 doi:10.20944/preprints202005.0471.v2

Hyperbolic rules of the oligomer cooperative organization of eukaryotic and prokaryotic genomes

Sergey V. Petoukhov

Mechanical Engineering Research Institute of Russian Academy of Sciences.
Russia, 101990, Moscow, M. Kharitonievskiy pereulok, 4, http://eng.imash.ru/,
info@imash.ru

Comment: Some elements of this article were presented by the author in his keynote speeches at the following conferences: the International Belgrade Bioinformatics Conference 2018 (Belgrade, Serbia, 18-22 June 2018, http://belbi.bg.ac.rs/); the 3rd International Conference on Computer Science, Engineering and Education Applications (Kiev, Ukraine, 21-22 January 2020). Also an author's presentation with elements of this article was done at the 6th International Conference in Code Biology (Friedrichsdorf, Germany, 3-7 June 2019, http://www.codebiology.org/conferences/Friedrichsdorf2019/).

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 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 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 are shown to the analysis of the following long sequences: 1) amino acid sequences in long proteins like the protein Titin; 2) phonetic sequences of long Russan literary texts (for checking of thoughts of many authors that phonetic organization of human languages is deeply connected with the genetic language). The topics of the algebraic harmony in living bodies and of the quantum-information approach in biology are discussed.

Key words. DNA oligomers, harmonic progression, hyperbolic rules, matrices, tensor product, quantum informatics, oligomer sums method, genomes, genes, viruses, proteins, long Russian texts, phonetic sequences.

Contents

 doi:10.20944/preprints202005.0471.v2

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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

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, Gumbel, Karpuzoglu, Petoukhov, 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 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, Gumbel, Karpuzoglu, Petoukhov, 2019; Prabhu, 1993; Rapoport, Trifonov, 2012; Rosandic, Vlahovic, Gluncic, Paar, 2016; Shporer, Chor, Rosset, and Horn, 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 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 (these results are briefly described in the published author's letter [Petoukhov, 2020d]. In addition, this second version of this article additionally shows that the oligomer sums method can be usefully applied to the analysis not only genomic sequences of nucleotides but also to the analysis of the following long biological sequences: 1) amino acid sequences of long proteins (the example of the protein Titin is presented); 2) phonetic sequences representing long Russian literary novels by L.N. Tolstoy, F.M. Dostoevsky, A.S. Pushkin (for checking of thoughts of many authors that phonetic organization of linguistic languages is deeply connected with the genetic language; the Russian alphabet has a one-to-one correspondence between letters and phonemes, and by this reason, long Russian literary texts are appropriate for such checking).

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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

which are also called *n*-plets, where *n* refers to the oligomer length. Each of nucleotide sequences in eukaryotic and prokaryotic genomes can be considered 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. The article describes the numerical analysis of sets of *n*-plets, which belong 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, one should calculate phenomenological quantities S_A, S_T, S_C, and S_G of monomers A, T, C, and G correspondingly in the considered nucleotide sequence. In the human chromosome № 1, the following quantities exist: $S_A = 67070277$, $S_T = 67244164$, $S_C = 48055043$, $S_G = 48111528$;
- Secondly, to construct the oligomer sums sequences, one should calculate the total amounts $\Sigma_{A,n,1}$, $\Sigma_{T,n,1}$, $\Sigma_{C,n,1}$, and $\Sigma_{G,n,1}$ of *n*-plets in equivalence classes of A_1-oligomers, T_1-oligomers, C_1-oligomers, and G_1-oligomers under $n = 1, 2, 3, 4, \ldots$ (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, 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. 2.1, middle. Deviations of phenomenological quantities $\Sigma_{A,n,1}$ from model values S_A/n lie in the range -0.030%÷0.024%, that is, they comprise only one-hundredths of a percent (Fig. 2.1, right). Initial data on this

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

chromosome were taken in the GenBank: https://www.ncbi.nlm.nih.gov/nuccore/NC_000001.11.

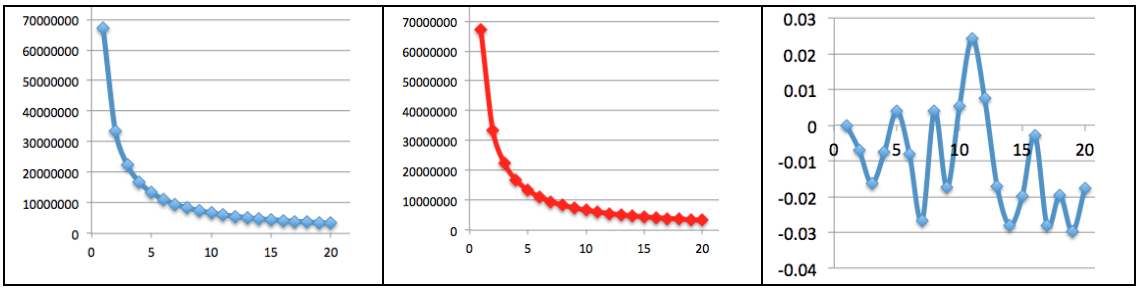


Fig. 2.1. The graphs of data for the case of the OS-sequences of *n*-plets from the class A_1-oligomers of the human chromosome №1. In these graphs, the abscissa axis represents the values n = 1, 2, 3, …, 20. **Left**: the ordinate axis represents the set of phenomenological total amounts $\Sigma_{A,n,1}$ of *n*-plets beginning with the nucleotide A. **Middle**: the ordinate axis represents modeling values 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.

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. 2.1, 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. 2.2, at left), and with the nucleotide C (Fig. 2.2, at middle), and with the nucleotide G (Fig. 2.2, 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:

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

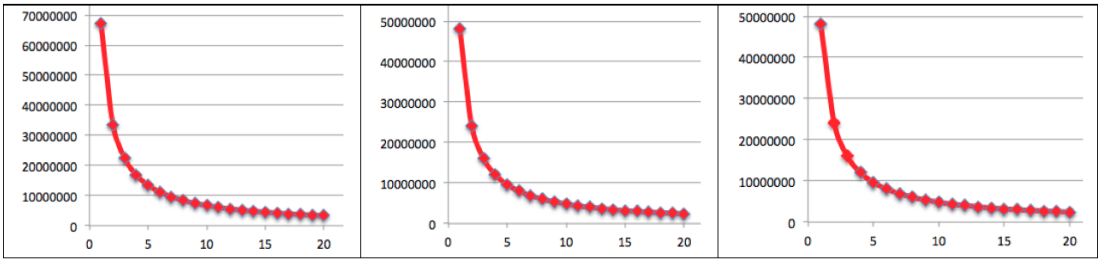


Fig. 2.2. Additional graph data to the OS-representation of the human chromosome №1. The abscissa axes represent the values n = 1, 2, 3, …, 20. The

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

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.

Fig. 2.3 shows real and model 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,…, 20. The model values of the 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 real 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.

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. 2.3. Real 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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

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%).

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):

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

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):

$$y = 1/n, \qquad (2.3)$$

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)

$$1/1, 1/2, 1/3, 1/4, 1/5,\ldots, 1/n \qquad (2.4)$$

is known long ago as a harmonic progression (or a 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 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, Knuth, Parashnik, 1994, p. 276]; this, in particular, connects the harmonic progression (2.4) with the Weber-Fechner logarithmic law, which is the main psychophysical law and dictates informatic peculiarities for all inherited sensory channels - vision, hearing, smell, etc ,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

whose organs (eyes, ears, nose, etc.) very differ each other in appearance. It testifies that genetic and different psychophysical levels of inherited biological informatics are structurally intercorrelated 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. 2.3 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):

$$\Sigma_{A,n,1} \approx \Sigma_{T,n,1}, \qquad \Sigma_{C,n,1} \approx \Sigma_{G,n,1} \tag{2.5}$$

In the particular case at $n = 1$, expressions (2.4) 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 (2.4), is a certain generalization of the 2nd 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.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

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 added that, in the same human chromosome № 1, the comparison of amounts of *n*-plets, consisting only of nucleotides of the same kind, 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 67070277, the amount of the doublet AA - 10952057, the amount of the triplet AAA – 2837038, the amount of the tetraplet AAAA – 856207, 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,... but not very large in comparison with the length of the DNA sequence. Fig. 2.4 shows general confirmational results of studying all 24 human chromosomes by the OS-method under n =1, 2, 3, …, 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.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

№	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. 2.4. 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, …, 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.

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):

$$S_A/(nS),\ S_T/(nS),\ S_C/(nS),\ S_G/(nS) \qquad (2.6)$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

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. 2.5 shows regarding the first main members S_A/S, S_T/S, S_C/S, and S_G/S of these hyperbolic sequences.

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	0.197	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. 2.5. 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 $S = S_A+S_T+S_C+S_G$. Harmonic means of the values S_A/S, S_T/S, S_C/S, and S_G/S in each chromosome are also indicated.

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 Sections 4-7 present appropriate results for some eukaryotic genomes. These results allow proposing the third hyperbolic (or harmonic) rule on 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, … is not too large compared to the full length of the nucleotide sequence).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

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 $x_1, x_2, \ldots, x_n$, is defined to be

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

Knowing two neighboring members of the harmonic progression, one can calculate its next member. 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 values of the harmonic mean in some figures of the article for comparison analysis of OS-sequences in different nucleotide sequences (Fig. 2.5 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 sequence S, S/2, S/3,…, S/n. Each member of this sequence is the sum of the four OS-sequences S_A/n, S_T/n, S_C/n, and S_G/n (2.8):

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

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.

Below Sections 4-7 represents the results, which have been obtained on the basis of the analysis of very different genomes by the OS-method and which testify in favor that the formulated hyperbolic rules have a general genomic significance. But previously the next Section will explain the matrix-algebraic approach, which has led the author to discover these rules.

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

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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

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, 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 = 1, G = T = 0). In such tables, all letters, doublets, and triplets automatically occupy their strictly individual places (Fig. 3.1).

	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. 3.1. 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).

These three tables (Fig. 3.1) 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. 3.1. 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. 3.1, 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.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

The genetic code is called a "degenerate code" because 64 triplets 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. 3.1) 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} seconds). 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.

Fig. 3.2 shows the real repetition and location of amino acids and stop-codons in 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).

	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. 3.2. 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. 3.1) for the Vertebrate Mitochondrial Code. The symbol “Stop” refers to stop-codons.

The location and repetition of all amino acids and stop-codons in the matrix of 64 triplets have the following algebraic feature (Fig. 3.2):

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

Bisymmetric (2*2)-matrices [a, b; b, a] are well known in algebra as matrix representations of two-dimensional hypercomplex numbers called hyperbolic

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

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 (3.1):

$$(x+jy)+(u+jv)=(x+u)+j(y+v); \quad (x+jy)(u+jv)=(xu+yv)+j(xv+yu) \qquad (3.1)$$

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. 3.3 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.

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

Fig. 3.3. The decomposition of the bisymmetric matrix [a, b; b, a] into two sparse matrices representing real and imaginary units of hyperbolic numbers correspondingly.

Regarding the hyperbolas from the hyperbolic rules of the previous Section (Figs. 2.1, 2.2, etc.), it can be noted that 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 bisymmetric matrix [a, b; b, a] representing a special form of hyperbolic numbers (the hyperbolic rotation is known in the special theory of relativity under the name of the Lorentz transformation).

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. 3.4) 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.

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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

Fig. 3.4. The numeric analog of the symbolic (8*8)-matrix of amino acids and stop-codons from Fig. 3.2 for the case of representing each of amino acids by numbers of its carbon atoms (stop-codons are conditionally represented by zero).

The connection of the genetic code with hyperbolic numbers supplements the following statement of the author, presented in a number of his 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 the Earth in a very weakened and distorted form. But, magicallly, 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.

4. The hyperbolic rules in all chromosomes of a fruit fly *Drosophila melanogaster*

This and upcoming Sections 5-7 are 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. Represented tabular data 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 article).

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 -

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

https://www.ncbi.nlm.nih.gov/genome/?term=drosophila+melanogaster. Resulting data in Fig. 4.1 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, …, 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. 4.1. 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 monomers 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, …, 20 (the model values are taken as 100%).

Fig. 4.2 shows data of normalized OS-sequences for all chromosomes of *Drosophila melanogaster*.

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. 4.2. 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.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

5. The hyperbolic rules in all chromosomes of a nematode *Caenorhabditis elegans*

The Section represents data about results of the analysis of single-stranded DNA sequences of the complete set of chromosomes of 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 six chromosomes. All initial data are taken from the GenBank (https://www.ncbi.nlm.nih.gov/genome?term=caenorhabditis%20elegans).

Resulting data in Fig. 5.1 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, …, 20. In all shown cases, the deviations of real sequences from model hyperbolic (harmonic) progressions are less than 0.5% 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 %
1	4835939	-0.144 ÷0.319	4848450	-0.160 ÷0.294	2695890	-0.487 ÷0.327	2692155	-0.498 ÷0.218
2	4878209	-0.196 ÷0.421	4869734	-0.229 ÷0.109	2769232	-0.256 ÷0.492	2762246	-0.253 ÷0.257
3	4444681	-0.139 ÷0.157	4423618	-0.269 ÷0.156	2449158	-0.451 ÷0.303	2466344	-0.173 ÷0.362
4	5711043	-0.106 ÷0.229	5730974	-0.253 ÷0.177	3034784	-0.393 ÷0.219	3017028	-0.199 ÷0.414
5	6750403	-0.145 ÷0.124	6760297	-0.164 ÷0.203	3712075	-0.222 ÷0.575	3701405	-0.418 ÷0.286
X	5747200	-0.256 ÷0.120	5734092	-0.166 ÷0.167	3119741	-0.156 ÷0.340	3117909	-0.272 ÷0.256

Fig. 5.1. The results of the analysis of all chromosomes of *Caenorhabditis elegans* 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 monomers A, T, C, and G in appropriate chromosomes. The columns "Range %" show deviations of real series 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, …, 20 (the model values are taken as 100%).

Fig. 5.2 shows data of normalized OS-sequences for all chromosomes of *Caenorhabditis elegans.*

Chrom	S_A/S	S_T/S	S_C/S	S_G/S	Harmonic mean
1	0.3208	0.3217	0.1789	0.1786	0.230
2	0.3193	0.3187	0.1812	0.1808	0.231
3	0.3225	0.3209	0.1777	0.1789	0.229
4	0.3265	0.3276	0.1735	0.1725	0.226
5	0.3226	0.3231	0.1774	0.1769	0.229
X	0.3244	0.3236	0.1761	0.1760	0.228

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

Fig. 5.2. Data of normalized OS-sequences S_A/nS, S_T/nS, S_C/nS, and S_G/nS of all chromosomes of *Caenorhabditis elegans* 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.

6. The hyperbolic rules in all chromosomes of a house mouse *Mus musculus*

The Section represents data about results of the analysis of single-stranded DNA sequences of the complete set of chromosomes of the laboratory mouse *Mus musculus*, which is a major model organism for basic mammalian biology, human disease, and genome evolution. All initial data are taken from the GenBank https://www.ncbi.nlm.nih.gov/genome?term=mus%20musculus.

Resulting data in Fig. 6.1 confirm that - for all the chromosomes - the model harmonic 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, …, 20. In all shown cases, the deviations of real sequences from model hyperbolic progressions are significantly less than 0.5% 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 %
1	56530182	-0.051 ÷0.054	56416289	-0.126 ÷0.067	39495313	-0.044 ÷0.121	39467408	-0.150 ÷0.107
2	51600126	-0.099 ÷0.076	51679955	-0.060 ÷0.063	37504114	-0.041 ÷0.111	37542456	-0.110 ÷0.036
3	46503996	-0.041 ÷0.063	46631177	-0.057 ÷0.092	31603703	-0.151 ÷0.120	31659979	-0.063 ÷0.075
4	43821952	-0.050 ÷0.081	43922197	-0.056 ÷0.076	32146231	-0.097 ÷0.062	32165231	-0.050 ÷0.090
5	42488105	-0.060 ÷0.059	42515761	-0.063 ÷0.050	31456650	-0.033 ÷0.085	31459158	-0.218 ÷0.052
6	42843713	-0.122 ÷0.082	42886213	-0.107 ÷0.048	30315703	-0.028 ÷0.108	30290914	-0.083 ÷0.094
7	40271749	-0.105 ÷0.070	40509547	-0.110 ÷0.079	30554235	-0.089 ÷0.081	30519876	-0.019 ÷0.125
8	36224525	-0.043 ÷0.083	36167473	-0.119 ÷0.127	26616967	-0.147 ÷0.104	26602467	-0.123 ÷0.081
9	34722476	-0.079 ÷0.066	34694585	-0.088 ÷0.066	25880876	-0.154 ÷0.185	25859081	-0.114 ÷0.095
10	37185184	-0.066 ÷0.068	37273294	-0.046 ÷0.142	26277876	-0.099 ÷0.096	26331308	-0.112 ÷0.123
11	33401283	-0.069 ÷0.049	33317397	-0.121 ÷0.049	26022668	-0.057 ÷0.177	26004597	-0.061 ÷0.081
12	33897029	-0.111 ÷0.069	34225639	-0.105 ÷0.144	24374340	-0.061÷ 0.094	24425412	-0.115 ÷0.109
13	34255191	-0.119 ÷0.088	34115119	-0.064 ÷0.082	24377641	-0.079 ÷0.119	24373242	-0.138 ÷0.090

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

14	35695406	-0.135 ÷0.027	35758968	-0.086 ÷0.115	24980458	-0.057 ÷0.135	25007278	-0.099 ÷0.124
15	29177752	-0.135 ÷0.069	29244798	-0.028 ÷0.092	21121081	-0.050 ÷0.180	21109684	-0.095 ÷0.088
16	28035438	-0.054 ÷0.115	28084677	-0.067 ÷0.099	19439086	-0.125 ÷0.023	19460557	-0.170 ÷0.003
17	26251160	-0.075 ÷0.097	26302830	-0.051 ÷0.079	19586991	-0.085 ÷0.060	19566481	-0.096 ÷0.063
18	25615329	-0.064 ÷0.075	25597990	-0.103 ÷0.054	18095575	-0.099 ÷0.159	18143740	-0.129 ÷0.089
19	16732680	-0.097 ÷0.094	16602953	-0.193 ÷0.076	12449343	-0.181 ÷0.201	12420880	-0.098 ÷0.225
X	49660944	-0.069 ÷0.049	49651848	-0.102 ÷0.052	32081377	-0.049 ÷0.093	32093826	-0.126 ÷0.131
Y	26842991	-0.084 ÷0.166	27013719	-0.072 ÷0.107	17175367	-0.221 ÷0.171	17092621	-0.163 ÷0.171

Fig. 6.1. The results of the analysis of all chromosomes of a house mouse *Mus musculus* 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 monomers A, T, C, and G in appropriate chromosomes. The columns "Range %" show deviations of real sequences from the model hyperbolic sequences $H_{A,1}(n) = S_A/n$, $H_{T,1}(n) = S_T/n$, $H_{C,1}(n) = S_C/n$, and $H_{G,1}(n) = S_G/n$ at $n = 1, 2, 3, \ldots, 20$ (the model values are taken as 100%).

Fig. 6.2 shows data of normalized OS-sequences for all chromosomes of *Mus musculus*.

Chrom	S_A/S	S_T/S	S_C/S	S_G/S	Harmonic mean
1	0.2946	0.2940	0.2058	0.2057	0.242
2	0.2894	0.2898	0.2103	0.2105	0.244
3	0.2973	0.2982	0.2021	0.2024	0.241
4	0.2882	0.2889	0.2114	0.2115	0.244
5	0.2872	0.2874	0.2127	0.2127	0.244
6	0.2928	0.2931	0.2072	0.2070	0.243
7	0.2839	0.2856	0.2154	0.2151	0.245
8	0.2884	0.2879	0.2119	0.2118	0.244
9	0.2866	0.2864	0.2136	0.2134	0.245
10	0.2926	0.2933	0.2068	0.2072	0.243
11	0.2813	0.2806	0.2191	0.2190	0.246
12	0.2899	0.2927	0.2085	0.2089	0.243
13	0.2925	0.2913	0.2081	0.2081	0.243
14	0.2939	0.2945	0.2057	0.2059	0.242
15	0.2899	0.2905	0.2098	0.2097	0.244
16	0.2950	0.2956	0.2046	0.2048	0.242
17	0.2862	0.2868	0.2136	0.2134	0.245
18	0.2929	0.2927	0.2069	0.2075	0.243
19	0.2875	0.2852	0.2139	0.2134	0.245
X	0.3038	0.3037	0.1962	0.1963	0.239
Y	0.3046	0.3065	0.1949	0.1940	0.238

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

Fig. 6.2. Data of normalized OS-sequences S_A/nS, S_T/nS, S_C/nS, and S_G/nS of all chromosomes of a house mouse *Mus musculus* 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.

7. The hyperbolic rules in all chromosomes of a plant *Arabidopsis thaliana*

One more model organism is a plant *Arabidopsis thaliana*. This small flowering plant 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. The Section represents data about results of the analysis of single-stranded DNA sequences of the complete set of 5 chromosomes of this plant by the oligomer sums method. All initial data about the chromosomes were taken from the GenBank - https://www.ncbi.nlm.nih.gov/genome/4, the column RefSeq).

Resulting data in Fig. 7.1 confirm that - for all the chromosomes - the model harmonic 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, …, 20. In all shown cases, the deviations of real sequences from model harmonic progressions are less than 0.6% 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 %
1	9709674	-0.275 ÷0.103	9697113	-0.140 ÷0.209	5435374	-0.130 ÷0.303	5421151	-0.186 ÷0.296
2	6315641	-0.035 ÷0.198	6316348	-0.256 ÷0.162	3542973	-0.406 ÷0.252	3520766	-0.148 ÷0.478
3	7484757	-0.121 ÷0.101	7448059	-0.141 ÷0.238	4258333	-0.283 ÷0.200	4262704	-0.206 ÷0.121
4	5940546	-0.155 ÷0.239	5914038	-0.109 ÷0.238	3371349	-0.222 ÷0.333	3356091	-0.293 ÷0.161
5	8621974	-0.123 ÷0.184	8652238	-0.213 ÷0.204	4832253	-0.515 ÷0.073	4858759	-0.132 ÷0.407

Fig. 7.1. The results of the analysis of all chromosomes of a plant *Arabidopsis thailana* 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 monomers 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, …, 20 (the model values are taken as 100%).

Fig. 7.2 shows data of normalized OS-sequences for all chromosomes of *Arabidopsis thailana.*

Chrom	S_A/S	S_T/S	S_C/S	S_G/S	Harmonic mean
1	0.3208	0.3204	0.1796	0.1791	0.230
2	0.3207	0.3207	0.1799	0.1788	0.230
3	0.3191	0.3176	0.1816	0.1817	0.231
4	0.3197	0.3183	0.1814	0.1806	0.231
5	0.3197	0.3209	0.1792	0.1802	0.230

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

Fig. 7.2. Data of normalized OS-sequences S_A/nS, S_T/nS, S_C/nS, and S_G/nS of all chromosomes of *Arabidopsis thailana* 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.

8. 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 sum 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. 8.1-8.6 show some results of the analysis - by the oligomer sums method - of the nucleotide sequence of 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 26373 nucleotides A, 19569 nucleotides T, 17097 nucleotides C, and 18901 nucleotides G, that is $S_A = 26373$, $S_T = 19569$, $S_C = 17097$, and $S_G = 18901$ 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 (26373 and 19569), 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.

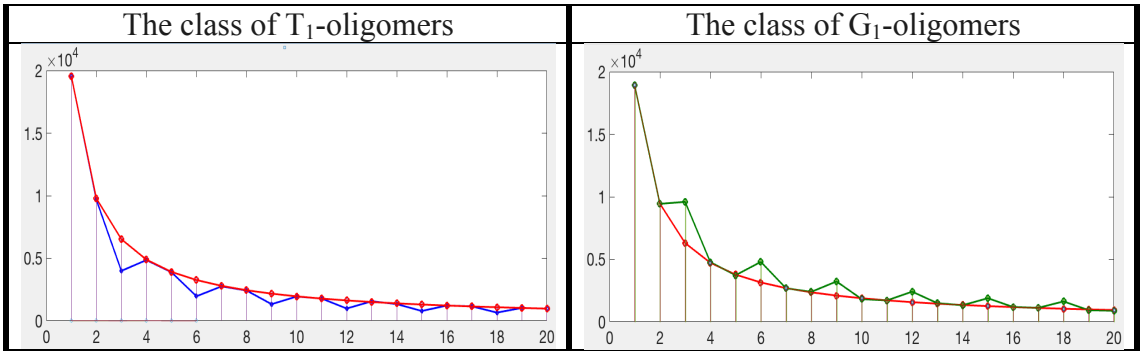


Fig. 8.1. Graphical representations of the results of the analysis - by the oligomer sums method – of the human *TNT* gene. The OS-sequences of its total amounts of *n*-plets, which start with the nucleotide T (left) and the nucleotide G (right), are shown. The red lines refer to model hyperbolic progressions S_T/n and S_G/n correspondingly, where $S_T = 19569$ and $S_G = 18901$ 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.

Fig. 8.1 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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

for cases of 3*m*-plets (here *m* = 1, 2, 3,…). Correspondingly these significant deviations can be called «triplet-deviations».

Fig. 8.2 shows the graph, which unites both graphs from Fig. 8.1 and demonstrates a few interesting features of the highly regular series of these triplet-deviations.

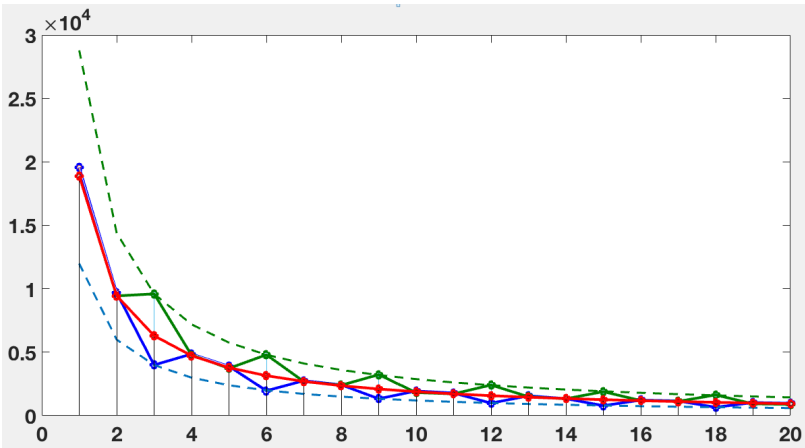


Fig. 8.2. The graph, uniting two graphs from Fig. 8.1 for the *TNT* gene, is shown. The blue dot line and the green dot lines correspond to those additional hyperbolic progressions 11979/*n* and 28788/*n*, which model real total amounts of 3*m*-plets. Other parts of this united graph are the same as in Fig. 8.1.

Firstly, one can see in Fig. 8.2 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*;
- 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. 8.2. 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 11979 and 28788.

Similar triplet-deviations exist in the OS-representations not only of the *TNT* 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-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

deviations in different species can be useful to study the secrets of the genetic system and biological evolution.

Fig. 8.2 demonstrated the highly regular rhythmic triplet-deviations for n = 1, 2, 3, …, 20, but similar rhythmic triplet-deviations exist in a much wider range of values n. Fig. 8.3 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 *TNT* 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,…, 20, and n = 1, 2, 3, .., 100.

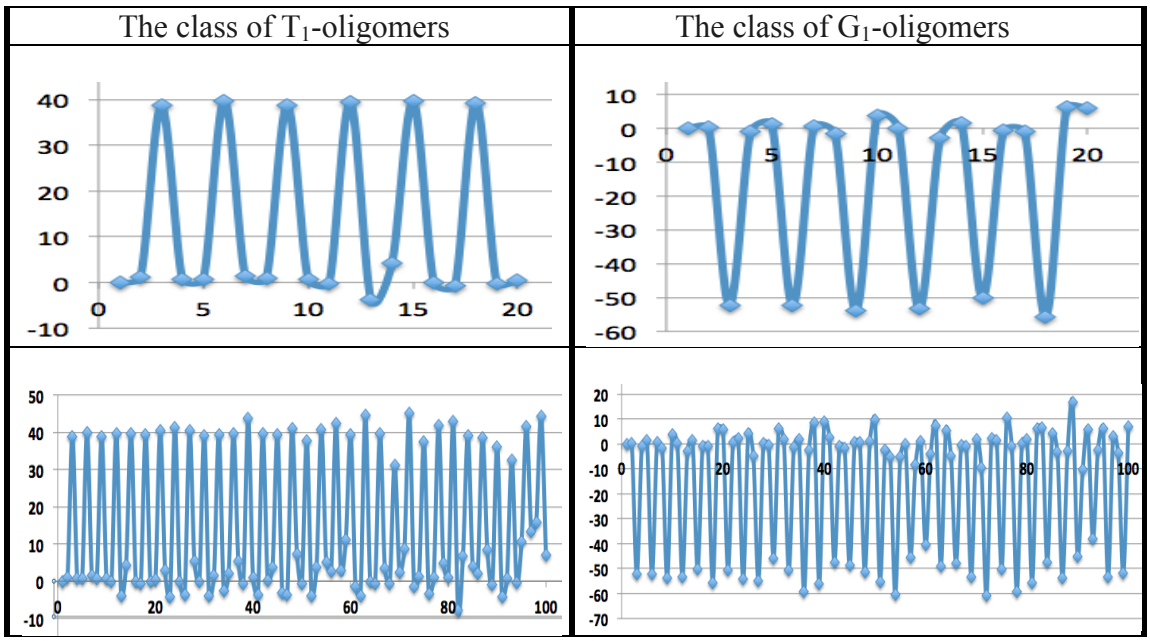


Fig. 8.3. Percentage representations of highly regular rhythmic sequences of the triplet-deviations of the real amounts of n-plets, which belong to classes of T_1-, and G_1-oligomers, from the appropriate model hyperbolic values 19569/n and 18901/n in the *TNT* gene. Here n = 1, 2, 3, …, 20 (upper row) and n = 1, 2, 3,…, 100 (bottom row) as shown at the abscissa axes. The ordinate axes show percentages of the deviations (the model values are taken as 100%).

The nucleobases T and G are keto-nucleobases. Figs. 8.2 and 8.3 draw attention to the phenomenon of long-range correlations in the *TNT* 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. 8.4 shows the OS-sequences of the total amounts of n-plets, which start with two other nucleotides A and C in the *TNT* gene. This gene contains 26373 nucleotides A and 17097 nucleotides C; correspondingly S_A = 26373 and S_C = 17097 for the model hyperbolic progressions (2.2).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

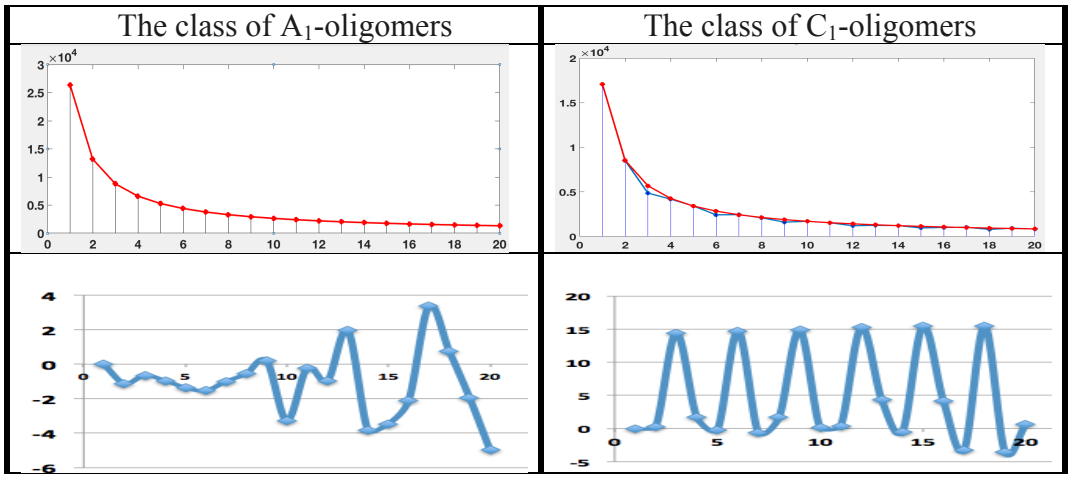


Fig. 8.4. Graphical representations of the results of the analysis - by the oligomer sums method – of the human *TNT* gene regarding the sequences of the total amounts of *n*-plets, which start with the nucleotide A (left) and the nucleotide C (right). Here n = 1, 2, 3,…, 20 (at the absciss axes). **Upper 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 3*m*-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%.

One can see in Fig. 8.4 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. 8.4, the red line of model values covers the line of real values.

Fig. 8.5 shows the numeric results of the analysis of the *TNT* gene by the oligomer sums method.

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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

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. 8.5. 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 *TNT* gene 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 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%).

The coordinated deviations of all four OS-sequences from their model harmonic progressions can be conveniently represented by the general sequence of harmonic mean values, which are calculated for their four corresponding members at each fixed n. Fig. 8.6 shows such a sequence for the human *TNT* gene. One can see the very regular rhythmic nature of this general sequence of harmonic mean values, expressively reflecting the phenomenon of agreed triplet-deviations under $3m$-plets in this gene.

