the human chromosome №1 in three cases of locations of attributive nucleotides in its *n*-plets: 1) at the first position in *n*-plets (data on the appropriate OS-sequences are taken from Fig. 3); 2) at the second position; 3) at the third position. One can see from the shown results that the differences Δ% of the corresponding members of these three OS-sequences from each other are less than 0.1%, that is these OS-sequences are practically identical. These differences were calculated for each *n* by formulas Δ% = 100 (1 - Pos1/Pos2)% and Δ% = 100 (1 - Pos1/Pos3)% where Pos1, Pos2, and Pos3 refer to values indicated in the rows Pos. 1, Pos. 2, and Pos. 3. Here the results are presented only for $n = 2, 3, 4, \ldots, 10$ but similar situations of practical coincidences of the corresponding members of the considered OS-sequences are also true for larger *n*.

These predictions about the oligomer cooperative organization and their confirmations in eukaryotic and prokaryotic genomes give a significant extension to the hyperbolic rule №1 regarding the hyperbolic-like OS-sequences of the total amounts of *n*-plets, which have the same attributive nucleotide at their *k*th position (not only in their first position). These results and the extended rules additionally open up the deep

n	1	2	3	4	5	6	7	8	9	10
A										
Pos. 1	67070277	33537501	22360413	16768845	13413532	11179286	9584038	8383461	7453552	6706672
Pos. 2	-	33532776	22353979	16767465	13413514	11174459	9578118	8383936	7452356	6704047
Δ%		0.014	0.029	0.008	0.000	0.043	0.062	-0.006	0.016	0.039
T										
Pos. 1	67244164	33620498	22412993	16808862	13445360	11207274	9606748	8405040	7470145	6724359
Pos. 2	-	33623666	22411166	16811071	13445910	11206100	9610249	8405351	7472348	6724456
Δ%		0.009	-0.008	0.013	0.004	-0.010	0.036	0.004	0.029	0.001
C										
Pos. 1	48055043	24024903	16012711	12013624	9612227	8005708	6865944	6008215	5336968	4803919
Pos. 2	-	24030140	16021444	12015843	9615911	8012553	6865662	6005986	5338638	4808410
Δ%		0.022	0.055	0.018	0.038	0.085	-0.004	-0.037	0.031	0.093
G										
Pos. 1	48111528	24057606	16040889	12028924	9625086	8021235	6869132	6013412	5348337	4813156
Pos. 2	-	24053922	16040412	12025875	9620866	8020389	6871831	6014853	5345656	4811187
Δ%		-0.015	-0.003	-0.025	-0.044	-0.011	0.039	0.024	-0.050	-0.041

Fig. 30. The comparison of the OS-sequences of the total amounts of *n*-plets, which have the nucleotide N (A, T, C, or G) at their first position (the row "Pos. 1") and at their second position (the row "Pos. 2") in the human chromosome №1. Δ% shows the percentage of differences between the corresponding total amounts of *n*-plets from each other. The comparison begins with doublets, since there is no second position in monoplets.

n	1	2	3	4	5	6	7	8	9	10
A										
Pos. 1	67070277	33537501	22360413	16768845	13413532	11179286	9584038	8383461	7453552	6706672
Pos. 3	-	-	22355885	16768656	13414900	11178695	9578685	8383657	7450656	6710255
Δ%			0.020	0.001	-0.010	0.005	0.056	-0.002	0.039	-0.053
T										
Pos. 1	67244164	33620498	22412993	16808862	13445360	11207274	9606748	8405040	7470145	6724359
Pos. 3	-	-	22420005	16811636	13448900	11208158	9604848	8406144	7472996	6723773
Δ%			-0.031	-0.017	-0.026	-0.008	0.020	-0.013	-0.038	0.009
C										
Pos. 1	48055043	24024903	16012711	12013624	9612227	8005708	6865944	6008215	5336968	4803919
Pos. 3	-	-	16020888	12011279	9611721	8010304	6867877	6005835	5342246	4803498
Δ%			-0.051	0.020	0.005	-0.057	-0.028	0.040	-0.099	0.009
G										
Pos. 1	48111528	24057606	16040889	12028924	9625086	8021235	6869132	6013412	5348337	4813156
Pos. 3	-	-	16030227	12028682	9620676	8016348	6874449	6014493	5343102	4810570
Δ%			0.066	0.002	0.046	0.061	-0.077	-0.018	0.098	0.054

Fig. 31. The comparison of the OS-sequences of the total amounts of *n*-plets, which have the nucleotide N (A, T, C, or G) at their first position (the row "Pos. 1") and at their third position (the row "Pos. 3") in the human chromosome №1. Δ% shows the percentage of differences of the corresponding total amounts of *n*-plets from each other. The comparison begins with triplets since there is no third position in monoplets and doublets.

connections of genomic sequences with the harmonic progression (2.4) and discover new aspects of the algebraic harmony of living bodies.

Another large bunch of predictions about genomic sequences is given by the quantum-information model for quantitative interrelations of different *n*-plets, which start from the same doublet, or from the same triplet, etc. The model predicts, in particular, that the amount S_2 of any of 16 doublets NN is algebra-harmonically interrelated with the total amounts S_3, S_4, S_5, …of oligomers in the following sets: 4 triplets, which start with this attributive doublet NN; 16 tetraplets, which start with this attributive doublet NN; 64 pentaplets, which start with this attributive doublet NN; and so on. This interrelation is again based on the harmonic progression (2.4). More precisely, according to the model prediction, the ratios of these total amounts S_2/S_3, S_2/S_4, S_2/S_5, …should be correspondingly equal to the ratios of the second member 1/2 of the harmonic progression (2.4) to its subsequent members 1/3, 1/4, 1/5, … that is to values 3/2, 4/2, 5/2, ….

