# Algebraic rules for the percentage composition of oligomers in genomes 

Sergey V. Petoukhov ${ }^{1}$<br>${ }^{1}$-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 data of this article were presented by the author at the following conferences: the Fourth International Conference of Artificial Intelligence, Medical Engineering, Education (03-04 October 2020, Moscow, Russia, http://www.icics.net/conf/2020/AIMEE2020/index.html) and the Second International Symposium on Computer Science, Digital Economy and Intelligent Systems (18-20 December 2020, Moscow, Russia, http://www.icics.net/conf/2020/CSDEIS2020/index.html).


#### Abstract

The article presents the author's results of studying hidden rules of structural organizations of long DNA sequences in eukaryotic and prokaryotic genomes. The results concern some rules of percentages (or probabilities) of n-plets in genomes. To reveal such rules, the author uses a tensor family of matrix representations of interrelated DNA-alphabets of 4 nucleotides, 16 doublets, 64 triplets, and 256 tetraplets. If percentages of each of these n-plets in tested genomic DNA-texts are disposed into appropriate cells of appropriate matrices, unexpected rules of invariance of total sums of their percentages in certain tetra-groupings of nplets are revealed. The author connects the received results about these genomic percentages rules with a supposition of P. Jordan, who is one of the creators of quantum mechanics and quantum biology, that life's missing laws are the rules of chance and probability of the quantum world. Algebraic features of the genomic matrices of percentages of n-plets are analyzed and discussed. The received results can be used for further development of quantum biology.


Keywords: DNA sequences, genomes, percentages of n-plets, binary-oppositions, tensor family of matrices, tetra-groupings, quantum biology.

## 1. Introduction.

The article continues publications [Petoukhov, 2020a-c] of author's results of studying hidden rules of structural organizations of long DNA nucleotide sequences (that is, DNA-texts) in eukaryotic and prokaryotic genomes.

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. Besides this, 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 percentages of n-plets (monoplets, doublets, triplets, etc., that is, oligomers with
lengths n ) in long DNA sequences is important for discovering hidden biological laws and for developing quantum biology. In his previous articles [Petoukhov, 2020a-c], the author described the universal hyperbolic rules of the oligomer cooperative organization of DNA nucleotide sequences in eukaryotic and prokariotic genomes. The formulated rules concerned total amounts of certain classess of n-plets. In this new preprint, the author focuses on searching possible rules of probabilities (or percentages) of n-plets in genomes in line with the mentioned supposition of Jordan about existence of such rules.

This research uses a well-known fact of binary-oppositional features of DNA nucleotides (adenine A, thymine T, cyrosine C, and guanine G ), which allows a constructing a family of square tables for DNA alphabets of 4 nucleotides, 16 doublets, 64 triplets, $\ldots, 4^{n} n$-plets. Each of n-plets occupies its strong individual place in this family of tables, which form a tensor family of square matrices. Any DNA sequence of nucleotides (for example, CAGGTACAT...) can be represented as a sequence of oligomers of a fixed length $n$ (for example, as a sequence of triplets CAG-GTA-CAT-...) and one can calculate the percentage content of each of the nplets in this special representation of the DNA sequence as a chain of n-plets. By placing the calculated percentage of each n-plet into the cell occupied by this n-plet in the appropriate square matrix, we obtain the numerical matrices of probabilities of all n-plets in the given DNA-text.

Analysis of this family of probability matrices for n-plets reveals hidden regularities in the structural organization of the studied genomic DNA-texts. Below these regularities for cases $\mathrm{n}=1,2,3,4$ are described and discussed [Petoukhov, 2020a-c].

## 2. The matrix representation of the DNA alphabets on the basis of binaryoppositional traits of nucleotides.

As it is known, the DNA alphabet of 4 nucleotides A, T, C, and G is endowed with a system of binary-opposition traits or indicators [Fimmel, Petoukhov, 2020; Petoukhov, 2008; Petoukhov, 2008; Petoukhov, He, 2010; Stambuk, 1999]:

1) two of these molecules are purines with two rings (A and G), and the other two are pyrimidines with one ring ( C and T ). In terms of these oppositional indicators, $\mathrm{C}=\mathrm{T}=1, \mathrm{~A}=\mathrm{G}=0$;
2) the two letters are keto molecules ( T and G ), and the other two - amino molecules ( C and A ). In terms of these oppositional indicators, $\mathrm{C}=\mathrm{A}=1, \mathrm{~T}=\mathrm{G}=0$.

In the DNA alphabet of 4 nucleotides, each of the letters $\mathrm{C}, \mathrm{A}, \mathrm{T}$, and G is uniquely determined by its named binary indicators. With this in mind, it is convenient to present sets of 4 DNA nucleotides, their 16 doublets and 64 triplets in the form of square tables, the columns of which are numbered with binary indicators "pyrimidine or purine" ( $\mathrm{C}=\mathrm{T}=1, \mathrm{~A}=\mathrm{G}=0$ ), and the rows are numbered with binary indicators "amino or keto " $(\mathrm{C}=\mathrm{A}=1, \mathrm{~T}=\mathrm{G}=0)$. In such tables, all 4 nucleotides, 16 doublets and 64 triplets of DNA automatically occupy their individual places in the strict order (Fig. 2.1).

|  | 1 | 0 |
| :---: | :---: | :---: |
| 1 | C | A |
| 0 | T | G |


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


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


|  | 1111 | 1110 | 1101 | 1100 | 1011 | 1010 | 1001 | 1000 | 0111 | 0110 | 0101 | 0100 | 0011 | 0010 | 0001 | 0000 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 1111 | CCCC | CCCA | CCAC | CCAA | CACC | CACA | CAAC | CAAA | ACCC | ACCA | ACAC | ACAA | AACC | AACA | AAAC | AAAA |
| 1110 | CCCT | CCCG | CCAT | CCAG | CACT | CACG | CAAT | CAAG | ACCT | ACCG | ACAT | ACAG | AACT | AACG | AAAT | AAAG |
| 1101 | CCTC | CCTA | CCGC | CCGA | CATC | CATA | CAGC | CAGA | ACTC | ACTA | ACGC | ACGA | AATC | AATA | AAGC | AAGA |
| 1100 | CCTT | CCTG | CCGT | CCGG | CATT | CATG | CAGT | CAGG | ACTT | ACTG | ACGT | ACGG | AATT | AATG | AAGT | AAGG |
| 1011 | CTCC | CTCA | CTAC | CTAA | CGCC | CGCA | CGAC | CGAA | ATCC | ATCA | ATAC | ATAA | AGCC | AGCA | AGAC | AGAA |
| 1010 | CTCT | CTCG | CTAT | CTAG | CGCT | CGCG | CGAT | CGAG | ATCT | ATCG | ATAT | ATAG | AGCT | AGCG | AGAT | AGAG |
| 1001 | CTTC | CTTA | CTGC | CTGA | CGTC | CGTA | CGGC | CGGA | ATTC | ATTA | ATGC | ATGA | AGTC | AGTA | AGGC | AGGA |
| 1000 | CTT | CTTG | CTGT | CTG | CGTT | CGTG | CGGT | CGGG | ATTT | ATTG | ATGT | ATGG | AGTT | AGTG | AGGT | AGGG |
| 0111 | TCCC | TCCA | TCAC | TCAA | TACC | TACA | TAAC | TAAA | GCCC | GCCA | GCAC | GCAA | GACC | GACA | GAAC | GAAA |
| 0110 | TCCT | TCCG | TCAT | TCAG | TACT | TACG | TAAT | TAAG | GCCT | GCCG | GCAT | GCAG | GACT | GACG | GAAT | GAAG |
| 0101 | TCTC | TCTA | TCGC | TCGA | TATC | TATA | TAGC | TAGA | GCTC | GCTA | GCGC | GCGA | GATC | GATA | GAGC | GAGA |
| 0100 | TCTT | TCTG | TCGT | TCGG | TATT | TATG | TAGT | TAGG | GCTT | GCTG | GCGT | GCGG | GATT | GATG | GAGT | GAGG |
| 0011 | TTCC | TTCA | TTAC | TTAA | TGCC | TGCA | TGAC | TGAA | GTCC | GTCA | GTAC | GTAA | GGCC | GGCA | GGAC | GGAA |
| 0010 | TTCT | TTCG | TTAT | TTAG | TGCT | TGCG | TGAT | TGAG | GTCT | GTCG | GTAT | GTAG | GGCT | GGCG | GGAT | GGAG |
| 0001 | TTTC | TTTA | TTGC | TTGA | TGTC | TGTA | TGGC | TGGA | GTTC | GTTA | GTGC | GTGA | GGTC | GGTA | GGGC | GGGA |
| 0000 | TTTT | TTTG | TTGT | TTGG | TGTT | TGTG | TGGT | TGGG | GTTT | GTTG | GTGT | GTGG | GGTT | GGTG | GGGT | GGGG |

Fig. 2.1. The square tables of the DNA-alphabets of 4 nucleotides, 16 doublets, 64 triplets, and 256 tetraplets, which are constructed by the method of binary numbering of their rows and columns and which are members of a tensor family of matrices $[\mathrm{C}, \mathrm{A} ; \mathrm{T}, \mathrm{G}]^{(\mathrm{n})}$ under $\mathrm{n}=1,2,3,4$ (see explanations in the text). $\rightarrow$

These four tables are not simple tables but they form a single tensor family of matrices: the second, the third, and the fourth tensor powers of the $(2 * 2)$-matrix [C, A; T, G] automatically give this ( $4 * 4$ )-matrix of 16 doublets, this ( $8 * 8$ )-matrix of 64 triplets, and this (16*16)-matrix of 256 tetraplets (Fig. 1). Using the same method of binary numbering of rows and columns of square matrices of DNA alphabets of n-plets, one can similarly construct square tables of 1024 pentaplets, and so on. These new tables will also be members of the unified tensor family of symbolic matrices $[\mathrm{C}, \mathrm{A} ; \mathrm{T}, \mathrm{G}]^{(\mathrm{n})}$ for values $\mathrm{n}=5,6, \ldots$

