

# ИРРАЦИОНАЛЬНОСТЬ ГЕНЕТИЧЕСКОГО КОДА

## Макаров Л.М. Email: Makarov1169@scientifictext.ru

*Макаров Леонид Михайлович – кандидат технических наук, профессор,  
кафедра конструирования и производств радиоэлектронных средств,  
Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича,  
г. Санкт-Петербург*

**Аннотация:** *представлена логическая система суждений, устанавливающих иррациональность генетического кода и вычислительных процедур создания количественной оценки сложности генома различных живых организмов, на поле понятий молекулярной генетики и теории информации, регламентирующих формат синтеза нуклеотидной цепи событий в бинарном виде с использованием производящей функции. Формализовано понятие вариативности генома живого организма, представленного набором нуклеотидов в информационном источнике NCBI, обеспечивающее создание набора сравнительных оценок видового подобия различных организмов.*

**Ключевые слова:** *вычислительная оценка сложности генома, генетический код.*

## IRRATIONALITY OF THE GENETIC CODE

### Makarov L.M.

*Makarov Leonid Mikhailovich - PhD in System analysis, Associate Professor,  
DEPARTMENT OF DESIGNING AND PRODUCTION OF RADIO-ELECTRONIC MEANS,  
ST.-PETERSBURG STATE UNIVERSITY OF TELECOMMUNICATIONS OF THE PROF. M.A. BONCH-BRUYEVICH,  
ST.-PETERSBURG*

**Abstract:** *a logical system of judgments establishing irrationality of genetic code and computational procedures for creating quantitative estimation of genome complexity of various living organisms is presented on the field of concepts of molecular genetics and information theory regulating format of synthesis of nucleotide chain of events in binary form using producing function. The concept of variation of the genome of a living organism, represented by a set of nucleotides in the information source NCBI, is formalized, providing creation of a set of comparative estimates of species similarity of different organisms.*

**Keywords:** *computational estimation of genotype complexity, genetic code.*

УДК 519.61 + 539.1

Формализм понятий и определений об объектах материального мира, создается средствами математики. Математика, как научная дисциплина, способствует созданию идеализированных объектов – моделей реального мира Природы. Используя различные аксиоматические системы, создаются общие представления о реальных объектах, процессах и явлениях. Так, например, аксиоматическая система Пиано, воспроизводимая на девяти логически связанных постулатах, отчетливо проявляется в многих естественнонаучных направлениях. В частности, выделяя в этой системе тезис о порядке исчисления событий некоторого процесса, устанавливается логическая связь прошлого и настоящего времени свершения элементарного события. В теории информации этот тезис расширяется понятиями об информационной емкости серии событий, создаваемых на основе простых элементов – знаков. Применяя определенную систему знаков, декларируется возможность создания различных информационных транспарантов, являющихся примерными описаниями или моделями реальных процессов, явлений или объектов Природы [1,3].

Создавая искусственную категорию объектов различными инструментами, в том числе и средствами математики, создаются новые понятия об объектах процессах и явлениях Природы. Так, например, в области связи – передачи данных, сформировалась устойчивая доминанта об иррациональной основе аналогового сигнала, воплощающего модель с бесконечной информационной емкостью. Переход от аналогового к цифровому сигналу всегда обладает свойством неполноты описания. В этом контексте уместно отметить, что теоретические разработки в области техники, в широком смысле этого понятия, через некоторое время становились актуальными и в сфере биологических исследований. Так для молекулярной генетики обнаружение нуклеотидов воспроизвело серию понятий о дискретности генетическом коде, которые следует рассматривать как источник, инициализирующий формирование концепции о иррациональных свойствах набора нуклеотидов.

Отсутствие возможности полностью отождествлять искусственные и природные объекты компенсируется наличием приобретения новых знаний. Чем совершеннее создаются описания и модели объектов Природы, тем больше потенциал знаний. В этом отношении увеличение потенциала знаний можно рассматривать как иррациональный процесс, в котором лишь приближенно прослеживается сочетание раз-

личных научных воззрений и концепций. Творчество, как иррациональный синтез современных научных суждений часто лишь приближенно связывает прошлые и настоящие прагматические задачи, находящиеся на поле понятий разных научных дисциплин.