Fig. 32 presents the confirmation of this model prediction by the comparison of the amount S_2 of each of 16 doublets to the total amounts S_3, S_4, S_5 of *n*-plets ($n = 3, 4, 5$), which start with this doublet, in the human chromosome №1.

The rows in the left part of Fig. 32 shows very different numeric series of total amounts, which are individual in each of rows. But the right part shows that in each row its amounts are interrelated identically based on the numeric series of the ratios 1.5, 2.0, and 2.5, which serves here as a general invariant for the cases of all 16 doublets. But this sequence of ratios exists in the harmonic progression (2.4): 1, 1/2, 1/3, 1/4, 1/5, …, where the ratios of its second member 1/2 to its third, fourth and fifth members (that is, 1/3, 1/4, and 1/5) give this series 3/2,

DOUBLETS	TRIPLETS	TETRAPLETS	PENTAPLETS	S_2/S_3	S_2/S_4	S_2/S_5
$S_2 = \Sigma$(**AA**)	$S_3 = \Sigma(AAN)_4$	$S_4 = \Sigma(AANN)_{16}$	$S_5 = \Sigma(AANNN)_{64}$			
10952057	7300222	5476855	4381298	**1.50**	**2.00**	**2.50**
$S_2 = \Sigma$(**AT**)	$S_3 = \Sigma(ATN)_4$	$S_4 = \Sigma(ATNN)_{16}$	$S_5 = \Sigma(ATNNN)_{64}$			
8561194	5706906	4280647	3420561	**1.50**	**2.00**	**2.50**
$S_2 = \Sigma$(**AC**)	$S_3 = \Sigma(ACN)_4$	$S_4 = \Sigma(ACNN)_{16}$	$S_5 = \Sigma(ACNNN)_{64}$			
5799729	3868541	2899991	2322063	**1.50**	**2.00**	**2.50**
$S_2 = \Sigma$(**AG**)	$S_3 = \Sigma(AGN)_4$	$S_4 = \Sigma(AGNN)_{16}$	$S_5 = \Sigma(AGNNN)_{64}$			
8224510	5484720	4111320	3289579	**1.50**	**2.00**	**2.50**
$S_2 = \Sigma$(**TA**)	$S_3 = \Sigma(TAN)_4$	$S_4 = \Sigma(TANN)_{16}$	$S_5 = \Sigma(TANNN)_{64}$			
7274275	4849731	3636741	2909412	**1.50**	**2.00**	**2.50**
$S_2 = \Sigma$(**TT**)	$S_3 = \Sigma(TTN)_4$	$S_4 = \Sigma(TTNN)_{16}$	$S_5 = \Sigma(TTNNN)_{64}$			
11026157	7346507	5511908	4409900	**1.50**	**2.00**	**2.50**
$S_2 = \Sigma$(**TC**)	$S_3 = \Sigma(TCN)_4$	$S_4 = \Sigma(TCNN)_{16}$	$S_5 = \Sigma(TCNNN)_{64}$			
6923689	4617788	3461837	2768794	**1.50**	**2.00**	**2.50**
$S_2 = \Sigma$(**TG**)	$S_3 = \Sigma(TGN)_4$	$S_4 = \Sigma(TGNN)_{16}$	$S_5 = \Sigma(TGNNN)_{64}$			
8396349	5598933	4198342	3357218	**1.50**	**2.00**	**2.50**
$S_2 = \Sigma$(**CA**)	$S_3 = \Sigma(CAN)_4$	$S_4 = \Sigma(CANN)_{16}$	$S_5 = \Sigma(CANNN)_{64}$			
8382478	5591208	4191829	3354600	**1.50**	**2.00**	**2.50**
$S_2 = \Sigma$(**CT**)	$S_3 = \Sigma(CTN)_4$	$S_4 = \Sigma(CTNN)_{16}$	$S_5 = \Sigma(CTNNN)_{64}$			
8221421	5477836	4111963	3289510	**1.50**	**2.00**	**2.50**
$S_2 = \Sigma$(**CC**)	$S_3 = \Sigma(CCN)_4$	$S_4 = \Sigma(CCNN)_{16}$	$S_5 = \Sigma(CCNNN)_{64}$			
6233384	4153642	3117570	2492824	**1.50**	**2.00**	**2.50**
$S_2 = \Sigma$(**CG**)	$S_3 = \Sigma(CGN)_4$	$S_4 = \Sigma(CGNN)_{16}$	$S_5 = \Sigma(CGNNN)_{64}$			
1187593	789995	592235	475262	**1.50**	**2.01**	**2.50**
$S_2 = \Sigma$(**GA**)	$S_3 = \Sigma(GAN)_4$	$S_4 = \Sigma(GANN)_{16}$	$S_5 = \Sigma(GANNN)_{64}$			
6923938	4612792	3462012	2768171	**1.50**	**2.00**	**2.50**
$S_2 = \Sigma$(**GT**)	$S_3 = \Sigma(GTN)_4$	$S_4 = \Sigma(GTNN)_{16}$	$S_5 = \Sigma(GTNNN)_{64}$			
5814874	3879880	2906516	2325903	**1.50**	**2.00**	**2.50**
$S_2 = \Sigma$(**GC**)	$S_3 = \Sigma(GCN)_4$	$S_4 = \Sigma(GCNN)_{16}$	$S_5 = \Sigma(GCNNN)_{64}$			
5073325	3381454	2536422	2032200	**1.50**	**2.00**	**2.50**
$S_2 = \Sigma$(**GG**)	$S_3 = \Sigma(GGN)_4$	$S_4 = \Sigma(GGNN)_{16}$	$S_5 = \Sigma(GGNNN)_{64}$			
6245451	4166742	3123944	2498784	**1.50**	**2.00**	**2.50**