The tensor family of matrices $[\mathrm{C}, \mathrm{A} ; \mathrm{T}, \mathrm{G}]^{(\mathrm{n})}$ was first used by the author for a comparative analysis of the percentage of different $n$-plets in the DNA-texts of various genomes. Let us explain our analytical approach using a specific example of the DNA of the first human chromosome, which contains a sequence of about 250 million nucleotides $\mathrm{C}, \mathrm{A}, \mathrm{T}$, and G (initial data on this chromosome were taken in the GenBank: https://www.ncbi.nlm.nih.gov/nuccore/NC_000001.11). 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 nucleotides A, T, C, G, which are 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 resulting values of percentages of separate nucleotides in the analyzed genomic sequences.
At the first step of the author's approach, percents of each of the nucleotides $\mathrm{C}, \mathrm{A}, \mathrm{T}$, and G in this chromosome are calculated: $\% \mathrm{C} \approx 0.2085, \% \mathrm{G} \approx 0.2089$, $\% \mathrm{~A}$ $\approx 0.2910, \% \mathrm{~T} \approx 0.2917$ (here percents are shown in fractions of one, and their values are rounded to the fourth decimal place). These percent values are used to be indicated in appropriate cells of the matrix of nucleotides [C, A; T, G] instead of nucleotide symbols for receiving a numeric matrix of nucleotides percents [Fig. 2]. Here and below, percentages are rounded to the fourth decimal place.

| C | A |
| :---: | :---: |
| T | G |$\rightarrow$| $\% \mathrm{C}$ | $\% \mathrm{~A}$ |
| :--- | :--- |
| $\% \mathrm{~T}$ | $\% \mathrm{G}$ |$=$| 0.2085 | 0.2910 |
| :--- | :--- |
| 0.2918 | 0.2087 |

Fig. 2.2. The transformation of the symbolic matrix of 4 nucleotides into numeric matrix of nucleotides percents in the case of human chromosome №1.

One can note that $\% \mathrm{C} \approx \% \mathrm{G}$ and $\% \mathrm{~A} \approx \% \mathrm{~T}$ in accordance with the second Chargaff's rule [Albrecht-Buehler, 2006; Chargaff, 1971; Prahbu, 1993].

At the second step of the described approach, the DNA-text of the analyzed chromosome is represented as a text of doublets (for example, the text TAACCCTA... is represented as TA-AC-CC-TA-...) and percents of each of 16 doublets are calculated. Then these percents are indicated in appropriate cells of the (4*4)-matrix [C, A; T, G] ${ }^{(2)}$ shown in Fig. 2.1. Fig. 2.3 presents the resulting matrix of percents of 16 doublets.

| \%CC | \%CA | \%AC | \%AA |
| :--- | :--- | :--- | :--- |
| $\% \mathrm{CT}$ | \%CG | \%AT | \%AG |
| \%TC | \%TA | \%GC | \%GA |
| \%TT | \%TG | \%GT | \%GG |
| 0.05409 | 0.07274 | 0.05033 | 0.09504 |
| 0.07134 | 0.01031 | 0.07429 | 0.07137 |
| 0.06008 | 0.06312 | 0.04402 | 0.06008 |
| 0.09568 | 0.07286 | 0.05046 | 0.05419 |

Fig. 2.3. The matrix of percents of the 16 doublets in the DNA-sequence of doublets in the human chromosome №1.

At the third step of the described approach, the DNA-text of the analyzed chromosome is represented as a text of triplets (for example, the text TAACCCTAG... is represented as TAA-CCC-TAG-...) and percents of each of 64 triplets are calculated. Then these percents are indicated in appropriate cells of the $(8 * 8)$-matrix $[\mathrm{C}, \mathrm{A} ; \mathrm{T}, \mathrm{G}]^{(3)}$ shown in Fig. 2.1. The resulting matrix of percents of 64 triplets is presented in Fig. 2.4.

| \%CCC | \%CCA | \%CAC | \%CAA | \%ACC | \%ACA | \%AAC | \%A |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| \%ССТ | \%CCG | \%CAT | \%CAG | \% ACT | \%ACG | \%AAT | \%AAG |
| \%CTC | \%CTA | \%CGC | \%CGA | \%ATC | \%ATA | \%AGC | \%AGA |
| \%CTT | \%CTG | \%CGT | \%CGG | \%ATT | \%ATG | \%AGT | \%AGG |
| \%TCC | \%TCA | \%TAC | \%TAA | \%GCC | \%GCA | \%GAC | \%GAA |
| \%TCT | \%TCG | \%TAT | \%TAG | \%GCT | \%GCG | \%GAT | \%GAG |
| \%TTC | \%TTA | \%TGC | \%TGA | \%GTC | \%GTA | \%GGC | \%GGA |
| \%TTT | \%TTG | \%TGT | \%TGG | \%GTT | \%GTG | \%GGT | \%GGG |


| 0.0138 | 0.0188 | 0.0152 | 0.0186 | 0.0118 | 0.0198 | 0.0145 | 0.0369 |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| 0.0185 | 0.0029 | 0.0179 | 0.0210 | 0.0162 | 0.0025 | 0.0238 | 0.0199 |
| 0.0176 | 0.0127 | 0.0025 | 0.0023 | 0.0132 | 0.0194 | 0.0144 | 0.0224 |
| 0.0201 | 0.0209 | 0.0026 | 0.0029 | 0.0239 | 0.0178 | 0.0161 | 0.0185 |
| 0.0159 | 0.0196 | 0.0110 | 0.0199 | 0.0125 | 0.0146 | 0.0096 | 0.0196 |
| 0.0223 | 0.0023 | 0.0194 | 0.0128 | 0.0144 | 0.0025 | 0.0133 | 0.0176 |
| 0.0197 | 0.0198 | 0.0146 | 0.0195 | 0.0096 | 0.0112 | 0.0126 | 0.0160 |
| 0.0372 | 0.0188 | 0.0199 | 0.0190 | 0.0145 | 0.0153 | 0.0119 | 0.0138 |

Fig. 2.4. The matrix of percents of the 64 triplets in the DNA-sequence of triplets in the human chromosome №1.

At the fourth step of the described approach, the DNA-text of the analyzed chromosome is represented as a text of tetraplers (such as TAAC-CCTA-...) and percents of each of 256 tetraplets are calculated. Then these percents are indicated in appropriate cells of the ( $16^{*} 16$ )-matrix $[\mathrm{C}, \mathrm{A} ; \mathrm{T}, \mathrm{G}]^{(4)}$ shown in Fig. 2.1. The resulting matrix of percents of 256 tetraplets is presented in Fig. 2.5.

| . 0033 | . 0055 | . 0042 | . 0044 | . 0040 | . 0056 | . 0032 | . 0070 | . 0030 | . 0042 | . 0040 | . 0053 | . 0032 | . 0059 | . 0055 | . 0149 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| . 0041 | . 0010 | . 0044 | . 0058 | . 0047 | . 0010 | . 0040 | . 0044 | . 0041 | . 0005 | . 0051 | . 0054 | . 0047 | . 0006 | . 0095 | . 0071 |
| . 0050 | . 0029 | . 0008 | . 000 | . 003 | . 0039 | . 004 | . 0059 | . 003 | . 003 | . 000 | . 0006 | . 0040 | . 0070 | . 0037 | . 0066 |
| . 0048 | . 005 | . 000 | . 000 | . 005 | . 00 | . 00 | 迷 | . 00 | . 00 | . 00 | . 0007 | . 00 | . 00 | . 00 | . 0047 |
| . 0 | . 0 | . 0 | . 0 | . 0 | . 0 | . 0 | . 0 | . 0 | . 00 | . 00 | . 00 | . 00 | . 00 | . 0039 | . 0078 |
| . 0058 | . 0009 | . 0035 | . 0028 | . 000 | . 0003 | . 0005 | . 0008 | . 0049 | . 0005 | . 0064 | . 0035 | . 0046 | . 000 | . 0049 | . 0059 |
| . 00 | . 0036 | . 004 | . 005 | . 00 | . 0004 | . 000 | . 00 | . 004 | . 005 | . 003 | . 0050 | . 0031 | . 0037 | . 0047 | . 0057 |
| . 0073 | . 004 | . 005 | . 005 | . 00 | . 0010 | . 000 | . 0010 | . 009 | . 004 | . 005 | . 004 | . 004 | . 00 | . 0040 | . 0041 |
| . 0 | . 0 | . 0 | . 0 | . 0 | . | . 0 | . 0078 | . 0 | . 0041 | . 0 | 8 | . 0023 | . 0037 | . 0029 | . 0073 |
| . 0 | . 0006 | . 0 | . 005 | . 003 | . 0 | . 0 |  | . 004 | . 0008 | . 003 | 0046 | . 0030 | . 00 | . 0046 | 8 |
| . 0057 | . 0040 | . 0005 | . 0006 | . 003 | . 0055 | . 0026 | . 0041 | . 003 | . 0027 | . 0006 | . 0005 | . 0025 | . 0031 | . 0032 | . 0057 |
| . 0068 | . 0058 | . 000 | . 0006 | . 007 | . 0039 | . 0032 | . 0029 | . 003 | . 0048 | . 0006 | . 0008 | . 0041 | . 0036 | . 0036 | . 0051 |
| . 00 | . 006 | . 00 | . 0063 | . 0 | . 0049 | . 0 | . 0061 | . 0023 | . 0 | . 0017 | . 0035 | . 0031 | . 0042 | . 0023 | 052 |
| . 0077 | . 0006 | . 0063 | . 0038 | . 0049 | . 0006 | . 0045 | . 0057 | . 0038 | . 0004 | . 0032 | . 0027 | . 0042 | . 0010 | . 0034 | . 0052 |
| . 0073 | . 0078 | . 0038 | . 0051 | . 0037 | . 0046 | . 0042 | . 0051 | . 0030 | . 0028 | . 0029 | . 0042 | . 0023 | . 0024 | . 0030 | . 0046 |
| . 0150 | . 0072 | . 005 | . 0045 | . 0059 | . 0057 | . 0043 | . 0054 | . 0055 | . 0032 | . 0040 | . 0042 | . 0032 | . 0039 | . 0030 | . 0033 |

Fig. 2.5. The matrix of percents of the 256 tetraplets in the DNA-sequence of tetraplets in the human chromosome №1.

