

# 1. Biological phenomena and prime numbers

Marinos Spiliopoulos<sup>1</sup>

<sup>1</sup>Biology Professor

June 9, 2020

As we know, numbers are involved in the phenomena of life and beyond, in various ways. Biomathematics and Bioinformatics nowadays are thriving and will have greater development in the future. Some number in particular have a lot of interest and give a special charm to both the inanimate and the living matter. For example, the  $p$  and  $f$  numbers are involved in many biological procedures in different ways. Specifically,  $p = 3.14159\dots$ , i.e. the quotient of the circumference of a circle to its diameter, is revealed in the formulas of the potential difference in the cell membranes, in the flow of liquid in vessels, etc. The  $f = 1.618$  aka golden ratio appears in various proportions of the human body, in the way leaves are assembled in plants etc. The Fibonacci sequence is correlated to  $f$ , when calculating rabbit offspring. The sequence and thus the  $\varphi$ , are connected to the logarithmic spiral that we meet at the shape of the shells of land or marine organisms such as snails, nautilus etc.. Something remarkable with the  $f$  is the following: If the core volumes, the cytoplasm and the cell are consecutive geometric progression conditions it is known that

$$\frac{V_{kp}}{V_p} = \varphi \text{ (cell-core constant).}$$

Because this may apply for each moment of a cell's lifecycle, given that volumes can be multiplied by a coefficient (effective volume) we end up with the relationship

$$\frac{V_{kp}}{V_p} = n [?] \varphi$$

The disorder of the relationship between the volumes of nucleus-cytoplasm is of great interested in cases of malignancies, tumors, etc. If somebody looks through a microscope, for instance, cells and their core, he will notice a huge difference between leukemic and normal cells. Another case is the number  $e = 2,711828$ , the base of natural logarithms. There is no science field which is not correlated with this number.

In this paper we will deal more thoroughly with the prime numbers in biology. The prime numbers are integer positive numbers divisible only by themselves and the unit 1, eg. 2, 3, 5, 7, 11, 13,17... The primes that differ by 2 are called twin primes. Euclid was the first who proved that the number of primes is infinite. Today we know that the greater the interval between the numbers is, the more prime sparsity we have. On the other hand, complex numbers (all positive integers that are not primes) result as a raw product of primes, e.g.  $70 = 2 \times 5 \times 7$  etc. Therefore any positive integer greater than 1 is either a prime or a product of primes. Mathematicians for their own reasons, consider the number 1 neither a prime nor a complex number. Prime numbers, are for Mathematics what for Chemistry and Physics are atoms of matter: the building blocks upon which all other numbers are built, the "atoms" that compose the Mathematics' universe. Someone said that mathematicians love prime numbers, such as chemists love atoms and biologists love genes. The interest in prime numbers is big and despite the era, it remains at the forefront of scientific activity. Moreover, there are major unresolved problems over time regarding prime numbers (Riemann conjecture, Goldbach conjecture,

conjecture of the twin primes etc.). Also, there are no mathematical formulas known to compute the next prime number, and how many prime numbers are smaller than a specific number. Two very characteristic sayings of great scientists about the importance of the above are the following: Hilbert, a great German mathematician had said that "if it was possible after 1000 years from my death to come back to life, the first thing I would ask would be if Riemann's conjecture has been resolved." This said, knowing, as all mathematicians, that the problem is extremely difficult and the solution of it will lead to numerous solutions of other mathematical problems.

Paul Erdos, a Hungarian great mathematician who dealt with prime numbers, just before his death said: "It will take one million years until we understand prime numbers."