Fig. 32. The comparison of total amounts $S_2 = \Sigma(NN)$ of each of 16 doublets NN to the total amounts S_3 of 4 triplets, S_4 of 16 tetraplets, and S_5 of 64 pentaplets, which start with such attributive doublet NN, is shown for the human chromosome №1. The left part of the table indicates the values of the corresponding total amounts. The right part contains appropriate values of the ratios S_2/S_3, S_2/S_4, and S_2/S_5, which are equal to the same magnitudes 1.5, 2.0, and 2.5 for the cases of all 16 doublets. Here N refers to any of nucleotides A, T, C, and G.

4/2, and 5/2. Similar results are true for all other human chromosomes and for all those genomes, which were analyzed by the author.

The model predicts similarly the following numeric interconnections in the complete genomic sequences:

- The amount S_3 of any of 64 triplets NNN is algebra-harmonically interrelated with the total amounts S_4, S_5, S_6, … of oligomers in the following sets: 64 tetraplets, which start with this attributive triplet NNN; 256 pentaplets, which start with this attributive triplet NNN; 1024 six-plets, which start with this attributive triplet NNN; …. The ratios of these total amounts S_3/S_4, S_3/S_5, S_3/S_6, …should be correspondingly equal to the ratios of the third member 1/3 of the harmonic progression (2.4) to its subsequent members 1/4, 1/5, 1/6, …, that is to values 4/3, 5/3, 6/3, …
- The amount S_4 of any of 256 tetraplets NNNN is algebra-harmonically interrelated with the total amounts of S_5, S_6, S_7, …of oligomers in the following sets: 256 pentaplets, which start with this attributive tetraplet NNNN; 1024 six-plets, which start with this attributive tetraplets NNNN; 4906 seven-plets, which start with this attributive tetraplets NNNN, …. The rations of these total amounts S_4/S_5, S_4/S_6, S_4/S_7, … should be correspondingly equal to the ratios of the fourth member 1/4 of the harmonic progression (2.4) to its subsequent members 1/5, 1/6, 1/7, …, that is to values 5/4, 6/4, 7/4, …
- And so on (the length of attributive oligomers NN…N in the considered sets of *n*-plets should not be too large compared to the full length of the genomic sequence).

Similar model predictions exist not only for the listed cases, when the considered attributive nucleotides, or attributive doublets, or attributive triplets, etc. occupy the first positions in *n*-plets of the considered sets, but also for cases when these attributive nucleotides or oligomers occupy there the second positions, or the third positions, etc (see corresponding rules about collective probabilities in oligomer tetra-groups for cases of locations of attributive oligomers in different positions of *n*-plets in the article (Petoukhov, 2018b)).

Most of the long list of predictions, stemming from this quantum information model, is still awaiting their checking through analysis of various genomes. So far, the author has conducted only a relatively small number of checks of such predictions and has not found a single case of a phenomenological refutation of these predictions. The author will be grateful to those members of the scientific community who will try to find in the full-length sequences of different genomes such cases where these model predictions are not fulfilled.

These and other confirmed predictions of the model enlarge significantly the list of hyperbolic rules in genomes and lead to new tools and opportunities to study genetic structures. The obtained phenomenological data and the set of confirmed predictions of the quantum-information model testify that the eukaryotic and prokaryotic genomes represent a regular algebraic fractal-like net with important participation of the harmonic progression (2.4) in interconnections of its parts. This allows us to say about the algebraic harmony in living bodies. In theoretical biology, the quantum-information model has appeared, which allows predicting with high accuracy a large number of quantitative interconnections between different kinds and sets of oligomers in eukaryotic and prokaryotic genomes (predictions “at the tip of the pen”).

12. Regarding the application of the oligomer sums method to long protein sequences

Till now we considered applications of the oligomer sums method to the analysis of long single-stranded DNA sequences of nucleotides. Such DNA sequences consist of 4 kinds of nucleotides, and corresponding 4 equivalency classes of A_1-, T_1-, C_1-, G_1-oligomers are analyzed. This Section discusses opportunities to apply this method for the similar revealing of possible algebra-harmonic features of primary structures of sequences of 20 amino acids in long proteins.

Each long sequence of amino acids (for example, ArgSerThrGlyPheLysLeuSer MetAla …) can be also represented in the form of fragmented sequences of different kinds: as a sequence of monomers (Arg-Ser-Thr-Gly-Phe-Lys-Leu-Ser-Met-Ala- …), or as a sequence of amino acid doublets (ArgSer-ThrGly-PheLys-LeuSer-MetAla- …), or as a sequence of amino acid triplets (ArgSerThr-GlyPheLys-LeuSerMet- …) , and so on. Analyzing above long DNA sequences of nucleotides, which consist of 4 kinds of nucleotides A, T, C, and G, we considered 4 equivalency classes of A_1-, T_1-, C_1-, G_1-oligomers. By analogy, in the case of sequences of 20 types of amino acids, we will analyze 20 equivalency classes, each of which is defined by corresponding amino acid and combines all oligomers, which start with this amino acid. For example, the amino acid Ala defines the equivalency class of Ala_1-oligomers, which includes all *n*-plets starting with this amino acid: the set of Ala_1-doublets contains all 20 doublets, which start with the Ala (AlaAla, AlaArg, AlaAsn, …, AlaCys); the set of Ala_1-triplets contains all 400 triplets, which start with the Ala (AlaAlaAla, AlaAlaArg, …., AlaCysCys), and so on.