## 3. Tetra-groupings of the percentage composition of $\mathbf{n}$-plets.

At first glance, the set of percent in the resulting matrices (Figs. 2.3-2.5) is quite chaotic. It has the following features regarding the percent of separate $n$-plets:

- Percent of presented n-plets significantly depend on the order of letters in them. For example, the percent of doublets CG and GC, having the same letter composition, differ several times: $\% \mathrm{CG}=0.0103$, and $\% \mathrm{GC}=0.0440$. Similarly, the percent of triplets of the same letter composition CAT, CTA, ACT, ATC, TCA, TAC are significantly different: \%CAT=0.0179, \%СТА $=0.0127, \quad \% \mathrm{ACT}=0.0162, \quad \% \mathrm{ATC}=0.0132, \quad \% \mathrm{TCA}=0.0196$, $\% \mathrm{TAC}=0.0110$, and so on;
- Accordingly, the numerical percent matrices for doublets, triplets, and tetraplets (Figs. 2.3-2.5) are not tensor powers of the nucleotide percent matrix (Fig. 2.2). The percentages of doublets, triplets and tetraplets in the
chromosome are not equal at all to the product of the percentages of separate nucleotides in their composition: $\% \mathrm{C}=0.2085, \% \mathrm{G}=0.2087$, $\% \mathrm{~A}=0.2910$, $\% \mathrm{~T}=0.2917$.
But unexpectedly these values $\% \mathrm{C}=0.2085, \% \mathrm{G}=0.2087$, $\% \mathrm{~A}=0.2910$, $\% \mathrm{~T}=0.2917$ showed themselves in the block organization of percentages as one can calculate from data of percent matrices in Figs. 2.3-2.5:
- The total sum $\Sigma \% \mathrm{CN}$ of percentages of all 4 doublets CN (hereinafter, the symbol N denotes any of the nucleotides $\mathrm{A}, \mathrm{T}, \mathrm{C}$, and G ), which start with the nucleotide C , is equal to $\% \mathrm{C}$, that is, $\Sigma \% \mathrm{CN} \approx \% \mathrm{CC}+\% \mathrm{CA}+\% \mathrm{CT}+\% \mathrm{CG} \approx$ $0.0541+0.0727+0.0713+0.0103 \approx 0.2085 \approx \% \mathrm{C}$;
- The total sum $\Sigma \% N C$ of percentages of all 4 doublets NC, which have the nucleotide C at their second positions, is practically equal to $\% \mathrm{C}$, that is, $\Sigma \% \mathrm{NC} \approx \% \mathrm{CC}+\% \mathrm{AC}+\% \mathrm{TC}+\% \mathrm{GC}=0.0541+0.0503+0.0601+0.0440 \approx$ $0.2085 \approx \% \mathrm{C}$ as well;
- The total sum $\Sigma \%$ CNN of percentages of all 16 triplets CNN, which have the nucleotide C at their first position, is practically equal to $\% \mathrm{C}$, that is, $\Sigma \% \mathrm{CNN}$ $\approx 0.0284 \approx \% \mathrm{C}$ as well;
- The total sum $\Sigma \%$ NCN of percentages of all 16 triplets NCN, which have the nucleotide C at their second position, is practically equal to $\% \mathrm{C}$, that is, $\Sigma \% \mathrm{NCN} \approx 0.0285 \approx \% \mathrm{C}$ as well;
- The total sum $\Sigma \%$ NNC of percentages of all 16 triplets NNC, which have the nucleotide C at their third position, is practically equal to $\% \mathrm{C}$, that is, $\Sigma \% \mathrm{NNC}$ $\approx 0.0285 \approx \% \mathrm{C}$ as well;
- The total sum $\Sigma \%$ CNNN of percentages of all 64 tetraplets CNNN, which have the nucleotide C at their first position, is practically equal to $\% \mathrm{C}$, that is, $\Sigma \%$ CNNN $\approx 0.0285 \approx \% \mathrm{C}$ as well;
- The total sum $\Sigma \%$ NCNN of percentages of all 64 tetraplets NCNN, which have the nucleotide C at their second position, is practically equal to $\% \mathrm{C}$, that is, $\Sigma \%$ NCNN $\approx 0.0285 \approx \% \mathrm{C}$ as well;
- The total sum $\Sigma \%$ NNCN of percentages of all 64 tetraplets NNCN, which have the nucleotide C at their third position, is practically equal to $\% \mathrm{C}$, that is, $\Sigma \%$ NNCN $\approx 0.0285 \approx \%$ C as well;
- The total sum $\Sigma \%$ NNNC of percentages of all 64 tetraplets NNNC, which have the nucleotide C at their fourth position, is practically equal to $\% \mathrm{C}$, that is, $\Sigma \% \mathrm{NNCN} \approx 0.0285 \approx \% \mathrm{C}$ as well.
Similar equalities turn out to be valid also for the total sums $\Sigma$ of the considered n-plets with nucleotides A, T, G at the analogical positions, as shown in Fig. 3.1.

| $\% \mathrm{C} \approx 0.2085$ | $\% \mathrm{G} \approx 0.2087$ | $\% \mathrm{~A} \approx 0.2910$ | $\% \mathrm{~T} \approx 0.2918$ |
| ---: | ---: | ---: | ---: |
| $\Sigma \% \mathrm{CN} \approx 0.2085$ | $\Sigma \% \mathrm{GN} \approx 0.2088$ | $\Sigma \% \mathrm{AN} \approx 0.2910$ | $\Sigma \% \mathrm{TN} \approx 0.2917$ |
| $\Sigma \% \mathrm{NC} \approx 0.2085$ | $\Sigma \% \mathrm{NG} \approx 0.2087$ | $\Sigma \% \mathrm{NA} \approx 0.2910$ | $\Sigma \% \mathrm{NT} \approx 0.2918$ |
| $\Sigma \% \mathrm{CNN} \approx 0.2084$ | $\Sigma \% \mathrm{GNN} \approx 0.2088$ | $\Sigma \% \mathrm{ANN} \approx 0.2910$ | $\Sigma \% \mathrm{TNN} \approx 0.2917$ |
| $\Sigma \% \mathrm{NCN} \approx 0.2085$ | $\Sigma \% \mathrm{NGN} \approx 0.2088$ | $\Sigma \% \mathrm{NAN} \approx 0.2910$ | $\Sigma \% \mathrm{NTN} \approx 0.2917$ |
| $\Sigma \% \mathrm{NNC} \approx 0.2085$ | $\Sigma \% \mathrm{NNG} \approx 0.2087$ | $\Sigma \% \mathrm{NNA} \approx 0.2910$ | $\Sigma \% \mathrm{NNT} \approx 0.2918$ |
| $\Sigma \% \mathrm{CNNN} \approx 0.2085$ | $\Sigma \% \mathrm{GNNN} \approx 0.2088$ | $\Sigma \% \mathrm{ANNN} \approx 0.2910$ | $\Sigma \% \mathrm{TNNN} \approx 0.2917$ |
| $\Sigma \% \mathrm{NCNN} \approx 0.2085$ | $\Sigma \% \mathrm{NGNN} \approx 0.2087$ | $\Sigma \% \mathrm{NANN} \approx 0.2910$ | $\Sigma \% \mathrm{NTNN} \approx 0.2918$ |
| $\Sigma \% \mathrm{NNCN} \approx 0.2085$ | $\Sigma \% \mathrm{NNGN} \approx 0.2088$ | $\Sigma \% \mathrm{NNAN} \approx 0.2910$ | $\Sigma \% \mathrm{NNTN} \approx 0.2918$ |
| $\Sigma \% \mathrm{NNNC} \approx 0.2085$ | $\Sigma \% \mathrm{NNNG} \approx 0.2087$ | $\Sigma \% \mathrm{NNNA} \approx 0.2910$ | $\Sigma \% \mathrm{NNNT} \approx 0.2918$ |

Fig. 3.1. Percentages of nucleotides C, G, A, T, and the sum $\Sigma$ of percent of n-plets with these nucleotides at the certain positions for the case of DNA of the first human chromosome. The symbol N denotes any of the nucleotides.

