International Conference on Combinatorial Methods and Probability Models.
A Conference in Memory of Professor Charalambos Charalambides

Dedicated to Babis

Ourania Chryssaphinou

Janouary 1971
to
December 2010

A Military Dictatorship there was in Greece (Apr. 1967-Jul.1974)

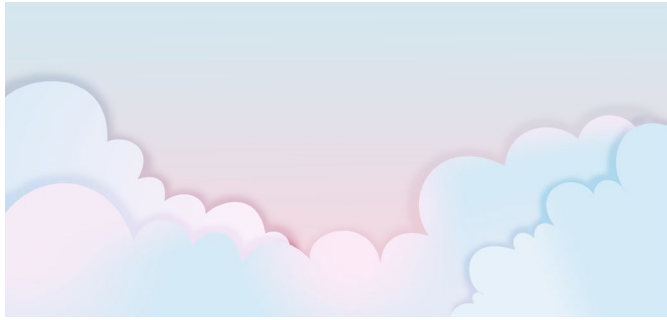
1972

(The ministry of Education invited the Academic Staff to Hilton Hotel to give them relative instructions!)

2008

- **Congress of GSI in Samos**

Since our life begins by chance and our death is sure the game is unfair. Thus, we decided to speak about life.



Applications of Probability models and Combinatorics to Biological Sequences

In **1869** DNA (Deoxyribose Nucleic Acid - Δεοξυριβονουκλεικό Οξύ) was detected in cell nucleus by Swiss chemist Friedrich Miescher and a new research period for Biology was started.

(The first known use of the word DNA was in 1944).
The first known use of the word *biology* was in 1799. It is derived from the Greek words /bios/ meaning /life/ and /logos/ meaning /study/ and is defined as the science of life and living organisms) .

Until **1944** it was not known that DNA kept the mystery of genetic code of the organisms.

Statistics was the basic branch of Mathematics which had an immediate relationship with Biology given the one to the other a lot of problems

Prof. Karl Pearson (1857-1936) the English mathematician and Biomathematician with his Research group and especially with G. Udny Yule developed and applied the following sections of Statistics:

A. Frequences Distributions

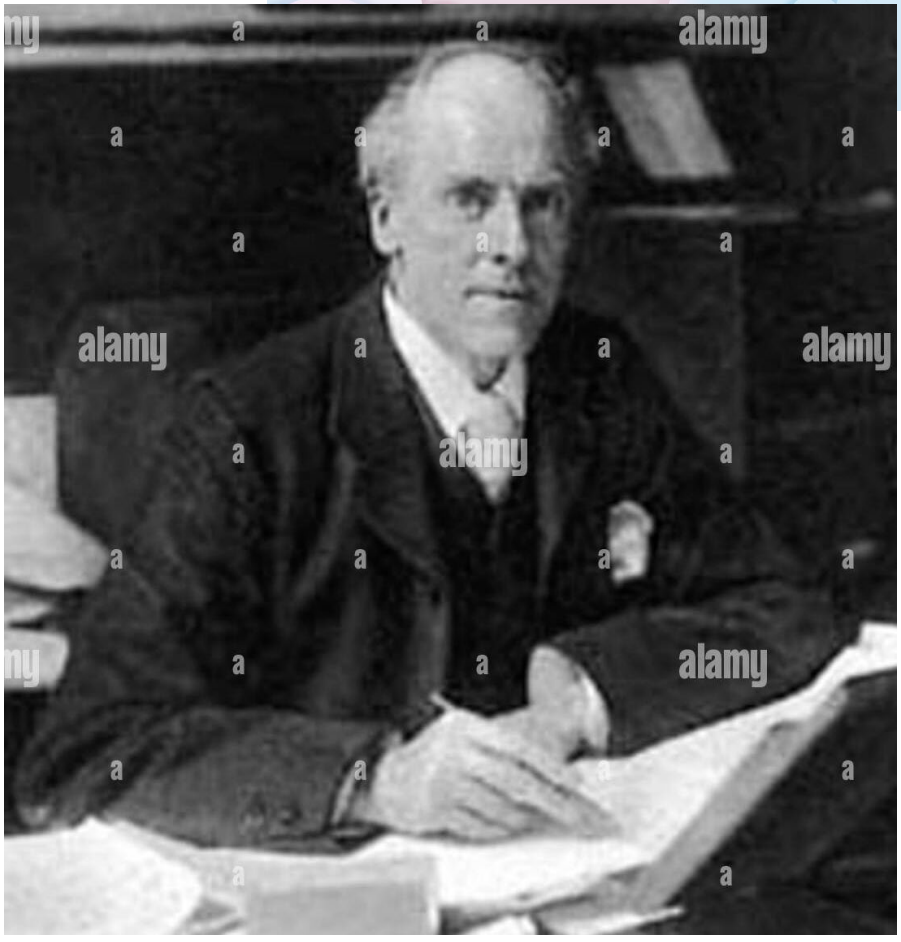
B. Correlation Coefficient

C. Samping Theories

D. Human heredity and General Theory

E. Mathematical Tables

F. Biometric (not for human being)

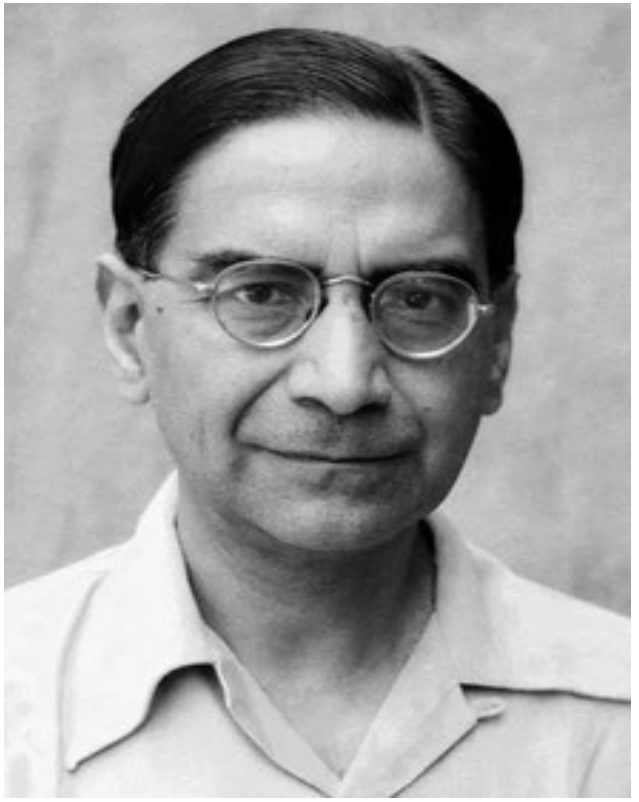


In the sequel, many statisticians worked on Biological problems.
We refer:
Prof. Ronald A. Fisher (1890-1962), who developed the theory of Analysis of Variance as well as many others statistical methods.
According to Prof. C. R. Rao (1920-2023) R. A. Fisher was the Founder of Modern Statistics. In 1950 Fisher published the article:

"Gene Frequencies in a Cline Determined by Selection and Diffusion".

It was the first serious application of the electronic delay storage automatic calculator machine (EDSAC) of the University of Cambridge.

(Computer science began to be established as a distinct academic discipline in the 1950s and early 1960s)

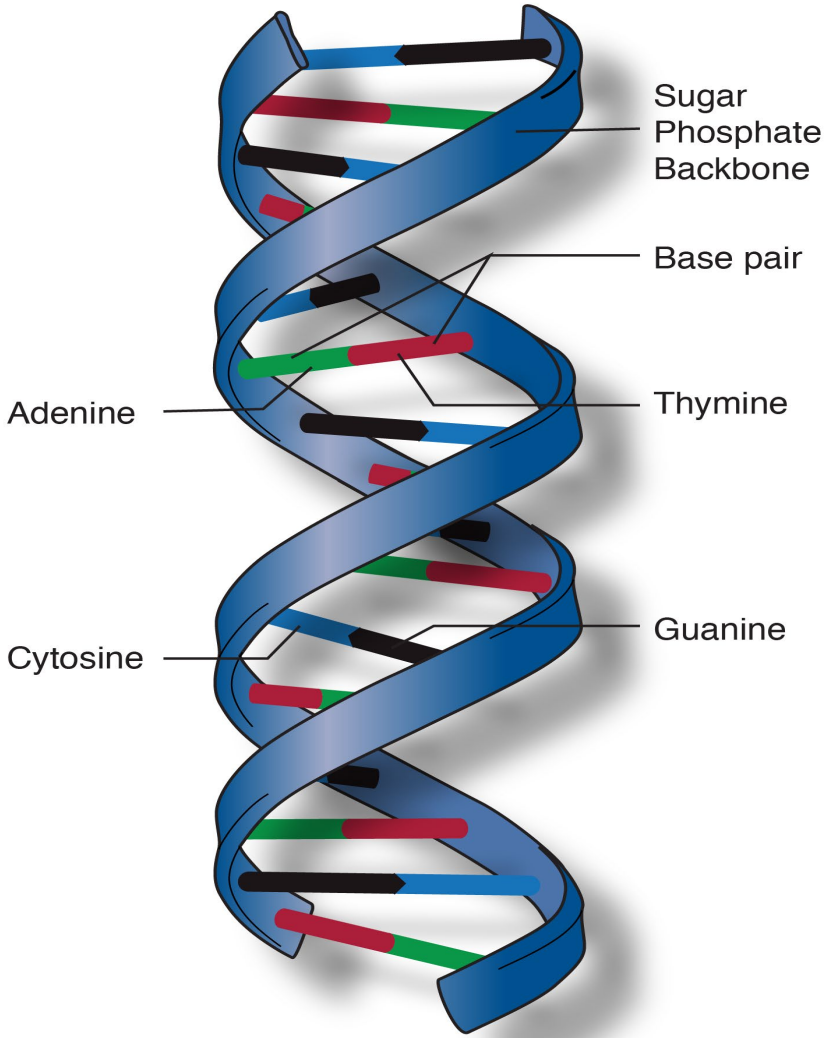


Pro. Fisher was a close friend and collaborator of Prof. P. C. Mahalanobis (1893– 1972), who in 1933, one year after the establishment of the "Indian Statistical Institute" , founded the journal Sankhya, along the lines of Prof. K. Pearson' s *Biometrika*.