The application of the oligomer sums method to the analysis of any long amino acid sequence and their 20 classes of the oligomer equivalency is as follows (by analogy with the above-described application of the method to analyze long nucleotide sequences and their 4 classes of the oligomer equivalency):

- Firstly, a considered amino acid sequence is represented in the form of a set of its fragmented sequences of oligomers (that is, fragments) of certain lengths $n = 1, 2, 3, \ldots$;
- Secondly, phenomenological quantities of each of 20 types of amino acids are calculated in the considered sequence;
- Thirdly, in each of the fragmented representations of the amino acid sequence under $n = 2, 3, 4, \ldots$, for any of the 20 classes of the

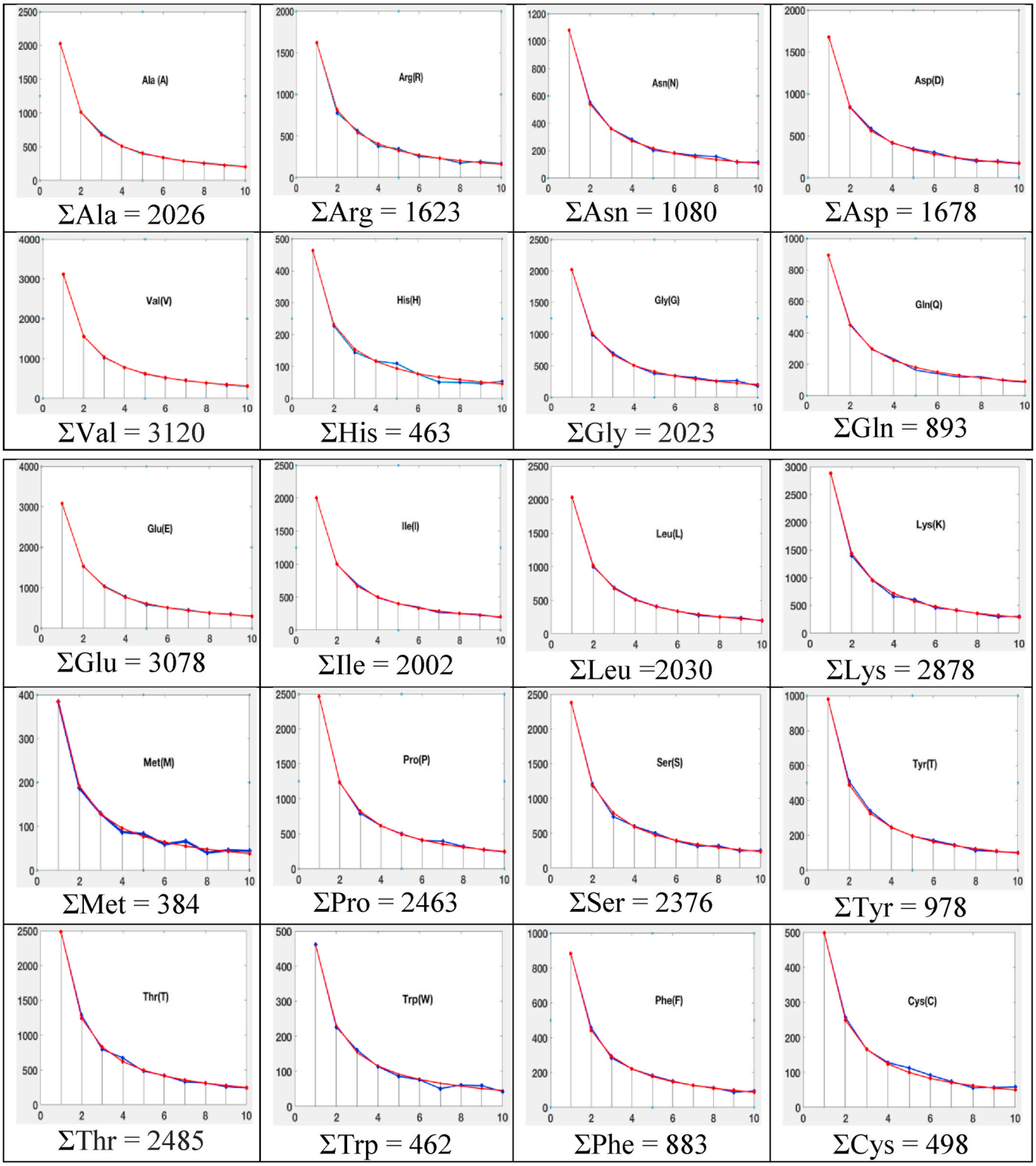


Fig. 33. Graphs of analysis results of the human protein Titin by the oligomer sums method for each of 20 equivalency classes, which are defined by its 20 types of amino acids. Each graph shows a sequence (in blue) of real total amounts of *n*-plets, which start with this amino acid, and also a model hyperbolic sequence Σ/n (in red), where Σ refers to a number of this amino acid ($n = 1, 2, \ldots, 10$). The abscissa axes show the values *n*; the ordinate axes show total amounts of the corresponding *n*-plets, which start with this amino acid. Initial data on this protein are taken on the site https://www.ncbi.nlm.nih.gov/protein/ACN81321.1. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

n	1	2	3	4	5	6	7	8	9	10
Ala										
Real	2026	1016	698	506	394	343	287	261	232	206
Model	2026	1013	675	506.5	405	338	289	253	225	203
Δ%	-0.3	-3.4	0.1	2.8	-1.6	0.8	-3.1	-3.1	-1.7	-0.3
Arg										
Real	1623	777	564	379	346	254	234	177	192	170
Model	1623	812	541	406	325	271	232	203	180	162
Δ%	0	4.3	-4.3	6.6	-6.6	6.1	-0.9	12.8	-6.5	-4.7