Briefly speaking, the following relations (1) hold true - with high level of accuracy - regarding percentages of the nucleotides $\mathrm{C}, \mathrm{G}, \mathrm{A}$, and T , and the considered n-plets in the human chromosome №1:

$$
\begin{align*}
& \% \mathrm{C} \approx \% \mathrm{CN} \approx \Sigma \% \mathrm{NC} \approx \Sigma \% \mathrm{CNN} \approx \Sigma \% \mathrm{NCN} \approx \Sigma \% \mathrm{NNC} \approx \\
& \Sigma \% \mathrm{CNNN} \approx \Sigma \% \mathrm{NCNN} \approx \Sigma \% \mathrm{NNCN} \approx \Sigma \% \mathrm{NNNC} ; \\
& \% \mathrm{G} \approx \Sigma \% \mathrm{GN} \approx \Sigma \% \mathrm{NG} \approx \Sigma \% \mathrm{GNN} \approx \Sigma \% \mathrm{NGN} \approx \Sigma \% \mathrm{NNG} \approx \\
& \Sigma \% \mathrm{GNNN} \approx \Sigma \% \mathrm{NGNN} \approx \Sigma \% \mathrm{NNGN} \approx \Sigma \% \mathrm{NNNG} ; \\
& \% \mathrm{~A} \approx \Sigma \% \mathrm{AN} \approx \Sigma \% \mathrm{NA} \approx \Sigma \% \mathrm{ANN} \approx \Sigma \% \mathrm{NAN} \approx \Sigma \% \mathrm{NNA} \approx \\
& \Sigma \% \mathrm{ANNN} \approx \Sigma \% \mathrm{NANN} \approx \Sigma \% \mathrm{NNAN} \approx \Sigma \% \mathrm{NNNA} ; \\
& \% \mathrm{~T} \approx \Sigma \% \mathrm{TN} \approx \Sigma \% \mathrm{NT} \approx \Sigma \% \mathrm{TNN} \approx \Sigma \% \mathrm{NTN} \approx \Sigma \% \mathrm{NNT} \approx \\
& \Sigma \% \mathrm{TNNN} \approx \Sigma \% \mathrm{NTNN} \approx \Sigma \% \mathrm{NNTN} \approx \Sigma \% \mathrm{NNNT} \tag{1}
\end{align*}
$$

These equalities (1) can also be written in the form (2) of the equality of 4-dimensional vectors of genomic percentages, the coordinates of which are the percentages of individual nucleotides or the considered $n$-plets with nucleotides at certain positions:

```
\([\% \mathrm{~A}, \% \mathrm{~T}, \% \mathrm{C}, \% \mathrm{G}] \approx[\Sigma \% \mathrm{AN}, \Sigma \% \mathrm{TN}, \Sigma \% \mathrm{CN}, \Sigma \% \mathrm{GN}] \approx\)
[ \(\Sigma \% \mathrm{NA}, \Sigma \% \mathrm{NT}, \Sigma \% \mathrm{NC}, \Sigma \% \mathrm{NG}] \approx\)
[ \(\Sigma \% \mathrm{ANN}, \Sigma \% \mathrm{TNN}, \Sigma \% \mathrm{CNN}, \Sigma \% \mathrm{GNN}] \approx\)
[ \(\Sigma \%\) NAN, \(\Sigma \%\) NTN, \(\Sigma \%\) NCN, \(\Sigma \%\) NGN \(] \approx\)
[ \(\Sigma \% \mathrm{ANN}, \Sigma \% \mathrm{TNN}, \Sigma \% \mathrm{CNN}, \Sigma \% \mathrm{GNN}] \approx\)
[ \(\Sigma \%\) ANNN, \(\Sigma \%\) TNNN, \(\Sigma \%\) CNNN, \(\Sigma \%\) GNNN] \(\approx\)
[ \(\Sigma \%\) NANN, \(\Sigma \%\) NTNN, \(\Sigma \%\) NCNN, \(\Sigma \% N G N N] \approx\)
[ \(\Sigma \%\) NNAN, \(\Sigma \%\) NNTN, \(\Sigma \%\) NNCN, \(\Sigma \%\) NNGN] \(\approx\)
[ \(\Sigma \%\) NNNA, \(\Sigma \%\) NNNT, \(\Sigma \%\) NNNC, \(\Sigma \%\) NNNG].
```

Knowing the percentages of nucleotides $\% \mathrm{~A}, \% \mathrm{~T}, \% \mathrm{C}$, and $\% \mathrm{G}$, it is possible to predict with high accuracy the sums of percentages of $n$-plets of the noted classes. The ability of such predictions on the basis of equalities (1) or (2) holds not only for the considered human chromosome №1 but also for many eukaryotic and prokaryotic genomes, which were analyzed by the author till now. One should note that percentages of nucleotides $\% \mathrm{~A}, \% \mathrm{~T}, \% \mathrm{C}$, and $\% \mathrm{G}$ can be essentially different in various genomes. (Appendix I contains one of many possible examples of percent matrices related to the genome of bacteria Bradyrhizobium japonicum where $\% \mathrm{~A} \approx 0.1819, \% \mathrm{~T} \approx 0.1815, \% \mathrm{C} \approx 0.3184$, and $\% \mathrm{G} \approx 0.3182$ in contrast to the considered case of the human chromosome \#1). This indicates a universal cooperative organization of n-plets in genomic DNA-texts, which is reflected in very special block-mosaic structures of the percent matrices of n-plets (Fig. 2.3-2.5).

The four columns in Fig. 3.1 show that in all considered cases of percentages of four kinds of nucleotides and percent sums of $n$-plets, which have these nucleotides as their positional attributes, there exist regular tetra-subsets of percentages (or percentage tetra-groupings).

It can be noted that the sum of the percentages (that is, the probabilities) of $n$ plets in each of these four subsets can be interpreted as the square of the length of a vector whose components are equal to the square roots of the probabilities of the corresponding n-plets (these square roots can be considered as amplitudes of the probabilities of $n$-plets). For example, the sum $\% \mathrm{CC}+\% \mathrm{CG}+\% \mathrm{CA}+\% \mathrm{CT}$ is the square of the length of the 4 -dimensional vector $\mathrm{V}_{\mathrm{CN}}=[\sqrt{\% \mathrm{CC}}, \sqrt{\% \mathrm{CG}}, \sqrt{\% \mathrm{CA}}, \sqrt{\% \mathrm{CT}}]$. From this point of view, equalities (2) mean the constancy of the length of the corresponding $2^{n}$-dimensional vectors, whose coordinates are the amplitudes of the probabilities of the corresponding n-plets. This metric approach allows for developing new methods of comparative vector analysis in genetics, which are now being studied in our laboratory.

## 4. DNA epi-chains and the tetra-grouping matrices of percentages of n-plets.

This Section presents some results of the analogical study of percentages of n-plets in special subsequences of long nucleotide sequences in single-stranded DNA. These subsequences are termed «DNA epi-chains» [Petoukhov, 2019, 2020a-c]. The author's initial results testify that the above described equalities (1) and (2) of total sums of percentages of n -plets hold for these epi-chains as well.

By definition, in a nucleotide sequence $\mathrm{N}_{1}$ of any DNA strand $\mathrm{N}_{1}$ (Fig. 4.1a) with sequentially numbered nucleotides 1, 2, 3, 4, ... (Fig. 4.1a), epi-chains of different orders $k$ are such subsequences that contain only those nucleotides, whose numeration differ from each other by natural number $k=1,2,3,4, \ldots$. For example, in any single-stranded DNA, one can consider its epi-chain of the second order $\mathrm{N}_{2}$, in which its nucleotide sequence numbers differ by $k=2$ : an epi-chain $\mathrm{N}_{2}$ contains nucleotides with numerations 1, 3, 5, ... (Fig. 4.1b). By analogy, an epi-chain of the third order $\mathrm{N}_{3}$ is connected with $\mathrm{k}=3$ and contains a subsequence of nucleotides with numerations 1, 4, 7, 10, ... (Fig. 4.1d).


Fig. 4.1. Schematic representations of a single-stranded DNA and its initial epi-chains of numerated nucleotides, denoted by black circles. a, a sequence $N_{1}$ of numerated nucleotides of the DNA strand. $\mathbf{b}$, an epi-chain of the second order $\mathrm{N}_{2}$ having nucleotides with numbers 1-3-5-7-... . c, an epi-chain of the third order $\mathrm{N}_{3}$ nucleotides numbers 1-4-7-10-... .

Each genomic DNA epi-chain of $k$-th order (if $k=2,3,4, \ldots$. ) 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 text differences, matrices of percentages of n-plets in these very new nucleotide sequences contain analogical tetra-groupings with the same equalities (1) and (2) for total sums of corresponding n-plets (at this stage of the research, the author studied the percent matrices of epi-chains only in cases of epi-chains with relatively small orders $k$ ).

To illustrate this result, let's consider the percent matrices for the second, third and fourth order epi-chains ( $k=2 \mathrm{~b} 3 \mathrm{~b} 4$ ) in the DNA of the first human chromosome (Fig. 4.2).

| $\% \mathrm{CC}$ | $\% \mathrm{CA}$ | $\% \mathrm{AC}$ | $\% \mathrm{AA}$ |
| :--- | :--- | :--- | :--- |
| $\% \mathrm{CT}$ | $\% \mathrm{CG}$ | $\% \mathrm{AT}$ | $\% \mathrm{AG}$ |
| $\% \mathrm{TC}$ | $\% \mathrm{TA}$ | $\% \mathrm{GC}$ | $\% \mathrm{GA}$ |
| $\% \mathrm{TT}$ | $\% \mathrm{TG}$ | $\% \mathrm{GT}$ | $\% \mathrm{GG}$ |


| 0.05409 | 0.07274 | 0.05033 | 0.09504 |
| :--- | :--- | :--- | :--- |
| 0.07134 | 0.01031 | 0.07429 | 0.07137 |
| 0.06008 | 0.06312 | 0.04402 | 0.06008 |
| 0.09568 | 0.07286 | 0.05046 | 0.05419 |


| Epi-chain of the 2nd order |  |  |  | Epi-chain of the 3rd order |  |  |  | Epi-chain of the 4th order |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 0.0492 | 0.0524 | 0.0538 | 0.0985 | 0.0487 | 0.0562 | 0.0579 | 0.0912 | 0.0487 | 0.0558 | 0.0574 | 0.0940 |
| 0.0591 | 0.0478 | 0.0800 | 0.0588 | 0.0571 | 0.0464 | 0.0849 | 0.0570 | 0.0566 | 0.0475 | 0.0831 | 0.0565 |
| 0.0612 | 0.0788 | 0.0443 | 0.0613 | 0.0608 | 0.0826 | 0.0411 | 0.061 | 0.0579 | 0.0835 | 0.0445 | 0.0579 |
| 0.0988 | 0.0530 | 0.0539 | 0.0492 | 0.0918 | 0.0565 | 0.0578 | 0.0489 | 0.0943 | 0.0561 | 0.0577 | 0.0487 |

Fig. 4.2. Matrices of percentages of 16 doublets in the single-stranded DNA of the human chromosome \#1. Upper row: percentages matrices of the DNA $(k=1)$. Bottom row: percent matrices for the DNA epi-chains of the second, third, and fourth orders ( $k=2,3,4$ ).