Оперируя представлениями о порядке следования знаков, создаются информационные постулаты о семантике, информационной емкости сообщения. Естественнонаучные дисциплины, в основе интересов которых находятся прагматические задачи о принципах синтеза живой материи, также используют информационное поле понятий. Здесь в качестве знаков, образующих базис процессов жизнеобеспечения организмов, представлены нуклеотиды, которые являются основной структурной единицей нуклеиновых кислот.

Нуклеотид – азотистое основание, моно мерное звено. Набор звеньев является основой нуклеиновых кислот, которые разделяются на рибонуклеиновые (РНК) и дезоксирибонуклеиновые (ДНК). Бинарный характер такого деления нуклеотидов сложился эволюционно. Первыми в Природных процессах появились рибонуклеиновые кислоты, которые представлены нуклеотидами: аденином (А), гуанином (G), цитозином (С) и урацилом (U). Дезоксирибонуклеиновые кислоты являются продуктом эволюционного развития живых организмов и представлены нуклеотидами: аденином (А), гуанином (G), цитозином (С) и тимином (Т). В общем понимании информационного базиса нуклеотиды РНК и ДНК отличаются только одним элементом: урацилом и тимином. Несмотря на минимальное различие, нуклеотиды демонстрируют значительные отличия по форме. Первичная по происхождению РНК позиционируется сферическими образованиями, а ДНК реализуется в виде спиралей. Одновременное участие нуклеотидов в синтезе живой материи консолидирует форматы, которые по внешним признакам напоминают галактические спиральные структуры. Три последовательно соединенных нуклеотида образуют триплет. Сочетания триплетов обладают иррациональной знаковой информацией о строении многочисленных белков, сосредоточенных в хромосомах.

Хромосома – информационный стек, содержащий построчную запись нуклеотидов, реализующих процессы синтеза различных клеток, необходимых для организации жизненно важных функций организма. Используя термины теории информации, можно указать на наличие у человека 23 информационных стеков, что соответствует количеству непарных хромосом. Кроме этого набора присутствует две парные хромосомы, которые определяют пол. Длина хромосом оценивается в специальных единицах – мегабазах.

Мегабаза (megabase, Mb) – набор из 1 миллиона пар нуклеотидов в линейной цепи нуклеиновой кислоты. Это количество эквивалентно одной килобазе: 1 kb=1 000 пар нуклеотидов и обозначается 1000 kb (kilobase) в РНК и одноцепочечной ДНК или пар нуклеотидов (п.н.) в двух цепочечной ДНК.

Сопоставление записей генетической информации в хромосоме и, например, текста на естественном языке, указывает на наличие подобия, которое проявляется, в первую очередь, линейной схемой развертки событий – последовательной записи элементов текста (стека) или нуклеотидов.

Для хромосомы развертка событий задается однозначно в последовательности элементов - нуклеотидов (G; T; C; A), представленной в информационном источнике [2] в формате fasta. Это исторически сложившийся формат построчной записи нуклеотидов от 50 до 160 элементов в строке. Количественное описание строчной записи нуклеотидов создается на численных показателях, по выражению:

$$M = \sum[G, T, C, A]$$

$$m_G = G; m_T = T; m_C = C; m_A = A$$

$$l_G = M - G; l_T = M - T; l_C = M - C; l_A = M - A \quad (1)$$

Где M – размер массива нуклеотидов.