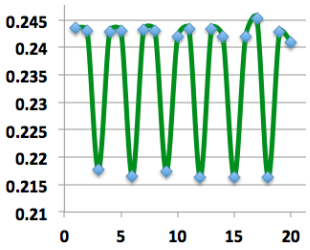


Fig. 8.6. The sequence of harmonic mean values of agreed deviations of all four OS-sequences from their model harmonic progressions $H_{A,1}(n) = 26373/n$, $H_{T,1}(n) = 19569/n$, $H_{C,1}(n) = 17097/n$, and $H_{G,1}(n) = 18901/n$ in the human *TNT* gene. $n = 1, 2, \ldots, 20$ are plotted along the abscissa axes. The ordinate axes show harmonic mean values.

Let us show briefly, for comparison, also the OS-representations of two long human genes: *NEB* gene and *SYNE1* gene.

Figs. 8.7 and 8.8 show graphs with the results of the *NEB* gene by the oligomer sums method. Initial data on this gene were taken from

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

https://www.ncbi.nlm.nih.gov/nuccore/X83957. This gene contains 7071 nucleotides A, 4478 nucleotides T, 4578 nucleotides C, and 4754 nucleotides G, that is S_A = 7071, S_T = 4478, S_C = 4578, and S_G = 4754 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 (7071 and 4478), that is, the second Chargaff's rule on their approximate equality in long sequences is not fulfilled here since this nucleotide sequence is not enough long for the Chargaff's rule.

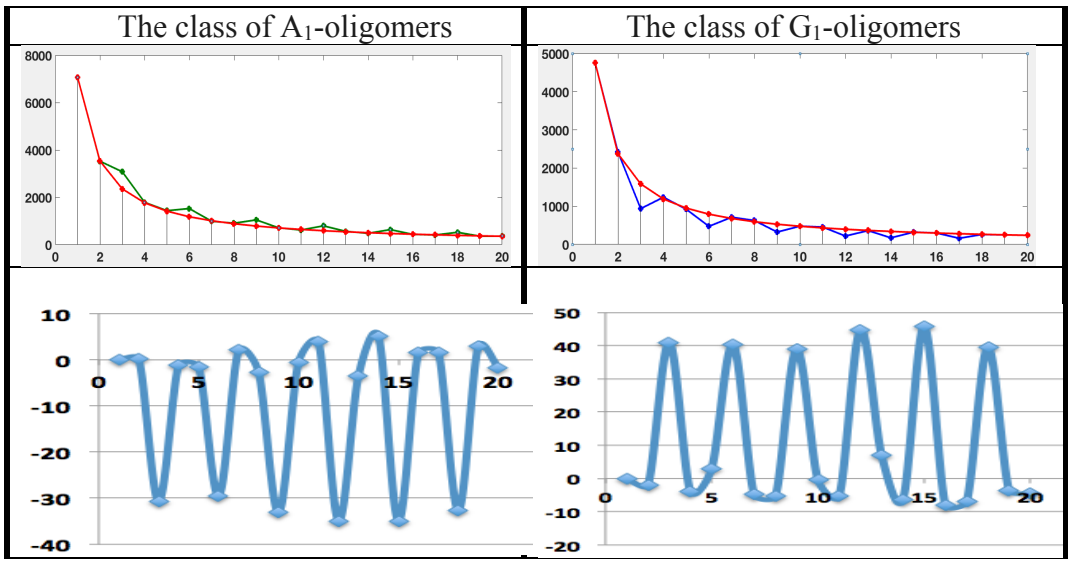
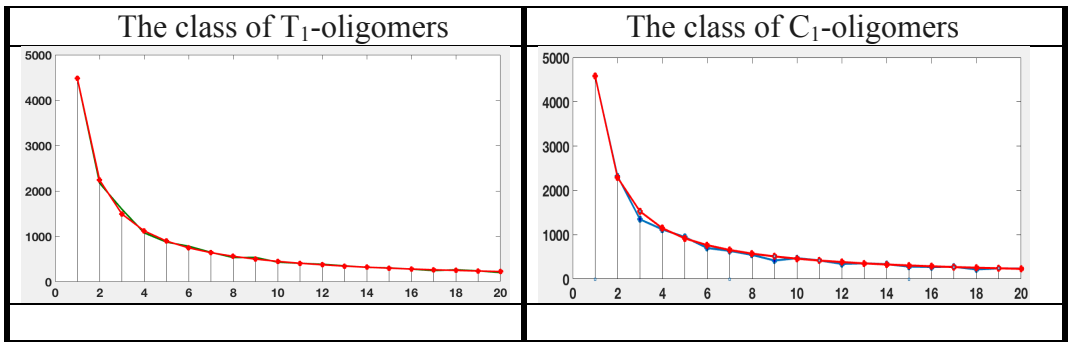


Fig. 8.7. Graphical representations of the results of the analysis - by the oligomer sums method – of the human *NEB* gene: the sequences of the total amounts of *n*-plets, which start with the nucleotide A (left) and the nucleotide G (right) are shown. Here n = 1, 2, 3,…, 20 (at the absciss axes). **Upper row**: the red lines refer to model hyperbolic progressions S_A/n = 7071/n (left) and S_G/n = 4754/n (right). The ordinate axes show the total amounts of appropriate *n*-plets. The highly regular sequences of the significant triplet-deviations at 3*m*-plets shown by the green line (left) and the blue line (right). **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 7071/n and 4754/n (the ordinate axes show these percentages). The model values are taken as 100%.

Fig. 8.8. additionally draws attention to the phenomenon of long-range correlations in the *NEB* gene between sequences of the triplet-deviations in classes of A_1-, and G_1-oligomers: the triplet-deviations in these sequences happen in opposite directions.



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

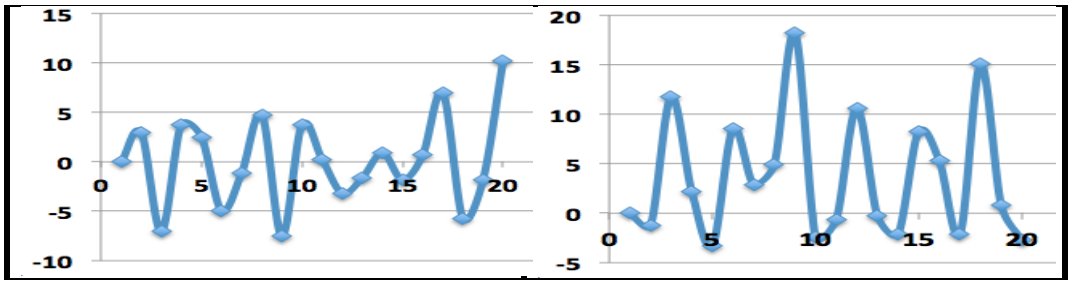


Fig. 8.8. Graphical representations of the results of the analysis - by the oligomer sums method – of the human *NEB* gene: the sequences of the total amounts of *n*-plets, which start with the nucleotide T (left) and the nucleotide C (right) are shown. Here $n = 1, 2, 3,\ldots, 20$ (at the absciss axes). **Upper row**: the red lines refer to model hyperbolic progressions $S_T/n = 4478/n$ (left) and $S_C/n = 4578/n$ (right). The ordinate axes show the total amounts of appropriate *n*-plets. The weakly regular sequences of the significant triplet-deviations at 3*m*-plets shown by the green line (left) and the blue line (right). **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 $4478/n$ and $4578/n$ (the ordinate axes show these percentages). The model values are taken as 100%.

By analogy with Fig. 8.6, Fig. 8.9 shows the sequence of harmonic mean values of agreed deviations of all four OS-sequences from their model harmonic progressions for the case of the human *NEB* gene. One can see the very regular rhythmic nature of this general sequence of harmonic mean values, reflecting the phenomenon of agreed triplet-deviations under 3*m*-plets in this gene.

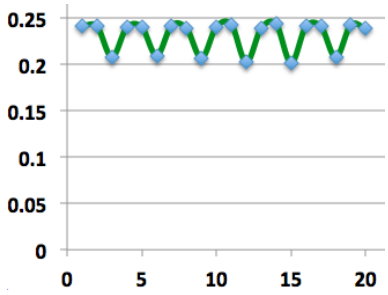


Fig. 8.9. The sequence of harmonic mean values of agreed deviations of all four OS-sequences from their model harmonic progressions in the human *NEB* gene. $n = 1, 2, \ldots, 20$ are plotted along the abscissa axes. The ordinate axis shows harmonic mean values.

Figs. 8.10 and 8.11 show graphs with the results of the human *SYNE1* gene by the oligomer sums method. Initial data on this gene were taken from https://www.ncbi.nlm.nih.gov/nuccore/NM_182961. This gene contains 8697 nucleotides A, 6032 nucleotides T, 5940 nucleotides C, and 7039 nucleotides G, that is $S_A = 8697$, $S_T = 6032$, $S_C = 5940$, and $S_G = 7039$ 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, as are the amounts of nucleotides C and G, and therefore the second Chargaff's rule for long nucleotide sequences is not satisfied here since this nucleotide sequence is not enough long.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

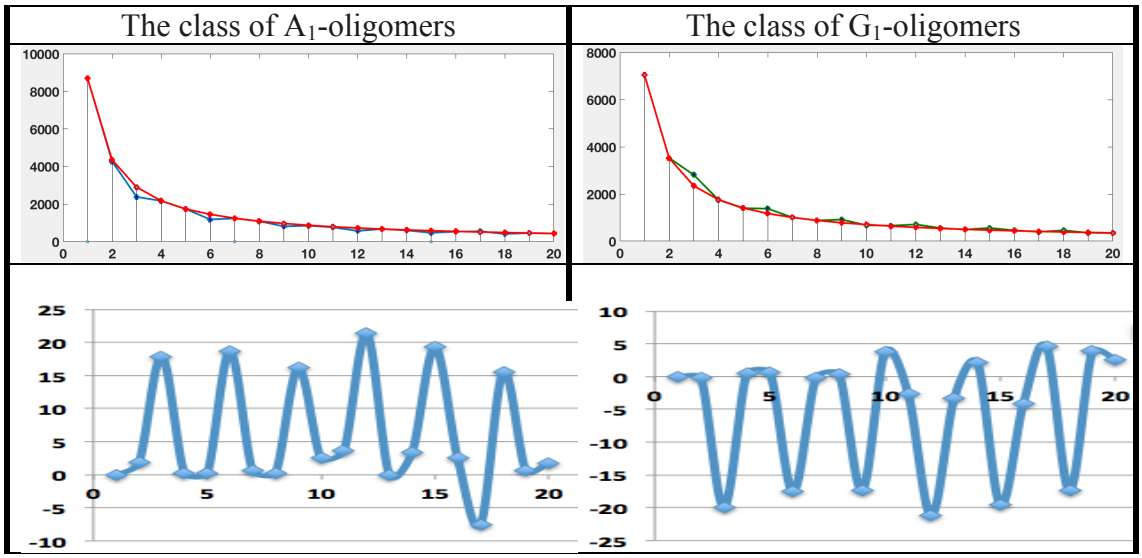
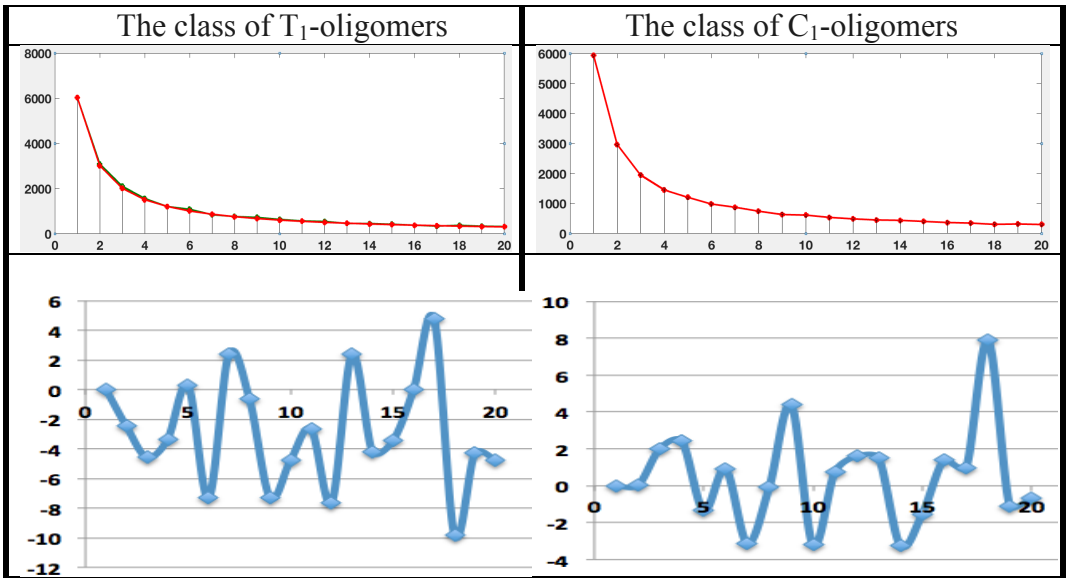


Fig. 8.10. Graphical representations of the results of the analysis - by the oligomer sums method – of the human *SYNE1* gene: the sequences of the total amounts of *n*-plets, which start with the nucleotide A (left) and the nucleotide G (right) are shown. Here n = 1, 2, 3,…, 20 (at the abscis axes). **Upper row**: the red lines refer to model hyperbolic progressions $S_A/n = 8697/n$ (left) and $S_G/n = 7039/n$ (right). The ordinate axes show the total amounts of appropriate *n*-plets. The highly regular sequences of the significant triplet-deviations at 3*m*-plets shown by the blue line (left) and the green line (right). **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 $7071/n$ and $4754/n$ (the ordinate axes show these percentages). The model values are taken as 100%.

Fig. 8.10 additionally draws attention to the phenomenon of long-range correlations in the *SYNE1* gene between sequences of the triplet-deviations in classes of A_1-, and G_1-oligomers: the triplet-deviations in these sequences happen in opposite directions.



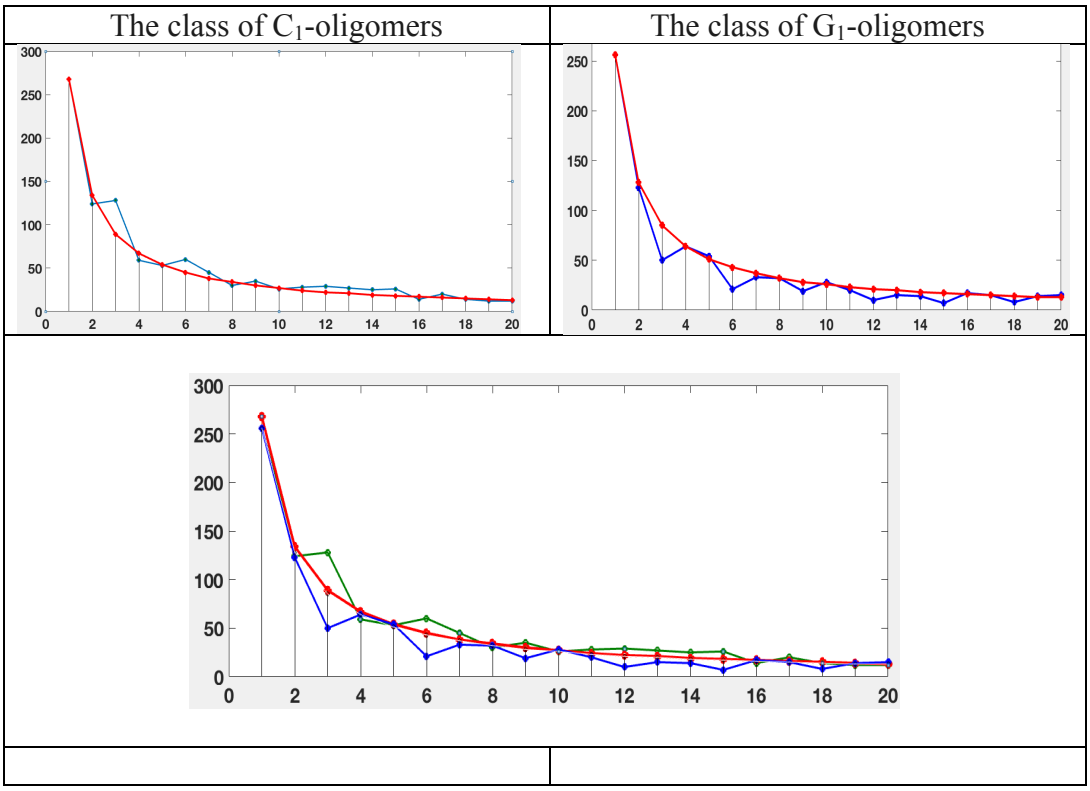
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

Fig. 8.11. Graphical representations of the results of the analysis - by the oligomer sums method – of the human *SYNE1* gene: the sequences of the total amounts of *n*-plets, which start with the nucleotide T (at left) and the nucleotide C (at right) are shown. Here n = 1, 2, 3,…, 20 (at the abscis axes). **Upper row**: the red lines refer to model hyperbolic progressions $S_T/n = 6032/n$ (left) and $S_C/n = 5940/n$ (right). The ordinate axes show the total amounts of appropriate *n*-plets. The triplet-deviations in both of these classes are small in magnitude, and therefore, on these graphs, the model hyperbolic progressions (in red) practically hide under themselves the sequences of real total amounts of 3*m*-plets. **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 $6032/n$ and $5940/n$ (the ordinate axes show these percentages). The model values are taken as 100%.

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.

One of the interesting topics for comparative analysis of genes by the oligomer sums method relates to the structure of histones, which is highly conservative in evolution. Histones are highly basic proteins found in eukaryotic cell nuclei that pack and order the DNA into structural units called nucleosomes. Histones are the chief protein components of chromatin, acting as spools around which DNA winds, and playing a role in gene regulation.

Figs. 8.12 and 8.13 show results of the analysis of one of the short genes of human histones by the OS-method (this gene was randomly selected from multiple histone genes for analysis): H.sapiens H1.1 gene for histone H1, 1034 bp DNA (GenBank: X57130.1, https://www.ncbi.nlm.nih.gov/nuccore/X57130.1). The results confirm the implementation of the hyperbolic (harmonic) rule № 1 for this gene.



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

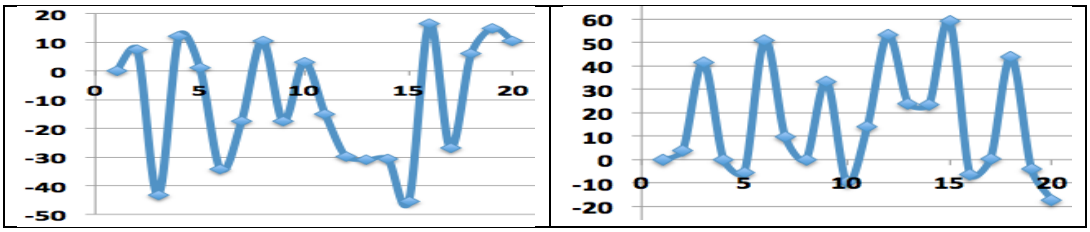
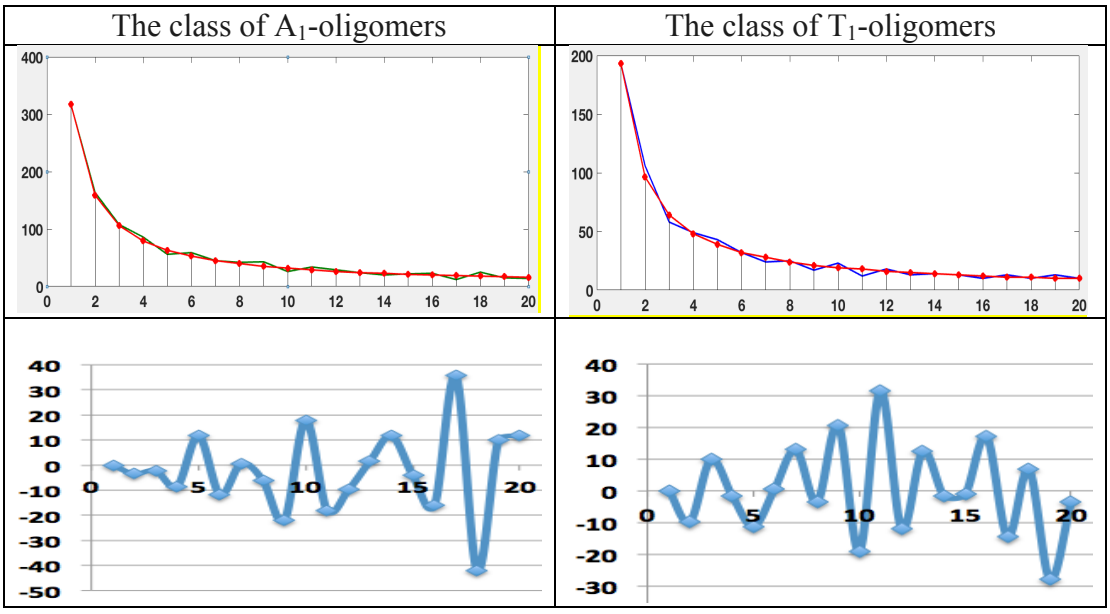


Fig. 8.12. Graphical representations of the results of the analysis - by the oligomer sums method – of the human histone *H1.1* gene regarding its classes of C_1- and G_1-oligomers. The sequences of its total amounts of *n*-plets, which start with the nucleotide C (left) and the nucleotide G (right), jointly with model hyperbolic progressions are shown. **Top graphic row**: the red lines refer to model hyperbolic progressions S_C/n and S_G/n correspondingly, where $S_C = 268$ and $S_G = 256$ are the quantities of nucleotides C and G in the gene; $n = 1, 2, 3, \ldots, 20$ as shown at the abscissa axes. The green line (left) and the blue 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. **Middle graphic row**: the graph combining both graphs from the top row. **Bottom row**: graphs indicate the sequences of percent deviations of the real total amounts of *n*-plets, which start with appropriate nucleotides C and G in the gene, from the model hyperbolic values S_C/n (left) and S_G/n (right) under $n = 1, 2, 3, \ldots, 20$ (at the absciss axis). The ordinate axis indicates values of percent deviations. The model values are taken as 100%.

One can see in Fig. 8.12. the existence in this short histone gene some analog of those triplet-deviations related to 3*m*-plets that were described above for long genes and shown in Figs. 8.1-8.10. In particular, the correlation exists in this short gene between two sequences of the triplet-deviations in the considered classes of C_1- and G_1-oligomers: the triplet-deviations in these sequences happen in opposite directions.

Fig. 8.13 shows the results of a similar analysis of the same histone gene regarding its classes of A_1- and T_1-oligomers.



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

Fig. 8.13. Graphical representations of the results of the analysis - by the oligomer sums method – of the human histone *H1.1* gene regarding its classes of A_1- and T_1-oligomers. **Top graphic row**: the red lines with red dots refer to model hyperbolic progressions S_A/n and S_T/n correspondingly, where $S_A = 317$ and $S_T = 193$ are the quantities of nucleotides A and T; $n = 1, 2, 3, \ldots, 20$ as shown at the abscissa axes. The green line (left) and the blue 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. **Bottom row**: graphs indicate the sequences of percent deviations of the real total amounts of *n*-plets, which start with appropriate nucleotides A and T in the gene, from the model hyperbolic values S_A/n and S_T/n under $n = 1, 2, 3, \ldots, 20$ (at the abscis axis). The ordinate axis indicates values of percent deviations. The model values are taken as 100%.

One can see from Fig. 8.13 that rhythmic deviations in the classes of A_1- and T_1-oligomers are less regular and stable than in the classes of C_1- and G_1-oligomers in Fig. 8.12. To clarify the general picture of such properties of histone genes, systematic studies of a wide set of histone genes in their OS-representations are required.

By analogy with Figs. 8.6 and 8.9, Fig. 8.14 shows the sequence of harmonic mean values of agreed deviations of all four OS-sequences from their model harmonic progressions for the case of the human histone *H1.1* gene. One can see the regular rhythmic nature of this general sequence of harmonic mean values, reflecting the phenomenon of agreed triplet-deviations under 3*m*-plets in this gene.

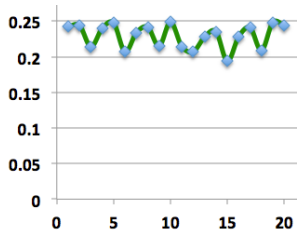


Fig. 8.14. The sequence of harmonic mean values of agreed deviations of all four OS-sequences from their model harmonic progressions in the human histone *H1.1* gene. $n = 1, 2, \ldots, 20$ are plotted along the abscissa axes. The ordinate axis shows harmonic mean values.

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 shown in this Section for individual genes.

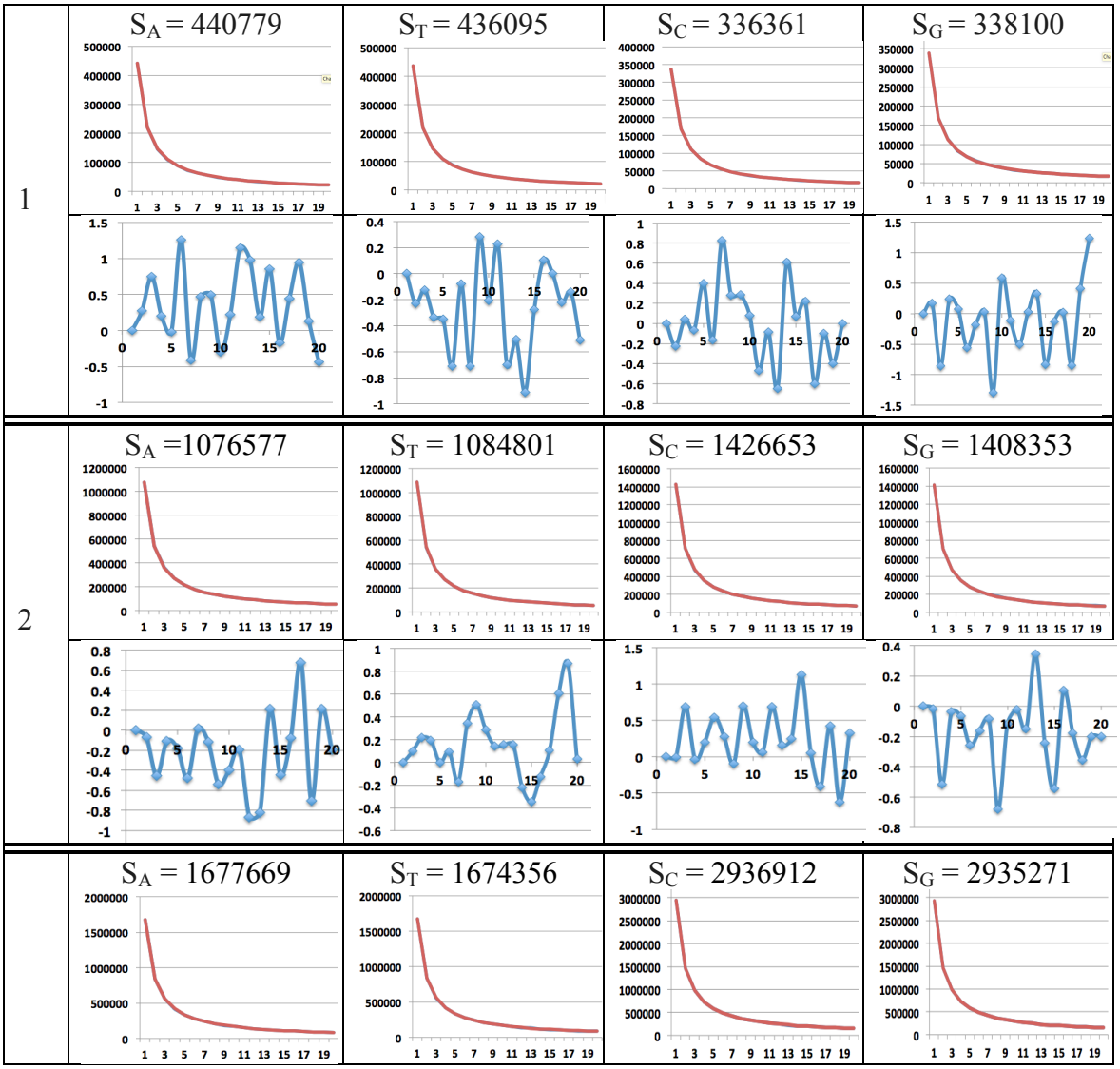
9. The hyperbolic rules in bacterial genomes of different groups both from Bacteria and Archaea.

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 together 19 genomes were used: Aquifex aeolicus, Acidobacteria*

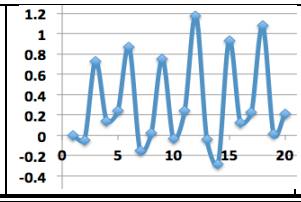
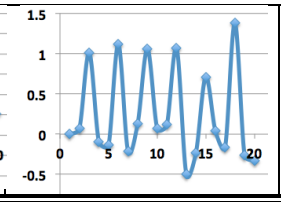
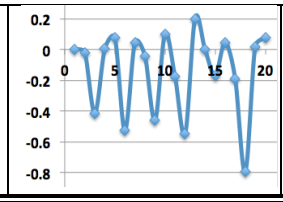
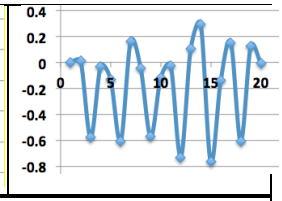
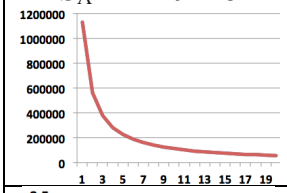
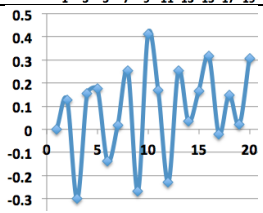
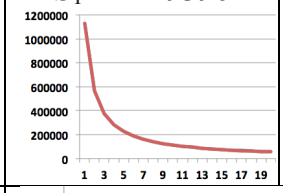
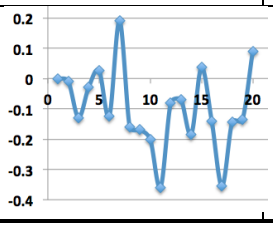
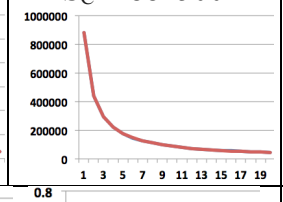
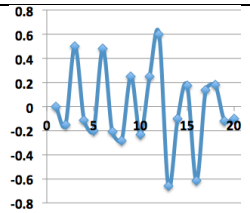
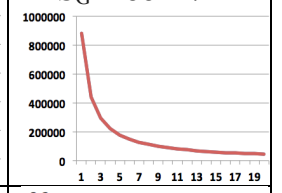
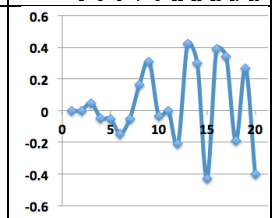
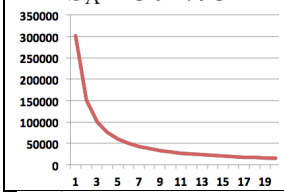
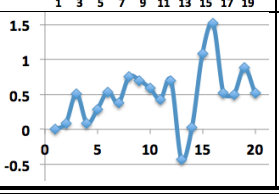
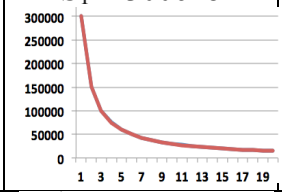
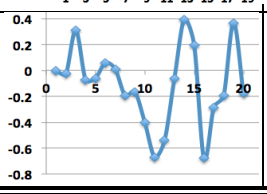
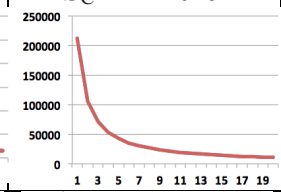
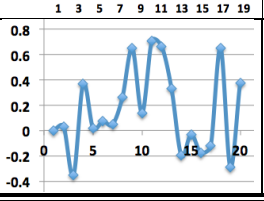
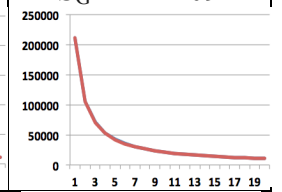
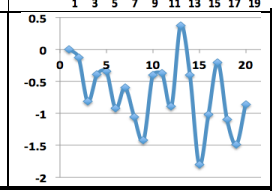
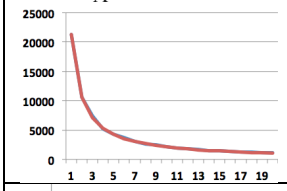
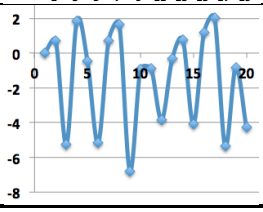
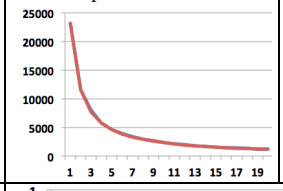
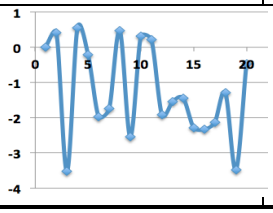
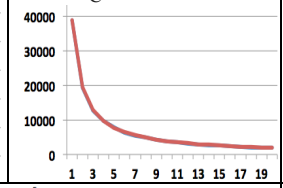
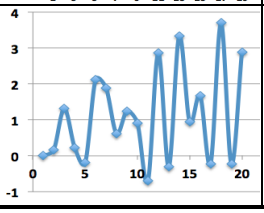
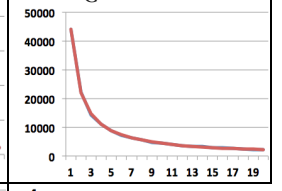
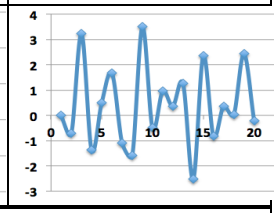
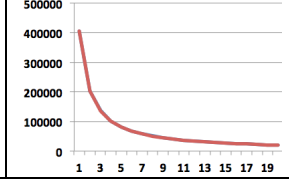
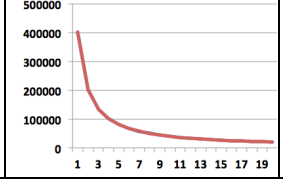
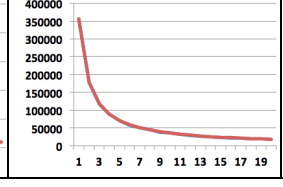
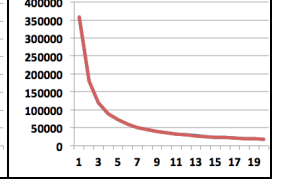
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

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".