One can see from data in Fig. 4.2 that different percentages matrices contain essential different percentages in corresponding separate cells. For example, $\% \mathrm{CG}=$ 0.0103 in the complete single-stranded DNA; $\% \mathrm{CG}=0.0478$ in the epi-chain of the $2^{\text {nd }}$ order; $\% \mathrm{CG}=0.0464$ in the epi-chain of the $3^{\text {rd }}$ order; $\% \mathrm{CG}=0.0475$ in the epi-chain of the $4^{\text {th }}$ order. But in each of these four percentages matrices the equalities (2) are realized with high level of accuracy:

$$
\begin{align*}
& {[\% \mathrm{~A}, \% \mathrm{~T}, \% \mathrm{C}, \% \mathrm{G}] \approx[\Sigma \% \mathrm{AN}, \Sigma \% \mathrm{TN}, \Sigma \% \mathrm{CN}, \Sigma \% \mathrm{GN}] \approx} \\
& {[\Sigma \% \mathrm{NA}, \Sigma \% \mathrm{NT}, \Sigma \% \mathrm{NC}, \Sigma \% \mathrm{NG}] \approx[0.2910,0.2918,0.2085,0.2087]} \tag{3}
\end{align*}
$$

It illustrates that the total sums of the percentages in each of the tetragroupings of n-plets practically do not depend on the percent values of individual nplets in these sums. This resembles the phenomenon of perceiving a musical melody, which can be reproduced in different frequency ranges of octaves, that is, under significantly changing the frequency of the sound of each of its note elements, but despite these changes, the melody remains generally recognizable. Many such phenomena of perception, in which there is relative independence of the integral form from its constituent individual components, are studied in Gestalt psychology. The described universal regularities in the preservation of total percentages in tetragroupings of $n$-plets relatively regardless of the percentage of individual $n$-plets in genomic DNA allow the author to develop gestalt genetics. Gestalt genetics is
interrelated in some degree with Gestalt psychology, which studies the corresponding genetically inherited properties of our brain regarding the perception of the environment.

## 5. Percentages matrices of n-plets and algebras of hyperbolic numbers

This section shows that each of the vectors of percentages sums of n-plets in the equalities (2) corresponds to such mosaical matrix, which is connected with wellknown algebras of $2^{\text {n }}$-dimensional hyperbolic numbers (hyperbolic numbers are termed also as double numbers, Lorentz numbers, split-complex numbers and perplex numbers). Consequently the arrangements of the described tetra-groupings of $n$-plets inside these $\left(2^{n} * 2^{\mathrm{n}}\right)$-matrices of percentages obey special algebraic rules. It is additionally interesting since, as it is known, structures of some genetically inherited biological phenomena are related to $2^{\text {n }}$-dimensional hyperbolic numbers [Petoukhov, 2020b]. 2-dimensional hyperbolic numbers form algebra over the field of real numbers [Harkin, Harkin, 2004; Kantor, Solodovnikov, 1989].

To demonstrate these connections, we will take initially from equalities (2) the vector of percentage sums of 16 doublets [ $\Sigma \% \mathrm{AN}, \Sigma \% \mathrm{TN}, \Sigma \% \mathrm{CN}, \Sigma \% \mathrm{GN}$ ]; its coordinates correspond to summary percentages of tetra-groupings of doublets, which start with one of nucleotides C, G, A, or T. Let us denone matrix cells, which are occupied by 8 doublets CN and GN, by number +1 , and matrix cells, which are occupied by 8 doublets AN and AT, by number -1 . (These two groupings of doublets occupy appropriate quadrants along diagonals of the matrix as it it shown in Figs. 2.1 and 2.3). Fig. 5.1 gives a result of these denotations, which transform a symbolic matrix of doublets into numeric matrix $\mathrm{M}_{2}$ with entries +1 and -1 .

| $\% \mathrm{CC}$ | \%CA | \%AC | \%AA |
| :--- | :--- | :--- | :--- |
| $\% \mathrm{CT}$ | $\% \mathrm{CG}$ | $\% \mathrm{AT}$ | $\% \mathrm{AG}$ |
| $\% \mathrm{TC}$ | $\% \mathrm{TA}$ | $\% \mathrm{GC}$ | $\% \mathrm{GA}$ |
| $\% \mathrm{TT}$ | $\% \mathrm{TG}$ | $\% \mathrm{GT}$ | $\% \mathrm{GG}$ |$\quad \rightarrow \quad \mathbf{M}_{\mathbf{2 1}}=$| +1 | +1 | -1 | -1 |
| :---: | :---: | :---: | :---: |
| +1 | +1 | -1 | -1 |
| -1 | -1 | +1 | +1 |
| -1 | -1 | +1 | +1 |

Fig. 5.1. Transformation of the symbolic (4*4)-matrix of 16 doublets into the numeric matrix $\mathrm{M}_{21}$ which shows the arrangements of cells, occupied by doublets CN and GN (black cells) and also by AN and TN (white cells).

The recevied matrix can be decomposed into sum of four sparse matrices $\mathrm{e}_{0}$, $e_{1}, e_{2}, e_{3}: M_{21}=e_{0}+e_{1}+e_{2}+e_{3}$ (Fig. 5.2). The set of sparse matrices turns out to be closed under the multiplication operation: the product of any two matrices of this set by each other gives a matrix of the same set. This defines the multiplication table of these matrices (Fig. 5.2), which is known as the multiplication table of the basis units of algebra of 4-dimensional hyperbolic numbers (or algebra of 4-dimensional hyperbolic matrions [Petoukhov, 2008; Petoukhov, He, 2010]). It should be noted that the decomposition of the matrix $\mathrm{M}_{21}$ was done on the basis of the known method of the dyadic-shift decompositions [Petoukhov, 2008; Petoukhov, He, 2010], which is used in this section below for other matrices as well.

$$
\begin{gathered}
\mathrm{M}_{21}=\mathrm{e}_{0}+\mathrm{e}_{1} \\
+\mathrm{e}_{2}+\mathrm{e}_{3}= \\
\hline
\end{gathered}
$$



Fig. 5.2. The decomposition of the matrix $M_{21}=e_{0}+e_{1}+e_{2}+e_{3}$ from Fig. 5.1 and the multiplication table (at right) of these sparse matrices $\mathrm{e}_{0}, \mathrm{e}_{1}, \mathrm{e}_{2}$, and $\mathrm{e}_{3}$. The bold box in the multiplication table refers the subalgebra of 2-dimensional hyperbolic numbers.

By analogy, one can take from the equalities (2) another vector [ $\Sigma \% \mathrm{NA}, \Sigma \% \mathrm{NT}, \Sigma \% \mathrm{NC}, \Sigma \% \mathrm{NG}]$; its coordinates correspond to summary percentages of those tetra-groupings of 16 doublets, which have nucleotides C, G, A, and T at their second positions and occupy a new set of cells in the percentages matrix of doublets. Let us denote matrix cells, which are occupied by 8 doublets NC and NG, by number +1 , and cells, which are occupied by of 8 doublets NA and NT, by number -1 . Fig. 5.3 shows a numeric matrix $\mathrm{M}_{22}$, which appears in the result and which has a mosaic resembling a chessboard.

| $\% \mathrm{CC}$ | \%CA | \%AC | \%AA |
| :--- | :--- | :--- | :--- |
| $\% \mathrm{CT}$ | \%CG | \%AT | $\% \mathrm{AG}$ |
| $\% \mathrm{TC}$ | $\% \mathrm{TA}$ | $\% \mathrm{GC}$ | $\% \mathrm{GA}$ |
| $\% \mathrm{TT}$ | $\% \mathrm{TG}$ | $\% \mathrm{GT}$ | $\% \mathrm{GG}$ |$\quad \rightarrow \quad \mathbf{M}_{22}=$| +1 | -1 | +1 | -1 |
| :--- | :--- | :--- | :--- |
| -1 | +1 | -1 | +1 |
| +1 | -1 | +1 | -1 |
| -1 | +1 | -1 | +1 |

Fig. 5.3. Transformation of the symbolic matrix of doublets into the numeric matrix $\mathrm{M}_{22}$, which shows the arrangements of cells, occupied by doublets NC and NG (black cells) and also by NA and NT (white cells).

The recevied matrix can be decomposed into sum of four new sparse matrices $e_{0}, e_{1}, e_{2}, e_{3}: M_{22}=e_{0}+e_{1}+e_{2}+e_{3}$ (Fig. 5.4). The set of these sparse matrices turns out to be closed under the multiplication as well: the product of any two matrices of this set by each other gives a matrix of the same set. This defines the multiplication table of these matrices (Fig. 5.4), which coincides with the multiplication table shown in Fig. 5.2. It means that both matrices $\mathrm{M}_{12}$ and $\mathrm{M}_{22}$ are different matrix representations of 4-dimensional hyperbolic number with unit coordinates.

Fig. 5.4. The decomposition of the matrix $M_{22}=e_{0}+e_{1}+e_{2}+e_{3}$ from Fig. 5.3 and the multiplication table (at right) of these sparse matrices $\mathrm{e}_{0}, \mathrm{e}_{1}, \mathrm{e}_{2}$, and $\mathrm{e}_{3}$. The bold box in the multiplication table refers the subalgebra of 2-dimensional hyperbolic numbers.