Fig. 34. Examples of numeric data about OS-sequences concerning two equivalency classes of Ala_1-oligomers and Arg_1-oligomers in the human protein Titin. Graphic representations of corresponding OS-sequences are shown in Fig. 33 at the very top.

oligomeric equivalency, the total amount Σ of its defining amino acid and also total amounts of all those *n*-plets ($n = 2, 3, 4, \ldots$), that have this acid in their first position (or in other fixed position), are calculated;

- The sequence of these phenomenological amounts is compared with the model hyperbolic sequence Σ/n of this equivalency class, where $n = 1, 2, 3, \ldots$

Let us explain the proposed application of the OS-method by an example of the analysis of the primary amino acid sequence of the protein Titin, which is one of the longest proteins. Titin is important in the contraction of striated muscle fibers and is the third most abundant protein in the muscle (after myosin and actin). Below some results of the author's analysis of the human protein Titin by the OS-method are presented. Fig. 33 shows 20 graphs demonstrating the OS-sequences for each of 20 amino acids combined in the single general amino acid sequence of the Titin. Each of these 20 graphs presents data for one of the types of amino acids and shows number Σ of this amino acid in Titin and also two sequences: one of them (in blue) corresponds to the sequence of the real total amounts of those *n*-plets, which start with this amino acid, and the second sequence (in red) corresponds to the model hyperbolic sequence Σ/n (here $n = 1, 2, 3, \ldots, 10$).

One can see from Fig. 33 that, in the protein Titin, for each of all 20 amino acids its sequence (in blue) of phenomenological values of total amounts of those *n*-plets, which start with this amino acid, approximately coincides with the corresponding model hyperbolic sequence Σ/n (in red) or slightly fluctuates around it. In the considered case of Titin, the accuracy of the coincidence of the sequences of phenomenological and model values is lower than in the case of genomes described above. This seems to be due to the relatively short length of the titin amino acid sequence compared to the lengths of genomic nucleotide sequences. The graphs in Fig. 33 show that the largest deviations of the sequences of real values from sequences of model values occur in cases of amino acids, whose number is minimal: the number of amino acids His is 463, Met - 384, Trp - 462, Cys - 498. Moreover, the deviations of the phenomenological values of oligomer sums from model values are relatively small for small values $n = 2, 3$, but with an increase in the length of oligomers at $n = 4, 5, \ldots, 10$, these deviations can increase (the number of corresponding *n*-plets decreases with increasing *n*).

Fig. 34 gives examples of phenomenological and model numeric values for the classes Ala_1-and Arg_1-oligomers from the first graphs in Fig. 33.

The study of the amino acid sequences of long proteins by this OS-method should be continued to allow the comparative analysis of various proteins.

Some concluding remarks

As is known, mutations and the pressure of natural selection influence the genomic sequences of nucleotides. For these reasons, one can assume that as a result of many millions of years of biological evolution, genomic sequences, due to various influences, receive a completely random structure as a whole. This article provides pieces of evidence that, despite mutations, the pressure of natural selection, and other evolutionary factors, the nucleotide sequences of the eukaryotic and prokaryotic genomes have universal algebraic invariants. One can believe that the algebraic unity of living organisms is found (this should be tested further and further on more and more number of genomes). New mathematical tools and approaches for an in-depth study of this genetic world and its evolution appear.

The discovery of the algebraic genomic invariants gives new knowledge about the unity of the world of all living organisms and the features of biological evolution. This concerns additionally the problem of the origin of life, since the following natural question arises: where and how did these genomic algebraic invariants come from, which are expressed in the described hyperbolic (harmonic) rules and related to the quantum-information model if they exist even in the genomes of archaea and bacteria? The received results are interesting also for discussions concerning various well-known theories of biological evolution: Darwinism, nomogenesis, orthogenesis, etc. Some of these results are briefly described in the author's letter (Petoukhov, 2020d).

Living matter appears as an algebraic-harmonic entity. One can separately note the result on that the cooperative system of oligomers in the genomes are associated with the harmonic progression, which is widely known in connection with musical harmony and the frequency system of musical overtones. But the harmonic progression is important not only in music. At least from the time of the Pythagorean doctrine of the aesthetics of proportions, the following idea exists: "*the aesthetic principle is the same in every art; only the material differs*" (Schumann, 1969). In light of this, architecture has long been interpreted as frozen music, and music as dynamic architecture. Additional speculation about the possible genetic basis of some aesthetic parallels in various arts arises though the very idea of the connection between the feeling of beauty and the genetic system is not new: it is reflected, for example, in the title of the article "Beauty is in the genes of the beholder" about a connection of some parameters of the DNA double helix with the golden section (Harel et al., 1986). These problems are discussed at the International interdisciplinary seminar "Algebraic Biology and Theory of Systems" in Moscow (Petoukhov, Tolokonnikov, 2020).

The genomic invariants, described in the article, are connected with hyperbolic sequences and transformations of hyperbolic rotations that shift the hyperbolic sequence along with itself. Hyperbolic rotations, which are also called Lorentz transformations and known in the special theory of relativity, draw attention to the structural connection of genetic phenomena with the hyperbolic geometry of the Minkowski plane. One of the well-known models of two-dimensional hyperbolic geometry is the Poincaré disk model, also called the conformal disk model. The Poincaré disk model is connected with split-quaternions by J. Cockle and seems to be interesting for studying some genetic structures and inherited physiological phenomena as it was mentioned in previous author's publications on matrix genetics (see, for example (Petoukhov, 2012)).