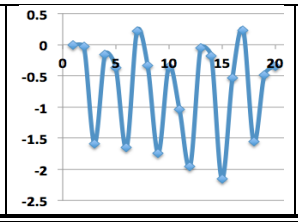
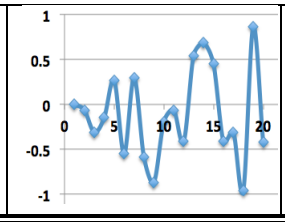
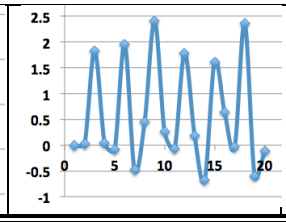
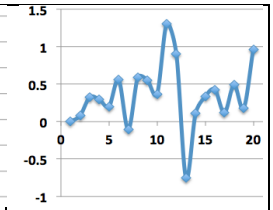
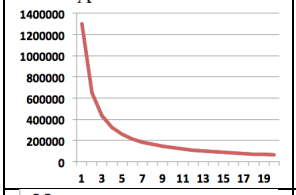
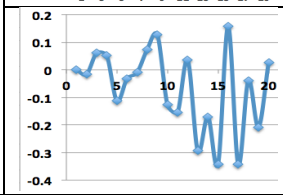
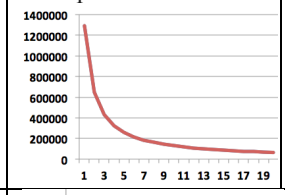
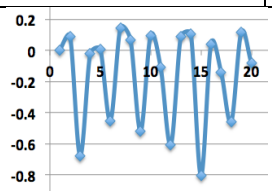
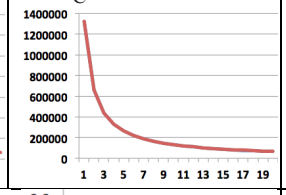
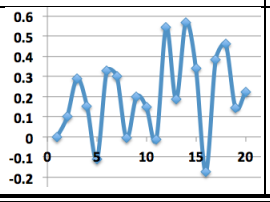
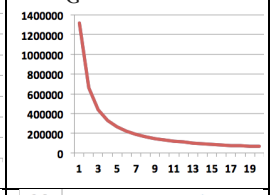
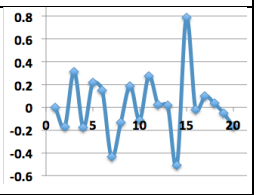
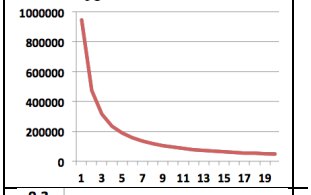
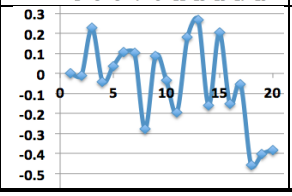
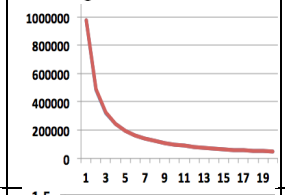
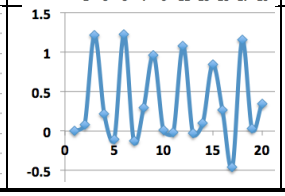
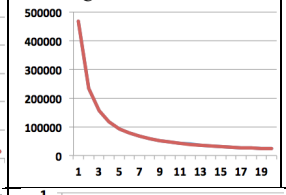
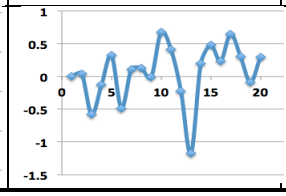
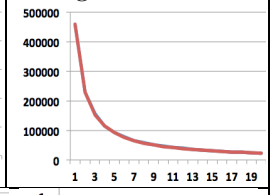
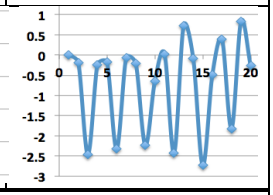
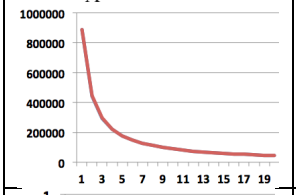
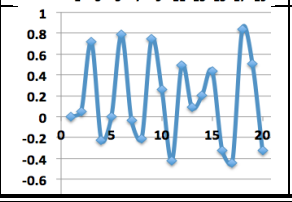
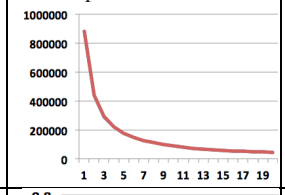
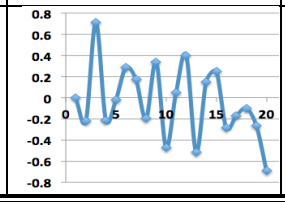
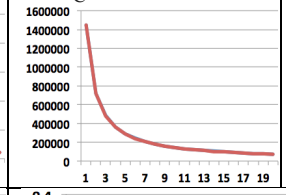
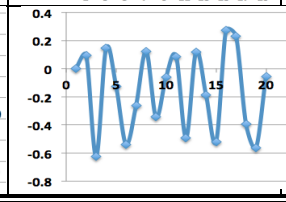
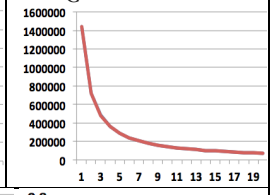
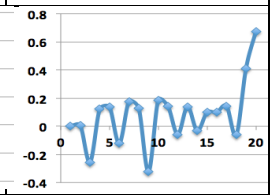
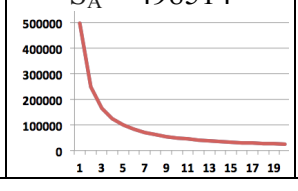
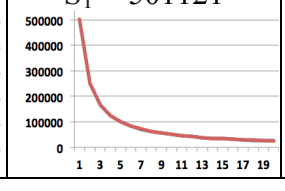
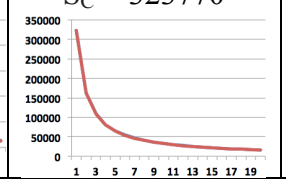
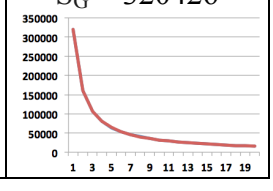
Fig. 9.1 shows the results of the analysis of these prokaryotic genomes by the oligomer sums method. These results demonstrate that the hyperbolic rule No. 1 is fulfilled for all the listed genomes of prokaryotes: the model hyperbolic progressions $H_{A,1}(n) = S_A/n$, $H_{T,1}(n) = S_T/n$, $H_{C,1}(n) = S_C/n$, 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, …, 20. Because of this coincidence, the model hyperbolic progressions, which are represented by red lines in the graphs of Fig. 9.1, 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).



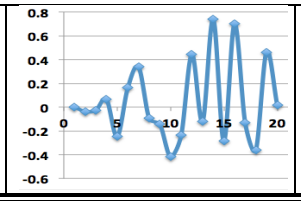
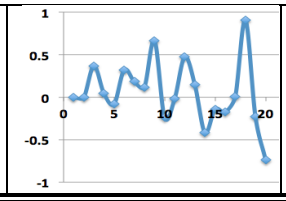
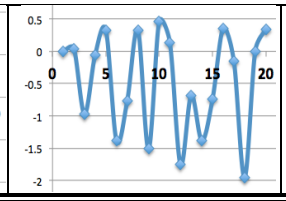
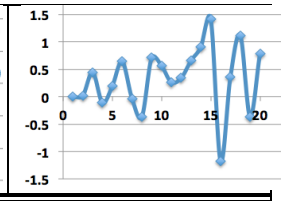
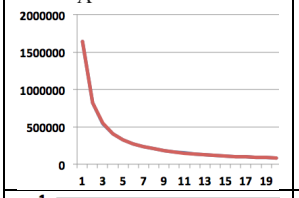
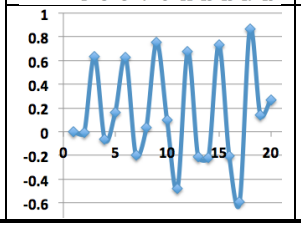
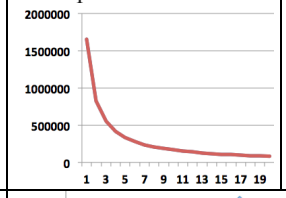
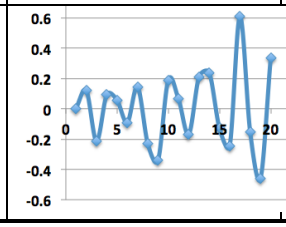
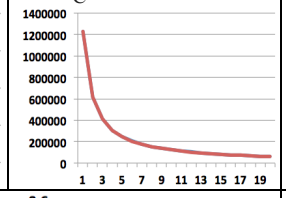
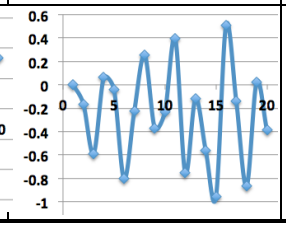
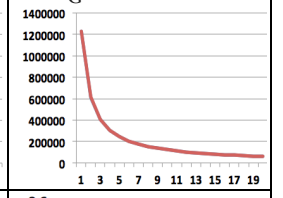
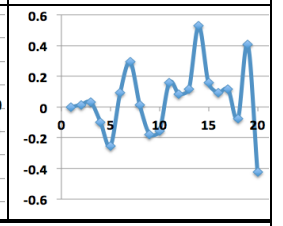
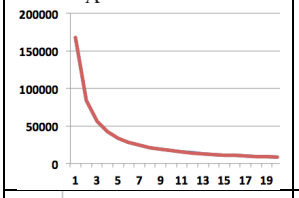
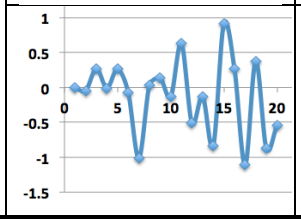
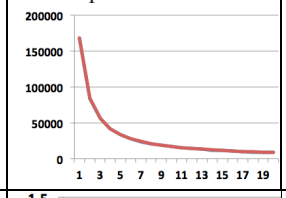
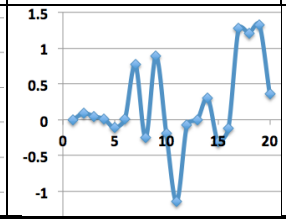
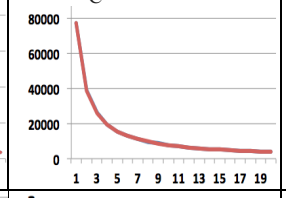
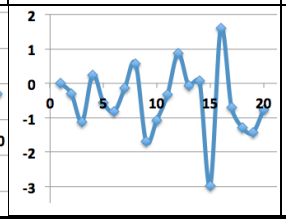
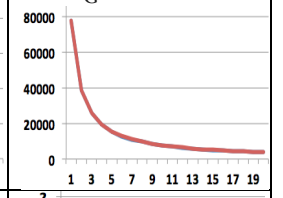
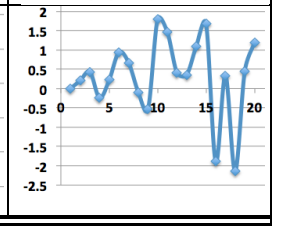
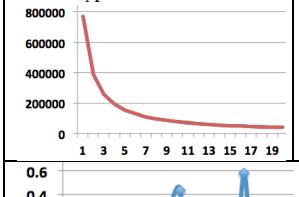
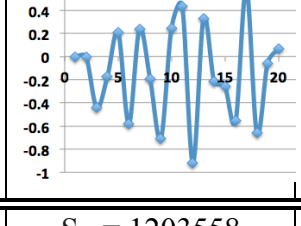
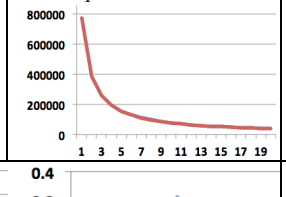
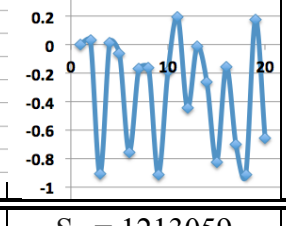
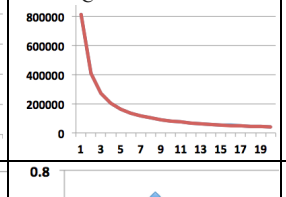
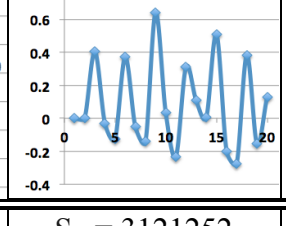
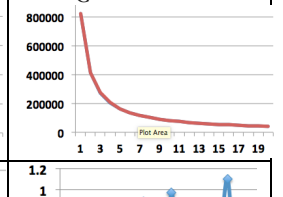
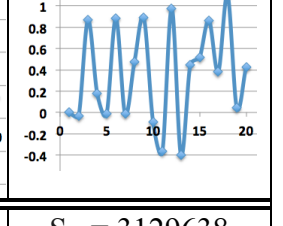
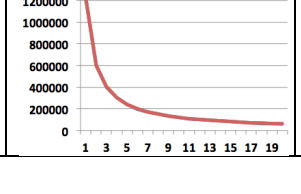
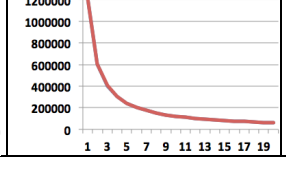
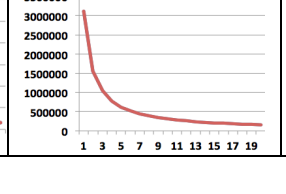
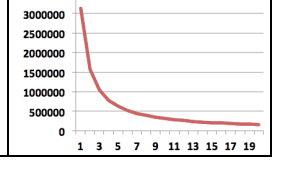
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

3				
4	$S_A = 1129118$  	$S_T = 1129396$  	$S_C = 882500$  	$S_G = 884272$  
5	$S_A = 301793$  	$S_T = 300618$  	$S_C = 212019$  	$S_G = 211409$  
6	$S_A = 21274$  	$S_T = 23172$  	$S_C = 38842$  	$S_G = 44089$  
7	$S_A = 405227$ 	$S_T = 402383$ 	$S_C = 355663$ 	$S_G = 358014$ 

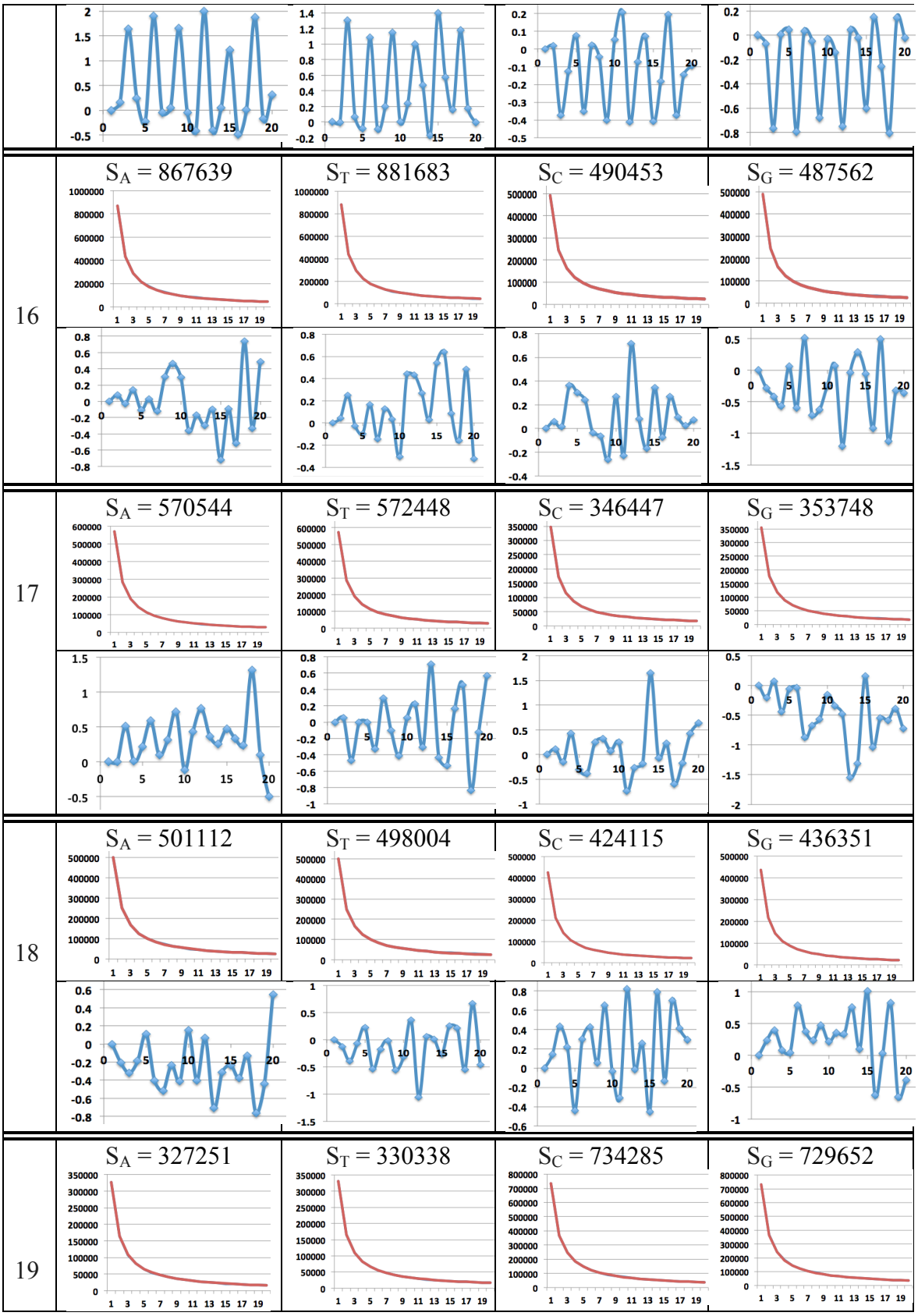
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

				
8	$S_A = 1297551$  	$S_T = 1293044$  	$S_C = 1321325$  	$S_G = 1319228$  
9	$S_A = 945771$  	$S_T = 975318$  	$S_C = 468718$  	$S_G = 458213$  
10	$S_A = 887941$  	$S_T = 882586$  	$S_C = 1444547$  	$S_G = 1443945$  
	$S_A = 498514$ 	$S_T = 501121$ 	$S_C = 323770$ 	$S_G = 320426$ 

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

11				
12	$S_A = 1638004$  	$S_T = 1658700$  	$S_C = 1228410$  	$S_G = 1226378$  
13	$S_A = 167981$  	$S_T = 167983$  	$S_C = 77361$  	$S_G = 77560$  
14	$S_A = 772747$  	$S_T = 770484$  	$S_C = 812772$  	$S_G = 823297$  
15	$S_A = 1203558$ 	$S_T = 1213059$ 	$S_C = 3121252$ 	$S_G = 3129638$ 

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

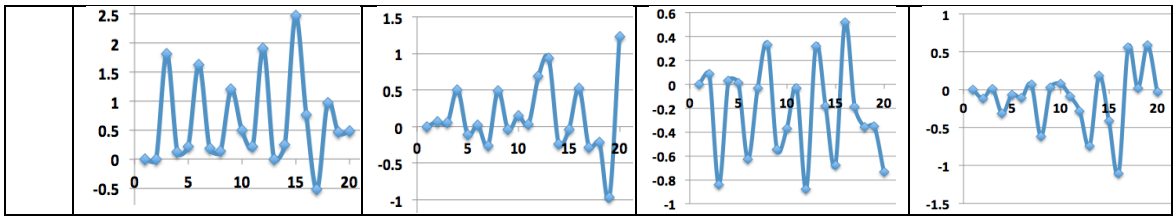


Fig. 9.1. Graphical representations of the results of the analysis - by the oligomer sums method – of 19 bacterial genomes of Bacteria and Archaea mentioned in [Rapoport, Trifonov, 2012, p. 2]. For each of genomes two rows of resulting data are shown at n = 1, 2, …, 20 plotted along the abscissa axes: 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 real values (the ordinate axes show appropriate values); the bottom blue lines show in percent slight alternating deviations of real values from model values. The left column indicates numbers denoted the genomes as explained in the text.

The genomes are enumerated in Fig. 9.1 by numbers 1-19:

1) Aquifex aeolicus VF5, complete genome, 1551335 bp, accession AE000657, version AE000657.1, https://www.ncbi.nlm.nih.gov/nuccore/AE000657.1?report=genbank ;
2) Acidobacteria bacterium KBS 146 M015DRAFT_scf7180000000004_quiver.1_C, whole genome shotgun sequence, 4996384 bp, accession JHVA01000001, https://www.ncbi.nlm.nih.gov/nuccore/JHVA01000001.1?report=genbank;
3) Bradyrhizobium japonicum strain E109, complete genome, 9224208 bp, accession CP010313, https://www.ncbi.nlm.nih.gov/nuccore/CP010313.1?report=genbank ;
4) Bacillus subtilis strain UD1022, complete genome, 4025326 bp, accession CP011534, https://www.ncbi.nlm.nih.gov/nuccore/CP011534.1?report=genbank;
5) Chlamydia trachomatis strain QH111L, complete genome, 1025839 bp, accession CP018052, https://www.ncbi.nlm.nih.gov/nuccore/CP018052.1?report=genbank;
6) Chromobacterium violaceum strain LK30 1, whole genome shotgun sequence, 127377 bp, accession LDUX01000001 version LDUX01000001.1, https://www.ncbi.nlm.nih.gov/nuccore/LDUX01000001.1?report=genbank;
7) Dehalococcoides mccartyi strain CG3, complete genome, NCBI Reference Sequence: NZ_CP013074.1, 1521287 bp, https://www.ncbi.nlm.nih.gov/nuccore/NZ_CP013074.1?report=genbank;
8) Escherichia coli CFT073, complete genome, GenBank: AE014075.1, 5231428 bp, https://www.ncbi.nlm.nih.gov/nuccore/AE014075.1?report=genbank;
9) Flavobacterium psychrophilum JIP02/86, complete genome, 2860382 bp, accession NC_009613, https://www.ncbi.nlm.nih.gov/nuccore/NC_009613.3;
10) Gloeobacter violaceus PCC 7421 DNA, complete genome, GenBank: BA000045.2, 4659019 bp, accession BA000045 AP006568-AP006583 version BA000045.2, https://www.ncbi.nlm.nih.gov/nuccore/BA000045.2?report=genbank;
11) Helicobacter pilory, NCBI Reference Sequence: NC_000921.1, complete genome, 1643831 bp, accession NC_000921 NZ_AE001440-NZ_AE001571

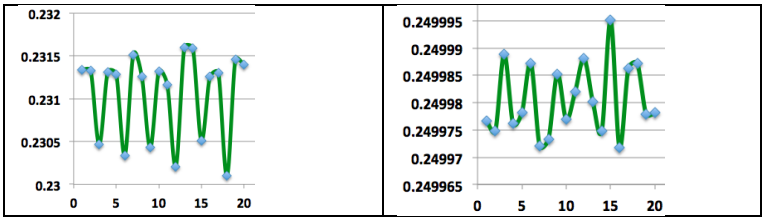
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

version NC_000921.1, https://www.ncbi.nlm.nih.gov/nuccore/NC_000921.1;

12) Methanosarcina acetivorans str. C2A, complete genome, 5751492 bp, accession AE010299 AE010656-AE011189 version AE010299.1, https://www.ncbi.nlm.nih.gov/nuccore/AE01029;
13) Nanoarchaeum equitans Kin4-M, complete genome, 490885 bp, accession AE017199 AACL01000000 AACL01000001 version AE017199.1, https://www.ncbi.nlm.nih.gov/nuccore/AE017199.1?report=genbank;
14) Syntrophus aciditrophicus SB, complete genome, 3179300 bp, accession CP000252, https://www.ncbi.nlm.nih.gov/nuccore/CP000252.1?report=genbank;
15) Streptomyces coelicolor A3(2) complete genome, 8667507 bp, accession AL645882, https://www.ncbi.nlm.nih.gov/nuccore/AL645882.2?report=genbank;
16) Sulfolobus solfataricus strain SULA, complete genome, 2727337 bp, accession CP011057, https://www.ncbi.nlm.nih.gov/nuccore/CP011057.1?report=genbank;
17) Treponema denticola SP33 supercont1.1, whole genome shotgun sequence, NCBI Reference Sequence: NZ_KB442453.1, 1850823 bp, accession NZ_KB442453 NZ_AGDZ01000000 version NZ_KB442453.1, https://www.ncbi.nlm.nih.gov/nuccore/NZ_KB442453.1?report=genbank;
18) Thermotoga maritima strain Tma200, complete genome, 1859582 bp, accession CP010967, https://www.ncbi.nlm.nih.gov/nuccore/CP010967.1?report=genbank;
19) Thermus thermophilus DNA, complete genome, strain: TMY, 2121526 bp, accession AP017920, https://www.ncbi.nlm.nih.gov/nuccore/AP017920.1?report=genbank

One can see from Fig. 9.1 that in some prokaryotic genomes (for example in №№ 3, 7, 9, and 15) the alternating small deviations of real values from model values are systematic and related to 3*m*-plets; it seems to be analogous to the much stronger triplet-deviations described above for human genes in Figs. 8.1-8.7. Can a sign of the presence of such triplet-deviations in the genomes of some bacteria serve as a criterion for the selection of bacterial species for genetic engineering problems? It is one of many new questions arisen due to the discovery of the represented hyperbolic rules and the applications of the oligomer sums method.

Fig. 9.2 shows examples of sequences of the harmonic mean values for two of these bacterial genomes. One can see triplet-deviations in these sequences at dots corresponding to 3*m*-plets.



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

Fig. 9.2. The sequences of harmonic mean values of agreed deviations of all four OS-sequences from their model harmonic progressions in the genomes of Bradyrhizobium japonicum strain E109 (left) and Escherichia coli CFT073 (right). n = 1, 2, …, 20 are plotted along the abscissa axes. The ordinate axes show harmonic mean values.

10. 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. The initial data on the genomes were taken from the GenBank. Figs. 10.1-10.8 show the results of their analysis.

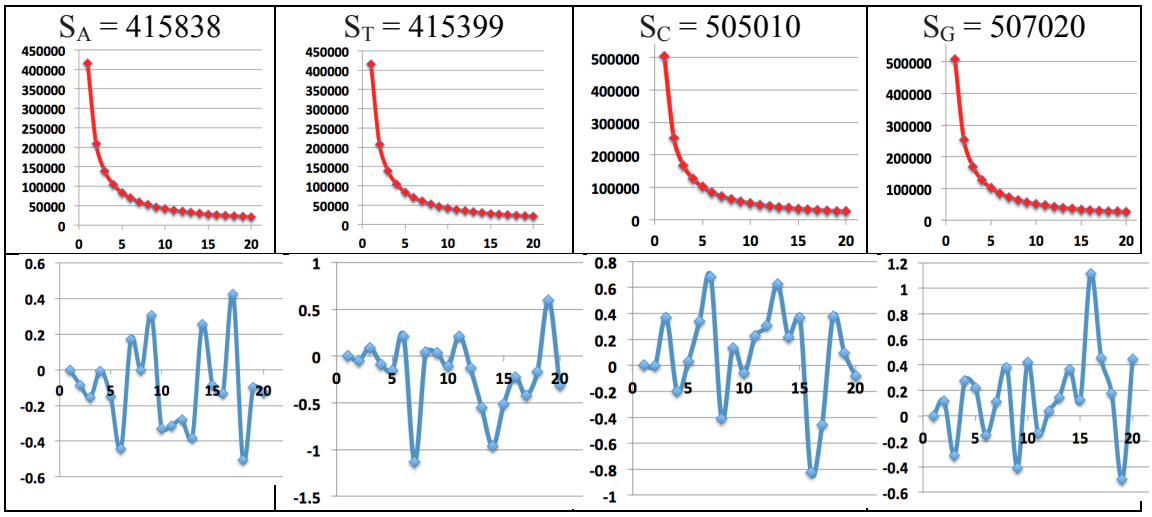
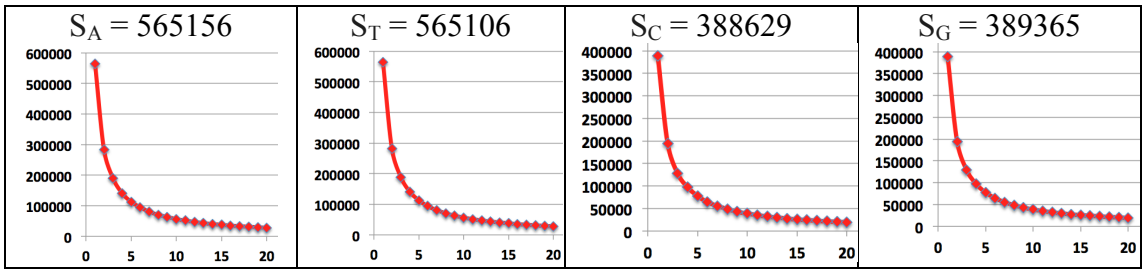


Fig. 10.1. The results of the analysis - by the oligomer sums method – the extremophile *Pyrolobus fumarii 1A*, complete genome, 1843267 bp (this extremophile lives in submarine hydrothermal vents), https://www.ncbi.nlm.nih.gov/nuccore/NC_015931.1. All abscissa axes show the values n = 1, 2, …, 20. The top row demonstrates 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 real values (the ordinate axes show appropriate values). The bottom row show in percent slight alternating deviations of real values of the OS-sequences from model values.



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

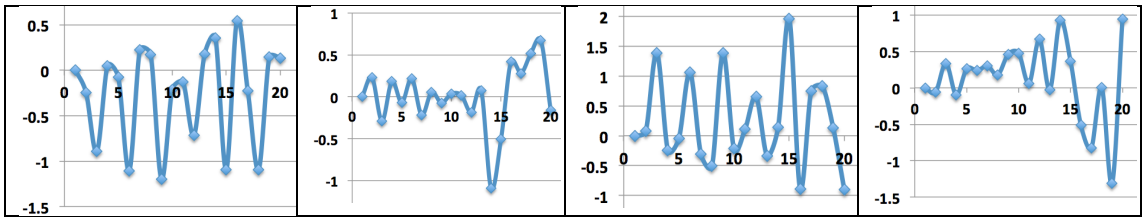


Fig. 10.2. The results of the analysis - by the oligomer sums method – the extremophile *Pyrococcus furiosus DSM 3638*, complete genome, 1908256 (this extremophile lives in submarine hydrothermal vents), *https://www.ncbi.nlm.nih.gov/nuccore/NC_003413.1*
The explanation of these graphs is identical to the explanation to Fig. 10.1.