В таком случае для каждого нуклеотида, входящего в состав произвольно выбранной строки избранного стека, с общим количеством элементов M, имеем набор показателей:

$$a = \frac{m}{M}; b = \frac{l}{M} \quad (2)$$

Где  $m = [G, T, C, A]$  и  $l = M - m$

Полагаем, что при условии  $a + b = 1$  параметр b характеризует семантическую составляющую линейной нуклеотидной цепи событий. Понимание линейной схемы развития событий в синтезе последующих поколений нуклеотидов учитывает наличие мутаций. Такое суждение отчетливо проявляется на части хромосомы – гене, представленном относительно короткой линейной записью набора нуклеотидов. Однако, вариабельность фрагментов хромосомы (генов), в определенном смысле, не оказывает существенного влияния на видовой образ живого организма. Такой живой организм по-прежнему обладает хорошо различимыми отличиями от других живых организмов. Полагаем, что такая схема модификации характеризует «горизонтальную мутацию гена». Декларируя наличие этого свойства живого организма, констатируем наличие производящей функции, осуществляющей модификацию линейной цепи нуклеотидов. Зададим производящую функцию в виде [4]:

$$x(\varphi) = a + b(\varphi) * \cos(\varphi) \quad (3)$$

$$y(\varphi) = a + b(\varphi) * \sin(\varphi)$$

Где  $\varphi = \frac{2\pi}{b}$  естественным образом характеризует иррациональность нуклеотидной цепи, а, следовательно, и генетического кода.

Расширяя представления о вариации нуклеотидов в хромосоме, аналогично вводится понятие о «вертикальной мутации гена», наблюдаемой на разных участках хромосомы при сравнении с разными видами живых организмов.

Очевидно, что выделение нуклеотидов в структуре стека – хромосоме, декларирует наличие составных элементов, количество которых определяется размером рассматриваемого фрагмента хромосомы (M). Иначе говоря, выделенный, некоторым образом, фрагмент хромосомы, можно идентифицировать в качестве гена. В таком случае в единой линейной развертке событий создания нуклеотидной цепи - стека, на основе четырех нуклеотидов, как элементов, проявляется возможность воспроизведения цифрового кода, иллюстрация которого представлена в таблице 1.

Таблица 1. Разложение стека нуклеотидов на элементарные события

	<b>C</b>	<b>T</b>	<b>A</b>	<b>T</b>	<b>G</b>
<b>G</b>	0	0	0	0	1
<b>T</b>	0	1	0	1	0
<b>C</b>	1	0	0	0	0
<b>A</b>	0	0	1	0	0

Естественным следствием построения набора кодовых состояний, по всей установленной длине стека событий, является наличие частичных сумм (m), для которых справедливо равенство:

$$m_G + m_T + m_C + m_A = M \quad (4)$$

В таком случае бинарное позиционирование элементов стека (хромосомы), в виде линейной последовательности нуклеотидов, порождает уникальность генетического кода. Формирование такого понятия создается традиционным методом генетического исследования с известным термином «рамки считывания», которое устанавливает положение первого нуклеотида в первом инициализирующем кодоне и фиксируется последним нуклеотидом терминирующего кодона. Принимая во внимание, что генетический код создается триплетом нуклеотидов, в генетических исследованиях фиксируется три положения рамки считывания.

Следуя этим правилам, инициализируем процедуру создания стека нуклеотидов, естественно с учетом извлекаемого из информационного источника [2] набора нуклеотидов. Алгоритм развертки событий генетического кода представим в виде:

$$\tau_{m_i} = m_i \Delta\varphi_{m_i} \quad (5)$$

Где  $\tau_{m_i}$  - позиционный код развертки событий;

$m \rightarrow [G, T, C, A] = m_G, m_T, m_C, m_A$  – бинарный код одного из нуклеотидов;  $\Delta\varphi_{m_i}$  - позиция фиксации кода / радианы;  $0 < i < M$

Тогда выражение (3) запишем в виде:

$$\begin{aligned} x(\tau) &= a + b(\tau_{m_i}) * \cos(\tau_{m_i}) \\ y(\tau) &= a + b(\tau_{m_i}) * \sin(\tau_{m_i}) \end{aligned} \quad (6)$$

Частичный модельный образ живого организма воспроизводится с использованием избранного набора нуклеотидов на установленной хромосоме. Такой метод позволяет установить принадлежность выбранного организма к определенному виду живых организмов. Метод сравнения нуклеотидных наборов широко развит в генетике. В такой процедуре рассматриваются фрагменты хромосом одного или нескольких организмов. С позиции математики, выбор фрагментов хромосом для сравнения может осуществляться произвольно, но при этом необходимо соблюдать соответствие размерностей привлекаемых для анализа нуклеотидных массивов.