By analogy, one can take from the equalities (2) another vector [ $\Sigma \%$ ANN, $\Sigma \% \mathrm{TNN}, \Sigma \% \mathrm{CNN}, \Sigma \% \mathrm{GNN}]$; its coordinates correspond to summary percentages of those tetra-groupings of 64 triplets, which start with nucleotides C, G, A, and T. Let us denote marix cells, which are occupied by 32 triplets CNN and GNN, by number +1 , and matrix cells, which are occupied by 32 triplets ANN and TNN, by number -1 . These two groupings of triplets ocupy appropriate quadrants of the matrix as it it shown in Figs. 2.1 and 2.4. Fig. 5.5 shows a result of these denotations, which lead to a numeric matrix $\mathrm{M}_{31}$ with entries +1 and -1 .

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


$\mathrm{M}_{31}=$| +1 | +1 | +1 | +1 | -1 | -1 | -1 | -1 |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| +1 | +1 | +1 | +1 | -1 | -1 | -1 | -1 |
| +1 | +1 | +1 | +1 | -1 | -1 | -1 | -1 |
| +1 | +1 | +1 | +1 | -1 | -1 | -1 | -1 |
| -1 | -1 | -1 | -1 | +1 | +1 | +1 | +1 |
| -1 | -1 | -1 | -1 | +1 | +1 | +1 | +1 |
| -1 | -1 | -1 | -1 | +1 | +1 | +1 | +1 |
| -1 | -1 | -1 | -1 | +1 | +1 | +1 | +1 |

Fig. 5.5. Transformation of the symbolic (8*8)-matrix of 64 triplets into the numeric matrix $\mathrm{M}_{31}$, which shows the arrangements of cells, occupied by triplets CNN and GNN (black cells) and also by triplets ANN and TNN (white cells).

The recevied matrix $\mathrm{M}_{31}$ can be decomposed by the method of the dyadic-shift decompositions into sum of 8 sparse matrices: $M_{31}=e_{0}+e_{1}+e_{2}+e_{3}+e_{4}+e_{5}+e_{6}+$ $e_{7}$ (Fig. 5.6). The set of these 8 sparse matrices turns out to be closed under the multiplication operation and defines the multiplication table of these matrices (Fig. 5.6), which is known as the multiplication table of the basis units of algebra of 8 -dimensional hyperbolic numbers (or 8-dimensional hyperbolic matrions [Petoukhov, 2008; Petoukhov, He, 2010]).
$\mathrm{M}_{31}=\mathrm{e}_{0}+\mathrm{e}_{1}+$ $\mathrm{e}_{2}+\mathrm{e}_{3}+\mathrm{e}_{4}+\mathrm{e}_{5}+$ $\mathrm{e}_{6}+\mathrm{e}_{7}=$

| 10000000 |  | 01000000 |  | 00100000 |
| :---: | :---: | :---: | :---: | :---: |
| 01000000 |  | 10000000 |  | 00010000 |
| 00100000 |  | 00010000 |  | 10000000 |
| 00010000 | + | 00100000 | + | 01000000 |
| 00001000 |  | 00000100 |  | 00000010 |
| 00000100 |  | 00001000 |  | 00000001 |
| 00000010 |  | 00000001 |  | 00001000 |
| 00000001 |  | 00000010 |  | 00000100 |
| 00010000 |  | 0000-1000 |  | 00000-100 |
| 00100000 |  | 00000-100 |  | 0000-1000 |
| 01000000 |  | 000000-10 |  | $0000000-1$ |
| 10000000 | + | $0000000-1$ | + | 000000-10 |
| 00000001 |  | $-10000000$ |  | 0-1000000 |
| 00000010 |  | 0-1000000 |  | $-10000000$ |
| 00000100 |  | 00-100000 |  | 000-10000 |
| 00001000 |  | 000-10000 |  | 00-100000 |
| 000000-10 |  | $0000000-1$ |  |  |
| $0000000-1$ |  | 000000-10 |  |  |
| 0000-1000 |  | 00000-100 |  |  |
| 00000-100 | + | 0000-1000 |  |  |
| 00-100000 |  | 000-10000 |  |  |
| 000-10000 |  | 00-100000 |  |  |
| -10000000 |  | 0-1000000 |  |  |
| 0-1000000 |  | -10000000 |  |  |


| $*$ | $\mathbf{e}_{0}$ | $\mathbf{e}_{1}$ | $\mathbf{e}_{2}$ | $\mathbf{e}_{3}$ | $\mathbf{e}_{4}$ | $\mathbf{e}_{5}$ | $\mathbf{e}_{6}$ | $\mathbf{e}_{7}$ |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mathbf{e}_{0}$ | $\mathrm{e}_{0}$ | $\mathrm{e}_{1}$ | $\mathrm{e}_{2}$ | $\mathrm{e}_{3}$ | $\mathrm{e}_{4}$ | $\mathrm{e}_{5}$ | $\mathrm{e}_{6}$ | $\mathrm{e}_{7}$ |
| $\mathbf{e}_{1}$ | $\mathrm{e}_{1}$ | $\mathrm{e}_{0}$ | $\mathrm{e}_{3}$ | $\mathrm{e}_{2}$ | $\mathrm{e}_{5}$ | $\mathrm{e}_{4}$ | $\mathrm{e}_{7}$ | $\mathrm{e}_{6}$ |
| $\mathbf{e}_{2}$ | $\mathrm{e}_{2}$ | $\mathrm{e}_{3}$ | $\mathrm{e}_{0}$ | $\mathrm{e}_{1}$ | $\mathrm{e}_{6}$ | $\mathrm{e}_{7}$ | $\mathrm{e}_{4}$ | $\mathrm{e}_{5}$ |
| $\mathbf{e}_{3}$ | $\mathrm{e}_{3}$ | $\mathrm{e}_{2}$ | $\mathrm{e}_{1}$ | $\mathrm{e}_{0}$ | $\mathrm{e}_{7}$ | $\mathrm{e}_{6}$ | $\mathrm{e}_{5}$ | $\mathrm{e}_{4}$ |
| $\mathbf{e}_{4}$ | $\mathrm{e}_{4}$ | $\mathrm{e}_{5}$ | $\mathrm{e}_{6}$ | $\mathrm{e}_{7}$ | $\mathrm{e}_{0}$ | $\mathrm{e}_{1}$ | $\mathrm{e}_{2}$ | $\mathrm{e}_{3}$ |
| $\mathbf{e}_{5}$ | $\mathrm{e}_{5}$ | $\mathrm{e}_{4}$ | $\mathrm{e}_{7}$ | $\mathrm{e}_{6}$ | $\mathrm{e}_{1}$ | $\mathrm{e}_{0}$ | $\mathrm{e}_{3}$ | $\mathrm{e}_{2}$ |
| $\mathbf{e}_{6}$ | $\mathrm{e}_{6}$ | $\mathrm{e}_{7}$ | $\mathrm{e}_{4}$ | $\mathrm{e}_{5}$ | $\mathrm{e}_{2}$ | $\mathrm{e}_{3}$ | $\mathrm{e}_{0}$ | $\mathrm{e}_{1}$ |
| $\mathbf{e}_{7}$ | $\mathrm{e}_{7}$ | $\mathrm{e}_{6}$ | $\mathrm{e}_{5}$ | $\mathrm{e}_{4}$ | $\mathrm{e}_{3}$ | $\mathrm{e}_{2}$ | $\mathrm{e}_{1}$ | $\mathrm{e}_{0}$ |

Fig. 5.6. The decomposition of the ( $8 * 8$ )-matrix $\mathrm{M}_{32}$ of 64 triplets from Fig. 5.5 and the multiplication table (at bottom) of the sparse matrices $\mathrm{e}_{0}, \mathrm{e}_{1}, \ldots, \mathrm{e}_{7}$, which are shown above. The bold boxes in the multiplication table refer the subalgebras of 2-dimensional and 4-dimensional hyperbolic numbers.

By analogy, one can take from the equalities (2) another vector [ $\Sigma \% \mathrm{NAN}$, $\Sigma \% \mathrm{NTN}, \Sigma \% \mathrm{NCN}, \Sigma \% \mathrm{NGN}]$; its coordinates correspond to summary percentages of those tetra-groupings of 64 triplets, which have nucleotides $\mathrm{C}, \mathrm{G}, \mathrm{A}$, and T at their
second positions and occupy a new set of cells in the percentages matrix of triplets. Let us denote marix cells, which are occupied by 32 triplets NCN and NGN, by number +1 , and matrix cells, which are occupied by 32 triplets NAN and NTN, by number -1 . Fig. 5.7 shows a result of such denotations, which lead to a numeric matrix $\mathrm{M}_{31}$ with entries +1 and -1 .

$M_{32}=$| +1 | +1 | -1 | -1 | +1 | +1 | -1 | -1 |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| +1 | +1 | -1 | -1 | +1 | +1 | -1 | -1 |
| -1 | -1 | +1 | +1 | -1 | -1 | +1 | +1 |
| -1 | -1 | +1 | +1 | -1 | -1 | +1 | +1 |
| +1 | +1 | -1 | -1 | +1 | +1 | -1 | -1 |
| +1 | +1 | -1 | -1 | +1 | +1 | -1 | -1 |
| -1 | -1 | +1 | +1 | -1 | -1 | +1 | +1 |
| -1 | -1 | +1 | +1 | -1 | -1 | +1 | +1 |

Fig. 5.7. Transformation of the symbolic (8*8)-matrix of 64 triplets (Figs. 2.1 and 2.4) into the numeric matrix $\mathrm{M}_{32}$, which shows the arrangements of cells, occupied by 32 triplets NCN and NGN (black cells) and also by 32 triplets NAN and NTN (white cells).