Living organisms are informational entities, in which everything is subordinate to the task of reliably transmitting genetic information to descendants. All inherited physiological systems, as parts of a whole organism, must be structurally coupled with a genetic code for transmission to descendants in encoded form. The question on a possible deep connection of physiology and brain functioning with principles of quantum informatics is considered in publications on many authors (Abbott et al., 2008; Altaisky, Filatov, 2001; Fimmel, Petoukhov, 2020; Igamberdiev, 1993, 2004; Matsuno, Paton, 2000; Patel, 2001a-c; Penrose, 1996; Petoukhov, 2018a, 2019b). The results presented in this article give new essential materials to this perspective direction of thoughts. For such thoughts about possible connections of brain activities with the mathematics of quantum mechanics, these oligomer sums method, algebra-harmonic hyperbolic rules, and the mentioned author's quantum-information model give new effective research instruments and phenomenological materials.

Researchers of the genetic system study the Nature system of storage, processing, and transmission of information, which has no direct analogies in modern science and technology, but which is studied on the basis of analogies with their achievements. The disclosure of informational patents of living nature can make an important contribution to scientific and technological progress.

It should be noted that the genomic hyperbolic rules are cardinally different from well-known hyperbolic Zipf's law. Zipf's law was originally formulated in terms of quantitative linguistics, stating that given some corpus of natural language utterances, the frequency of any word is inversely proportional to its rank in the frequency table (see, for example (Fagan, Gençay, 2010; Manin, 2013a,b)). In linguistics and other fields, Zipf's law speaks on the frequency of encounter of separate words or other separate objects. In contrast, the hyperbolic rules of the genomes focus on OS-sequences of the total amounts of *n*-plets and the genomic tetra-entanglement, that is, on the relative number of not separate oligomers, but the whole sums of sets of different *n*-plets distributed inside the genomic sequence, where each separate nucleotide is a part of many oligomers set existing simultaneously (each nucleotide is a distributed participant of many members of the appropriate genomic OS-sequence at once and makes a contribution to each of them). From the quantum-information model, OS-sequences serve as quantum-information characteristics of genomic sequences.

The proposed oligomer sums method and the quantum-information model give new opportunities to study genetic systems and the inherited algebra-harmonic organization of living bodies. The modern situation in the theoretic field of genetic informatics, where many millions of nucleotide sequences are described, can be characterized by the following citation: "*We are in the position of Johann Kepler when he first began looking for patterns in the volumes of data that Tycho Brahe had spent his life accumulating. We have the program that runs the cellular machinery, but we know very little about how to read it.*" (Fickett and Burks, 1989). Kepler did not make his astronomic observations, but he found – in the huge astronomic data of Tycho Brahe - his Kepler's laws of symmetric movements of planets relative to the Sun along ellipses. The author is convinced that further studies of symmetries in genetic and other physiological structures will reveal many more wonderful secrets of living matter.

The presented study is a continuation of the author's researches on symmetries in biological objects described in his publications (see References below). This study further illustrates the effectiveness of symmetry analysis in natural systems. No wonder the theory of symmetries is one of the foundations of modern mathematical natural science. The presented results reveal the existence of a new broad class of symmetries in eukaryotic and prokaryotic genomes. They are connected with previous rules of a generalized symmetry for collective probabilities of subalphabets of *n*-plets in long DNA sequences, which were described by the author in the article (Petoukhov, 2018b) and whose importance were noted in the article "Petoukhov's rules on symmetries in long DNA-texts" (Darvas, 2018). In this article, the head of the International Institute "Symmetrion" (Budapest, Hungary) proposed to launch a corresponding international project: "*Now, Petoukhov's above rules of symmetries are candidates for the role of universal rules of long DNA-texts in living bodies. Further researches are needed to determine the degree of universality of these rules. Taking into account the huge number of species and long DNA-texts to be tested in these relations, I propose to launch an international project to study these genetic symmetries. Symmetrion initiates and can take part as a center of such an international project*" (Darvas, 2018).

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, M. He, Z.B. Hu, Yu.I. Manin, I.V. Stepanyan, V.I. Svirin and G.K. Tolokonnikov for their collaboration. The author is also grateful to the members of the International Symmetry Association (Budapest, http://isa.symmetry.hu/) and the International Seminar "Algebraic Biology and System Theory" (Moscow, https://www.youtube.com/channel/UC8JLsuRzzPsRiHwrwEjMCtw) for discussing the author's researches in the field of matrix genetics and algebraic biology.

References

Abbott, D., Davies, P.C.W., Pati, A.K. (Eds.), 2008. Quantum Aspects of Life. foreword by Sir Roger Penrose, 13: 978-1-84816-253-2.

Albrecht-Buehler, G., 2006. Asymptotically increasing compliance of genomes withChargaff's second parity rules through inversions and inverted transpositions. Proc. Natl. Acad. Sci. U.S.A. 103 (47), 17828–17833. https://doi.org/10.1073/pnas.0605553103.

Altaisky, M.V., Filatov, F.P., 2001. Genetic Information and Quantum Gas. Submitted on 22.06.2001. quant-Ph/0106123v1.

Amico, L., Fazio, R., Osterloh, A., Vedral, V., 2008. Entanglement in many-body systems. Rev. Mod. Phys. 80, 517.

Chargaff, E., 1971. Preface to a Grammar of Biology: a hundred years of nucleic acid research. Science 172, 637–642.