Процедура формирования идентификационного показателя (D) организма, на основе избранного массива генов [2], представленного производящей функцией (6), с учетом известной методологии «трех рамок считывания», реализуется по выражению:

$$D = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (x_3 - x_2)^2 + (y_3 - y_2)^2} \quad (7)$$

Очевидно, такая процедура естественным образом применима и на случай сравнения двух видов живых организмов. Представленные для анализа генома выражения (6, 7) позволяют в иллюстративной форме показать широкие возможности построения оценок подобия для различных видов живых организмов.

Таблица 2. Сравнительные оценки подобия организмов (Хромосома Y, фрагмент 1736-1737)

Организм	Идентификатор D	Оценка подобия с организмом человека D %
Человек	1186,377	
Шимпанзе	1167,253	1,612
Мышь	1214,667	-2,385
Крыса	1090,443	8,0863
Макака	1118,637	5,7098
Свинья	1154,688	2,6711

Повторим расчеты для другого фрагмента хромосомы Y. Полученные результаты, представленные в таблице 2 и таблице 3, позволяют констатировать, что идентичные по номеру следования в хромосоме Y наборы нуклеотидов, для разных живых организмов, создают регламентированный набор оценок D, как при «горизонтальной модификации», так и при «вертикальной модификации» линейной цепи нуклеотидов.

Таблица 3. Сравнительные оценки подобия организмов (Хромосома Y, фрагмент 1836-1837)

Организм	Идентификатор D	Оценка подобия с организмом человека D %
Человек	1230,518	
Шимпанзе	1140,119	7,3464
Мышь	1189,11	3,3651
Крыса	1060,49	13,818
Макака	1088,404	11,549
Свинья	1123,956	8,6599

Очевидно, что независимо от выбора фрагмента сравнения на хромосоме, для представленного набора организмов, отчетливо, можно сказать, посредством численных оценок, просматривается иерархическая сложность живых организмов. При этом наблюдаем, что наборы оценок, для разных фрагментов хромосомы, имеют отличия. И этот факт является существенным аргументом, в пользу наличия свойств иррациональности генетического кода, наличия бесконечной изменчивости. Фактически это указывает на бесконечность возможностей численного позиционирования кода – нуклеотидного набора, что декларирует отсутствие возможности обнаружить в Природе два абсолютно одинаковых объекта. В этом смысле полученный результат исследования, созданный на основе математического описания набора нуклеотидных наборов, воспроизводит естественную Природную идею создания дискретных и аналоговых – иррациональных описаний объектов живой материи. Полученные знания о дуализме этих технологий творения Природой объектов окружающего мира, способствует укреплению научного потенциала знаний в естествознании.

#### Список литературы / References

1. Макаров Л.М. Формализм вычисления оценки эмерджентности Наука, техника и образование, 2020. № 1 (65). DOI: 10.24411/2312-8267-2020-10101.
2. [Электронный ресурс]. Режим доступа: <https://www.ncbi.nlm.nih.gov/genome/gdv/> (дата обращения: 25.05.2020).
3. Макаров Л.М. Информационная энтропия International scientific review of the problems and prospects of modern science and education Collection of scientific articles LXVII International correspondence scientific and practical conference, 2020. DOI: 10.24411/2542-0798-2020-16702.
4. Макаров Л.М., Поздняков А.В. Фрактальный образ генома коронавируса Covid-19 В сборнике: International scientific review of the problems and prospects of modern science and education Collection of scientific articles LXIX International correspondence scientific and practical conference, 2020. С. 6-10. DOI: 10.24411/2542-0798-2020-16902.