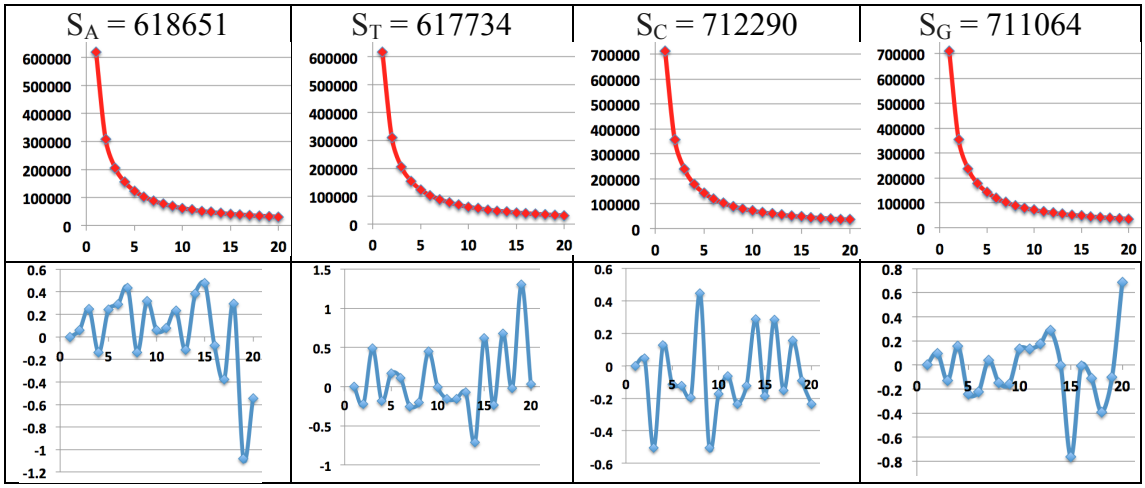


Fig. 10.3. The results of the analysis - by the oligomer sums method – the extremophile *Synechococcus lividus PCC 6715* chromosome, complete genome, 2659739 bp (this extremophile lives in low temperature conditions), https://www.ncbi.nlm.nih.gov/nuccore/NZ_CP018092.1). The explanation of these graphs is identical to the explanation to Fig. 10.1.

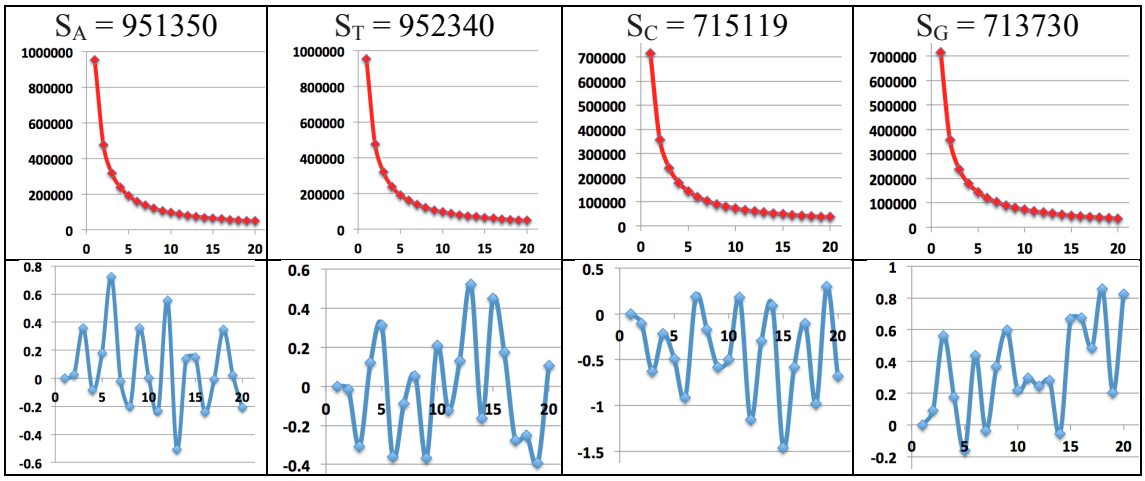


Fig. 10.4. The results of the analysis - by the oligomer sums method – the extremophile *Psychrobacter alimentarius strain PAMC 27889* chromosome,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

complete genome, 3332539 bp (this extremophile lives in soda lakes), https://www.ncbi.nlm.nih.gov/nuccore/NZ_CP014945.1. The explanation of these graphs is identical to the explanation to Fig. 10.1.

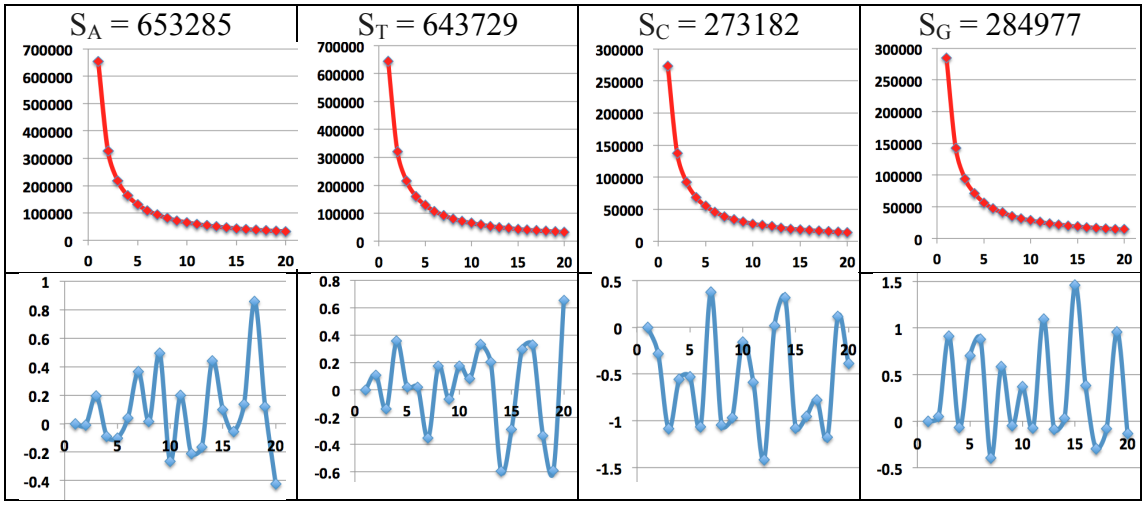


Fig. 10.5. The results of the analysis - by the oligomer sums method – the extremophile *Clostridium paradoxum* JW-YL-7 = DSM 7308 strain JW-YL-7 ctg1, whole genome shotgun sequence, 1855173 bp (this extremophile lives in volcanic springs, acid mine drainage), https://www.ncbi.nlm.nih.gov/nuccore/LSFY01000001.1The explanation of these graphs is identical to the explanation to Fig. 10.1.

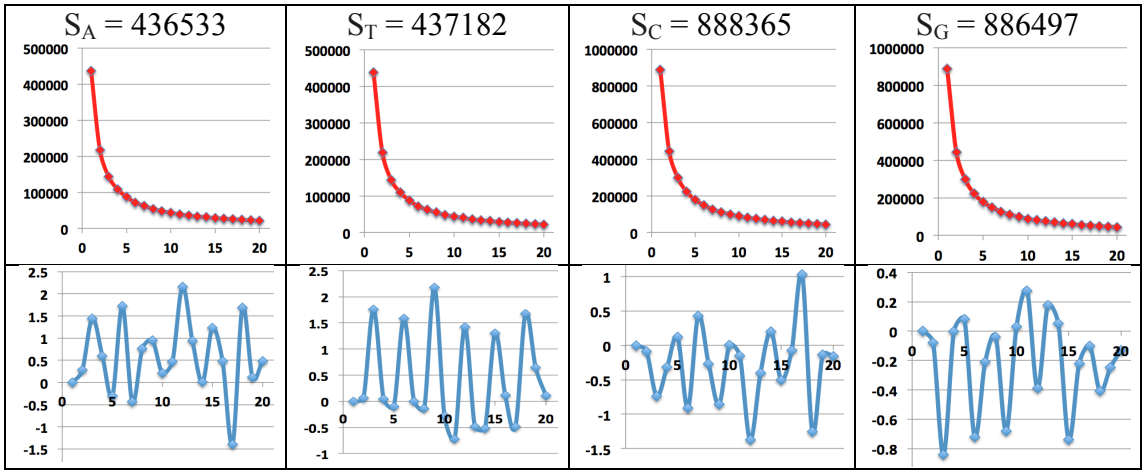


Fig. 10.6. The results of the analysis - by the oligomer sums method – the extremophile *Deinococcus radiodurans R1* chromosome 1, complete sequence, 2648638 bp, (this extremophile lives in conditions of cosmic rays, X-rays, radioactive decay), https://www.ncbi.nlm.nih.gov/nuccore/NC_001263.1. The explanation of these graphs is identical to the explanation to Fig. 10.1.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

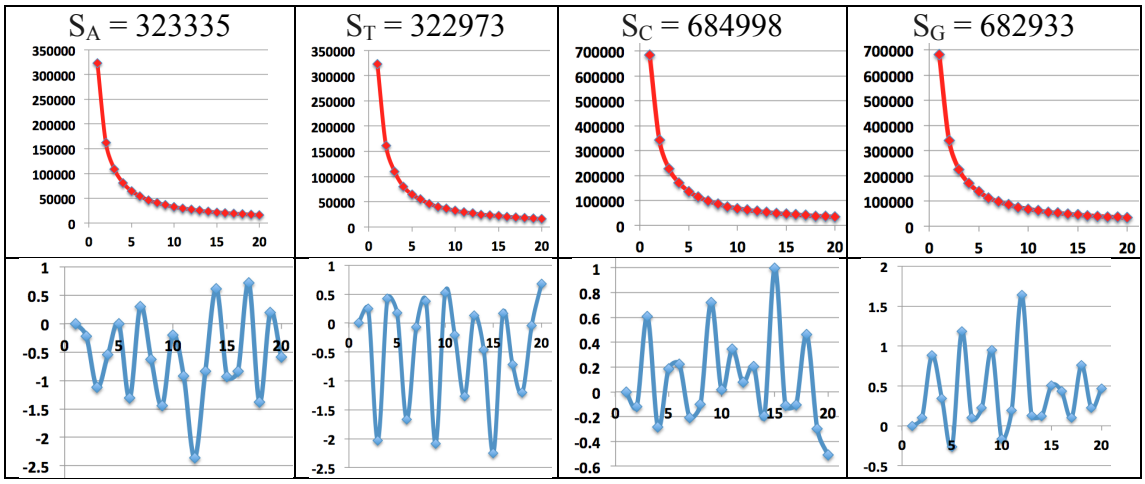


Fig. 10.7. The results of analysis - by the oligomer sums method – the extremophile *Halobacterium sp. NRC-1*, complete genome, 2014239 bp (this extremophile lives in conditions of high salt concentration), https://www.ncbi.nlm.nih.gov/nuccore/NC_002607.1. The explanation of these graphs is identical to the explanation to Fig. 10.1.

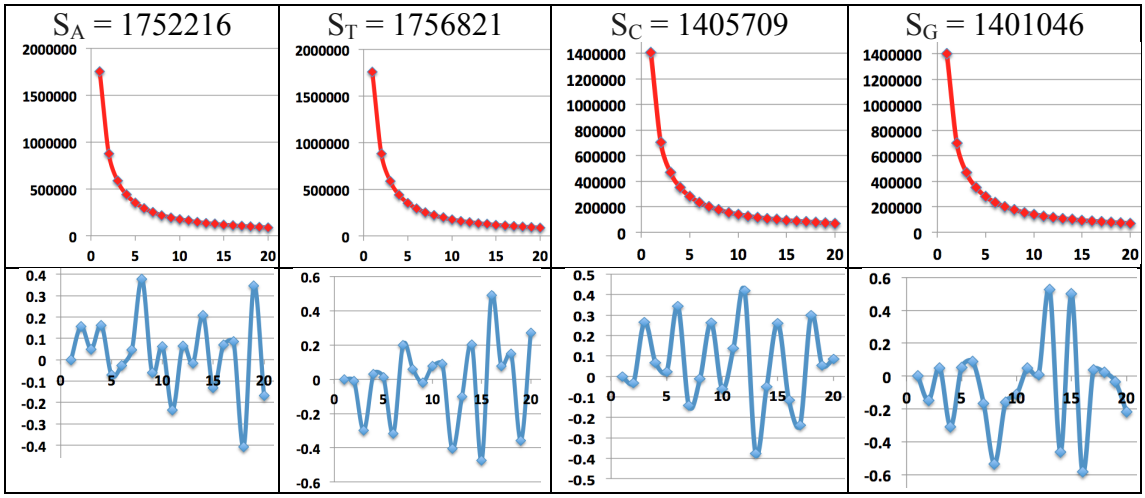


Fig. 10.8. The results of the analysis - by the oligomer sums method – the extremophile *Chroococcidiopsis thermalis PCC 7203*, complete genome, 6315792 bp, (this extremophile lives in conditions of desiccation), https://www.ncbi.nlm.nih.gov/nuccore/NC_019695.1. The explanation of these graphs is identical to the explanation to Fig. 10.1.

The resulting data in Figs. 10.1- 10.8 shows the fulfillment of the hyperbolic rule №1 of oligomeric sums for all studied and presented extremophiles. The extremal living conditions of these microorganisms do not affect the subordination of their genomes to the described hyperbolic (harmonic) rules of algebraic invariance, which are true for the genomes of other prokaryotes and eukaryotes.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

11. 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 (Figs. 11.1-11.4).

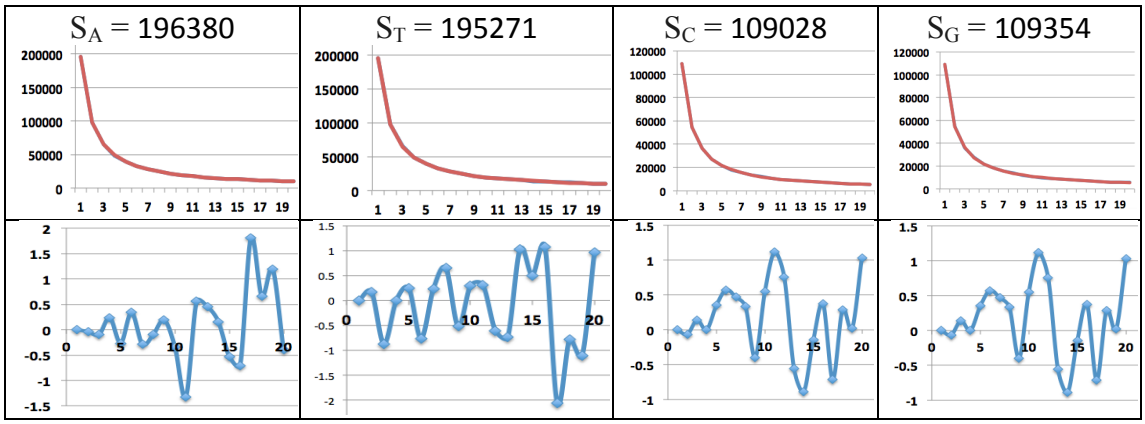


Fig. 11.1. The results of the analysis - by the oligomer sums method – the giant virus Pithovirus sibericum isolate P1084-T, complete genome, 610033 bp, NCBI Reference Sequence: NC_023423.1 https://www.ncbi.nlm.nih.gov/nuccore/NC_023423.1. All abscissa axes show the values n = 1, 2, …, 20. The top row demonstrates 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 real values (the ordinate axes show appropriate values). The bottom row show in percent slight alternating deviations of real values of the OS-sequences from model values.

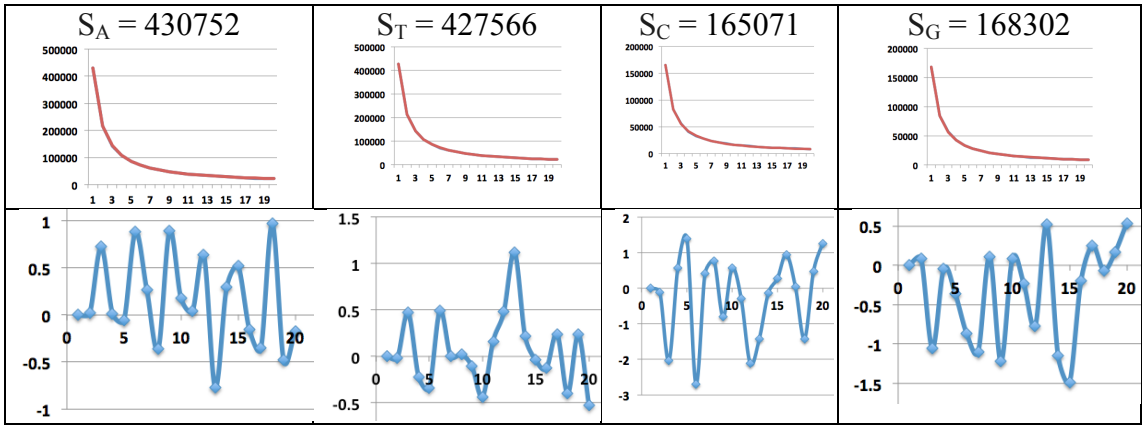


Fig. 11.2. The results of the analysis - by the oligomer sums method – the giant virus Acanthamoeba castellanii mamavirus strain Hal-V, complete genome, 1191693 bp, GenBank: JF801956.1, https://www.ncbi.nlm.nih.gov/nuccore/JF801956.1. The explanation of these graphs is identical to the explanation to Fig. 11.1.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

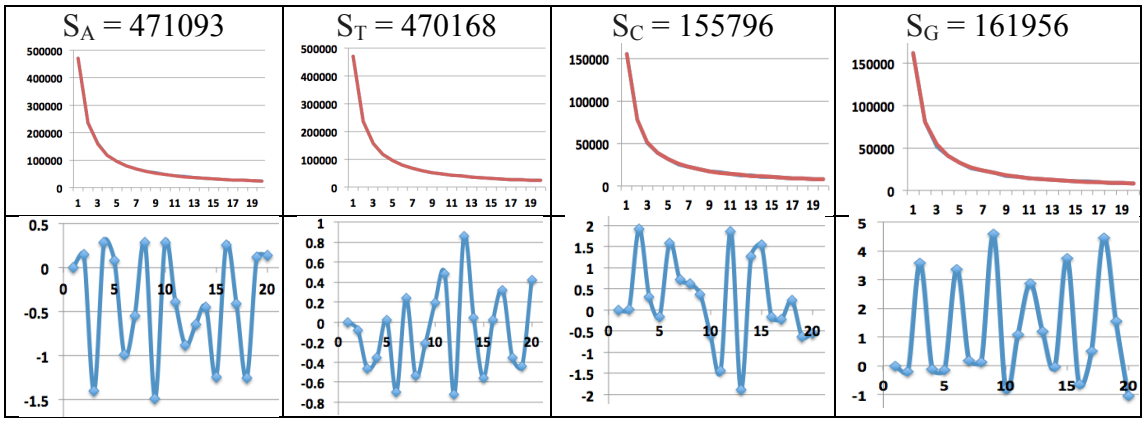


Fig. 11.3. The results of the analysis - by the oligomer sums method – the giant virus Megavirus chiliensis, complete genome, 1259197 bp, NCBI Reference Sequence: NC_016072.1, https://www.ncbi.nlm.nih.gov/nuccore/NC_016072.1. The explanation of these graphs is identical to the explanation to Fig. 11.1.

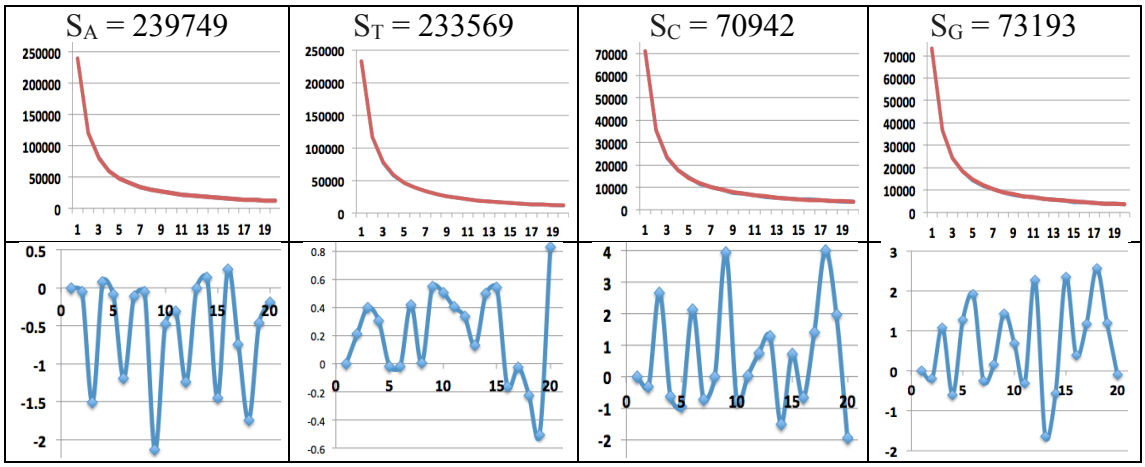


Fig. 11.4. The results of the analysis - by the oligomer sums method – the giant virus Cafeteria roenbergensis virus BV-PW1, complete genome, 617453 bp, NCBI Reference Sequence: NC_014637.1, https://www.ncbi.nlm.nih.gov/nuccore/NC_014637.1. The explanation of these graphs is identical to the explanation to Fig. 11.1.

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

11. 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, 29903 bp ss-RNA linear VRL 18-MAR-2020.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

Figs. 12.1-12.5 show some results of such an analysis of the virus.

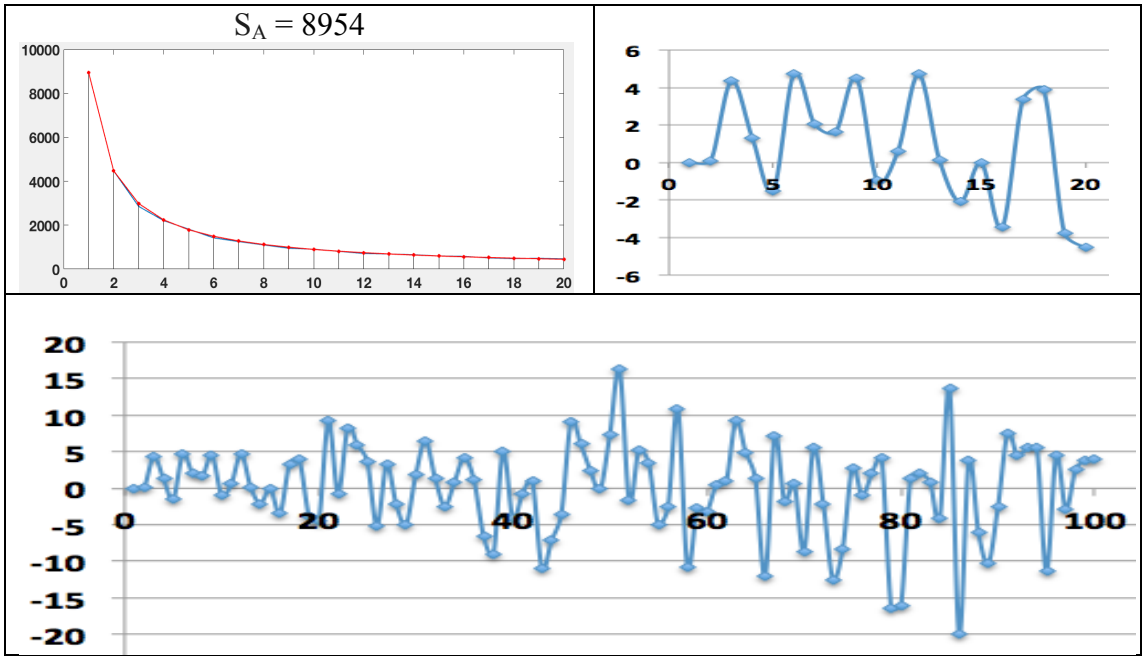
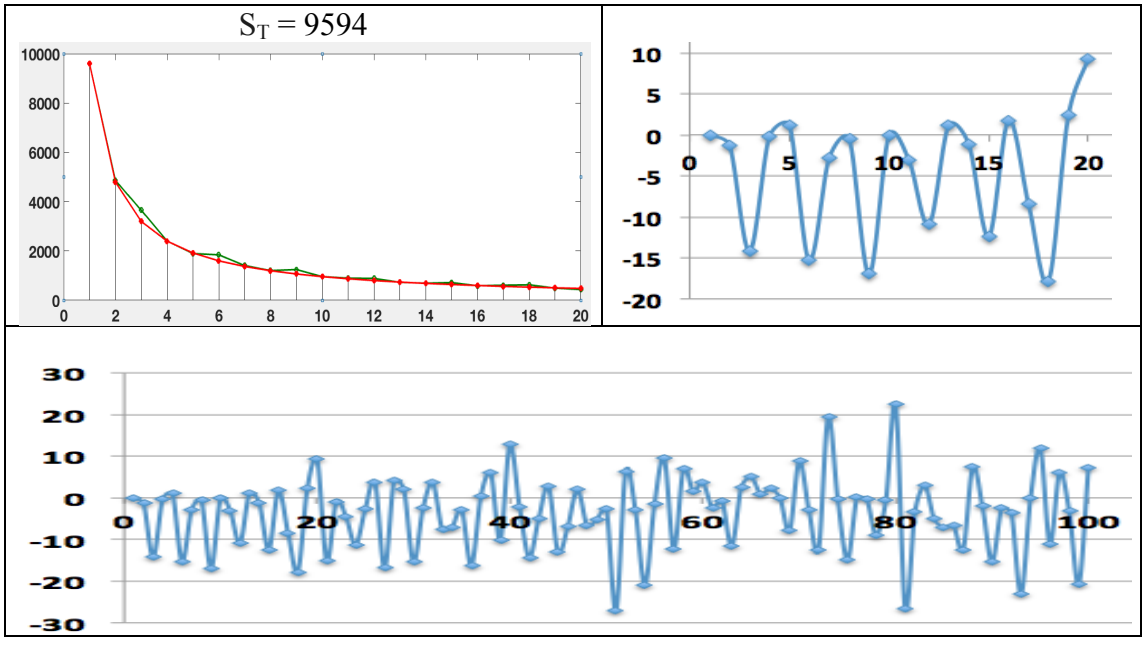


Fig. 12.1. The graphs for the case of the OS-sequences of n-plets from the class A_1-oligomers of the coronavirus 2 isolate Wuhan-Hu-1, complete genome, GenBank: MN908947.3, LOCUS MN908947, 29903 bp. In these graphs, the abscissa axis represents the values n = 1, 2, 3, …, 20 (in top row) and n = 1, 2, 3, …, 100 (in bottom row). **Top left**: the ordinate axis represents the set of phenomenological total amounts $\Sigma_{A,n,1}$ of n-plets beginning with the nucleotide A. **Top right, and bottom**: deviations of real OS-sequences $\Sigma_{A,n,1}$ from model hyperbolic progressions S_A/n = 8954/n in percentages.



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

Fig. 12.2. The graphs for the case of the OS-sequences of *n*-plets from the class T_1-oligomers of the coronavirus 2 isolate Wuhan-Hu-1, complete genome, GenBank: MN908947.3, LOCUS MN908947, 29903 bp. In these graphs, the abscissa axis represents the values n = 1, 2, 3, …, 20 (in top row) and n = 1, 2, 3, …, 100 (in bottom row). **Top left**: the ordinate axis represents the set of phenomenological total amounts $\Sigma_{T,n,1}$ of *n*-plets beginning with the nucleotide T. **Top right, and bottom**: deviations of the real OS-sequence $\Sigma_{T,n,1}$ from the model hyperbolic progression S_T/n = 9594/n in percentages.

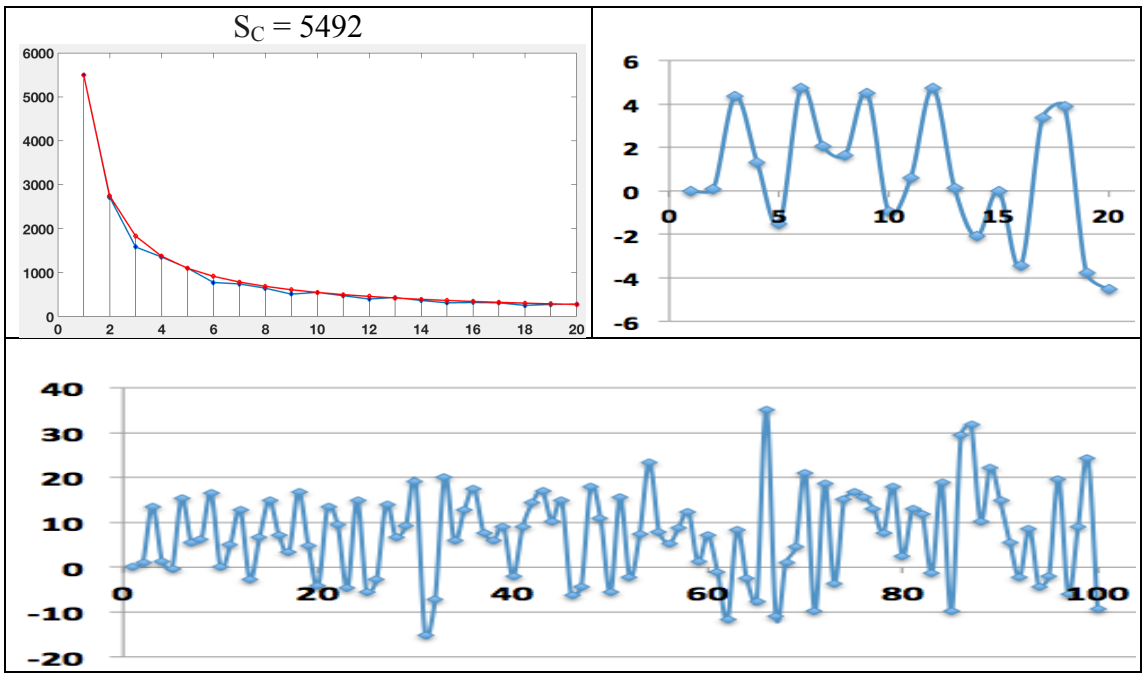
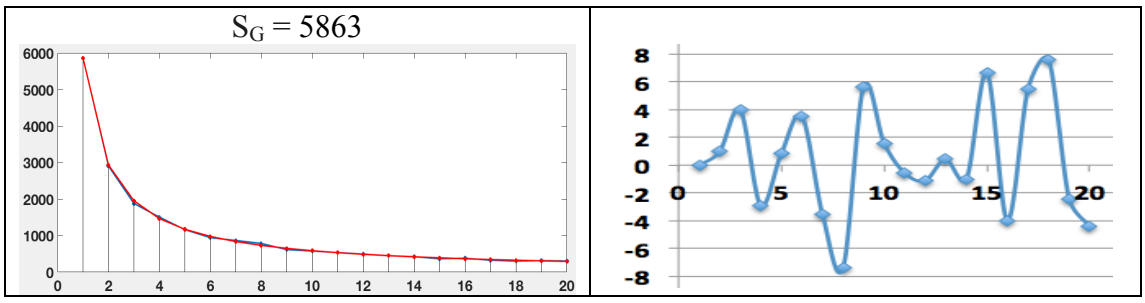


Fig. 12.3. The graphs for the case of the OS-sequences of *n*-plets from the class C_1-oligomers of the coronavirus 2 isolate Wuhan-Hu-1, complete genome, GenBank: MN908947.3, LOCUS MN908947, 29903 bp. In these graphs, the abscissa axis represents the values n = 1, 2, 3, …, 20 (in top row) and n = 1, 2, 3, …, 100 (in bottom row). **Top left**: the ordinate axis represents the set of phenomenological total amounts $\Sigma_{C,n,1}$ of *n*-plets beginning with the nucleotide C. **Top right, and bottom**: deviations of the real OS-sequences $\Sigma_{C,n,1}$ from the model hyperbolic progression S_C/n = 5492/n in percentages.



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

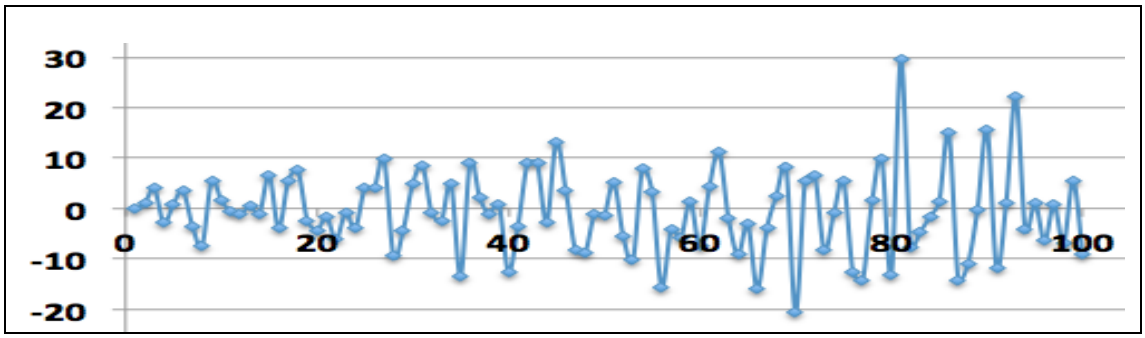


Fig. 12.4. The graphs for the case of the OS-sequences of *n*-plets from the class G_1-oligomers of the coronavirus 2 isolate Wuhan-Hu-1, complete genome, GenBank: MN908947.3, LOCUS MN908947, 29903 bp. In these graphs, the abscissa axis represents the values n = 1, 2, 3, …, 20 (in top row) and n = 1, 2, 3, …, 100 (in bottom row). **Top left**: the ordinate axis represents the set of phenomenological total amounts $\Sigma_{G,n,1}$ of *n*-plets beginning with the nucleotide G. **Top right, and bottom**: deviations of the real OS-sequence $\Sigma_{G,n,1}$ from the model hyperbolic progression S_G/n = 5863/n in percentages.