Conway, J.H., Guy, R.K., 1995. The Book of Number. N.-Y., Copernicus, ISBN 0-387-97993-X.

Darvas, G., 2018. Petoukhov's rules on symmetries in long DNA-texts. Symmetry: Culture and Science 29 (Number 2), 318–320. https://doi.org/10.26830/symmetry_2018_2_318, 10.26830/symmetry_2018_2_318. http://journal-scs.symmetry.hu/abstract/?pid=673.

Fagan, S., Gençay, R., 2010. An introduction to textual econometrics. In: Ullah, Aman, Giles, David E.A. (Eds.), Handbook of Empirical Economics and Finance. *CRC Press*, ISBN 9781420070361, pp. 133–153.

Fickett, J.W., Burks, C., 1989. Development of a database for nucleotide sequences. –. In: Waterman, M.S. (Ed.), Mathematical Methods for DNA Sequences. CRC Press, Inc., Florida, pp. 1–34.

Fimmel, E., Petoukhov, S.V., 2020. Development of models of quantum biology based on the tensor product of matrices. In: Hu, Z., Petoukhov, S., He, M. (Eds.), Advances in Artificial Systems for Medicine and Education III. AIMEE 2019, Advances in Intelligent Systems and Computing, vol. 1126. Springer, Cham, pp. 126–135. https://doi.org/10.1007/978-3-030-39162-1_12.

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

Frank-Kamenetskii, M.D., 1988. The Most Important Molecule. Nauka, Moscow (in Russian).

Graham, R.L., Knuth, D.E., Parashnik, O., 1994. Concrete Mathematics. A Foundations For Computer Science. Addison-Wesley, Massachusetts, ISBN 0-201-55802-5.

Gühne, O., Tóth, G., 2009. Entanglement detection. Phys. Rep. 474, 1–75.

Gusfield, D., 1997. Algorithms on Strings, Trees, and Sequences: Computer Science and Computational Biology. Cambridge University Press, p. 556.

Harel, D., Unger, R., Sussman, J.L., 1986. Beauty is in the genes of the beholder. *Trends in Bioch. Sc.* 11, 155–156.

Harkin, A.A., Harkin, J.B., 2004. Geometry of generalized complex numbers. Math. Mag. 77 (2), 118–129.

Horodecki, R., Horodecki, P., Horodecki, M., Horodecki, K., 2009. Quantum entanglement. Rev. Mod. Phys. 81, 865.

Igamberdiev, A.U., 1993. Quantum mechanical properties of biosystems: a framework for complexity, structural stability, and transformations. *Biosystems*, v. 31 (1), 65–73.

Igamberdiev, A.I., 2004. Quantum computation, non-demolition measurements, and reflective control in living systems. Bio*systems* 77, 47–56.

Jordan, P., 1932. Die Quantenmechanik und die Grundprobleme der Biologie und Psychologie. Naturwissenschaften 20, 815–821. https://doi.org/10.1007/BF01494844.

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

Kappraff, J., 2000. The arithmetic of Nichomachus of Gerasa and its applications to systems of proportions. Nexus Netw. J. 2 (4). Retrieved October 3, 2000, from. http://www.nexusjournal.com/Kappraff.html.

Kappraff, J., 2002. Beyond Measure: Essays in Nature, Myth, and Number. World Scientific, Singapore.

Kappraff, J., 2006. Anne bulckens' analysis of the proportions of the Parthenon. Symmetry: Culture and Science 17 (1–2), 91–96.

Manin, Yu.I., 2013a. Zipf's law and L. Levin's probability distributions. Preprint. 1301.0427v2, 19 pages.

Manin, Yu.I., 2013b. Complexity vs Energy: Theory of Computation and Theoretical Physics. Preprint. 1302.6695, 23 pages.

Matsuno, K., Paton, R.C., 2000. Is there a biology of quantum information? Bio*systems* 55, 39–46.

McConkey, E., 1993. Human Genetics: the Molecular Revolution. Jones and Barlett, Boston, MA.

McFadden, J., Al-Khalili, J., 12 December 2018. The origins of quantum biology. Proceedings of the Royal Society A 474 (2220), 1–13. https://doi.org/10.1098/rspa.2018.0674.