The received numeric matrix $\mathrm{M}_{32}$ can be also decomposed by the same method of the dyadic-shift decompositions into sum of new 8 sparse matrices (as it was made above for matrices $\mathrm{M}_{21}, \mathrm{M}_{22}$, and $\mathrm{M}_{31}$ ); the set of these 8 sparse matrices turns out to be closed in relation to the multiplication again and defines the same multiplication table of these matrices as in the case of the matrix $\mathrm{M}_{31}$ (see Fig. 5.6, at bottom).

One can take from the equalities (2) also another vector [ $\Sigma \% \mathrm{NNA}$, $\Sigma \%$ NNT, $\Sigma \%$ NNC, $\Sigma \% \mathrm{NNG}]$; its coordinates correspond to summary percentages of those tetra-groupings of 64 triplets, which have nucleotides C, G, A, and T at their third positions and occupy a new set of cells in the percentages matrix of triplets. Let us denote marix cells, which are occupied by 32 triplets NNC and NNG, by number +1 , and matrix cells, which are occupied by 32 triplets NNA and NNT, by number -1 . Fig. 5.8 shows a result of such denotations, which lead to a numeric matrix $\mathrm{M}_{33}$ with entries +1 and -1 .

$M_{33}=$| +1 | -1 | +1 | -1 | +1 | -1 | +1 | -1 |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| -1 | +1 | -1 | +1 | -1 | +1 | -1 | +1 |
| +1 | -1 | +1 | -1 | +1 | -1 | +1 | -1 |
| -1 | +1 | -1 | +1 | -1 | +1 | -1 | +1 |
| +1 | -1 | +1 | -1 | +1 | -1 | +1 | -1 |
| -1 | +1 | -1 | +1 | -1 | +1 | -1 | +1 |
| +1 | -1 | +1 | -1 | +1 | -1 | +1 | -1 |
| -1 | +1 | -1 | +1 | -1 | +1 | -1 | +1 |

Fig. 5.8. Transformation of the symbolic ( $8 * 8$ )-matrix of 64 triplets (from Figs. 2.1 and 2.4) into the numeric matrix $\mathrm{M}_{33}$, which shows the arrangements of cells, occupied by 32 triplets NNC and NNG (black cells) and also by 32 triplets NNA and NNT (white cells).

The received numeric matrix $\mathrm{M}_{33}$ can be also decomposed by the same method of the dyadic-shift decompositions into sum of new 8 sparse matrices (as it was made above for matrices $\mathrm{M}_{21}, \mathrm{M}_{22}, \mathrm{M}_{31}$, and $\mathrm{M}_{32}$ ); the set of these 8 sparse matrices turns out to be closed in relation to the multiplication again and defines the same multiplication table of these matrices as in the case of the matrix $\mathrm{M}_{31}$ (see Fig. 5.6, at bottom).

The same analysis with similar results can be done for tetra-groupings of matrix cells regarding 4-dimensional vectors of percentages sums of tetraplets from the equalities (2).

## Some concludinhg remarks.

The presented results of the study of the regularities in the distribution of the percentages of $n$-plets in long DNA-texts of various organisms are consistent with Jordan's claiming that life's missing laws are the rules of chance and probability of the quantum world [Jordan, 1932; McFadden, Al-Khalili, 2018]. The described author's results show existence of previously unknown genetic regularities. These results were obtained on the basis of new methods of analysis and modeling of DNA-texts, which are connected with mathematical formalisms of quantum mechanics and quantum informatics, algebras of multi-dimensional hypercomplex numbers, metric analysis anf the theory on noise-immune coding of informatics.

Considering the views of Jordan and Schrödinger about the dictatorial role of the structured informatics of genetic molecules for the whole organism [McFadden, Al-Khalili, 2018], it is natural to think that the structural features of DNA informatics leave their mark on all genetically inherited biological systems and phenomena. This is consistent with the fact that all physiological systems must be structurally aligned with genetic coding in order to be transmitted in genetically encoded form to offspring. This is also consistent with the point of view that the main task of living organisms is to transfer genetic information along the chain of generations.

The received results can be used, in particular, for developing our knowledge about principles of brain activities and about a relation between 'living' and nonliving matter. These themes are actively discussed in scientific community. For example, concerning a relation between 'living' and non-living matter, W. Pauli stated that the mental and the material domain are governed by common ordering principles, and can be understood as "complementary aspects of the same reality" [Pauli, 1994; Geesink, Meijer, 2016].

Regarding the metric features of biological phenomena, it can be noted that thoughts about metric spaces are spread by a number of authors even to mathematics as a high form of intellectual activity. For example, the book [Hofstadter, 1980, page 612] notes that a mathematician feels that in mathematics there is a certain metric that unites ideas - that all mathematics is a network of results that are interconnected by a huge number of connections; had we been able to introduce this highly developed sense of mathematical closeness - the mental metric of a mathematician - into the program, we could create a primitive artificial mathematician.

In accordance with these statements of Hofstadter, the book [Nalimov, 2015, p. 115] emphasizes: "In other words, artificial intelligence could be brought closer to mathematical thinking, if it were possible to realize the metric properties of the human thinking space ... We are ready to go further and say that consciousness itself is geometrically structured: existentially, a person is geometric ... In our minds, when
constructing texts through which we perceive the World, something very similar to what happens in morphogenesis happens. We are ready to see in the depths of consciousness the same geometric images that are revealed in morphogenesis".

In general, the presented results of the author's studies of the structural rules of genetic informatics give pieces of evidences in favor of effectivity of a model approach to living organisms as quantum-informational algebraic-harmonic essences on modular principles.

## Appendix I. An example of n-plets percentages in one of the bacterial genomes.

The Appendix presents data about percentages of $n$-plets in the rhizobacteria Bradyrhizobium japonicum strain E109, complete genome, 9224208 bp (initial data were taken from https://www.ncbi.nlm.nih.gov/nuccore/CP010313.1?report=genbank; the length of this sequence is more than 20 times shorter than the length of the sequence of the first human chromosome;). Figs. A1-A5 show corresponding matrices of percentages of 4 nucleotides, 16 doublets, 64 triplets and 256 tetraplets for this genome. These percentages matrices for the bacteria and the percentages matrices (Figs. 2.2-2.5) for human chromosome \#1 have similar tetra-groupings properties, which are expressed by the equalities (1) and (2).

| C | A |
| :---: | :---: |
| T | G |$\rightarrow$| $\% \mathrm{C}$ | $\% \mathrm{~A}$ |
| :--- | :--- |
| $\% \mathrm{~T}$ | $\% \mathrm{G}$ |$=$| 0.3184 | 0.1819 |
| :--- | :--- |
| 0.1815 | 0.3182 |

Fig. A.1. The transformation of the symbolic matrix of 4 nucleotides into numeric matrix of nucleotides percents in the case of the genomic DNA of the Bradyrhizobium japonicum strain E109.

| \%CC | \%CA | \%AC | \%AA |
| :---: | :---: | :---: | :---: |
| \%CT | \%CG | \%AT | \%AG |
| \%TC | \%TA | \%GC | \%GA |
| \%TT | \%TG | \%GT | \%GG |$=$| 0.0781 | 0.0587 | 0.0467 | 0.0367 |
| :--- | :--- | :--- | :--- |
| 0.0515 | 0.1302 | 0.0469 | 0.0517 |
| 0.0720 | 0.0145 | 0.1216 | 0.0719 |
| 0.0365 | 0.0584 | 0.0468 | 0.0779 |

Fig. A.2. The matrix of percents of the 16 doublets in the genomic DNA-sequence of doublets in the Bradyrhizobium japonicum strain E109 (compare with Fig. 2.3).

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


| 0.0131 | 0.0154 | 0.0136 | 0.0129 | 0.0145 | 0.0083 | 0.0097 | 0.0068 |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| 0.0148 | 0.0344 | 0.0145 | 0.0185 | 0.0053 | 0.0182 | 0.0075 | 0.0126 |
| 0.0174 | 0.0040 | 0.0446 | 0.0328 | 0.0205 | 0.0043 | 0.0205 | 0.0106 |
| 0.0126 | 0.0183 | 0.0185 | 0.0343 | 0.0075 | 0.0144 | 0.0052 | 0.0146 |
| 0.0149 | 0.0133 | 0.0044 | 0.0021 | 0.0353 | 0.0210 | 0.0195 | 0.0153 |
| 0.0105 | 0.0323 | 0.0043 | 0.0039 | 0.0210 | 0.0443 | 0.0209 | 0.0171 |
| 0.0151 | 0.0020 | 0.0206 | 0.0133 | 0.0195 | 0.0044 | 0.0356 | 0.0150 |
| 0.0069 | 0.0125 | 0.0082 | 0.0153 | 0.0098 | 0.0133 | 0.0148 | 0.0133 |

Fig. A.3. The matrix of percents of the 64 triplets in the genomic DNA-sequence of triplets in the Bradyrhizobium japonicum strain E109 (compare with Fig. 2.4).