In particular, Figs. 12.1-12.4 show that this virus in its OS-representations has under n = 3, 6, 9, ..., 3m such deviations of real values from model values, which resemble the triplet-deviations in human genes, which were described above in Figs. 8.1-8.10. 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 (Figs. 12.2 and 12.3) - 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.

By analogy with Figs. 8.6, 8.9, and 8.14, Fig. 12.5 shows the sequence of harmonic mean values of agreed deviations of all four OS-sequences from their model harmonic progressions for the case of the coronavirus 2 isolate Wuhan-Hu-1. Two cases are shown: for n = 1, 2, …, 20 and n = 1, 2, …, 100. One can see the regular rhythmic nature of this general sequence of harmonic mean values, reflecting the phenomenon of agreed triplet-deviations under 3m-plets in this coronavirus.

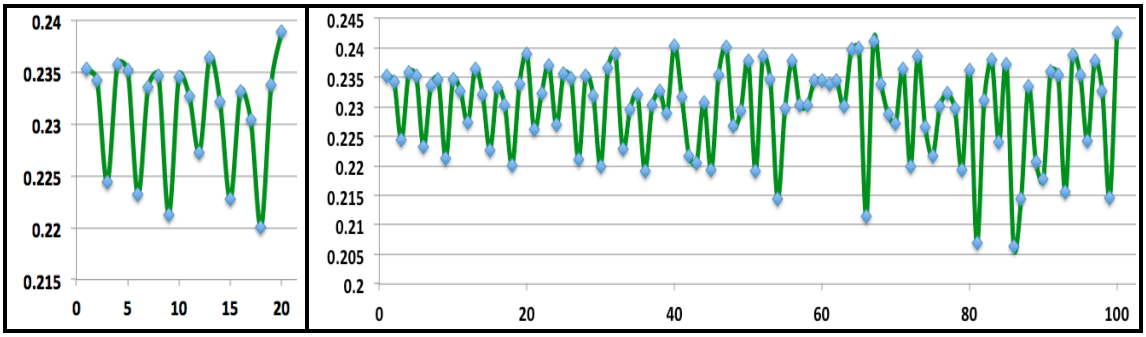


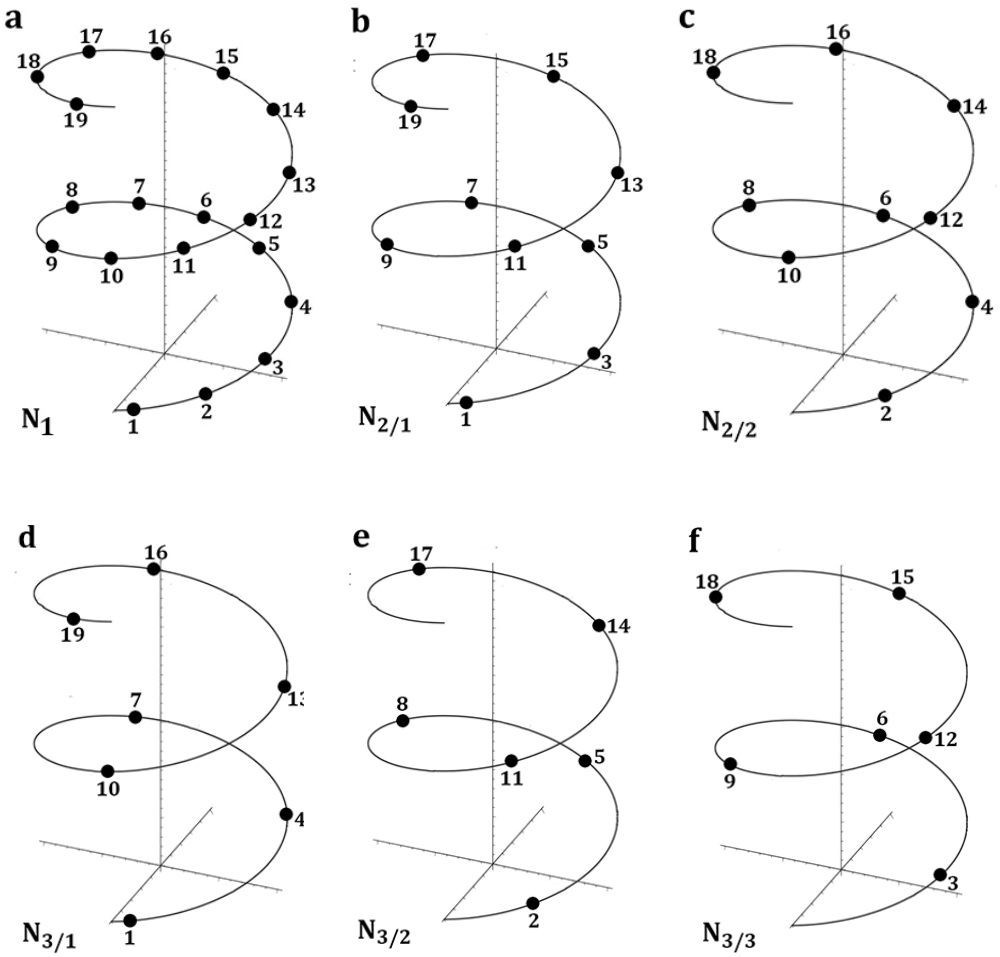
Fig. 12.5. The sequence of harmonic mean values of agreed deviations of all four OS-sequences from their model harmonic progressions in the coronavirus 2 isolate Wuhan-Hu-1. The ordinate axes show harmonic mean values. The left and right graphs show the cases of n = 1, 2, …, 20 and n = 1, 2, …, 100, which are plotted along the abscissa axes.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

13. 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. 13.1a), epi-chains of different orders n are those subsequences that contain only those nucleotides, whose numeration differ from each other by natural number n = 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 n = 2: the epi-chain $N_{2/1}$ contains nucleotides with odd numerations 1, 3, 5, … (Fig. 13.1b), and the epi-chain $N_{2/2}$ contains nucleotides with even numerations 2, 4, 6, ... (Fig. 13.1c). 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 n = 3: these epi-chains contain nucleotides with numerations 1, 4, 7, ... or 2, 5, 8, ... or 3, 6, 9, ... , respectively (Figs. 13.1d-f). The epi-chain of the first order N_1 coincides with the nucleotide sequence of the DNA strand (Fig. 13.1a).



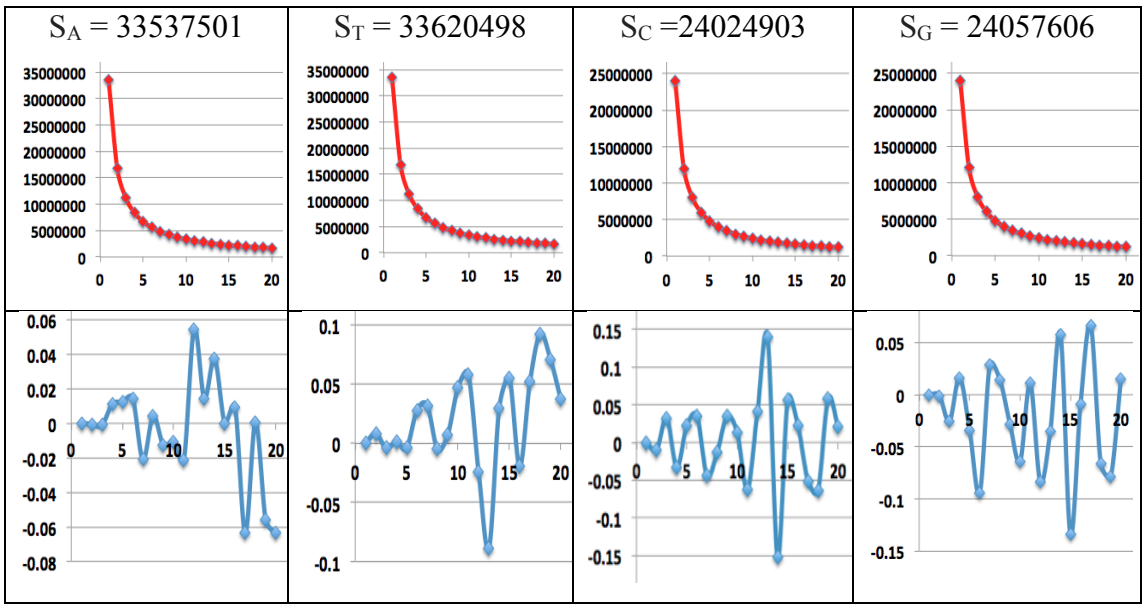
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

Fig. 13.1. 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.

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. 13.1a). 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. 13.1e).

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. Each DNA epi-chain of the order *k* (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 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. 13.2-13.6 explains 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. 2.1-2.3).



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

Fig. 13.2. 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. 13.1b), which consists of nucleotides with serial numerations 1-3-5-7-9-… in the DNA sequence of the human chromosome № 1. All abscissa axes show the values n = 1, 2, …, 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 show in percent slight alternating deviations of real values of the OS-sequences from model values.

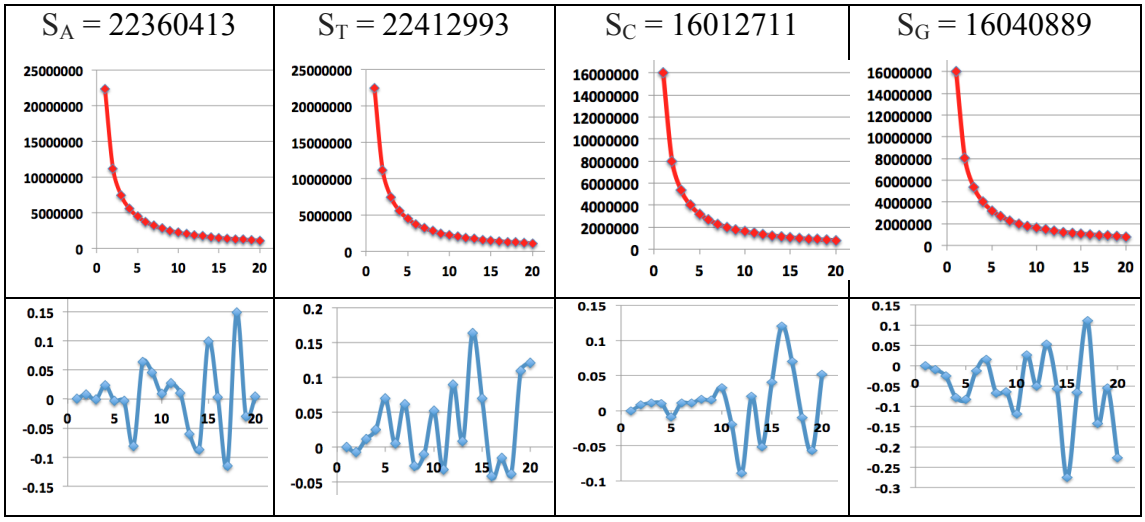
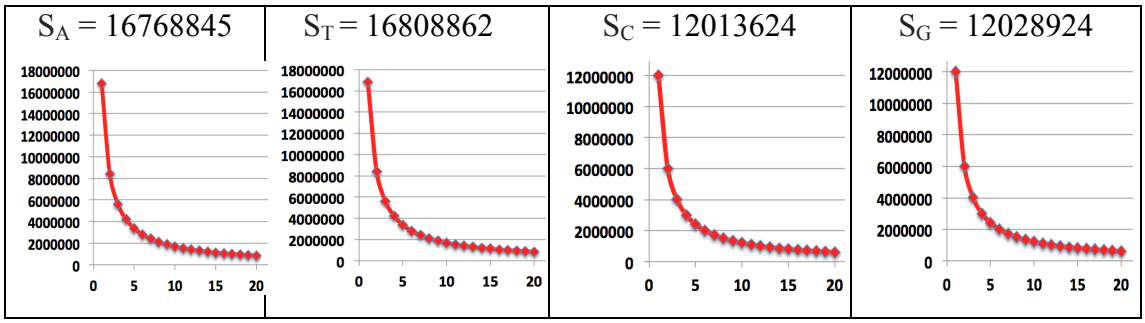


Fig. 13.3. The results of the analysis - by the oligomer sums method – the nucleotide sequence of the epi-chain of the third order $N_{3/1}$ (Fig. 13.1d), which consists of nucleotides with serial numerations 1-4-7-10-13-… in the DNA sequence of the human chromosome № 1. The top row demonstrates that 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 show in percent slight alternating deviations of real values of the OS-sequences from model values. All denotations are the same as in Fig. 13.2.



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

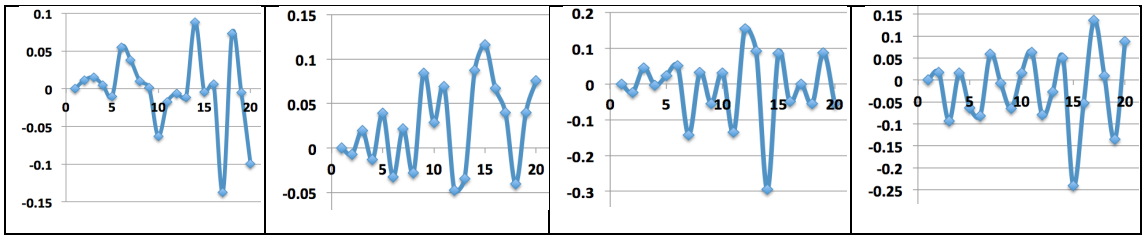


Fig. 13.4. The results of the analysis - by the oligomer sums method – the nucleotide sequence of the epi-chain of the 4th order $N_{4/1}$, which consists of nucleotides with serial numerations 1-5-9-13-… in the DNA sequence of the human chromosome № 1. The top row demonstrates that the model hyperbolic 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 show in percent slight alternating deviations of real values of the OS-sequences from model values. All denotations are the same as in Fig. 13.2.

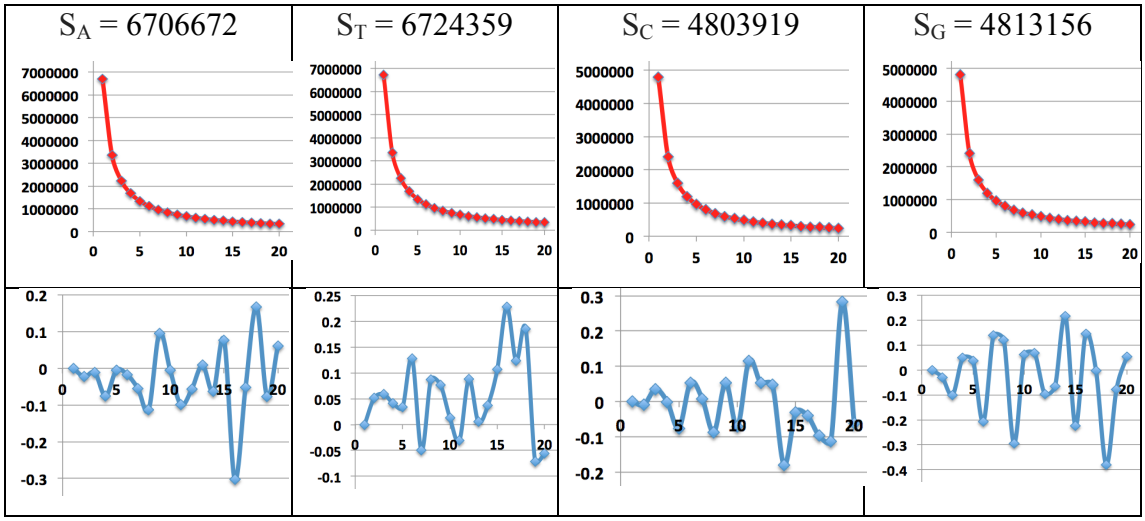


Fig. 13.5. The results of the analysis - by the oligomer sums method – the nucleotide sequence of the epi-chain of the 10th order $N_{10/1}$, which consists of nucleotides with serial numerations 1-11-21-31-41-… in the DNA sequence of the human chromosome № 1. The top row demonstrates that the model 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 show in percent slight alternating deviations of real values of the OS-sequences from model values. All denotations are the same as in Fig. 13.2.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

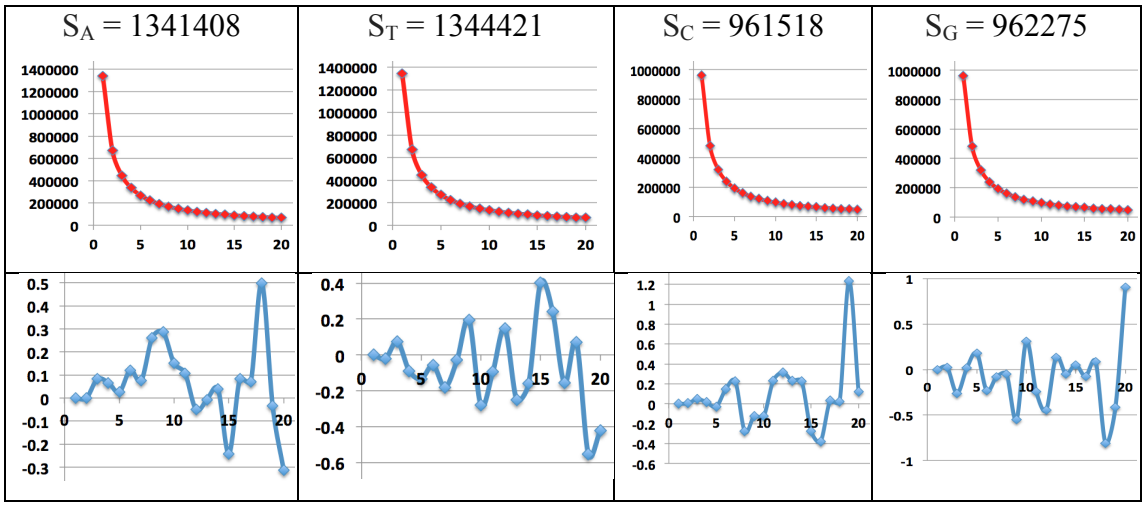


Fig. 13.6. The results of the analysis - by the oligomer sums method – the nucleotide sequence of the epi-chain of the 50th order $N_{50/1}$, which consists of nucleotides with serial numerations 1-51-101-151-201-… in the DNA sequence of the human chromosome № 1. The top row demonstrates that the model 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 show in percent slight alternating deviations of real values of the OS-sequences from model values. All denotations are the same as in Fig. 13.2.

Figs. 13.2-13.6 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, … is not too large compared to the full length of the nucleotide sequence).

Appendix I shows the numeric data represented in the graphs in Figs. 13.2.-13.6.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

14. 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 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:

$$S/n = \Sigma_{A,n,1} + \Sigma_{T,n,1} + \Sigma_{C,n,1} + \Sigma_{G,n,1} \qquad (14.1)$$

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 (14.1), is determined by the expression (14.2):

$$P_n(N_1) = \Sigma_{N,n,1}/(S/n) \approx (S_N/n)/(S/n) = S_N/S = P(N) \qquad (14.2)$$

The expression (14.2) 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 in Section 1 (Figs. 2.1-2.3), has the total amount of nucleotides $S = S_A+S_T+S_C+S_G = 67070277 + 67244164 + 48055043 + 48111528 = 230481012$. The probability P(A) of the nucleotide A is equal to $S_A/S = 67070277 / 230481012 \approx 0.2910$. From the data in Fig. 2.3, one can verify that, in this chromosome, the collective probabilities $P_n(A_1)$ of total amounts of *n*-plets (n = 2, 3, …, 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 $Pn(N_1)$ (where n = 1, 2, 3, …) 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 (14.2) allows us to construct a hyperbolic OS-sequence of the sums $\Sigma_{N,n,1}$ of n-plets (14.3):

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

$$\Sigma_{N,n,1} = P_n(N_1)*S/n \qquad (14.3)$$

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,... Given the expressions (14.2) and (14.3), 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 oligomeric 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, Petukhova, Svirin, 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,...

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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

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, Gross, Eisert, 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 the 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:

$$\mathbf{H} = \mathbf{H}_1 \otimes \mathbf{H}_2 \otimes \ldots \otimes \mathbf{H}_K \qquad (14.3)$$

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:

$$|\psi> = |\psi^{(1)}> \otimes |\psi^{(2)}> \otimes \ldots \otimes |\psi^{(K)}> \qquad (14.4)$$

If the vector cannot be expressed in this form (14.4), 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. 3.1). 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.

One can suppose that in eukaryotic and prokaryotic genomes we have some special case of multi-partite entangled states, but not in groups of many particles, but in genomic systems of many oligomers. 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 modelling the considered genetic sequences.

Let us turn to the above-mentioned author's model of properties of genomic sequences expressed by the expressions (14.2) and (14.3) [Petoukhov, 2018b; Petoukhov, Petukhova, Svirin, 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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

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 quntum-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.

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 $n = 2, 3, 4,$…? 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 $n = 3, 4, 5,$ …? 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 $n = k, k+1, k+2,$ … (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. 14.1 and 14.2 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. 2.3); 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 the formulas Δ% = 100(1 - Pos1/Pos2)% and Δ% = 100(1 - Pos1/Pos3)% where Pos1, Pos2, and Pos3 refer to values indicated in the rows

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

Pos. 1, Pos. 2, and Pos. 3. Here the results are presented only for n = 2, 3, 4, …, 10 but similar situations of practical coincidences of the corresponding members of the considered OS-sequences are also true for larger n.

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. 14.1. 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. 14.2. 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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

from each other. The comparison begins with triplets since there is no third position in monoplets and doublets.

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 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. 14.3 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.

DOUBLETS	TRIPLETS	TETRAPLETS	PENTAPLETS	S_2/S_3	S_2/S_4	S_2/S_5
$S_2 = \Sigma(\mathbf{AA})$	$S_3 = \Sigma(AAN)_4$	$S_4 = \Sigma(AANN)_{16}$	$S_5 = \Sigma(AANNN)_{64}$			
10952057	7300222	5476855	4381298	**1.50**	**2.00**	**2.50**
$S_2 = \Sigma(\mathbf{AT})$	$S_3 = \Sigma(ATN)_4$	$S_4 = \Sigma(ATNN)_{16}$	$S_5 = \Sigma(ATNNN)_{64}$			
8561194	5706906	4280647	3420561	**1.50**	**2.00**	**2.50**
$S_2 = \Sigma(\mathbf{AC})$	$S_3 = \Sigma(ACN)_4$	$S_4 = \Sigma(ACNN)_{16}$	$S_5 = \Sigma(ACNNN)_{64}$			
5799729	3868541	2899991	2322063	**1.50**	**2.00**	**2.50**
$S_2 = \Sigma(\mathbf{AG})$	$S_3 = \Sigma(AGN)_4$	$S_4 = \Sigma(AGNN)_{16}$	$S_5 = \Sigma(AGNNN)_{64}$			
8224510	5484720	4111320	3289579	**1.50**	**2.00**	**2.50**
$S_2 = \Sigma(\mathbf{TA})$	$S_3 = \Sigma(TAN)_4$	$S_4 = \Sigma(TANN)_{16}$	$S_5 = \Sigma(TANNN)_{64}$			
7274275	4849731	3636741	2909412	**1.50**	**2.00**	**2.50**
$S_2 = \Sigma(\mathbf{TT})$	$S_3 = \Sigma(TTN)_4$	$S_4 = \Sigma(TTNN)_{16}$	$S_5 = \Sigma(TTNNN)_{64}$			
11026157	7346507	5511908	4409900	**1.50**	**2.00**	**2.50**
$S_2 = \Sigma(\mathbf{TC})$	$S_3 = \Sigma(TCN)_4$	$S_4 = \Sigma(TCNN)_{16}$	$S_5 = \Sigma(TCNNN)_{64}$			
6923689	4617788	3461837	2768794	**1.50**	**2.00**	**2.50**
$S_2 = \Sigma(\mathbf{TG})$	$S_3 = \Sigma(TGN)_4$	$S_4 = \Sigma(TGNN)_{16}$	$S_5 = \Sigma(TGNNN)_{64}$			
8396349	5598933	4198342	3357218	**1.50**	**2.00**	**2.50**
$S_2 = \Sigma(\mathbf{CA})$	$S_3 = \Sigma(CAN)_4$	$S_4 = \Sigma(CANN)_{16}$	$S_5 = \Sigma(CANNN)_{64}$			
8382478	5591208	4191829	3354600	**1.50**	**2.00**	**2.50**
$S_2 = \Sigma(\mathbf{CT})$	$S_3 = \Sigma(CTN)_4$	$S_4 = \Sigma(CTNN)_{16}$	$S_5 = \Sigma(CTNNN)_{64}$			
8221421	5477836	4111963	3289510	**1.50**	**2.00**	**2.50**
$S_2 = \Sigma(\mathbf{CC})$	$S_3 = \Sigma(CCN)_4$	$S_4 = \Sigma(CCNN)_{16}$	$S_5 = \Sigma(CCNNN)_{64}$			
6233384	4153642	3117570	2492824	**1.50**	**2.00**	**2.50**

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

$S_2 = \Sigma$(**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. 14.3. 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.

The rows in the left part of Fig. 14.3 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, 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).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

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 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 one to predict with high accuracy a large number of quantitative interconnections between different kinds and sets of oligomers in eukaryotic and prokaryotic genomes.

15. 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.

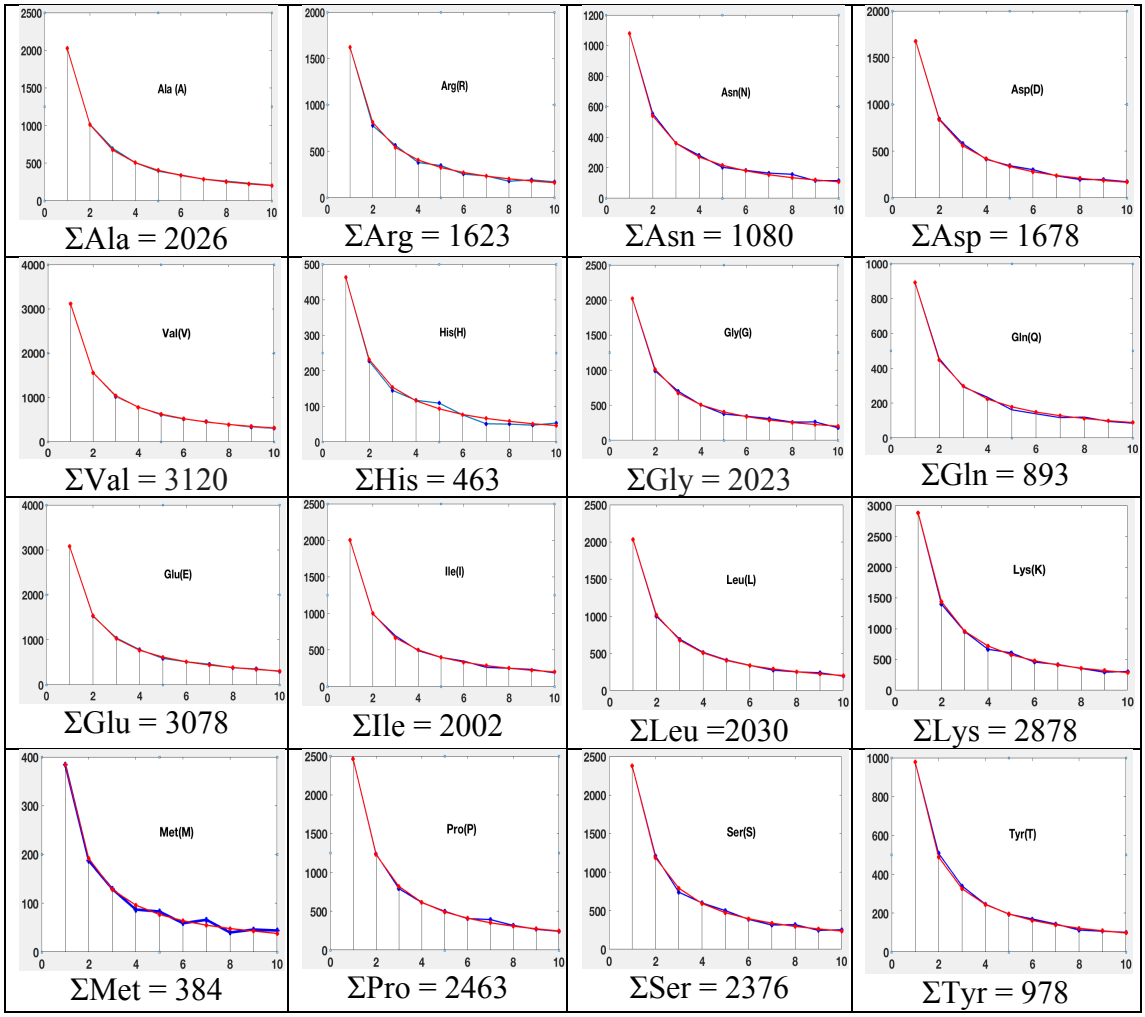
Each long sequence of amino acids (for example, ArgSerThrGlyPheLysLeuSer MetAla...) can be represented either as a sequence of monomers (Arg-Ser-Thr-Gly-Phe-Lys-Leu-Ser-Met-Ala-...), or as a sequence of amino acid doublets (ArgSer-ThrGly-PheLys-LeuSer-MetAla-...), or as a sequence of amino acid triplets (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 kinds 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, …, Ala Cys); the set of Ala_1-triplets contains all 400 triplets, which start with the Ala (AlaAlaAla, AlaAla Arg, …., AlaCysCys), and so on.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

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):

- For any of the 20 classes of the oligomer equivalency, the total amount Σ of its defining amino acid and total amounts of all those *n*-plets (*n* = 2, 3, 4, ...), 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, …

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. 15.1 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 kinds 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,…, 10).



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

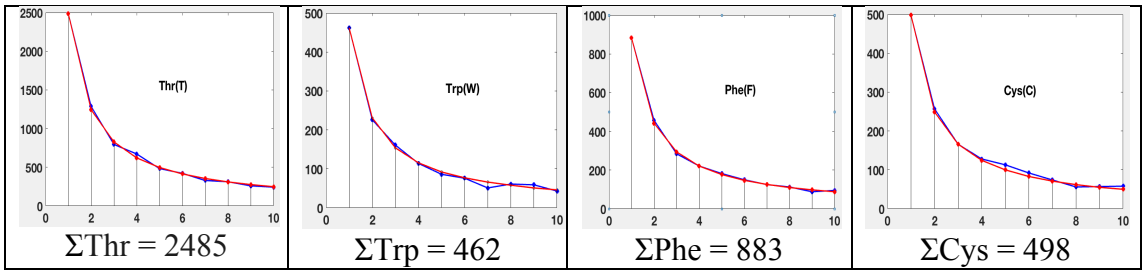


Fig. 15.1. Graphs of analysis results of the human protein Titin by the oligomer sums method for each of 20 equivalency classes, which are defined by its 20 kinds of amino acids. Each graph shows a sequence (in blue) of real total amounts of *n*-plets, which start with this amino acid, and also a model hyperbolic sequence Σ/*n* (in red), where Σ refers to a number of this amino acid (*n* = 1, 2,…, 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.

One can see from Fig. 15.1 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 real 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 the figure 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 real 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, ..., 10, these deviations can increase (the number of corresponding *n*-plets decreases with increasing *n*).

Fig. 15.2 gives examples of real and model numeric values for the classes Ala_1- and Arg_1-oligomers from the first graphs in Fig. 15.1.

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. 15.2. 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 presentations of corresponding OS-sequences are shown in Fig. 15.1 at the very top.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

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

16. Hyperbolic rules in phonetic sequences of long Russian literary texts revealed by the oligomer sums method

Leading experts in the field of structural linguistics have long believed that languages of human dialogue were formed not by random processes but by a continuation of genetic language or, are, at least, closely connected with genetic language, suggesting the compelling possibility that all organisms may utilize their genetic code in communication mechanisms. Analogies between systems of genetic and linguistic information are of wide and important scientific interest. Such direction of thoughts touches on the fundamental issues of intellectual activity and suggests that the principles of informational activity of the brain, reflected in human speech, did not arise from an empty place, but are a continuation of the principles of organization of genetic informatics. The Section is devoted to deep structural analogies between genomic nucleotide sequences and the phonetic features of long literary Russian texts related to the binary-oppositional phonetic structure of the Russian alphabet. These new analogies are connected with the described above hyperbolic rules and the harmonic progression (2.4).

One should note here the works by Roman Jakobson, one of the most famous linguistics experts and an author of an in-depth theory of binary phonetic oppositions in human languages [Jakobson, 1987, 1999; Jakobson, Fant, Halle, 1951; Jakobson, Halle, 1971; Jakobson, Waugh, 2002; Holenstein, 1975]. Jointly with F. Jacob, Nobel Prize winner in molecular genetics, and with other linguistic specialists holding the same views, Jakobson proposed that genetic language is the structural basis of linguistic languages [Jacob et al., 1968; Jakobson, 1985].