Let's have a look at primes' relations with biology, starting with some simple observations:

- a) The codons of the genetic code that encode amino acids are 61, while there are three (dating) that do not encode any amino acid. Each codon consists of three nucleotides.
- b) Among the 37 genes of mitochondrial DNA 13 encode proteins.
- c) The smallest form of life, a bacterium that belongs to the mycoplasma has 521 genes (a prime number).
- d) To nucleic acids (DNA, RNA), the orientation of the polynucleotide chain is 5' - 3'.
- e) Microtubules, which are structures of the cellular shell of the ciliums and flagellums of various eukaryotic cells, consist of 13 parallel protofibrils (with a 5nm diameter each) of alternating molecules of alpha and beta tubulin.
- f) 2 and 3 phosphate groups of nucleotides (ADP, ATP, etc.), 2, 3, 5, N atoms depending on the nitrogenous base of nucleotides, consist of 5 Carbon atoms therein.
- g) 5 radial symmetry in echinoderms, e.g. starfish.
- h) 23 chromosomes in human reproductive cell etc.
- i) In the genome of prokaryotic organisms, the genes for enzymes involved in a metabolic way, are organized in operons, i.e. groups that are subject to common control of their expression. The lactose operon has three structural genes, many other 2 or 3 structural genes, tryptophan has 5 etc.

In order to avoid any misunderstanding, we clarify that in many biological issues where numerical quantities are involved, there may be no prime numbers, but can be complex or decimal numbers. However, our interest in this paper focuses on prime numbers and their correlation with biological issues. We continue with examples of prime numbers that appear in natural selection issues and in evolution of organisms. A known example from the field of the animal kingdom is that two kinds of cicadas *Magicicada tredecim* and *Magicicada septendecim* live in the same environment with a lifecycle of 13 and 17 years respectively. Throughout their life, except the last year, they live in soil and feed on the juices of tree roots. During the last year of their life, for a few weeks, they transform from larvae to adult, occupy the forest, eat, lay eggs and die. But how these species evolved so that their cycle of life is actual prime numbers? An answer is that their mutual emergence to the forest takes too long and happens once in  $13 \times 17 = 221$  years. While if their life cycles were complex numbers, eg. 12 and 18 years in the above time they would have competed six times, as many are the common multiples of 12 and 18. As we can see, the prime numbers 13 and 17 are not something abstract and random, but the basis of their survival. Such examples are likely to exist in the animal and plant kingdom.

The birthday paradox is very interesting. It belongs to the probability theory and refers to a problem which by the common sense has an unlikely answer. The formulation of the problem is this: In a group of 23 people which is the possibility that two among these individuals have their birthdays on the same day? The obvious answer is  $23/365 = 0.063$ , i.e. 6%. Despite this, the mathematical solution is 50%. The possibility is even

100% when it refers to 367 people, including those who were born on February 29. 23 and 367 are prime numbers, and the reference to this paradox is used because it is related to people (biological entities) and a biological fact, their birth in a specific time period (birthday). It is worth mentioning that this paradox is the basis for one of the most common methods of cryptanalysis, in the corresponding field of computer science (cryptography).

Prime numbers are related to the molecular basis of apoptosis, or programmed cell death, which was initially studied in the filamentary worm *Caenorhabditis elegans*. Later, it was studied in other invertebrates and vertebrates. Only in the hermaphrodite filamentary worm, we have the best of apoptosis study system, to the point that is considered to be created for this purpose. Apoptosis is a fundamental biological process that is necessary for the removal of surplus cells during embryonic development, maintaining the homeostasis in the adult organism, maturation of cell populations of the hematopoietic and immune system and malignant transformation.

Typically the apoptosis is not a consequence of disfunction and therefore differs from necrosis, which is always a result of a harmful effect. In filamentary worm, during its development when 1090 cells are produced (hence 1091 total with the original zygote), of which 131 somatic cells are eliminated by the process of programmed cell death while 3 genes are responsible for it. 1091, 131 and 3 are prime numbers, and the above constitute an ascertainment, the causes of which are unknown for now. Maybe, in the future for the entire process of the phenomenon, emerge also a mathematical model.