| . 0023 | . 0025 | . 0033 | . 0030 | . 0052 | . 0017 | . 0035 | . 0021 | . 0022 | . 003 | . 0018 | . 0019 | . 0023 | . 0023 | . 0016 | . 0015 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| . 0026 | . 0057 | . 0038 | . 0055 | . 0014 | . 0051 | . 0027 | . 0045 | . 0029 | . 0058 | . 0024 | . 0022 | . 0015 | . 0036 | . 0019 | . 0018 |
| . 0048 | . 0011 | . 0119 | . 0082 | . 0066 | . 0011 | . 0081 | . 0030 | . 001 | . 0006 | . 0057 | . 0050 | . 0025 | . 0010 | . 0045 | . 0028 |
| . 0039 | . 0052 | . 0048 | . 0096 | . 0021 | . 0044 | . 0018 | . 0052 | . 0014 | . 0019 | . 0029 | . 0047 | . 0018 | . 0022 | . 0015 | . 0039 |
| . 0 | . 0 | . 00 | . 0 | . 01 | . 00 | . 0 | . 00 | . 00 | . 00 | . 00 | . 00 | . 00 | . 0043 | . 0026 | 6 |
| . 0 | . 0 | . 00 | . 0 | . 0 | . 0 | . 0 | . 0 | . 00 | . 00 | . 00 | . 0016 | . 0039 | . 0072 | . 0035 | . 0019 |
| . 0056 | . 0005 | . 0068 | . 003 | . 0 | . 0016 | . 0173 | . 0055 | . 002 | . 000 | . 0048 | . 0034 | . 0018 | . 0006 | . 0058 | . 0036 |
| . 0019 | . 0043 | . 0022 | . 005 | . 003 | . 005 | . 005 | . 005 | . 001 | . 002 | . 002 | . 003 | . 0015 | . 0013 | . 0030 | . 0027 |
| . 0 | . 00 | . 0 | . 0 | . 0 | . 0 | . 0 | . 0 | . 0 | . 0 | . 0 | . 0 | . 00 | . 0 | . 0042 | 9 |
| . 0035 | . 0057 | . 0 | . 0 | . 0 | . 0016 | . 0 | . 0 | . 005 | . 0173 | . 004 | . 0067 | . 0018 | . 0081 | . 0025 | . 0057 |
| . 00 | . 001 | . 0103 | . 00 | . 0017 | . 0004 | . 00 | . 0010 | . 006 | . 0013 | . 0167 | . 0102 | . 0098 | . 0017 | . 0070 | . 0040 |
| . 0027 | . 0030 | . 0050 | . 0082 | . 0011 | . 0011 | . 0006 | . 0010 | . 0045 | . 0080 | . 0058 | . 0118 | . 0024 | . 0066 | . 0014 | . 0047 |
| . 0036 | . 0031 | . 0007 | . 0004 | . 0065 | . 0033 | . 0035 | . 0031 | . 0037 | . 0036 | . 0010 | . 0007 | . 0094 | . 0065 | . 0037 | . 0035 |
| . 0026 | . 0058 | . 0006 | . 0004 | . 0043 | . 0069 | . 0044 | . 0024 | . 0026 | . 0095 | . 0010 | . 0016 | . 0053 | . 0141 | . 0041 | . 0036 |
| . 0028 | . 0005 | . 0047 | . 0030 | . 0036 | . 0009 | . 0061 | . 0032 | . 0040 | . 0005 | . 0048 | . 0034 | . 0059 | . 0011 | . 0059 | . 0027 |
| . 0015 | . 0021 | . 0020 | . 0029 | . 0024 | . 0016 | . 0036 | . 0026 | . 0016 | . 0034 | . 0018 | . 0033 | . 0023 | . 0052 | . 0021 | . 0024 |

Fig. A.4. The matrix of percents of the 256 tetraplets in the genomic DNA-sequence of tetraplets in the Bradyrhizobium japonicum strain E109 (compare with Fig. 2.5).

| $\% \mathrm{C} \approx 0.3184$ | $\% \mathrm{G} \approx 0.3182$ | $\% \mathrm{~A} \approx 0.1819$ | $\% \mathrm{~T} \approx 0.1815$ |
| ---: | ---: | ---: | ---: |
| $\Sigma \% \mathrm{CN} \approx 0.3185$ | $\Sigma \% \mathrm{GN} \approx 0.3182$ | $\Sigma \% \mathrm{AN} \approx 0.1820$ | $\Sigma \% \mathrm{TN} \approx 0.1814$ |
| $\Sigma \% \mathrm{NC} \approx 0.3183$ | $\Sigma \% \mathrm{NG} \approx 0.3182$ | $\Sigma \% \mathrm{NA} \approx 0.1818$ | $\Sigma \% \mathrm{NT} \approx 0.1816$ |
| $\Sigma \% \mathrm{CNN} \approx 0.3197$ | $\Sigma \% \mathrm{GNN} \approx 0.3201$ | $\Sigma \% \mathrm{ANN} \approx 0.1806$ | $\Sigma \% \mathrm{TNN} \approx 0.1797$ |
| $\Sigma \% \mathrm{NCN} \approx 0.3166$ | $\Sigma \% \mathrm{NGN} \approx 0.3173$ | $\Sigma \% \mathrm{NAN} \approx 0.1836$ | $\Sigma \% \mathrm{NTN} \approx 0.1825$ |
| $\Sigma \% \mathrm{NNC} \approx 0.3189$ | $\Sigma \% \mathrm{NNG} \approx 0.3173$ | $\Sigma \% \mathrm{NNA} \approx 0.1815$ | $\Sigma \% \mathrm{NNT} \approx 0.1824$ |
| $\Sigma \% \mathrm{CNNN} \approx 0.3184$ | $\Sigma \% \mathrm{GNNN} \approx 0.3183$ | $\Sigma \% \mathrm{ANNN} \approx 0.1816$ | $\Sigma \% \mathrm{TNNN} \approx 0.1817$ |
| $\Sigma \% \mathrm{NCNN} \approx 0.3184$ | $\Sigma \% \mathrm{NGNN} \approx 0.3185$ | $\Sigma \% \mathrm{NANN} \approx 0.1816$ | $\Sigma \% \mathrm{NTNN} \approx 0.1815$ |
| $\Sigma \% \mathrm{NNCN} \approx 0.3185$ | $\Sigma \% \mathrm{NNGN} \approx 0.3181$ | $\Sigma \% \mathrm{NNAN} \approx 0.1823$ | $\Sigma \% \mathrm{NNTN} \approx 0.1811$ |
| $\Sigma \% \mathrm{NNNC} \approx 0.3182$ | $\Sigma \% \mathrm{NNNG} \approx 0.3180$ | $\Sigma \%$ NNNA $\approx 0.1820$ | $\Sigma \% \mathrm{NNNT} \approx 0.1818$ |

Fig. A.5. Percentages of nucleotides C, G, A, T, and the sum $\Sigma$ of percent of n-plets with these nucleotides at the attributive positions for the case of the genomic DNA-text of the the Bradyrhizobium japonicum strain E109 (compare with Fig. 3.1).

## Acknowledgments

Some results of this paper have been possible due to a 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 selforganizing systems", where the author was a scientific chief from the Russian Academy of Sciences. The author is grateful to G. Darvas, E. Fimmel, A.A. Koblyakov, M. He, Z.B. Hu, Yu.I. Manin, I.V. Stepanyan, V.I. Svirin and G.K. Tolokonnikov for their collaboration.

## References.

Albrecht-Buehler G. Asymptotically increasing compliance of genomes with Chargaff's second parity rules through inversions and inverted transpositions. Proc Natl Acad Sci U S A. November 21; 103(47): 17828-17833 (2006).
Chargaff, E. Preface to a Grammar of Biology: A hundred years of nucleic acid research. - Science, 172, 637-642(1971).
Fimmel E., Petoukhov S.V. Development of Models of Quantum Biology Based on the Tensor Product of Matrices. In: Hu Z., Petoukhov S., He M. (eds). Advances in Intelligent Systems and Computing, v. 1126, p.126-135. Springer, Cham (2020), DOI https://doi.org/10.1007/978-3-030-39162-1_12.
Geesink H.J.H., Meijer D.K.F. Quantum wave information of life revealed. NeuroQuantology, vol. 14, issue 1 (March 2016), doi: 10.14704/nq.2016.14.1.911
Harkin A.A., Harkin J.B. Geometry of Generalized Complex Numbers. Mathematics Magazine, v. 77(2), p. 118-29 (2004).
Hofstadter D.R. Gödel, Escher, Bach. An Eternal Golden Braid. N.Y.: Vintage Books, 777 p. (1980).
Jordan P. Die Quantenmechanik und die Grundprobleme der Biologie und Psychologie. Naturwissenschaften 20, 815-821, 1932 (doi:10.1007/BF01494844)
Kantor I.L., Solodovnikov A.S. Hypercomplex numbers. Berlin, New York: Springer-Verlag (1989). ISBN 978-0-387-96980-0.
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.
Nalimov V.V. I am scattering thoughts (in Russian: Razbrasyvaiu mysli). Moscow, Center for Humanitarian Initiatives (2015). ISBN 978-5-98712-521-2].
Pauli W. Writings on Physics and Philosophy. Ed. by C.P. Enz and K. von Meyenn, Springer, Berlin, 1994.
Petoukhov S.V. Matrix genetics, algebrases of genetic code, noise immunity. Moscow, RCD, 316 p. (2008, in Russian). ISBN 978-5-93972-643-6.
Petoukhov S.V. Nucleotide Epi-Chains and New Nucleotide Probability Rules in Long DNA Sequences. Preprints 2019, 2019040011, 17 pages (doi: 10.20944/preprints201904.0011.v1), https://www.preprints.org/manuscript/201904.0011/v1
Petoukhov S.V. Hyperbolic Rules of the Cooperative Organization of Eukaryotic and Prokaryotic Genomes. Biosystems, 198, 104273 (2020a).
Petoukhov S.V. Modeling inherited physiological structures based on hyperbolic numbers, BioSystems (2020b), https://doi.org/10.1016/j.biosystems.2020.104285.
Petoukhov S.V. Hyperbolic Rules of the Oligomer Cooperative Organization of Eukaryotic and Prokaryotic Genomes. Preprints 2020, 2020050471 (2020c),
doi:10.20944/preprints202005.0471.v2, https://www.preprints.org/manuscript/202005 .0471/v2.
Petoukhov S.V., He M. Symmetrical Analysis Techniques for Genetic Systems and Bioinformatics: Advanced Patterns and Applications. Hershey, USA, IGI Global (2010).

Prabhu V. V. Symmetry observation in long nucleotide sequences. Nucleic Acids Res., 21, 2797-2800 (1993).
Stambuk N. Circular coding properties of gene and protein sequences. Croat. Chem. Acta, 72, pp. 999-1008 (1999).