According to Jakobson, all relations among linguistic phonemes are decomposed into a series of binary oppositions of elementary differential attributes (or traits). By analogy, the set of the four nucleotides ("letters" of DNA texts) of the genetic alphabet contains the binary sub-alphabets, which allow creating new mathematical models in molecular genetics [Petoukhov, 2017, 2018a]. As Jakobson wrote, the genetic code system is the basic simulator, which underlies all verbal codes of human languages. "*The heredity in itself is the fundamental form of communications … Perhaps, the bases of language structures, which are imposed on molecular communications, have been constructed by its structural principles directly*" [Jakobson, 1985, p. 396]. These questions had arisen to Jakobson as a consequence of his long-term research into the connections between linguistics, biology, and physics. Such connections were considered at a united seminar of physicists and linguists, organized by Niels Bohr and Roman Jakobson, jointly, at the Massachusetts Institute of Technology.

"*Jakobson reveals distinctly a binary opposition of sound attributes as underlying each system of phonemes... The subject of phonology has changed by him: the phonology considered phonemes (as the main subject) earlier, but now Jakobson has offered that distinctive attributes should be considered as "quantums" (or elementary units of language)… Jakobson was interested especially in the general analogies of language structures with the genetic code, and he considered these analogies as indubitable*" [Ivanov, 1985].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

This connection between linguistics and the genetic code interests many researchers, and some even perceive linguistic language as a living organism. In his book, “Linguistic Genetics”, Makovsky says: "*A look at language as a living organism, subject to the natural laws of nature, ascends to deep antiquity ... Research of nature, of disposition and of reasons of isomorphism between genetic and linguistic regularities is one of the most important fundamental problems for linguistics of our time*" [Makovsky, 1992].

Wanting to advance in the study of the relationship between genetic language and linguistic languages, let's concentrate on Jakobson's fundamental theory about a binary opposition of sound attributes as underlying each system of phonemes and about distinctive attributes considered as “quantums” (or elementary units of various languages). In the Russian alphabet, a one-to-one correspondence between the letters and the phonemes exists. For this reason, analyzing long literary Russian texts, researches can study their phonetic structures (only a few human languages have in their alphabets a one-to-one correspondence between the letters and the phonemes). Below the author presents the results of his studying the phonetic structures of long Russian literary texts by L.N.Tolstoy, F.M.Dostoevsky, A.S.Pushkin, and others.

The DNA alphabet has the two-level binary-oppositional structure: it contains two sub-alphabets of purines (A, G) and pyrimidines (T, C). Each sub-alphabet dichotomously divides into two sub-sub-alphabets according to the signs of 2 or 3 hydrogen bonds in the complementary nucleotide pairs A-T and C-G (Fig. 16.1, at left).

The Russian alphabet, like the DNA alphabet, is phonetically based on binary oppositions and divided into sub-alphabets of vowels and consonants. In turn, the sub-alphabet of vowels dichotomously divided into sub-sub-alphabets of long and iotated vowels, and the sub-alphabet of consonants - into sub-sub-alphabets of voiced and deaf consonants (Fig. 16.1, at right). The soft sign “ь” and the hard sign “ъ” in the Russian alphabet do not convey any sound, and therefore they are not taken into account in its phonologic structure.

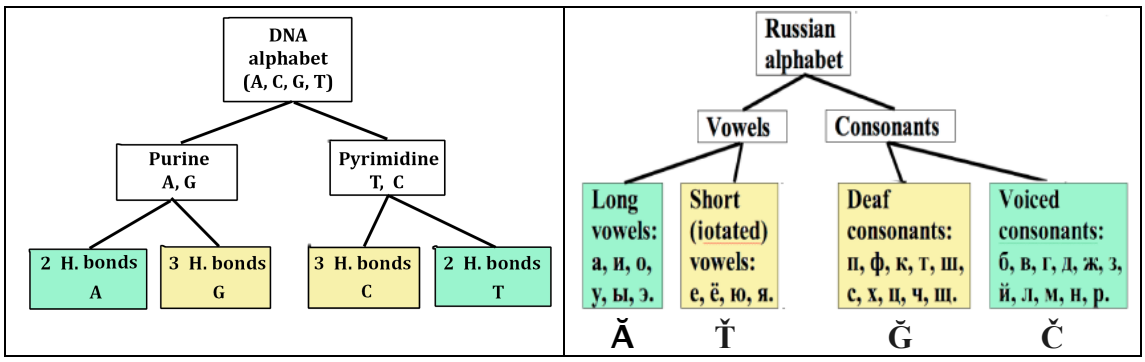


Fig. 16.1. The two-level binary-oppositional structure of the DNA alphabet of 4 nucleotides A, T, C, and G (at left) and the similar two-level binary-oppositional structure of the Russian alphabet, consisting of 4 phonetic classes denoted by symbols Ă, Ť, Č, and Ğ (at right), are shown.

As is well known, the Russian alphabet according to the phonetic features of its shown 31 phonetic letters consists of the following four classes of the phonetic equivalency of letters (Fig. 16.1, at right):

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

- The first class of the phonetic equivalency combines all letters, which are long vowels (а, и, о, у, ы, э). Any letter from this class will be denoted by the general symbol Ă, which we will call the phonetic monomer representing the first class;
- The second class of the phonetic equivalency combines all letters, which are short (iotated) vowels (е, ё, ю, я). Any letter from this class will be denoted by the general symbol Ť, which we will call the phonetic monomer representing the second class;
- The third class of the phonetic equivalency combines all letters, which are deaf consonants (п, ф, к, т, ш, с, х, ц, ч, щ). Any letter from this class will be denoted by the general symbol Ğ, which we will call the phonetic monomer representing the third class;
- The fourth class of the phonetic equivalency combines all letters, which are voiced consonants (б, в, г, д, ж, з, й, л, м, н, р). Any letter from this class will be denoted by the general symbol Č, which we will call the phonetic monomer representing the fourth class.

Leaving only these letters from the 4 classes in a studied literary text and replacing each letter by its appropriate symbol Ă, or Ť, or Ğ, or Č, we represent the literary text into a sequence of the phonetic monomers, for example, ČĂĞĞŤČŤĂĞ... We will call such sequence «the phonetic literary sequence" (or simply "the phonetic sequence") representing this literary text.

Below the author shows the results of studying the phonetic sequences, which represent long Russian literary texts, by the same oligomer sums method that was used above to study genomic sequences of nucleotides. These results reveal deep analogies of long genomic sequences and long phonetic literary sequences from the point of view of regularly interrelated oligomeric sums. Both of them are similarly related to the harmonic progression (2.4) and obey corresponding hyperbolic rules. Both of them show themselves as holistic cooperative essences, whose parts are interrelated by ratios of the algebraic harmony.

- 16.1. The analysis of the Russian novel «Anna Karenina» by L.N. Tolstoy

Let us start with an analysis of the Russian novel “Anna Karenina” by Leo Tolstoy (the original literary text was accessed from http://samolit.com/books/62/). This novel contains 1309047 phonetic letters from Fig. 16.1. Its phonetic literary sequence can be considered as a chain of 4 phonetic monomers Ă, Ť, Ğ, and Č; or as a chain of 16 phonetic doublets ĂĂ, ĂŤ, ĂĞ, …, ČČ; or as a chain of 64 phonetic triplets ĂĂĂ, ĂĂŤ, ĂĂĞ, ….; and so on. By analogy the analysis of genomic nucleotide sequences by the oligomer sums method (see Section 2), this phonetic literary sequence can be also analyzed by this method using the following steps:

- Firstly, one should calculate phenomenological quantities $S_{Ă}$, $S_{Ť}$, $S_{Č}$, and $S_{Ğ}$ of phonetic monomers Ă, Ť, Ğ, and Č in the considered phonetic literature sequence. In the phonetic sequence of the novel «Anna Karenina» the following quantities exist: $S_{Ă}$ = 419490, $S_{Ť}$ = 154800, $S_{Č}$ = 452716, and $S_{Ğ}$ = 282041.
- Secondly, to construct the phonetic oligomer sums sequences (or briefly, phonetic OS-sequences), one should calculate the total amounts $\Sigma_{Ă,n,1}$, $\Sigma_{Ť,n,1}$, $\Sigma_{Č,n,1}$, and $\Sigma_{Ğ,n,1}$ of phonetic n-plets in equivalence classes of $Ă_1$-oligomers, $Ť_1$-oligomers, $Č_1$-oligomers, and $Ğ_1$-oligomers at n = 1, 2, 3, 4, … (here, for

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

example, the symbol $\Sigma_{Ă,3,1}$ refers to the total amount of triplets, which start with the phonetic monomer Ă). These total amounts regarding each of the classes are members of the appropriate phonetic OS-sequence of the class.

- At the final step, such phenomenological amounts $\Sigma_{Ă,n,1}$, $\Sigma_{Ť,n,1}$, $\Sigma_{Č,n,1}$, and $\Sigma_{Ğ,n,1}$ are compared with their model values from the corresponding hyperbolic sequence $S_{Ă}/n$, or $S_{Ť}/n$, or $S_{Č}$, or $S_{Ğ}/n$.

Fig. 16.2 shows sequences of amounts $\Sigma_{Ă,n,1}$, $\Sigma_{Ť,n,1}$, $\Sigma_{Č,n,1}$, and $\Sigma_{Ğ,n,1}$ of all phonetic monomers, doublets, and other n-plets (at n =1, 2, 3, …, 10), representing the phonetic sequence of this famous Russian novel. The real amounts of phonetic n-plets are compared with their appropriate model values $S_{Ă}/n$, or $S_{Ť}/n$, or $S_{Č}$, or $S_{Ğ}/n$.

n	1	2	3	4	5	6	7	8	9	10
Ă										
Real	419490	209616	138897	104787	83861	69392	59842	52350	46457	41852
$S_{Ă}/n$	419490	209745	139830	104872.5	83898	69915	59927	52436	46610	41949
Δ%	0.00	0.06	0.67	0.08	0.04	0.75	0.14	0.16	0.33	0.23
Ť										
Real	154800	77480	51959	38958	31065	26088	22123	19525	17203	15467
$S_{Ť}/n$	154800	77400	51600	38700	30960	25800	22114	19350	17200	15480
Δ%	0.00	-0.10	-0.70	-0.67	-0.34	-1.12	-0.04	-0.90	-0.02	0.08
Č										
Real	452716	226508	151043	113054	90448	75588	64753	56664	50397	45424
$S_{Č}/n$	452716	226358	150905	113179	90543	75453	64674	56590	50302	45272
Δ%	0.00	-0.07	-0.09	0.11	0.11	-0.18	-0.12	-0.13	-0.19	-0.34
Ğ										
Real	282041	140919	94450	70462	56435	47106	40288	35091	31392	28161
$S_{Ğ}/n$	282041	141021	94014	70510	56408	47007	40292	35255	31338	28204
Δ%	0	0.07	-0.46	0.07	-0.05	-0.21	0.01	0.47	-0.17	0.15

Fig. 16.2. Real values and model values $S_{Ă}/n$, or $S_{Ť}/n$, or $S_{Č}$, or $S_{Ğ}/n$ (in red) in the case of the oligomer sums representations of the phonetic sequence of the Russian novel «Anna Karenina» by L. Tolstoy. Symbols Ă, Ť, Ğ, and Č refer to the phonetic monomers (Fig. 16.1, at right). The symbol Δ% denotes deviations of real values from model values in percent (the model values are taken as 100%).

One can see in Fig. 16.2 that the deviations of real values $\Sigma_{Ă,n,1}$, $\Sigma_{Ť,n,1}$, $\Sigma_{Č,n,1}$, and $\Sigma_{Ğ,n,1}$ of the phonetic oligomeric sums - from their model values in the corresponding hyperbolic sequences $S_{Ă}/n$, or $S_{Ť}/n$, or $S_{Č}$, or $S_{Ğ}/n$ - are small and lie within one percent. Thus, the deviations in the phonetic representations of the Russian text of the novel "Anna Karenina" have the same order of smallness as the deviations in the nucleotide sequences of eukaryotic and prokaryotic genomes shown above in Figs. 2.3, 2.4, 4.1, 5.1, 6.1, 7.1, 9.1, 10.1-10.8. In other words, long genomic nucleotide sequences and long phonetic sequences of this Russian novel are structurally akin and obey the similar hyperbolic rules related to the harmonic progression (2.4). This holds not only for the named Russian novel but also for other Russian long texts, whose phonetic literary sequences were analyzed by the oligomer sums method as it is described below. This gives evidence in favor of the deep algebra-harmonic relation of the phonetic structurization of the Russian language to the structurization of eukaryotic and prokaryotic genomes.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

The results of the analyses of phonetic representations of the Russian texts of not only the novel «Anna Karenina» but also of other long Russian literary works by L.N. Tolstoy, F.M. Dostoevsky, and A.S. Pushkin, which are described below, testify in favor of the following hyperbolic phonetic rule about interrelations of oligomer sums in phonetic representations of such Russian works:

The first hyperbolic phonetic rule:

- For any of the phonetic classes of $\check{A}_1$-oligomers, $\check{T}_1$-oligomers, $\check{C}_1$-oligomers, and $\check{G}_1$-oligomers in the phonetic representations of long Russian literary texts, the total amounts $\Sigma_{\check{N},n,1}(n)$ of their phonetic n-plets, corresponding different n, are interrelated each other through the hyperbolic expression $\Sigma_{\check{N},n,1} \approx S_{\check{N}}/n$ with a high level of accuracy (here $\check{N}$ denotes any of 4 phonetic monomers $\check{A}$, $\check{T}$, $\check{G}$, and $\check{C}$; $S_{\check{N}}$ denotes the number of the phonetic monomer $\check{N}$; $n = 1, 2, 3, 4, \ldots$ is not too large compared to the full length of the phonetic sequence). The phenomenological points with coordinates $[n, \Sigma_{\check{N},n,1}]$ practically lie on the hyperbola in its points $H_{\check{N},1} = S_{\check{N}}/n$.

This phonetic hyperbolic rule for long Russian texts is analogical to the first hyperbolic rule of eukaryotic and prokaryotic genomes formulated above in Section 2. Fig. 16.3 shows graphically the phonetic oligomer sums sequences $\Sigma_{\check{A},n,1}$, $\Sigma_{\check{T},n,1}$, $\Sigma_{\check{C},n,1}$, whose numeric data are given in Fig. 16.2.

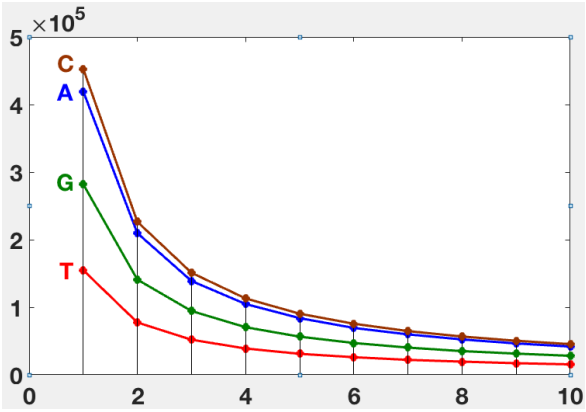


Fig. 16.3. Graphs of the hyperbolic-like sequences of the phonetic oligomer sum $\Sigma_{\check{A},n,1}$, $\Sigma_{\check{T},n,1}$, $\Sigma_{\check{C},n,1}$, and $\Sigma_{\check{G},n,1}$, which are shown numerically in Fig. 16.2 and which practically coincide with the model hyperbolic sequences $S_{\check{A}}/n$ (in blue), $S_{\check{T}}/n$ (in red), $S_{\check{C}}$,(in brown), and $S_{\check{G}}/n$ (in green).

Let us continue to list those deep algebra-structural analogies between phonetic sequences of long Russian texts and nucleotide sequences of eukaryotic and prokaryotic genomes, which are revealed by the oligomer sums method. As shown above in Figs. 13.1-13.6, the genomic DNA sequences have a fractal-like structure revealed by the oligomer sums method at the analysis of their epi-chains, i.e. shortened subsequences consisting of every second nucleotide, or every third nucleotide, or of every fourth nucleotide, and so on. By analogy, you can numerate members in the phonetic sequence $\check{D}_1$ of any Russian text by numbers 1, 2, 3, ... in their sequent order. Further, you can consider its phonetic epi-chains that is shortened subsequences consisting of every second member (the epi-chains $\check{D}_{2/1}$, $\check{D}_{2/2}$), or every third member (the epi-chains $\check{D}_{3/1}$, $\check{D}_{3/2}$, $\check{D}_{3/3}$), or of every fourth member, and so on (Fig. 16. 4). In the designation "$\check{D}_{k/m}$" of such phonetic epi-chains, the numerator "k" indicates the order of the phonetic epi-chain, and the denominator "m" indicates the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

numeration of the initial member of this epi-chain in the holistic phonetic sequence $Ď_1$. Fig. 16.4 shows examples of phonetic epi-chains $Ď_{2/1}$, $Ď_{3/1}$, $Ď_{4/1}$, and of their compositions.

$Ď_1$:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
$Ď_{2/1}$:	1		3		5		7		9		11		13		15		17
$Ď_{3/1}$:	1			4			7			10			13			16	
$Ď_{4/1}$:	1				5				9				13				17

Fig. 16.4. Compositions of the phonetic epi-chains $Ď_{2/1}$, $Ď_{3/1}$, $Ď_{4/1}$ compared to the composition of the full phonetic sequence $Ď_1$.

Figs. 16.5-16.7 show numeric and graphical results of the analysis of these phonetic epi-chains $Ď_{2/1}$, $Ď_{3/1}$, $Ď_{4/1}$ in the phonetic representation of the Russian novel «Anna Karenina» by the oligomer sums method. The full phonetic sequence $Ď_1$ contains 1309047 phonetic letters; its epi-chains $Ď_{2/1}$, $Ď_{3/1}$, $Ď_{4/1}$ contains correspondingly 654524, 436349, and 327262 phonetic letters. They all differ from each other not only in the number of members but also in the sequence of different phonetic oligomers in them. But one can see in 16.5-16.7 that in all these phonetic epi-chains $Ď_{2/1}$, $Ď_{3/1}$, $Ď_{4/1}$ the sequences of oligomeric sums are hyperbolic sequences with high accuracy. From this point of view, the considered phonetic epi-chains $Ď_{2/1}$, $Ď_{3/1}$, $Ď_{4/1}$ are practically no different from the full phonetic chain $Ď_1$(Figs. 16.2 and 16.3); the hyperbolic rule of phonetic sequences, associated with harmonic progression (2.4), is for them a general algebraic rule or a general algebraic invariant.

Similar results are valid for other epi-chains $Ď_{2/2}$, $Ď_{3/2}$, $Ď_{3/3}$, $Ď_{4/2}$, $Ď_{4/3}$, $Ď_{4/4}$, and not only for the novel «Anna Karenina», but also for all other long Russian texts by L.N. Tolstoy, F.M. Dostoevsky, and A.S. Pushkin, analyzed by the author and presented below. These results testify in favor of the following hyperbolic phonetic rule, which is an analoq of the fourth hyperbolic rule of eukaryotic and prokaryotic genomes formulated above in Section 13.

The second hyperbolic phonetic rule (about interrelations of phonetic oligomer sums in epi-chains of phonetic sequences of long Russian literary texts):

- In any of the phonetic sequences, representing long Russian literary texts, the first hyperbolic phonetic rule is fulfilled not only for oligomer sums of the full phonetic sequence but also for its epi-chains of the order k (where k = 2, 3, 4, … is not too large compared to the length of the full phonetic sequence).

n	1	2	3	4	5	6	7	8	9	10
Ă										
Real	209617	104788	69392	52351	41852	34802	29866	26297	23331	20924
$S_{Ă}/n$	209617	104809	69872	52404.25	41923	34936	29945	26202	23291	20962
Δ%	0	0.02	0.69	0.10	0.17	0.38	0.26	-0.36	-0.17	0.18
Ť										
Real	77480	38958	26088	19525	15467	13172	11037	9705	8591	7789
$S_{Ť}/n$	77480	38740	25827	19370	15496	12913	11069	9685	8609	7748
Δ%	0	-0.56	-1.01	-0.80	0.19	-2.00	0.29	-0.21	0.21	-0.53

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

Č										
Real	226508	113054	75588	56664	45424	37776	32434	28240	25258	22579
$S_{Č}/n$	226508	113254	75503	56627	45302	37751	32358	28314	25168	22651
Δ%	0.00	0.18	-0.11	-0.07	-0.27	-0.07	-0.23	0.26	-0.36	0.32
Ğ										
Real	140919	70462	47106	35091	28161	23337	20166	17573	15544	14160
$S_{Ğ}/n$	140919	70460	46973	35230	28184	23487	20131	17615	15658	14092
Δ%	0	0.00	-0.28	0.39	0.08	0.64	-0.17	0.24	0.73	-0.48

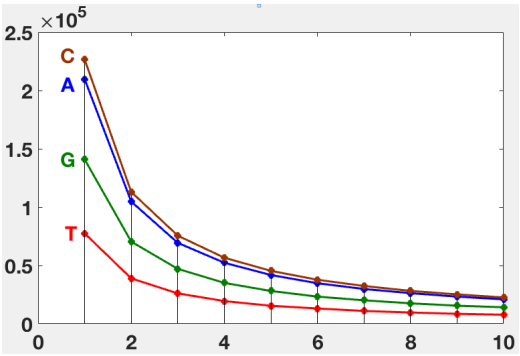


Fig. 16.5. Numeric and graphical representations of the series of the phonetic oligomer sum $\Sigma_{Ă,n,1}$ (in blue), $\Sigma_{Ť,n,1}$ (in red), $\Sigma_{Č,n,1}$ (in brown), and $\Sigma_{Ğ,n,1}$ (in green) in the epi-chain $Ď_{2/1}$ of the second order, which is the subsequence of the phonetic sequence $Ď_1$ representing the Russian text of the novel «Anna Karenina» by L.N.Tolstoy.

n	1	2	3	4	5	6	7	8	9	10
Ă										
Real	138897	69392	46457	34802	27650	23331	19807	17255	15585	13876
$S_{Ă}/n$	138897	69449	46299	34724.25	27779	23150	19842	17362	15433	13890
Δ%	0	0.08	-0.34	-0.22	0.47	-0.78	0.18	0.62	-0.98	0.10
Ť										
Real	51959	26088	17203	13172	10526	8591	7517	6657	5746	5257
$S_{Ť}/n$	51959	25979.5	17320	12990	10392	8660	7423	6495	5773	5196
Δ%	0	-0.42	0.67	-1.40	-1.29	0.79	-1.27	-2.50	0.47	-1.18
Č										
Real	151043	75588	50397	37776	30103	25258	21573	19009	16772	15059
$S_{Č}/n$	151043	75521.5	50348	37761	30209	25174	21578	18880	16783	15104
Δ%	0.00	-0.09	-0.10	-0.04	0.35	-0.33	0.02	-0.68	0.06	0.30
Ğ										
Real	94450	47106	31392	23337	18990	15544	13438	11622	10380	9442
$S_{Ğ}/n$	94450	47225	31483	23613	18890	15742	13493	11806	10494	9445
Δ%	0.00	0.25	0.29	1.17	-0.53	1.26	0.41	1.56	1.09	0.03

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

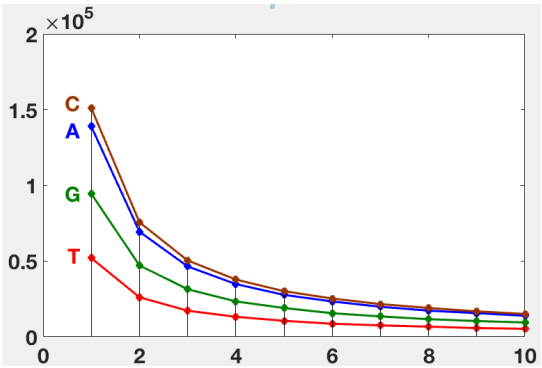


Fig. 16.6. Numeric and graphical representations of the series of phonetic oligomer sum $\Sigma_{\check{A},n,1}$ (in blue), $\Sigma_{\check{T},n,1}$ (in red), $\Sigma_{\check{C},n,1}$ (in brown), and $\Sigma_{\check{G},n,1}$ (in green) in the epi-chain $\check{D}_{3/1}$ of the third order, which is the subsequence of the phonetic sequence $\check{D}_1$ representing the Russian text of the novel «Anna Karenina» by L.N.Tolstoy.

n	1	2	3	4	5	6	7	8	9	10
Ă										
Real	104788	52351	34802	26297	20924	17255	14877	13008	11765	10485
$S_{\check{A}}/n$	104788	52394	34929	26197	20958	17465	14970	13099	11643	10479
Δ%	0	0.08	0.36	-0.38	0.16	1.20	0.62	0.69	-1.05	-0.06
Ť										
Real	38958	19525	13172	9705	7789	6657	5600	4894	4322	3930
$S_{\check{T}}/n$	38958	19479	12986	9740	7792	6493	5565	4870	4329	3896
Δ%	0.00	-0.24	-1.43	0.35	0.03	-2.53	-0.62	-0.50	0.15	-0.88
Č										
Real	113054	56664	37776	28240	22579	19009	16157	14237	12645	11253
$S_{\check{C}}/n$	113054	56527	37685	28264	22611	18842	16151	14132	12562	11305
Δ%	0	-0.24	-0.24	0.08	0.14	-0.88	-0.04	-0.74	-0.66	0.46
Ğ										
Real	70462	35091	23337	17573	14160	11622	10117	8768	7630	7058
$S_{\check{G}}/n$	70462	35231	23487	17616	14092	11744	10066	8808	7829	7046
Δ%	0	0.40	0.64	0.24	-0.48	1.04	-0.51	0.45	2.54	-0.17

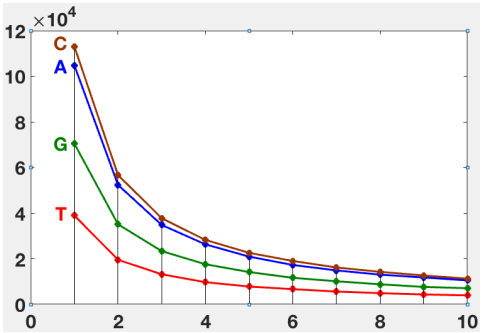


Fig. 16.7. Numeric and graphical representations of the series of phonetic oligomer sum $\Sigma_{\check{A},n,1}$ (in blue), $\Sigma_{\check{T},n,1}$ (in red), $\Sigma_{\check{C},n,1}$ (in brown), and $\Sigma_{\check{G},n,1}$ (in green) in the epi-chain $\check{D}_{4/1}$ of the fourth order, which is the subsequence of the phonetic sequence $\check{D}_1$ representing the Russian text of the novel «Anna Karenina» by L.N.Tolstoy.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

Many other deep analogies exist between oligomer sums organization of long genomic nucleotide sequences and phonetic sequences, presenting long Russian literary texts. But various connections of long genomic sequences with the harmonic progression (2.4) and appropriate hyperbolic rules were predicted by the mentioned quantum-information model [Petoukhov, 2018b; Petoukhov, Petukhova, Svirin, 2019], as it was noted above in Section 14. Correspondingly one should think that a similar quantum-information model is appropriate also for phonetic sequences, presenting long Russian literary texts. This idea is correct: many predictions of this author's model hold also for oligomer structures of the phonetic sequences representing long Russian literary texts. The author should here emphasize that he is talking about the similarity between the mathematical apparatus of quantum informatics and the mathematical apparatus of the cooperative organization of the considered long genomic and phonetic sequences, but not at all about physical quantum entanglement in these genomic and phonetic sequences.

One should remark that for reformulating the genomic quantum-information model into the phonetic-oriented quantum-information model, the computational basis states of the considered phonetic systems should be connected not with the binary-oppositional indicators (or molecular attributes) of the DNA bases A, T, C, and G, but with binary-oppositional elementary phonetic attributes, which were noted by R. Jakobson as the basis of all relations among linguistic phonemes (see the beginnings of this Section).

Let us show an example of the model predictions regarding the phonetic sequence representing the Russian novel „Anna Karenina". It was shown above in Section 14, that one of the model predictions concerns quantitative interrelations of different *n*-plets, which start from the same doublet, or from the same triplet, etc. The model predicted, in particular, that the amount S_2 of any of 16 doublets NN was algebra-harmonically interrelated with the total amounts S_3, S_4, S_5, … of nucleotide 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, equal to values 3/2, 4/2, 5/2, ….

Regarding the phonetic sequences of long Russian literary texts, this model predicts correspondingly that the amount S_2 of any of 16 phonetic doublets ŇŇ was algebra-harmonically interrelated with the total amounts S_3, S_4, S_5, … of phonetic oligomers in the following sets: 4 triplets, which start with this attributive phonetic doublet ŇŇ; 16 tetraplets, which start with this attributive doublet ŇŇ; 64 pentaplets, which start with this attributive doublet ŇŇ; 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 be equal to values 3/2, 4/2, 5/2, ….

The analysis of the phonetic sequence representing the text of the novel "Anna Karenina" completely confirmed this prediction as Fig. 16.8 shows by analogy with similar results for the genomic sequence in the human chromosome №1 (Fig. 14.3.).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

DOUBLETS	TRIPLETS	TETRAPLETS	PENTAPLETS	S_2/S_3	S_2/S_4	S_2/S_5
$S_2 = \Sigma(ĂĂ)$	$S_3 = \Sigma(ĂĂŇ)_4$	$S_4 = \Sigma(ĂĂŇŇ)_{16}$	$S_5 = \Sigma(ĂĂŇŇŇ)_{64}$			
13387	8918	6718	5435	**1.5**	**2.0**	**2.5**
$S_2 = \Sigma(ĂŤ)$	$S_3 = \Sigma(ĂŤŇ)_4$	$S_4 = \Sigma(ĂŤŇŇ)_{16}$	$S_5 = \Sigma(ĂŤŇŇŇ)_{64}$			
14447	9449	7307	5758	**1.5**	**2.0**	**2.5**
$S_2 = \Sigma(ĂČ)$	$S_3 = \Sigma(ĂČŇ)_4$	$S_4 = \Sigma(ĂČŇŇ)_{16}$	$S_5 = \Sigma(ĂČŇŇŇ)_{64}$			
116714	77660	58438	46642	**1.5**	**2.0**	**2.5**
$S_2 = \Sigma(ĂĞ)$	$S_3 = \Sigma(ĂĞŇ)_4$	$S_4 = \Sigma(ĂĞŇŇ)_{16}$	$S_5 = \Sigma(ĂĞŇŇŇ)_{64}$			
65068	42870	32324	26026	**1.5**	**2.0**	**2.5**
$S_2 = \Sigma(ŤĂ)$	$S_3 = \Sigma(ŤĂŇ)_4$	$S_4 = \Sigma(ŤĂŇŇ)_{16}$	$S_5 = \Sigma(ŤĂŇŇŇ)_{64}$			
5842	3932	2929	2297	**1.5**	**2.0**	**2.5**
$S_2 = \Sigma(ŤŤ)$	$S_3 = \Sigma(ŤŤŇ)_4$	$S_4 = \Sigma(ŤŤŇŇ)_{16}$	$S_5 = \Sigma(ŤŤŇŇŇ)_{64}$			
3627	2480	1820	1464	**1.5**	**2.0**	**2.5**
$S_2 = \Sigma(ŤČ)$	$S_3 = \Sigma(ŤČN)_4$	$S_4 = \Sigma(ŤČŇŇ)_{16}$	$S_5 = \Sigma(ŤČŇŇŇ)_{64}$			
44390	29806	22271	17877	**1.5**	**2.0**	**2.5**
$S_2 = \Sigma(ŤĞ)$	$S_3 = \Sigma(ŤĞŇ)_4$	$S_4 = \Sigma(ŤĞŇŇ)_{16}$	$S_5 = \Sigma(ŤĞŇŇŇ)_{64}$			
23621	15741	11938	9427	**1.5**	**2.0**	**2.5**
$S_2 = \Sigma(ČĂ)$	$S_3 = \Sigma(ČĂŇ)_4$	$S_4 = \Sigma(ČĂŇŇ)_{16}$	$S_5 = \Sigma(ČĂŇŇŇ)_{64}$			
126038	84179	62831	50474	**1.5**	**2.0**	**2.5**
$S_2 = \Sigma(ČŤ)$	$S_3 = \Sigma(ČŤŇ)_4$	$S_4 = \Sigma(ČŤŇŇ)_{16}$	$S_5 = \Sigma(ČŤŇŇŇ)_{64}$			
38951	26030	19477	15672	**1.5**	**2.0**	**2.5**
$S_2 = \Sigma(ČČ)$	$S_3 = \Sigma(ČČŇ)_4$	$S_4 = \Sigma(ČČŇŇ)_{16}$	$S_5 = \Sigma(ČČŇŇŇ)_{64}$			
37438	24916	18755	14719	**1.5**	**2.0**	**2.5**
$S_2 = \Sigma(ČĞ)$	$S_3 = \Sigma(ČĞŇ)_4$	$S_4 = \Sigma(ČĞŇŇ)_{16}$	$S_5 = \Sigma(ČĞŇŇŇ)_{64}$			
24081	15918	11991	9583	**1.5**	**2.0**	**2.5**
$S_2 = \Sigma(ĞĂ)$	$S_3 = \Sigma(ĞĂŇ)_4$	$S_4 = \Sigma(ĞĂŇŇ)_{16}$	$S_5 = \Sigma(ĞĂŇŇŇ)_{64}$			
64606	43428	32482	25899	**1.5**	**2.0**	**2.5**
$S_2 = \Sigma(ĞŤ)$	$S_3 = \Sigma(ĞŤŇ)_4$	$S_4 = \Sigma(ĞŤŇŇ)_{16}$	$S_5 = \Sigma(ĞŤŇŇŇ)_{64}$			
20295	13692	10085	8214	**1.5**	**2.0**	**2.5**
$S_2 = \Sigma(ĞČ)$	$S_3 = \Sigma(ĞČŇ)_4$	$S_4 = \Sigma(ĞČŇŇ)_{16}$	$S_5 = \Sigma(ĞČŇŇŇ)_{64}$			
27666	18370	13717	11045	**1.5**	**2.0**	**2.5**
$S_2 = \Sigma(ĞĞ)$	$S_3 = \Sigma(ĞĞŇ)_4$	$S_4 = \Sigma(ĞĞŇŇ)_{16}$	$S_5 = \Sigma(ĞĞŇŇŇ)_{64}$			
28352	18960	14178	11277	**1.5**	**2.0**	**2.5**

Fig. 16.8. The confirmation of the quantum-information model prediction by the comparison of the amount S_2 of each of 16 phonetic doublets to the total amounts S_3, S_4, S_5 of phonetic *n*-plets (*n* = 3, 4, 5), which start with this doublet, in the phonetic sequence, representing the Russian novel «Anna Karenina» by L.Tolstoy. Tabular data present total sums of each kind of phonetic oligomers: for example, the total sum $S_4 = \Sigma(ĂĂŇŇ)_{16}$ of all 16 phonetic tetraplets, which start with the phonetic doublet ĂĂ, is equal to 6718. The symbol Ň denotes any of phonetic monomers Ă, Ť, Ğ, and Č.