*Amino acids, proteins and prime numbers:* The Greek language is dynamic and mysterious. Perhaps it is the only language in which there is such a complete identification of signifier and signified. Amino acids are the building blocks of proteins. The word “protein” is derived from the Greek word “protos” (which means “first”), i.e. the importance of proteins to the structure and function of cells and organisms, is of high importance. Let’s see how they relate to prime numbers. Based on their chemical formula 5 amino acids have a molecular weight which is a prime number. Specifically: Methionine 149, Alanine 89, Leucine 131, Isoleucine 131, Tyrosine 181. The molecular weight of all the other amino acids is easily calculated and there is no need to be listed in a table. Because the interesting fact is not the molecular weights of individual amino acids, but the fact that we can “artificially” and very easily design peptides and proteins, which have a total molecular weight of a prime number, either as a simple sum of the molecular weights of the free aminoacids or when all water molecules moving away are removed, multiplied by 18 (molecular weight of water), or by other criteria. For example, we have the heptapeptide: Methionine - Glycine - Tryptophan - Aspartic acid - threonine - Glutamine - Cysteine. The sum of the molecular weights of these amino acids in a free state is:  $149 + 75 + 204 + 133 + 119 + 146 + 121 = 947$  (prime number). Subtracting  $6 \times 18 = 108$  (the water that is removed by condensation), we now have  $947 - 108 = 839$  (prime number). If we multiply the molecular weights of the amino acids to their corresponding position in the chain, i.e.,  $1 \times 149 + 2 \times 75 + 3 \times 204 + 4 \times 133 + 5 \times 119 + 6 \times 146 + 7 \times 121 = 3761$  (prime number). The above example is as we said “artificial”, while in the same way we can design a lot of others, with less or more amino acids. The point here is what happens in “real conditions” with various proteins found in cells and organisms. Surely the primate to a protein is its biological role, which is a function of the information that exists in the DNA, and its organizational levels (primary structure, secondary, tertiary, etc.) depending on this information. At first glance, it may seem as a useless information compared to prime numbers for that matter. I disagree with this point of view, especially now, that Bioinformatics has had a great development (in collaboration with the Proteomics course) and such studies and calculations can be made easily and quickly. By the “look” in this direction, a fixed pattern for various biological data may possibly not occur. Form of the protein chain in space, class or subclass of proteins based on the functionality of e.g. catalytic proteins (enzymes), etc. Some proteins that have specificity criterion, compared with those who have not etc .. Or better yet, something relative may occur to prime numbers in proteins with those present in DNA, which we will see just below. It is known that a connection is found between prime numbers and quantum physics, the mathematics of Chaos (mathematics of unpredictability) etc. so it is not impossible to find such a connection to Proteomics and Genomics. Besides the innovative ideas do no harm and if anyone disagrees he can follow what a philosopher

said: "Give me ideas to reject". Moreover biology is an exact science which follows the scientific method (observation, experiment, etc.) where something wrong easily gets confuted.

*DNA and prime numbers:*

DNA works, among others, as a living computer. Let's look at some specific aspects in this area, namely the issue related to prime numbers. Do not forget that even to computers prime numbers play an important role in control systems, cryptography and computer development. The nucleotides of adenine (A), thymine (T), guanine (G) and cytosine (C), which are the building blocks of DNA, having respective molecular weights of 331, 322, 347, and 307. The numbers 331, 347, 307 are primes while 322 is complex. With the symbols A, T, G, C simultaneously we denote the corresponding nucleotides, the respective bases of the nucleotides and their molecular weights. Based on the above we find that

$$A + T = 331 + 322 = 653 \text{ (1), which is a prime number.}$$

$$A + T + G + C = 331 + 322 + 347 + 307 = 1307 \text{ (2)}$$

which is also a prime number. Easily we see that

$$(G + T) - (A + C) = 31 \text{ (3)}$$

$$11G + T - A - C = 31$$

$$14G + 6T - 14A - 6C = 314$$

$$141G + 59T - 141A - 59C = 3141$$

$$1415G + 585T - 1415A - 585C = 31415$$

$$14159G + 5841T - 14159A - 5841C = 314159$$

Each connection gives  $\pi = 3.14159 \dots$  with one, two, three, etc. decimal digits multiplied respectively with 10, 100, 1000 etc. The sum of the coefficients