Nielsen, M.A., Chuang, I.L., 2010. In: Quantum Computation and Quantum Information. Cambridge Univ. Press, New York. https://doi.org/10.1017/CBO9780511976667, 10.1017/CBO9780511976667.
Patel, A., 2001a. Quantum algorithms and the genetic code. Pramana - J. Phys. 56 (2–3), 367–381 arXiv:quant-ph/0002037.
Patel, A., 2001b. Testing quantum dynamics in genetic information processing. J. Genet. 80 (1), 39–43.
Patel, A., 2001c. Why genetic information processing could have a quantum basis. – Journal of Biosciences 26 (2), 145–151.
Penrose, R., 1996. Shadows of the Mind: A Search for the Missing Science of Consciousness. - Oxford University Press, USA, p. 480.
Petoukhov, S.V., January 2016. The system-resonance approach in modeling genetic structures. Biosystems 139, 1–11 (January 2016).
Petoukhov, S.V., 2020c. Hyperbolic Numbers, Genetics and Musicology. In: Advances in Artificial Systems for Medicine and Education III. AIMEE 2019. Advances in Intelligent Systems and Computing, vol. 1126. Springer, Cham, pp. 195–207. https://doi.org/10.1007/978-3-030-39162-1_18.
Petoukhov, S.V., 1989. Non-Euclidean geometries and algorithms of living bodies. Comput. Math. Appl. 17 (4–6), 505–534. HYPERLINK "http://www.sciencedirect.com/science/article/pii/0898122189902484. http://www.sciencedirect.com/science/article/pii/0898122189902484.
Petoukhov, S.V., 2008. Matrix Genetics, Algebrases of Genetic Code, Noise Immunity. RCD, Moscow, p. 316 (in Russian).
Petoukhov, S.V., 2011. Matrix genetics and algebraic properties of the multi-level system of genetic alphabets. NeuroQuantology 9 (4), 60–81.
Petoukhov, S.V., 2012. Symmetries of the genetic code, hypercomplex numbers and genetic matrices with internal complementarities. Symmetry: Culture and Science 23 (3–4), 275–301. http://petoukhov.com/petoukhov_genetic_matrices_complementarities.pdf.
Petoukhov, S.V., 2017. Genetic coding and united-hypercomplex systems in the models of algebraic biology. Biosystems 158, 31–46. August 2017.
Petoukhov, S.V., 2018a. The genetic coding system and unitary matrices. Preprints2018201804013110.20944/preprints201804.0131.v2. http://www.preprints.org/manuscript/201804.0131/v2.
Petoukhov, S.V., 2018b. The Rules of Long DNA-Sequences and Tetra-Groups of Oligonucleotides. arXiv:1709.04943v5, 5th Version from 8 October 2018, p. 159.
Petoukhov, S.V., 2019a. Nucleotide epi-chains and new nucleotide probability rules in long DNA sequences. Preprints2019201904001110.20944/preprints201904.0011.v2. https://www.preprints.org/manuscript/201904.0011/v2https://www.preprints.org/manuscript/201904.0011/v2.
Petoukhov, S.V., 2019. Connections Between Long Genetic and Literary Texts. The Quantum-Algorithmic Modelling. In: Hu, Z. (Ed.), Advances in Computer Science for Engineering and Education II, pp. 534–543.
Petoukhov, S.V., 2020a. Hyperbolic numbers in modeling genetic phenomena. Preprints201908028420193610.20944/preprints201908.0284.v42020ahttps://www.preprints.org/manuscript/201908.0284/v4. .
Petoukhov, S.V., 2020b. The genetic code, algebraic codes and double numbers. Preprints2019201911030110.20944/preprints201911.0301.v2. https://www.preprints.org/manuscript/201911.0301/v2.
Petoukhov, S.V., 2020e. Hyperbolic Rules of the Oligomer Cooperative Organization of Eukaryotic and Prokaryotic Genomes. Preprints 2020, 2020050471, 95 pages. https://www.preprints.org/manuscript/202005.0471/v2.
Petoukhov, S.V., 2020d. Genomes symmetries and algebraic harmony in living bodies. Symmetry: Culture and Science 31 (2), 222–223. https://doi.org/10.26830/symmetry_2020_2_222, file:///Users/Sergej/Downloads/2020_2_222-223_Petoukhov-letter-to-the-Editor.pdf.
Petoukhov S.V.,.Petukhova E.S., Svirin V.I. Symmetries of DNA alphabets and quantum informational formalisms. Symmetry: Culture and Science302201916117910.26830/symmetry_2019_2_161http://petoukhov.com/PETOUKHOV%20GENETIC%20QUANTUM%20INFORMATIONAL%20MODEL%202019.pdf.
Petoukhov, S.V., He, M., 2010. Symmetrical Analysis Techniques For Genetic Systems and Bioinformatics: Advanced Patterns and Applications. IGI Global, USA.
Petoukhov, S.V., Petukhova, E.S., 2017. Symmetries in genetic systems and the concept of geno-logical coding. Information 8 (1), 2. https://doi.org/10.3390/info8010002, 2017a. http://www.mdpi.com/2078-2489/8/1/2/htm.
Petoukhov, S.V., Tolokonnikov, G.K., 2020. Algebraic biology and matrix genetics systems. In: Presentation at the international interdisciplinary seminar "Algebraic biology and theory of systems" 02/13/2020, Moscow, Russia (in Russian). https://www.youtube.com/watch?v=H2dNtvTMl1M&t=330s.
Petoukhov, S.V., Petukhova, E.S., 2017. Resonances and the quest for transdisciplinarity. In: Burgin, M., Hofkirchner, W. (Eds.), Information Studies and the Quest for Transdisciplinarity. World Scientific, pp. 467–487.
Prabhu, V.V., 1993. Symmetry observation in long nucleotide sequences. Nucleic Acids Res 21, 2797–2800.
Rapoport, A.E., Trifonov, E.N., 2012. Compensatory nature of Chargaff's second parity rule. *Journal of Biomolecular Structure and Dynamics*, November 1–13. https://doi.org/10.1080/07391102.2012.736757.
Rosandic, M., Vlahovic, I., Gluncic, M., Paar, V., 2016. Trinucleotide's quadruplet symmetries and natural symmetry law of DNA creation ensuing Chargaff's second parity rule. J. Biomol. Struct. Dyn. 34 (7), 1383–1394. https://doi.org/10.1080/07391102.2015.1080628.
Schumann, R., 1969. On Music and Musicians. Konrad Wolff., New York.
Shporer, S., Chor, B., Rosset, S., Horn, D., 2016. Inversion symmetry of DNA k-mer counts: validity and deviations. BMC Genom. 17 (1), 696.
Walter, M., Gross, D., Eisert, J., 2017. Multi-partite Entanglement arXiv:1612.02437.
Yamagishi, M.E.B., 2017. Mathematical Grammar of Biology. Springer International Publishing AG, Switzerland, ISBN 978-3-319-62689-5.