Numeric data in Fig. 16.8 show that really - for each of the considered 16 phonetic doublets - the ratios of the total amounts S_2/S_3, S_2/S_4, S_2/S_5, … are 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 be equal to values 3/2, 4/2, 5/2. Similar results hold for the phonetic sequences representing other long Russian literary texts by L.N.Tolstoy, F.M.Dostoevsky, A.S.Pushkin noted below.

The idea of a possible connection of intellectual brain activity with the principles of quantum mechanics and quantum informatics has long worried researchers. For example, an article [Bruza, Kitto, Nelson, McEvoy, 2009] presents a

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

quantum model of a word association system with word entanglement in human memory.

For such researches about possible connections of brain activities with the mathematics of quantum mechanics, the oligomer sums method, algebra-harmonic hyperbolic rules, and the mentioned author's quantum-information model give new effective research instruments and important phenomenological materials. In particular, they can be useful for developing new approaches to the creation of artificial intelligence, maximal resembling features of intellectual systems of eukaryotes and prokaryotes including humans. As is known, all living bodies have inborn intellectual-like systems providing solutions to many vital tasks: food search, rescue from predators, coordinated movements of body elements, etc. These inborn abilities for intellectual activities are connected with structural features of the genetic system.

- 16.2. The analysis of the Russian novel «War and Peace» by L.N. Tolstoy

Now we present results of the analysis of the phonetic sequence representing the Russian text of Tolstoy's novel «War and Peace», Book I, by the oligomer sums method. This text contains 1068479 phonetic letters, and it was taken from the web site http://samolit.com/books/64/. These results are shown in Figs. 16.9–16.13, and they are similar to the described results of the analysis of the novel "Anna Karenina" in all essential aspects.

Fig. 16.9 shows - in numeric and graphical forms - phonetic sequences of amounts $\Sigma_{Ă,n,1}$, $\Sigma_{Ť,n,1}$, $\Sigma_{Č,n,1}$, and $\Sigma_{Ğ,n,1}$ of all appropriate phonetic monomers, doublets, and other n-plets (at n =1, 2, 3, …, 10), representing the phonetic sequence of this famous Russian novel. The real amounts of phonetic n-plets are compared with their appropriate model values $S_{Ă}/n$, or $S_{Ť}/n$, or $S_{Č}$, or $S_{Ğ}/n$.

n	1	2	3	4	5	6	7	8	9	10
Ă	342653	170997	114314	85601	68755	56889	48985	42864	38177	34182
Real	342653	171327	114218	85663.25	68531	57109	48950	42832	38073	34265
$S_{Ă}/n$	0	0.19	-0.08	0.07	-0.33	0.38	-0.07	-0.08	-0.27	0.24
Δ%	342653	170997	114314	85601	68755	56889	48985	42864	38177	34182
Ť										
Real	120471	60122	40309	30177	23900	20240	17225	15015	13336	12013
$S_{Ť}/n$	120471	60235.5	40157	30118	24094	20079	17210	15059	13386	12047
Δ%	0.00	0.19	-0.38	-0.20	0.81	-0.80	-0.09	0.29	0.37	0.28
Č										
Real	375043	187598	125137	93713	75431	62716	53710	46871	41595	37645
$S_{Č}/n$	375043	187521.5	125014	93761	75009	62507	53578	46880	41671	37504
Δ%	0.00	-0.04	-0.10	0.05	-0.56	-0.33	-0.25	0.02	0.18	-0.38
Ğ										
Real	230312	115522	76399	57628	45609	38234	32719	28809	25611	23007
$S_{Ğ}/n$	230312	115156	76771	57578	46062	38385	32902	28789	25590	23031
Δ%	0	-0.32	0.48	-0.09	0.98	0.39	0.56	-0.07	-0.08	0.11

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

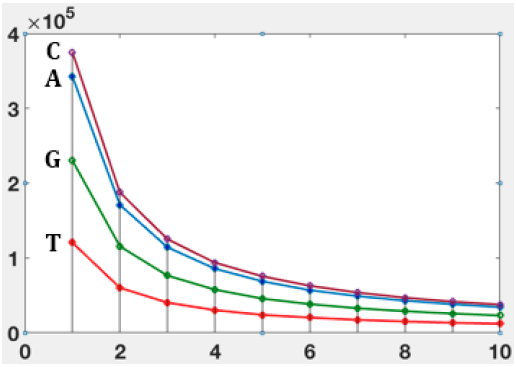
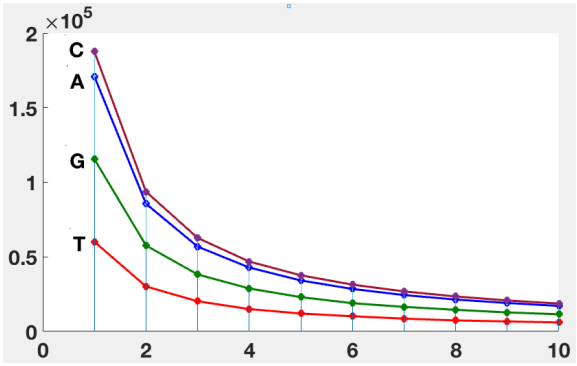


Fig. 16.9. Numeric and graphical representations of the sequences of phonetic oligomer sum $\Sigma_{\check{A},n,1}$ (in blue), $\Sigma_{\check{T},n,1}$ (in red), $\Sigma_{\check{C},n,1}$ (in brown), and $\Sigma_{\check{G},n,1}$ (in green) in the full phonetic sequence $\check{D}_1$, representing the Russian text of the novel «War and Peace».

Figs. 16.10-16.12 show results of the analysis of the phonetic epi-chains $\check{D}_{2/1}$, $\check{D}_{3/1}$, $\check{D}_{4/1}$ (see Fig. 16.4) in the phonetic representation of this Russian novel by the oligomer sums method (by analogy with Figs. 16.5-16.7 for the novel «Anna Karenina»).

n	1	2	3	4	5	6	7	8	9	10
Ă										
Real	170998	85602	56890	42865	34183	28519	24392	21411	19093	17109
$S_{\check{A}}/n$	170998	85499	56999	42749.5	34200	28500	24428	21375	19000	17100
Δ%	0	-0.12	0.19	-0.27	0.05	-0.07	0.15	-0.17	-0.49	-0.05
Ť										
Real	60122	30177	20240	15015	12013	10187	8586	7395	6716	6062
$S_{\check{T}}/n$	60122	30061	20041	15031	12024	10020	8589	7515	6680	6012
Δ%	0.00	-0.39	-0.99	0.10	0.09	-1.66	0.03	1.60	-0.54	-0.83
Č										
Real	187598	93713	62716	46871	37645	31384	26860	23460	20791	18696
$S_{\check{C}}/n$	187598	93799	62533	46900	37520	31266	26800	23450	20844	18760
Δ%	0	0.09	-0.29	0.06	-0.33	-0.38	-0.22	-0.04	0.26	0.34
Ğ										
Real	115522	57628	38234	28809	23007	18950	16482	14514	12760	11557
$S_{\check{G}}/n$	115522	57761	38507	28881	23104	19254	16503	14440	12836	11552
Δ%	0	0.23	0.71	0.25	0.42	1.58	0.13	-0.51	0.59	-0.04



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

Fig. 16.10. Numeric and graphical representations of the sequences of phonetic oligomer sums $\Sigma_{Ă,n,1}$ (in blue), $\Sigma_{Ť,n,1}$ (in red), $\Sigma_{Č,n,1}$ (in brown), and $\Sigma_{Ğ,n,1}$ (in green) in the epi-chain $Ď_{2/1}$ of the second order, which is the subsequence of the full phonetic sequence $Ď_1$ representing the Russian text of the novel «War and Peace».

n	1	2	3	4	5	6	7	8	9	10
Ă										
Real	114314	56890	38177	28519	22911	19093	16399	14318	12629	11273
$S_{Ă}/n$	114314	57157	38105	28578.5	22863	19052	16331	14289	12702	11431
Δ%	0	0.47	-0.19	0.21	-0.21	-0.21	-0.42	-0.20	0.57	1.39
Ť										
Real	40309	20240	13336	10187	7981	6716	5818	5028	4480	4044
$S_{Ť}/n$	40309	20154.5	13436	10077	8062	6718	5758	5039	4479	4031
Δ%	0	-0.42	0.75	-1.09	1.00	0.03	-1.03	0.21	-0.03	-0.32
Č										
Real	125138	62716	41595	31384	25065	20791	17834	15700	13906	12643
$S_{Č}/n$	125138	62569	41713	31285	25028	20856	17877	15642	13904	12514
Δ%	0	-0.23	0.28	-0.32	-0.15	0.31	0.24	-0.37	-0.01	-1.03
Ğ										
Real	76399	38234	25612	18950	15275	12760	10829	9474	8558	7656
$S_{Ğ}/n$	76399	38200	25466	19100	15280	12733	10914	9550	8489	7640
Δ%	0	-0.09	-0.57	0.78	0.03	-0.21	0.78	0.79	-0.82	-0.21

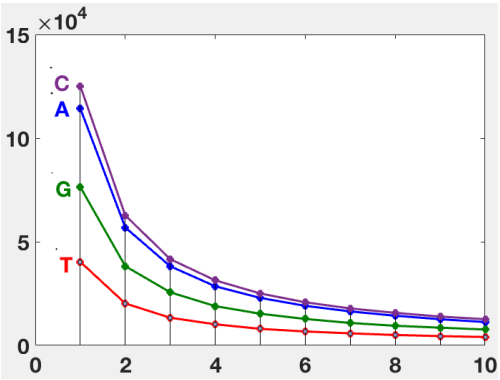


Fig. 16.11. Numeric and graphical representations of the sequences of phonetic oligomer sums $\Sigma_{Ă,n,1}$ (in blue), $\Sigma_{Ť,n,1}$ (in red), $\Sigma_{Č,n,1}$ (in brown), and $\Sigma_{Ğ,n,1}$ (in green) in the epi-chain $Ď_{3/1}$ of the third order, which is the subsequence of the full phonetic sequence $Ď_1$ representing the Russian text of the novel «War and Peace».

n	1	2	3	4	5	6	7	8	9	10
Ă										
Real	85602	42865	28519	21411	17109	14318	12241	10651	9586	8534
$S_{Ă}/n$	85602	42801	28534	21400.5	17120	14267	12229	10700	9511	8560
Δ%	0.00	-0.15	0.05	-0.05	0.07	-0.36	-0.10	0.46	-0.79	0.31
Ť										
Real	30177	15015	10187	7395	6062	5028	4231	3700	3378	3012
$S_{Ť}/n$	30177	15088.5	10059	7544	6035	5030	4311	3772	3353	3018
Δ%	0	0.49	-1.27	1.98	-0.44	0.03	1.86	1.91	-0.75	0.19

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

Č										
Real	93713	46871	31384	23460	18696	15700	13377	11812	10336	9381
$S_{Č}/n$	93713	46856.5	31238	23428	18743	15619	13388	11714	10413	9371
Δ%	0	-0.03	-0.47	-0.14	0.25	-0.52	0.08	-0.84	0.74	-0.10
Ğ										
Real	57628	28809	18950	14514	11557	9474	8311	7227	6380	5785
$S_{Ğ}/n$	57628	28814	19209	14407	11526	9605	8233	7204	6403	5763
Δ%	0	0.02	1.35	-0.74	-0.27	1.36	-0.95	-0.33	0.36	-0.39

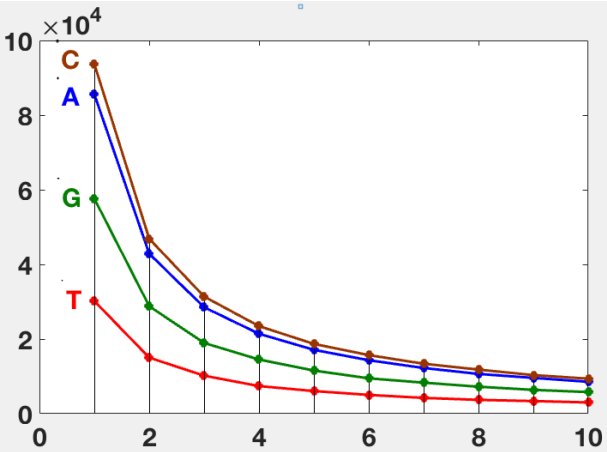


Fig. 16.12. Numeric and graphical representations of the sequences of phonetic oligomer sums $\Sigma_{Ă,n,1}$ (in blue), $\Sigma_{Ť,n,1}$ (in red), $\Sigma_{Č,n,1}$ (in brown), and $\Sigma_{Ğ,n,1}$ (in green) in the epi-chain $Ď_{4/1}$ of the fourth order, which is the subsequence of the full phonetic sequence $Ď_1$ representing the Russian text of the novel «War and Peace».

Now let us return to the prediction of the quantum-information model that the amount S_2 of any of 16 phonetic doublets ŇŇ was algebra-harmonically interrelated on the basis of the harmonic progression (2.4) with the total amounts S_3, S_4, S_5, … of phonetic oligomers in the following sets: 4 triplets, which start with this attributive phonetic doublet ŇŇ; 16 tetraplets, which start with this attributive doublet ŇŇ; 64 pentaplets, which start with this attributive doublet ŇŇ; and so on. 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, equal to values 3/2, 4/2, 5/2, … .

Numeric data in Fig. 16.13 show that - for each of the considered 16 phonetic doublets - the ratios of the total amounts S_2/S_3, S_2/S_4, S_2/S_5 are 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, they are equal to values 3/2, 4/2, 5/2 in the case of the novel “War and Piece” (by analogy with the novel “Anna Karenina” in Fig. 16.8).

DOUBLETS	TRIPLETS	TETRAPLETS	PENTAPLETS	S_2/S_3	S_2/S_4	S_2/S_5
$S_2 = Σ(ĂĂ)$	$S_3 = Σ(ĂĂŇ)_4$	$S_4 = Σ(ĂĂŇŇ)_{16}$	$S_5 = Σ(ĂĂŇŇŇ)_{64}$	**1.5**	**2.0**	**2.5**
9899	6683	5040	3988			
$S_2 = Σ(ĂŤ)$	$S_3 = Σ(ĂŤŇ)_4$	$S_4 = Σ(ĂŤŇŇ)_{16}$	$S_5 = Σ(ĂŤŇŇŇ)_{64}$	**1.5**	**2.0**	**2.5**
10528	7175	5340	5340			
$S_2 = Σ(ĂČ)$	$S_3 = Σ(ĂČŇ)_4$	$S_4 = Σ(ĂČŇŇ)_{16}$	$S_5 = Σ(ĂČŇŇŇ)_{64}$	**1.5**	**2.0**	**2.5**
95923	63779	47943	38202			

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

$S_2 = \Sigma(ĂĞ)$	$S_3 = \Sigma(ĂĞŇ)_4$	$S_4 = \Sigma(ĂĞŇŇ)_{16}$	$S_5 = \Sigma(ĂĞŇŇŇ)_{64}$			
54647	36677	27278	22284	**1.5**	**2.0**	**2.5**
$S_2 = \Sigma(ŤĂ)$	$S_3 = \Sigma(ŤĂŇ)_4$	$S_4 = \Sigma(ŤĂŇŇ)_{16}$	$S_5 = \Sigma(ŤĂŇŇŇ)_{64}$			
4236	2833	2109	1677	**1.5**	**2.0**	**2.5**
$S_2 = \Sigma(ŤŤ)$	$S_3 = \Sigma(ŤŤŇ)_4$	$S_4 = \Sigma(ŤŤŇŇ)_{16}$	$S_5 = \Sigma(ŤŤŇŇŇ)_{64}$			
2430	1595	1197	969	**1.5**	**2.0**	**2.5**
$S_2 = \Sigma(ŤČ)$	$S_3 = \Sigma(ŤČŇ)_4$	$S_4 = \Sigma(ŤČŇŇ)_{16}$	$S_5 = \Sigma(ŤČŇŇŇ)_{64}$			
36093	24260	18135	14208	**1.5**	**2.0**	**2.5**
$S_2 = \Sigma(ŤĞ)$	$S_3 = \Sigma(ŤĞŇ)_4$	$S_4 = \Sigma(ŤĞŇŇ)_{16}$	$S_5 = \Sigma(ŤĞŇŇŇ)_{64}$			
17363	11621	8736	7046	**1.5**	**2.0**	**2.5**
$S_2 = \Sigma(ČĂ)$	$S_3 = \Sigma(ČĂŇ)_4$	$S_4 = \Sigma(ČĂŇŇ)_{16}$	$S_5 = \Sigma(ČĂŇŇŇ)_{64}$			
103846	69288	51766	41636	**1.5**	**2.0**	**2.5**
$S_2 = \Sigma(ČŤ)$	$S_3 = \Sigma(ČŤŇ)_4$	$S_4 = \Sigma(ČŤŇŇ)_{16}$	$S_5 = \Sigma(ČŤŇŇŇ)_{64}$			
31266	20667	15675	12615	**1.5**	**2.0**	**2.5**
$S_2 = \Sigma(ČČ)$	$S_3 = \Sigma(ČČŇ)_4$	$S_4 = \Sigma(ČČŇŇ)_{16}$	$S_5 = \Sigma(ČČŇŇŇ)_{64}$			
31923	21436	16023	12826	**1.5**	**2.0**	**2.5**
$S_2 = \Sigma(ČĞ)$	$S_3 = \Sigma(ČĞŇ)_4$	$S_4 = \Sigma(ČĞŇŇ)_{16}$	$S_5 = \Sigma(ČĞŇŇŇ)_{64}$			
20563	13746	10249	8354	**1.5**	**2.0**	**2.5**
$S_2 = \Sigma(ĞĂ)$	$S_3 = \Sigma(ĞĂŇ)_4$	$S_4 = \Sigma(ĞĂŇŇ)_{16}$	$S_5 = \Sigma(ĞĂŇŇŇ)_{64}$			
53674	35082	26823	21174	**1.5**	**2.0**	**2.5**
$S_2 = \Sigma(ĞŤ)$	$S_3 = \Sigma(ĞŤŇ)_4$	$S_4 = \Sigma(ĞŤŇŇ)_{16}$	$S_5 = \Sigma(ĞŤŇŇŇ)_{64}$			
16125	10744	8100	6381	**1.5**	**2.0**	**2.5**
$S_2 = \Sigma(ĞČ)$	$S_3 = \Sigma(ĞČŇ)_4$	$S_4 = \Sigma(ĞČŇŇ)_{16}$	$S_5 = \Sigma(ĞČŇŇŇ)_{64}$			
23506	15788	11713	9274	**1.5**	**2.0**	**2.5**
$S_2 = \Sigma(ĞĞ)$	$S_3 = \Sigma(ĞĞŇ)_4$	$S_4 = \Sigma(ĞĞŇŇ)_{16}$	$S_5 = \Sigma(ĞĞŇŇŇ)_{64}$			
22217	14785	10992	8780	**1.5**	**2.0**	**2.5**

Fig. 16.13. The confirmation of the quantum-information model prediction by the comparison of the amount S_2 of each of 16 phonetic doublets to the total amounts S_3, S_4, S_5 of phonetic *n*-plets (*n* = 3, 4, 5), which start with this doublet, in the phonetic sequence, representing the Russian novel «War and Peace». Tabular data present total sums of each kind of phonetic oligomers: for example, the total sum $S_4 = \Sigma(ĂĂŇŇ)_{16}$ of all 16 phonetic tetraplets, which start with the phonetic doublet ĂĂ, is equal to 5040. The symbol Ň denotes any of phonetic monomers Ă, Ť, Ğ, and Č.

Similar results were received at such an analysis of other famous long Russian literary texts, some of which are presented in the next subsection.

- 16.3. The analysis of Russian novels by F.M. Dostoevsky and A.S.Pushkin

This subsection gives some results of the analysis - by the oligomer sums method - of the phonetic sequences representing a few long Russian literary works by F.M. Dostoevsky and A.S. Pushkin, as well as the Russian text of the Bible. Here, for brevity, the author shows only the initial data of the analysis of the named phonetic sequences without presenting many additional results that are similar to those shown above for Tolstoy's novels in Figs. 16.5-16.8, 16.10-16.13.

Fig. 16.14 gives numeric and graphic results of the named analysis of the phonetic sequence of the Russian text of the novel "Crime and Punishment" by Dostoevsky. The text contains 818099 phonetic letters, and it was taken from http://samolit.com/books/57/ .

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

n	1	2	3	4	5	6	7	8	9	10
Ă										
Real	258836	129490	86219	64656	51620	43194	36976	32292	28640	25813
$S_{Ă}/n$	258836	129418	86279	64709	51767	43139	36977	32355	28760	25884
Δ%	0.00	-0.06	0.07	0.08	0.28	-0.13	0.00	0.19	0.42	0.27
Ť										
Real	95575	47872	31935	24008	19116	15893	13569	12172	10699	9561
$S_{Ť}/n$	95575	47787.5	31858	23894	19115	15929	13654	11947	10619	9558
Δ%	0	-0.18	-0.24	-0.48	-0.01	0.23	0.62	-1.88	-0.75	-0.04
Č										
Real	279404	139555	93348	69630	56055	46631	40085	34827	30937	28035
$S_{Č}/n$	279404	139702	93135	69851	55881	46567	39915	34926	31045	27940
Δ%	0	0.11	-0.23	0.32	-0.31	-0.14	-0.43	0.28	0.35	-0.34
Ğ										
Real	184284	92132	61197	46230	36828	30631	26241	22971	20623	18400
$S_{Ğ}/n$	184284	92142	61428	46071	36857	30714	26326	23036	20476	18428
Δ%	0	0.01	0.38	-0.35	0.08	0.27	0.32	0.28	-0.72	0.15

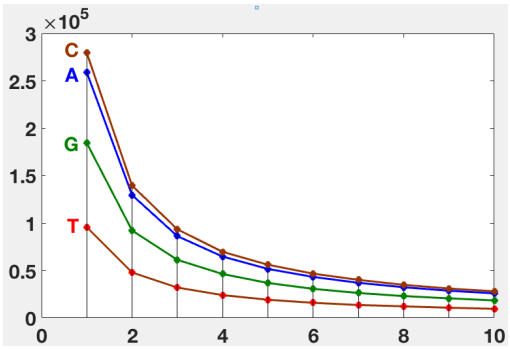


Fig. 16.14. The table shows real values and model values $S_{Ă}/n$, or $S_{Ť}/n$, or $S_{Č}$, or $S_{Ğ}/n$ (in red) in the case of the oligomer sums representations of the phonetic sequence of the Russian novel «Crime and Punishment» by F. Dostoevsky. Symbols Ă, Ť, Ğ, and Č refer to the phonetic monomers (Fig. 16.1, at right). The symbol Δ% denotes deviations of real values from model values in percent (the model values are taken as 100%). Graphs show the hyperbolic-like sequences of the phonetic oligomer sums $\Sigma_{Ă,n,1}$, $\Sigma_{Ť,n,1}$, $\Sigma_{Č,n,1}$, and $\Sigma_{Ğ,n,1}$, which practically coincide with the model hyperbolic sequences $S_{Ă}/n$ (in blue), $S_{Ť}/n$ (in red), $S_{Č}$,(in brown), and $S_{Ğ}/n$ (in green).

Fig. 16.15 gives numeric and graphic results of the named analysis of the phonetic sequence of the Russian text of the novel "Idiot" by Dostoevsky. The text contains 1001129 phonetic letters; it was taken from http://samolit.com/books/56/.

n	1	2	3	4	5	6	7	8	9	10
Ă										
Real	310571	155325	103572	77647	61814	51742	44356	38820	34503	30853
$S_{Ă}/n$	310571	155286	103524	77643	62114	51762	44367	38821	34508	31057
Δ%	0	-0.03	-0.05	-0.01	0.48	0.04	0.03	0.00	0.01	0.66
Ť										
Real	122981	61401	41058	30847	24641	20461	17617	15426	13705	12389
$S_{Ť}/n$	122981	61490.5	40994	30745	24596	20497	17569	15373	13665	12298
Δ%	0	0.15	-0.16	-0.33	-0.18	0.17	-0.27	-0.35	-0.30	-0.74

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

Č										
Real	344200	172087	114890	86125	69127	57432	49224	42913	38386	34655
$S_{Č}/n$	344200	172100	114733	86050	68840	57367	49171	43025	38244	34420
Δ%	0	0.01	-0.14	-0.09	-0.42	-0.11	-0.11	0.26	-0.37	-0.68
Ğ										
Real	223377	111751	74189	55663	44643	37219	31821	27982	24642	22215
$S_{Ğ}/n$	223377	111689	74459	55844	44675	37230	31911	27922	24820	22338
Δ%	0	-0.06	0.36	0.32	0.07	0.03	0.28	-0.21	0.72	0.55

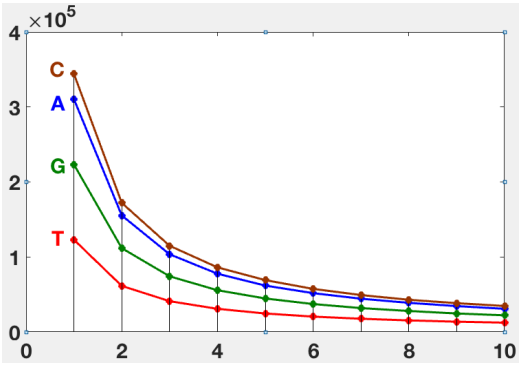


Fig. 16.5. The table shows real values and model values $S_{Ă}/n$, or $S_{Ť}/n$, or $S_{Č}$, or $S_{Ğ}/n$ (in red) in the case of the oligomer sums representations of the phonetic sequence of the Russian novel «Idiot» by F. Dostoevsky. Symbols Ă, Ť, Ğ, and Č refer to the phonetic monomers (Fig. 16.1, at right). The symbol Δ% denotes deviations of real values from model values in percent (the model values are taken as 100%). Graphs show the hyperbolic-like sequences of the phonetic oligomer sums $\Sigma_{Ă,n,1}$, $\Sigma_{Ť,n,1}$, $\Sigma_{Č,n,1}$, and $\Sigma_{Ğ,n,1}$, which practically coincide with the model hyperbolic sequences $S_{Ă}/n$ (in blue), $S_{Ť}/n$ (in red), $S_{Č}$,(in brown), and $S_{Ğ}/n$ (in green).

Fig. 16.16 gives numeric and graphic results of the named analysis of the phonetic sequence of the Russian text of the novel "Evgenij Onegin" by A.S. Pushkin. The text is relatively short and contains 107146 phonetic letters; it was taken from http://tululu.org/b57798/.

n	1	2	3	4	5	6	7	8	9	10
Ă										
Real	31725	15781	10683	7862	6436	5304	4561	3849	3536	3182
$S_{Ă}/n$	31725	15863	10575	7931.25	6345	5288	4532	3966	3525	3173
Δ%	0	0.51	-1.02	0.87	-1.43	-0.31	-0.64	2.94	-0.31	-0.30
Ť										
Real	12602	6380	4207	3211	2483	2114	1799	1599	1392	1263
$S_{Ť}/n$	12602	6301	4201	3151	2520	2100	1800	1575	1400	1260
Δ%	0.00	-1.25	-0.15	-1.92	1.48	-0.65	0.07	-1.51	0.59	-0.22
Č										
Real	40301	20225	13348	10126	8062	6680	5728	5105	4441	4065
$S_{Č}/n$	40301	20150.5	13434	10075	8060	6717	5757	5038	4478	4030
Δ%	0.00	-0.37	0.64	-0.50	-0.02	0.55	0.51	-1.34	0.82	-0.87
Ğ										
Real	22518	11187	7477	5587	4448	3759	3218	2840	2536	2204
$S_{Ğ}/n$	22518	11259	7506	5630	4504	3753	3217	2815	2502	2252
Δ%	0.00	0.64	0.39	0.75	1.23	-0.16	-0.04	-0.90	-1.36	2.12

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

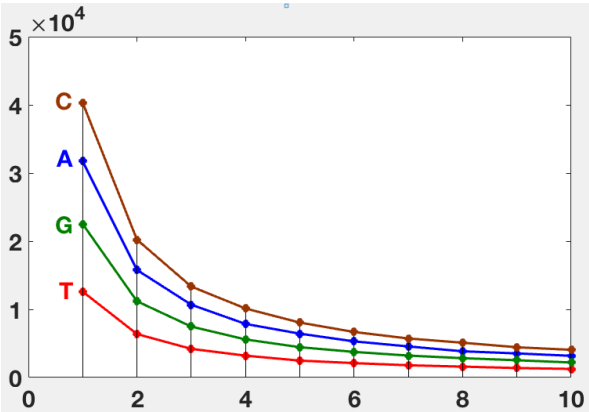


Fig. 16.16. The table shows real values and model values $S_{Ă}/n$, or $S_{Ť}/n$, or $S_{Č}$, or $S_{Ğ}/n$ (in red) in the case of the oligomer sums representations of the phonetic sequence of the Russian novel «Evgenij Onegin» by A.S. Pushkin. Symbols Ă, Ť, Ğ, and Č refer to the phonetic monomers (Fig. 16.1, at right). The symbol Δ% denotes deviations of real values from model values in percent (the model values are taken as 100%). Graphs show the hyperbolic-like sequences of the phonetic oligomer sums $\Sigma_{Ă,n,1}$, $\Sigma_{Ť,n,1}$, $\Sigma_{Č,n,1}$, and $\Sigma_{Ğ,n,1}$, which practically coincide with the model hyperbolic sequences $S_{Ă}/n$ (in blue), $S_{Ť}/n$ (in red), $S_{Č}$,(in brown), and $S_{Ğ}/n$ (in green).