$$AT - GC = 53 \text{ (4)}$$

$$A + T + C - G = 613 \text{ (5) and}$$

$$AT - GC + A + T + C = 1013 \text{ (6),}$$

which are all prime numbers. Artificially and by using a computer, one can create DNA chains that have a sum of molecular weights of nucleotides a prime number. Example: by the portion of a DNA chain containing scattered overall 29A, 59T, 131G, 107C, turns out a molecular weight of 106,903 which is a prime number. The numbers 29, 59, 131, 107 are selected and are prime in this example and the complementary part of the chain with a total nucleotide 29T, 59A, 131C, 107G has a molecular weight of 106,213, which is also a prime number. Of course, what matters for DNA is the sequence of nucleotides because it determines the sequence of the protein amino acids based on the genetic code. Only a small percentage of the genome encodes proteins. Nowadays with the mapping of the human genome and other organizations, with the help of Biomathematics and Bioinformatics easily we can make, among others, a research of information relations with prime numbers, in genes or in larger segments of DNA. I wonder what is the relation among evolution of genome information through natural selection and primes?

We finish with something that is very interesting: In summary and without doing many arithmetic operations, if we remove the link (5) by the relationship (6) we saw above, resulting

$$AT - GC + G = 400 = AT - GC + G = 10(GC) \text{ since } 347 - 307 = 40. \text{ Finally turns out the relation}$$

$$G = \frac{(A \cdot T + 10 \cdot C)}{(9 + C)} \text{ (7)}$$

A relation simple and very important linking the molecular weight of the nucleotide of guanine with the molecular weights of the other nucleotides in the DNA. But we do not stop here, because the relation (7) can often be a primes production 'machine', like: setting for C and G prime values we find a value for the AT that can be decomposed into prime factorization. Keeping A a prime value and T the product of the rest (complex number). Often we find that the sum  $A + T + G + C$  is a prime number, as in DNA. Example: with  $C = 47$ ,  $G = 173$  we find  $A = 419$  and  $T = 22$ . The sum of all of them is 661, and is a prime number. Another example of a large prime is:  $C = 311$ ,  $G = 106213$  (from previous example),  $A = 479$  and  $T = 70950$ . Their sum 177. 953 is a prime number. Often it is found that the number that turns out from the relation (4), i.e.  $AT - GC$  is prime. The symbols can be defined by  $C'$ ,  $G'$ ,  $D'$ ,  $T'$ , if we want to emphasize better differentiation.

The above process has two perspectives. First it is a case in which there are "hypothetical DNA molecules" where the molecular weights of the nucleotides are defined each time with different numbers, the three ( $G'$ ,  $C'$ ,  $D'$ ) being primes and one ( $T'$ ) complex, where their sum creates a new prime one, etc. This can be a useful tool in search of unknown prime numbers, both for mathematicians and amateur researchers working on groups, with a goal of finding even larger primes. Moreover, for those who use prime numbers for practical reasons, in control systems, cryptography, etc. The second view is the numbers derived from the relation (7), to express nucleotide numbers along a chain or a DNA segment. Those can be applied to RNA as well. The RNA is linked to the DNA from which it appears, and with the proteins produced by the encoded information. Last but not least, the first epigenetic alteration that was detected on DNA, was the methylation of cytosine. Methylation therein means the addition of a methyl group, i.e., a small chemical molecule consisting of one carbon atom and three hydrogen atoms, the total molecular weight of 15. The cytosine acquires a molecular weight of  $307 + 15 = 322$  and thymine. This mutation of cytosine which causes various phenomena (gene inactivation etc.), appears on one hand in the chemical structure (which is the most significant), but also in the equation of the molecular weight with another chemical basis of DNA. Finally, I hope this work to provide an opportunity to explore other unknown aspects in the relation of prime numbers and biological phenomena.