Fig. 16.17 gives numeric and graphic results of the named analysis of the phonetic sequence of the Russian text of the novel "Dubrovsky" by A.S. Pushkin. The text contains 106891 phonetic letters; it was taken from http://samolit.com/books/61/ .

n	1	2	3	4	5	6	7	8	9	10
Ă										
Real	34341	17293	11514	8570	6953	5735	4889	4238	3814	3493
$S_{Ă}/n$	34341	17171	11447	8585.25	6868	5724	4906	4293	3816	3434
Δ%	0	-0.71	-0.59	0.18	-1.23	-0.20	0.34	1.27	0.04	-1.72
Ť										
Real	12135	6071	4050	3085	2367	2065	1720	1540	1374	1148
$S_{Ť}/n$	12135	6067.5	4045	3034	2427	2023	1734	1517	1348	1214
Δ%	0.00	-0.06	-0.12	-1.69	2.47	-2.10	0.78	-1.52	-1.90	5.40
Č										
Real	37714	18772	12456	9436	7557	6200	5471	4778	4178	3766
$S_{Č}/n$	37714	18857	12571	9429	7543	6286	5388	4714	4190	3771
Δ%	0	0.45	0.92	-0.08	-0.19	1.36	-1.55	-1.35	0.30	0.14
Ğ										
Real	22701	11309	7610	5631	4501	3815	3190	2805	2510	2282
$S_{Ğ}/n$	22701	11351	7567	5675	4540	3784	3243	2838	2522	2270
Δ%	0	0.37	-0.57	0.78	0.86	-0.83	1.63	1.15	0.49	-0.52

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

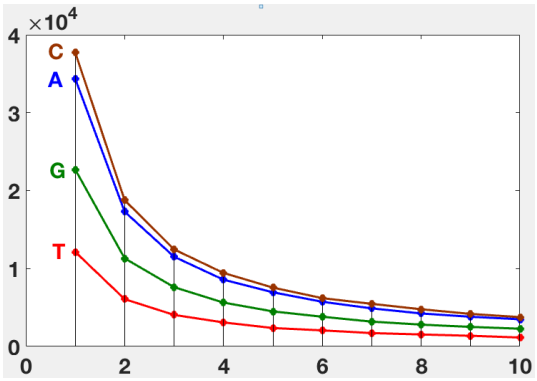


Fig. 16.17. The table shows real values and model values $S_{Ă}/n$, or $S_{Ť}/n$, or $S_{Č}$, or $S_{Ğ}/n$ (in red) in the case of the oligomer sums representations of the phonetic sequence of the Russian novel «Dubrovsky» by A.S. Pushkin. Symbols Ă, Ť, Ğ, and Č refer to the phonetic monomers (Fig. 16.1, at right). The symbol Δ% denotes deviations of real values from model values in percent (the model values are taken as 100%). Graphs show the hyperbolic-like sequences of the phonetic oligomer sums $\Sigma_{Ă,n,1}$, $\Sigma_{Ť,n,1}$, $\Sigma_{Č,n,1}$, and $\Sigma_{Ğ,n,1}$, which practically coincide with the model hyperbolic sequences $S_{Ă}/n$ (in blue), $S_{Ť}/n$ (in red), $S_{Č}$,(in brown), and $S_{Ğ}/n$ (in green).

Fig. 16.18 gives numeric and graphic results of the named analysis of the phonetic sequence of the Russian text of the Bible. The text contains 3122489 phonetic letters; it was taken from http://petoukhov.com/bible.zip.

n	1	2	3	4	5	6	7	8	9	10
Ă										
Real	1026290	513013	342965	256527	205099	171465	146474	128448	114178	102399
$S_{Ă}/n$	1026290	513145	342097	256572.5	205258	171048	146613	128286	114032	102629
Δ%	0	0.03	-0.25	0.02	0.08	-0.24	0.09	-0.13	-0.13	0.22
Ť										
Real	371375	185303	123591	92719	74407	61771	53044	46033	41276	37149
$S_{Ť}/n$	371375	185688	123792	92844	74275	61896	53054	46422	41264	37138
Δ%	0	0.21	0.16	0.13	-0.18	0.20	0.02	0.84	-0.03	-0.03
Č										
Real	1072094	536374	357539	268263	214546	179095	153361	134208	119333	107098
$S_{Č}/n$	1072094	536047	357365	268024	214419	178682	153156	134012	119122	107209
Δ%	0	-0.06	-0.05	-0.09	-0.06	-0.23	-0.13	-0.15	-0.18	0.10
Ğ										
Real	652730	326554	216734	163113	130445	108083	93190	81622	72156	65602
$S_{Ğ}/n$	652730	326365	217577	163183	130546	108788	93247	81591	72526	65273
Δ%	0	-0.06	0.39	0.04	0.08	0.65	0.06	-0.04	0.51	-0.50

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

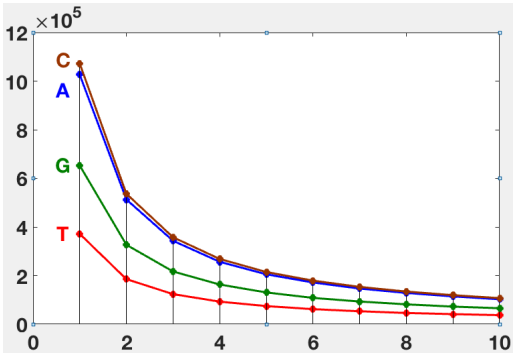


Fig. 16.18. The table shows real values and model values $S_{Ă}/n$, or $S_{Ť}/n$, or $S_{Č}$, or $S_{Ğ}/n$ (in red) in the case of the oligomer sums representations of the phonetic sequence of the Russian text of the Bible. Symbols Ă, Ť, Ğ, and Č refer to the phonetic monomers (Fig. 16.1, at right). The symbol Δ% denotes deviations of real values from model values in percent (the model values are taken as 100%). Graphs show the hyperbolic-like sequences of the phonetic oligomer sums $\Sigma_{Ă,n,1}$, $\Sigma_{Ť,n,1}$, $\Sigma_{Č,n,1}$, and $\Sigma_{Ğ,n,1}$, which practically coincide with the model hyperbolic sequences $S_{Ă}/n$ (in blue), $S_{Ť}/n$ (in red), $S_{Č}$,(in brown), and $S_{Ğ}/n$ (in green).

Let us explain additionally in more detail why we analyze long literary texts in Russian specifically. This explanation uses the well-known facts about linguistics and alphabet writing [Coulmas, 1996; https://en.wikipedia.org/wiki/Alphabet]. Alphabetical writing differs from pictographic (ideographic) writing, where signs denote concepts (Sumerian cuneiform writing), and from morphemic and logographic writing, where signs denote individual morphemes (Chinese writing) or words. The Russian alphabet belongs to the class of consonant vocal alphabets, where letters denote both vowels and consonants. When an alphabet is adopted or developed to represent a given language, an orthography generally comes into being, providing rules for the spelling of words in that language. In accordance with the principle on which alphabets are based, these rules will generally map letters of the alphabet to the phonemes (significant sounds) of the spoken language. In a perfectly phonemic orthography, there would be consistent one-to-one correspondence between the letters and the phonemes so that a writer could predict the spelling of a word given its pronunciation, and a speaker would always know the pronunciation of a word given its spelling, and vice versa. However this ideal is not usually achieved in practice; some languages (such as Spanish and Finnish) come close to it, while others (such as English) deviate from it to a much larger degree.

Languages may fail to achieve a one-to-one correspondence between letters and sounds in different ways. For example, a language may represent a given phoneme by a combination of letters rather than just a single letter. In this way, German uses the tetragraphs (four letters) "tsch" for the phoneme [tʃ] and (in a few borrowed words) "dsch" for [dʒ]. National languages sometimes elect to address the problem of dialects by simply associating the alphabet with the national standard. Some national languages like Russian, Finnish, Armenian, Turkish, Bulgarian have a very regular spelling system with a nearly one-to-one correspondence between letters and phonemes. French has silent letters. In English, the pronunciations of many words simply have to be memorized as they do not correspond to the spelling in a consistent way. For English, this is partly because the Great Vowel Shift occurred after the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

orthography was established, and because English has acquired a large number of loanwords at different times, retaining their original spelling at varying levels.

In the Russian alphabet, a one-to-one correspondence between the letters and the phonemes exists. For this reason, analyzing long literary Russian texts, researches can study their phonetic structures. In this way, the author has received interesting results described above. But he would very much like to continue the analysis - by the method of oligomeric sums - of the phonetic representations of long literary texts in all those languages, in which letters and phonemes are near to a one-to-one correspondence. The results of such an analysis can more fully show the deep connection of genetic and linguistic languages, testifying that linguistic languages are a continuation and a superstructure over the general biological language of eukaryotic and prokaryotic genomes.

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 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 world and its evolution appear. In particular, the oligomer sums method can be used for the analysis of amino acid primary sequences in long proteins (see an example in Section 15).

The discovery of the algebraic genomic invariants gives new knowledge about the unity of the world of all living organisms and about 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 published author's letter [Petoukhov, 2020d].

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

 doi:10.20944/preprints202005.0471.v2

inherited physiological systems as parts of a whole organism must be structurally coupled with a genetic code for transmission to descendants in encoded form. Therefore, inherited physiological macrostructures can bear the imprint of structural features of the genetic code. For this reason, structural analogies exist between the genetic system and the properties of inherited physiological systems, for example, the unified properties of different sensor systems, which are reflected in the main psychophysical Weber-Fechner law [Petoukhov, 2016, 2019b, 2020a-c]. These problems are discussed at the International interdisciplinary seminar "Algebraic Biology and Theory of Systems" in Moscow [Petoukhov, Tolokonnikov, 2020].

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, Davies, Pati, 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. In particular, these materials include results on such intellectual brain activity as the writing of long Russian literary texts, whose phonetic sequences obey the hyperbolic rules, which are similar to the hyperbolic rules of eukaryotic and prokaryotic genomes and correspond to the author's quantum-information model (see Section 16).

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]). 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,*

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

but we know very little about how to read it." [Fickett & 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 sub-alphabets 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].

Appendix I. Numeric data on some epi-chains of the human chromosome № 1.

This Appendix shows numeric data about epi-chains represented graphically above in Figs. 13.2.-13.6.

n	1	2	3	4	5	6	7	8	9	10
A										
Real	33537501	16768845	11179286	8383461	6706672	5588773	4792078	4192017	3726860	3354107
Model	33537501	16768751	11179167	8384375	6707500	5589584	4791072	4192188	3726389	3353750
Δ%	0	-0.001	-0.001	0.011	0.012	0.015	-0.021	0.004	-0.013	-0.011
T										
Real	33620498	16808862	11207274	8405040	6724359	5601854	4801395	4202773	3735327	3360459
Model	33620498	16810249	11206833	8405125	6724100	5603416	4802928	4202562	3735611	3362050
Δ%	0	0.008	-0.004	0.001	-0.004	0.028	0.032	-0.005	0.008	0.047
C										
Real	24024903	12013624	8005708	6008215	4803919	4002753	3433636	3003511	2668499	2402186
Model	24024903	12012451.5	8008301	6006226	4804981	4004151	3432129	3003113	2669434	2402490
Δ%	0	-0.010	0.032	-0.033	0.022	0.035	-0.044	-0.013	0.035	0.013
G										
Real	24057606	12028924	8021235	6013412	4813156	4013372	3435824	3006763	2673815	2407301
Model	24057606	12028803	8019202	6014402	4811521	4009601	3436801	3007201	2673067	2405761
Δ%	0	-0.001	-0.025	0.016	-0.034	-0.094	0.028	0.015	-0.028	-0.064

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

n	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
A										
Real	3049510	2793265	2579432	2394635	2235831	2095893	1974049	1863181	1766123	1677938
Model	3048864	2794792	2579808	2395536	2235833	2096094	1972794	1863195	1765132	1676875
Δ%	-0.021	0.055	0.015	0.038	0.000	0.010	-0.064	0.001	-0.056	-0.063
T										
Real	3054627	2802390	2588494	2400749	2240133	2101692	1976639	1866079	1768246	1680401
Model	3056409	2801708	2586192	2401464	2241367	2101281	1977676	1867805	1769500	1681025
Δ%	0.058	-0.024	-0.089	0.030	0.055	-0.020	0.052	0.092	0.071	0.037
C										
Real	2185450	2001255	1845496	1718676	1600752	1501210	1413949	1335576	1263741	1200991
Model	2184082	2002075	1848069	1716065	1601660	1501556	1413230	1334717	1264469	1201245
Δ%	-0.063	0.041	0.139	-0.152	0.057	0.023	-0.051	-0.064	0.058	0.021
G										
Real	2186817	2006465	1851228	1717404	1605990	1503735	1414212	1337414	1267181	1202700
Model	2187055	2004801	1850585	1718400	1603840	1503600	1415153	1336534	1266190	1202880
Δ%	0.011	-0.083	-0.035	0.058	-0.134	-0.009	0.067	-0.066	-0.078	0.015

Fig. I.1. 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. 13.1b), which consists of nucleotides with serial numerations 1-3-5-7-9-… in the DNA sequence of the human chromosome № 1. The table demonstrates that the model hyperbolic progressions S_A/n, S_T/n, S_C/n, S_G/n (in red) almost completely coincide with 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. Differences between the corresponding values in these numerical sequences are expressed by shown small percentage values Δ%.

n	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
A										
Real	**A** Real	22360413	11179286	7453552	5588773	4472245	3726860	3196917	2793265	2483348
Model	Model	22360413	11180207	7453471	5590103.25	4472083	3726736	3194345	2795052	2484490
Δ%	Δ%	0	0.008	-0.001	0.024	-0.004	-0.003	-0.081	0.064	0.046
T										
Real	22412993	11207274	7470145	5601854	4479492	3735327	3199876	2802390	2490602	2240133
Model	22412993	11206496.5	7470998	5603248	4482599	3735499	3201856	2801624	2490333	2241299
Δ%	0	-0.007	0.011	0.025	0.069	0.005	0.062	-0.027	-0.011	0.052
C										
Real	16012711	8005708	5336968	4002753	3202830	2668499	2287279	2001255	1778911	1600752
Model	16012711	8006355.5	5337570	4003178	3202542	2668785	2287530	2001589	1779190	1601271
Δ%	0	0.008	0.011	0.011	-0.009	0.011	0.011	0.017	0.016	0.032
G										
Real	16040889	8021235	5348337	4013372	3210839	2673815	2291215	2006465	1783466	1605990
Model	16040889	8020445	5346963	4010222	3208178	2673482	2291556	2005111	1782321	1604089
Δ%	0	-0.010	-0.026	-0.079	-0.083	-0.012	0.015	-0.068	-0.064	-0.119

n	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
A										
Real	2032220	1863181	1721074	1598554	1489212	1397489	1316829	1240400	1177210	1117975
Model	2032765	1863368	1720032	1597172	1490694	1397526	1315318	1242245	1176864	1118021
Δ%	0.027	0.010	-0.061	-0.087	0.099	0.003	-0.115	0.149	-0.029	0.004
T										
Real	2038200	1866079	1723940	1598312	1493164	1401402	1318617	1245654	1178340	1119290
Model	2037545	1867749	1724076	1600928	1494200	1400812	1318411	1245166	1179631	1120650
Δ%	-0.032	0.089	0.008	0.163	0.069	-0.042	-0.016	-0.039	0.109	0.121
C										
Real	1455982	1335576	1231496	1144356	1067083	999584	941265	889682	843256	800223
Model	1455701	1334393	1231747	1143765	1067514	1000794	941924	889595	842774	800636
Δ%	-0.019	-0.089	0.020	-0.052	0.040	0.121	0.070	-0.010	-0.057	0.052

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

G										
Real	1457871	1337414	1233259	1146423	1072344	1003210	942530	892426	844720	803863
Model	1458263	1336741	1233915	1145778	1069393	1002556	943582	891161	844257	802044
Δ%	0.027	-0.050	0.053	-0.056	-0.276	-0.065	0.111	-0.142	-0.055	-0.227

Fig. I.2. 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. 13.1d), which consists of nucleotides with serial numerations 1-4-7-10-13-… in the DNA sequence of the human chromosome № 1. The table demonstrates that the model hyperbolic progressions S_A/n, S_T/n, S_C/n, S_G/n (in red) almost completely coincide with 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. Differences between the corresponding values in these numerical sequences are expressed by shown small percentage values Δ%.

n	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
A										
Real	16768845	8383461	5588773	4192017	3354107	2793265	2394635	2095893	1863181	1677938
Model	16768845	8384423	5589615	4192211.25	3353769	2794808	2395549	2096106	1863205	1676885
Δ%	0	0.011	0.015	0.005	-0.010	0.055	0.038	0.010	0.001	-0.063
T										
Real	16808862	8405040	5601854	4202773	3360459	2802390	2400749	2101692	1866079	1680401
Model	16808862	8404431	5602954	4202216	3361772	2801477	2401266	2101108	1867651	1680886
Δ%	0	-0.007	0.020	-0.013	0.039	-0.033	0.022	-0.028	0.084	0.029
C										
Real	12013624	6008215	4002753	3003511	2402186	2001255	1718676	1501210	1335576	1200991
Model	12013624	6006812	4004541	3003406	2402725	2002271	1716232	1501703	1334847	1201362
Δ%	0	-0.023	0.045	-0.003	0.022	0.051	-0.142	0.033	-0.055	0.031
G										
Real	12028924	6013412	4013372	3006763	2407301	2006465	1717404	1503735	1337414	1202700
Model	12028924	6014462	4009641	3007231	2405785	2004821	1718418	1503616	1336547	1202892
Δ%	0	0.017	-0.093	0.016	-0.063	-0.082	0.059	-0.008	-0.065	0.016

n	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
A										
Real	1524710	1397489	1290062	1196717	1117975	1047993	987755	930924	882614	839279
Model	1524440	1397404	1289911	1197775	1117923	1048053	986403	931603	882571	838442
Δ%	-0.018	-0.006	-0.012	0.088	-0.005	0.006	-0.137	0.073	-0.005	-0.100
T										
Real	1527023	1401402	1293440	1199582	1119290	1049849	988367	934203	884323	839809
Model	1528078	1400739	1292989	1200633	1120591	1050554	988757	933826	884677	840443
Δ%	0.069	-0.047	-0.035	0.088	0.116	0.067	0.039	-0.040	0.040	0.075
C										
Real	1093622	999584	923273	860649	800223	751218	706684	667789	631744	601012
Model	1092148	1001135	924125	858116	800908	750852	706684	667424	632296	600681
Δ%	-0.135	0.155	0.092	-0.295	0.086	-0.049	0.000	-0.055	0.087	-0.055
G										
Real	12028924	6013412	4013372	3006763	2407301	2006465	1717404	1503735	1337414	1202700
Model	12028924	6014462	4009641	3007231	2405785	2004821	1718418	1503616	1336547	1202892
Δ%	0	0.017	-0.093	0.016	-0.063	-0.082	0.059	-0.008	-0.065	0.016

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

Fig. I.3. 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 table demonstrates that the model hyperbolic progressions S_A/n, S_T/n, S_C/n, S_G/n (in red) almost completely coincide with 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. Differences between the corresponding values in these numerical sequences are expressed by shown small percentage values Δ%.

n	1	2	3	4	5	6	7	8	9	10
A										
Real	6706672	3354107	2235831	1677938	1341408	1117975	958626	839279	744475	670703
Model	6706672	3353336	2235557	1676668	1341334	1117779	958096	838334	745186	670667
Δ%	0	-0.023	-0.012	-0.076	-0.005	-0.018	-0.055	-0.113	0.095	-0.005
T										
Real	6724359	3360459	2240133	1680401	1344421	1119290	961102	839809	746575	672348
Model	6724359	3362179.5	2241453	1681090	1344872	1120727	960623	840545	747151	672436
Δ%	0	0.051	0.059	0.041	0.034	0.128	-0.050	0.088	0.077	0.013
C										
Real	4803919	2402186	1600752	1200991	961518	800223	686222	601012	533486	480738
Model	4803919	2401959.5	1601306	1200980	960784	800653	686274	600490	533769	480392
Δ%	0	-0.009	0.035	-0.001	-0.076	0.054	0.008	-0.087	0.053	-0.072
G										
Real	4813156	2407301	1605990	1202700	962275	803863	686639	600918	536368	481023
Model	4813156	2406578	1604385	1203289	962631	802193	687594	601645	534795	481316
Δ%	0	-0.030	-0.100	0.049	0.037	-0.208	0.139	0.121	-0.294	0.061

n	11	12	13	14	15	16	17	18	19	20
A										
Real	610306	559209	515854	479353	446769	420435	394716	371969	353254	335131
Model	609697	558889	515898	479048	447111	419167	394510	372593	352983	335334
Δ%	-0.100	-0.057	0.008	-0.064	0.077	-0.303	-0.052	0.167	-0.077	0.060
T										
Real	611496	559871	517229	480135	447813	419315	395062	372883	354165	336406
Model	611305	560363	517258	480311	448291	420272	395551	373576	353914	336218
Δ%	-0.031	0.088	0.006	0.037	0.107	0.228	0.124	0.185	-0.071	-0.056
C										
Real	436216	400115	369357	343754	320358	300365	282859	267188	252122	240344
Model	436720	400327	369532	343137	320261	300245	282583	266884	252838	240196
Δ%	0.115	0.053	0.047	-0.180	-0.030	-0.040	-0.098	-0.114	0.283	-0.062
G										
Real	437262	401484	370485	343053	321595	300385	283136	268414	253519	240527
Model	437560	401096	370243	343797	320877	300822	283127	267398	253324	240658
Δ%	0.068	-0.097	-0.065	0.216	-0.224	0.145	-0.003	-0.380	-0.077	0.054

Fig. I.4. 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 table demonstrates that the model hyperbolic progressions S_A/n, S_T/n, S_C/n, S_G/n (in red) almost completely coincide with 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. Differences between the corresponding values in these numerical sequences are expressed by shown small percentage values Δ%.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

n	1	2	3	4	5	6	7	8	9	10
A										
Real	1341408	670703	446769	335131	268213	223299	191485	167237	148619	133939
Model	1341408	670704	447136	335352	268282	223568	191630	167676	149045	134141
Δ%	0	0.000	0.082	0.066	0.026	0.120	0.076	0.262	0.286	0.150
T										
Real	1344421	672348	447813	336406	269243	224202	192407	168101	149090	134818
Model	1344421	672210.5	448140	336105	268884	224070	192060	168053	149380	134442
Δ%	0	-0.020	0.073	-0.089	-0.133	-0.059	-0.181	-0.029	0.194	-0.280
C										
Real	961518	480738	320358	240344	192359	160018	137048	120522	106967	96272
Model	961518	480759	320506	240380	192304	160253	137360	120190	106835	96152
Δ%	0	0.004	0.046	0.015	-0.029	0.147	0.227	-0.276	-0.123	-0.125
G										
Real	962275	481023	321595	240527	192109	160749	137576	120343	107506	95930
Model	962275	481138	320758	240569	192455	160379	137468	120284	106919	96228
Δ%	0	0.024	-0.261	0.017	0.180	-0.231	-0.079	-0.049	-0.549	0.309

n	11	12	13	14	15	16	17	18	19	20
A										
Real	121816	111840	103193	95777	89643	83769	78850	74151	70625	67280
Model	121946	111784	103185	95815	89427	83838	78906	74523	70600	67070
Δ%	0.107	-0.050	-0.008	0.040	-0.241	0.082	0.071	0.499	-0.035	-0.313
T										
Real	122336	111872	103678	96184	89269	83822	79208	74638	71151	67505
Model	122220	112035	103417	96030	89628	84026	79084	74690	70759	67221
Δ%	-0.095	0.146	-0.252	-0.160	0.401	0.243	-0.157	0.070	-0.554	-0.422
C										
Real	87210	79875	73792	68526	64277	60322	56542	53407	49983	48018
Model	87411	80127	73963	68680	64101	60095	56560	53418	50606	48076
Δ%	0.230	0.314	0.231	0.224	-0.274	-0.378	0.032	0.020	1.231	0.120
G										
Real	87691	80548	73923	68766	64118	60188	56555	53892	50857	47678
Model	87480	80190	74021	68734	64152	60142	56604	53460	50646	48114
Δ%	-0.242	-0.447	0.133	-0.047	0.052	-0.076	0.087	-0.809	-0.417	0.906

Fig. I.5. 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 table demonstrates that the model hyperbolic progressions S_A/n, S_T/n, S_C/n, S_G/n (in red) almost completely coincide with 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. Differences between the corresponding values in these numerical sequences are expressed by shown small percentage values Δ%.

Fig. I.6 shows that normalized values 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 the human chromosome №1, that is, they are independent of the epi-chain order.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 June 2020 doi:10.20944/preprints202005.0471.v2

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. I.6. The normalized values $S_N/(S_A+S_T+S_C+S_G)$ 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. Here N refers to any nucleotide.

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 (2008). ISBN-13: 978-1-84816-253-2.

Albrecht-Buehler G. Asymptotically increasing compliance of genomes with Chargaff's second parity rules through inversions and inverted transpositions. *Proc Natl Acad Sci USA*, 103 (47), pp. 17828–17833 (2006), doi:10.1073/pnas.0605553103

Altaisky M.V., Filatov F.P. Genetic information and quantum gas. - arXiv:quant-ph/0106123v1, submitted on 22.06.2001 (2001).

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

Bruza P.D., Kitto K., Nelson D., McEvoy C. Extracting spooky-activation-at-a-distance from considerations of entanglement. https://arxiv.org/abs/0901.4375, 13 p., from 27.01.2009.

Chargaff, E. Preface to a Grammar of Biology: A hundred years of nucleic acid research. *Science*, 172, pp. 637-642 (1971).

Coulmas Fl. The Blackwell Encyclopedia of Writing Systems. Oxford: Blackwell Publishing (1996). ISBN 978-0-631-21481-6.

Darvas G. Petoukhov's rules on symmetries in long DNA-texts. *Symmetry: Culture and Science*, Vol. 29, Number 2, pp. 318-320 (2018),

https://doi.org/10.26830/symmetry_2018_2_318,
http://journal- scs.symmetry.hu/abstract/?pid=673.

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

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

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

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, p.126-135. Springer, Cham **DOI** https://doi.org/10.1007/978-3-030-39162-1_12

Frank-Kamenetskii M.D. *The most important molecule*. Moscow, Nauka (1988) (in Russian).

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

Gühne O., Tóth G. Entanglement detection. *Phys. Rep*., 474, pp. 1-75 (2009).

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

Harkin A.A., Harkin J.B. Geometry of Generalized Complex Numbers. *Mathematics Magazine*, v. 77(2), p. 118-29 (2004).

Holenstein E. *Roman Jakobson's Approach to Language: Phenomenological Structuralism*. Bloomington and London: Indiana University Press (1975)

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

Jacob F., Jakobson R., Levi-Strauss C., Heritier P.L. Vivre et parler. – «Lettre francaises», p. 1221-1222 (Feb. 1968).

Jakobson, R. *Language in Literature.* Cambridge: MIT Press (1987).

Jakobson, R.O. *Texts, Documents, Studies*. Moscow: RGGU (1999, in Russian).

Jakobson, R., Fant C.G., Halle M. *Preliminaries to Speech Analysis: The distinctive features and their correlates.* Cambridge, MA: The MIT Press (1951).

Jakobson R., Halle M. *Fundamentals of Language*. The Hague: Mouton (1971);

Jakobson R., Waugh L. *The Sound Shape of Language*. Walter de Gruyter (2002). ISBN 9783110889451.

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

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

Igamberdiev A.I. Quantum computation, non-demolition measurements, and reflective control in living systems. – *BioSystems*, v. 77, pp. 47–56 (2004).

Ivanov Viach.Vs. Linguistic way of Roman Jakobson. – In: *Jakobson R.O. Selected*

works. – M., Progress (1985) (in Russian).

Kantor I.L., Solodovnikov A.S. *Hypercomplex numbers*. Berlin, New York: Springer-Verlag (1989). ISBN 978-0-387-96980-0.

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

Kappraff J. *Beyond measure: essays in nature, myth, and number*. Singapore: World Scientific (2002).

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

Makovsky M.M. *Linguistic genetics*. M., Nauka (1992). (in Russian).

Matsuno K., Paton R.C. Is there a biology of quantum information? – *BioSystems*, 55, pp. 39–46 (2000).

McConkey E. *Human Genetics: The molecular Revolution*. Boston, MA: Jones and Barlett (1993).

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

Nielsen M.A., Chuang I.L. *Quantum Computation and Quantum Information*. New York: Cambridge Univ. Press. (2010). https://doi.org/10.1017/CBO9780511976667

Patel A. Quantum algorithms and the genetic code. Pramana – Journal of Physics, v. 56, 2-3, pp. 367-381 (2001a), arXiv:quant-ph/0002037

Patel A. Testing quantum dynamics in genetic information processing. – Journal of Genetics, v. 80, 1, pp. 39-43 (2001b).

Patel A. Why genetic information processing could have a quantum basis. – Journal of Biosciences, v. 26, 2, pp. 145-151 (2001c).

Penrose R. Shadows of the Mind: A Search for the Missing Science of Consciousness. - Oxford University Press, USA, 480 p. (1996).

Petoukhov S.V. *Matrix genetics, algebrases of genetic code, noise immunity*. Moscow, RCD, 316 p. (2008) (in Russian).

Petoukhov S.V. Matrix genetics and algebraic properties of the multi-level system of genetic alphabets. *Neuroquantology*, v. 9, №4, p. 60-81 (2011).

Petoukhov S.V. Symmetries of the genetic code, hypercomplex numbers and genetic matrices with internal complementarities. *Symmetry: Culture and Science*, 2012, vol. 23, № 3-4, p. 275-301. http://petoukhov.com/PETOUKHOV_GENETIC_MATRICES_COMPLEMENTARITIES.pdf.

Petoukhov S.V. The system-resonance approach in modeling genetic structures. *Biosystems*, v. 139, p. 1-11 (January 2016).

Petoukhov S.V. Genetic coding and united-hypercomplex systems in the models of algebraic biology. *Biosystems*, v. 158, August 2017, pp. 31–46 (2017).

Petoukhov S.V. The Genetic Coding System and Unitary Matrices. Preprints 2018, 2018040131, doi: 10.20944/preprints201804.0131.v2 (2018a). http://www.preprints.org/manuscript/201804.0131/v2

Petoukhov S.V. The rules of long DNA-sequences and tetra-groups of oligonucleotides. arXiv:1709.04943v5, 5th version from 8 October 2018, 159

pages (2018b).

Petoukhov S.V. Nucleotide Epi-Chains and New Nucleotide Probability Rules in Long DNA Sequences. *Preprints* **2019**, 2019040011 (doi: 10.20944/preprints201904.0011.v2), (2019a). https://www.preprints.org/manuscript/201904.0011/v2.

Petoukhov S.V. Connections Between Long Genetic and Literary Texts. The Quantum-Algorithmic Modelling. In: Hu Z., Petoukhov S., Dychka I., He M. (eds) *Advances in Computer Science for Engineering and Education II*. pp 534-543, ICCSEEA 2019. Advances in Intelligent Systems and Computing, v. 938. Springer, Cham. Online publication on 29 March 2019. (2019b). https://link.springer.com/chapter/10.1007/978-3-030- 16621-2_50#citeas .

Petoukhov S.V. Hyperbolic Numbers in Modeling Genetic Phenomena. *Preprints* 2019, 2019080284, 36 p. (2020a), doi: 10.20944/preprints201908.0284.v4. https://www.preprints.org/manuscript/201908.0284/v4

Petoukhov S.V. The Genetic Code, Algebraic Codes and Double Numbers. *Preprints* **2019**, 2019110301 (2020b) (doi:10.20944/preprints201911.0301.v2). https://www.preprints.org/manuscript/201911.0301/v2.

Petoukhov S.V. Hyperbolic Numbers, Genetics and Musicology. 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, p. 195-207. Springer, Cham (2020c). DOI https://doi.org/10.1007/978-3-030-39162-1_18

Petoukhov S.V. Genomes symmetries and algebraic harmony in living bodies. – *Symmetry: Culture and Science*, Vol. 31, No. 2, p. 222-223 (2020d) 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., He M. *Symmetrical Analysis Techniques for Genetic Systems and Bioinformatics: Advanced Patterns and Applications*. IGI Global, USA (2010). http://petoukhov.com/Petoukhov,%20He%20-%202010%20-%20Symmetrical%20Analysis%20Techniques%20for%20Genetic%20Systems%20and%20Bioinformatics.pdf.

Petoukhov S.V., Petukhova E.S. Symmetries in genetic systems and the concept of geno-logical coding. - *Information*, 2017, *8*(1), 2 (2017a). doi:10.3390/info8010002, http://www.mdpi.com/2078-2489/8/1/2/htm

Petoukhov S.V., Petukhova E.S. Resonances and the quest for transdisciplinarity. *Information Studies and the Quest for Transdisciplinarity*. Editors M. Burgin, W.Hofkirchner, World Scientific, p. p. 467-487 (2017b).

Petoukhov S.V., Petukhova E.S., Svirin V.I. Symmetries of DNA alphabets and quantum informational formalisms. *Symmetry: Culture and Science*, v. 30, No. 2, pp. 161-179 (2019), https://doi.org/10.26830/symmetry_2019_2_161, http://petoukhov.com/PETOUKHOV%20GENETIC%20QUANTUM%20INFORMATIONAL%20MODEL%202019.pdf

Petoukhov S.V., Tolokonnikov S.V. Algebraic biology and matrix genetics systems. – 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 .

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

Rapoport A.E., Trifonov E.N. Compensatory nature of Chargaff's second parity rule. *Journal of Biomolecular Structure and Dynamics*, November, 1-13 (2012), DOI:10.1080/07391102.2012.736757

Rosandic M., Vlahovic I., Gluncic M., and Paar V. Trinucleotide's quadruplet symmetries and natural symmetry law of DNA creation ensuing Chargaff's second parity rule. *Journal of Biomolecular Structure and Dynamics*, 34:7, pp.1383-1394, (2016), DOI: 10.1080/07391102.2015.1080628

Shporer S., Chor B., Rosset S. , and Horn D. Inversion symmetry of DNA k-mer counts: validity and deviations. *BMC Genomics*, 17(1): 696 (2016).

Walter M., Gross D., Eisert J. Multi-partite entanglement. (2017). arXiv:1612.02437

Yamagishi M.E.B. *Mathematical Grammar of Biology*. Switzerland, Springer International Publishing AG (2017). ISBN 978-3-319-62689-5.