The most significant breakthrough in Biology was the announced to the world in 1953 the discovery of the double – helix structure of DNA by Francis Crick (1916-2004) and James Watson (1928 -). Their research took place in Cambridge laboratories and published in the journal Nature.

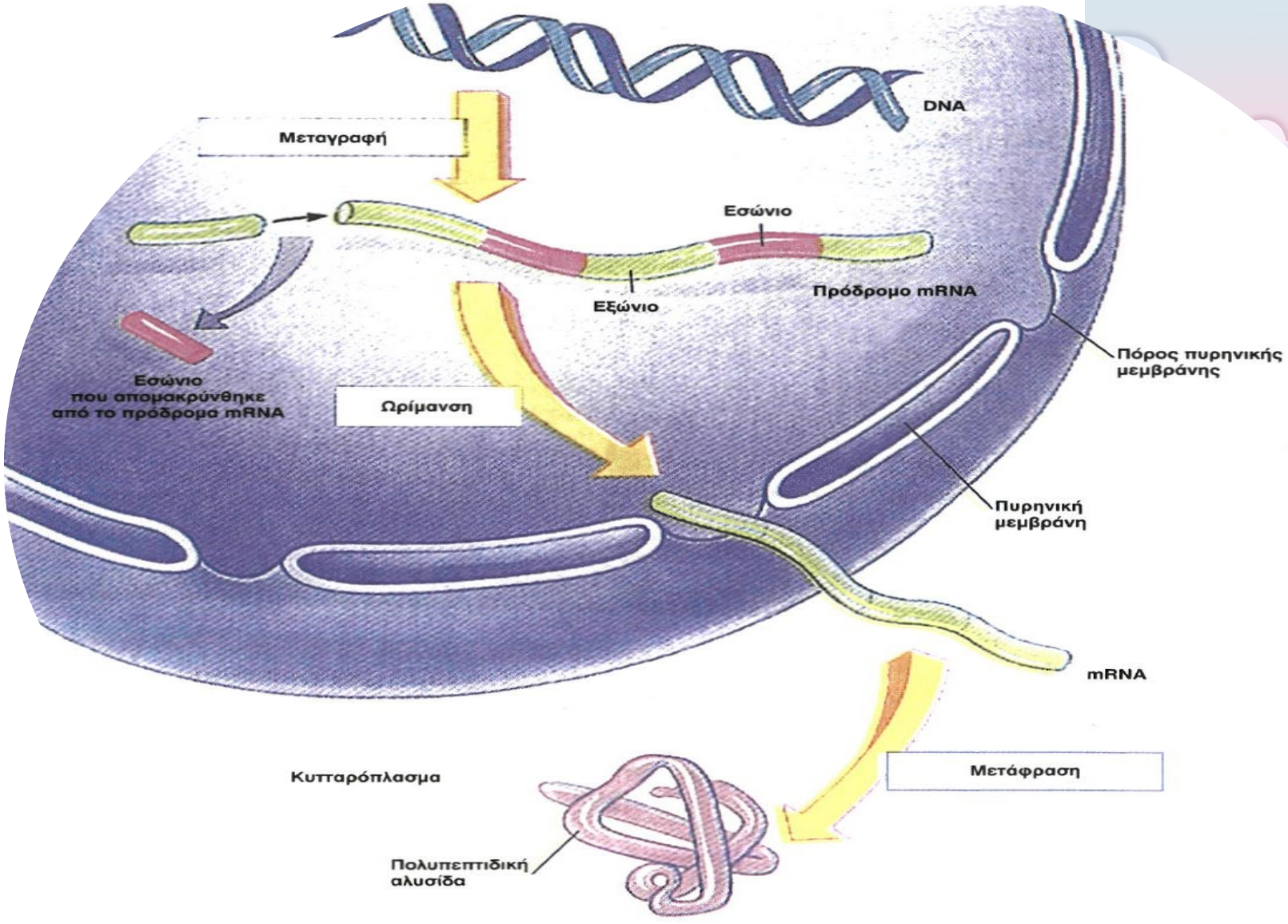
During the same period the researcers Rosalind Franklin (1920-1958) and Maurice Wilkins (1916-2004) were working on relative problems using X rays and crystallography at King's College London. Their results were of important significance for Crick and Watson.

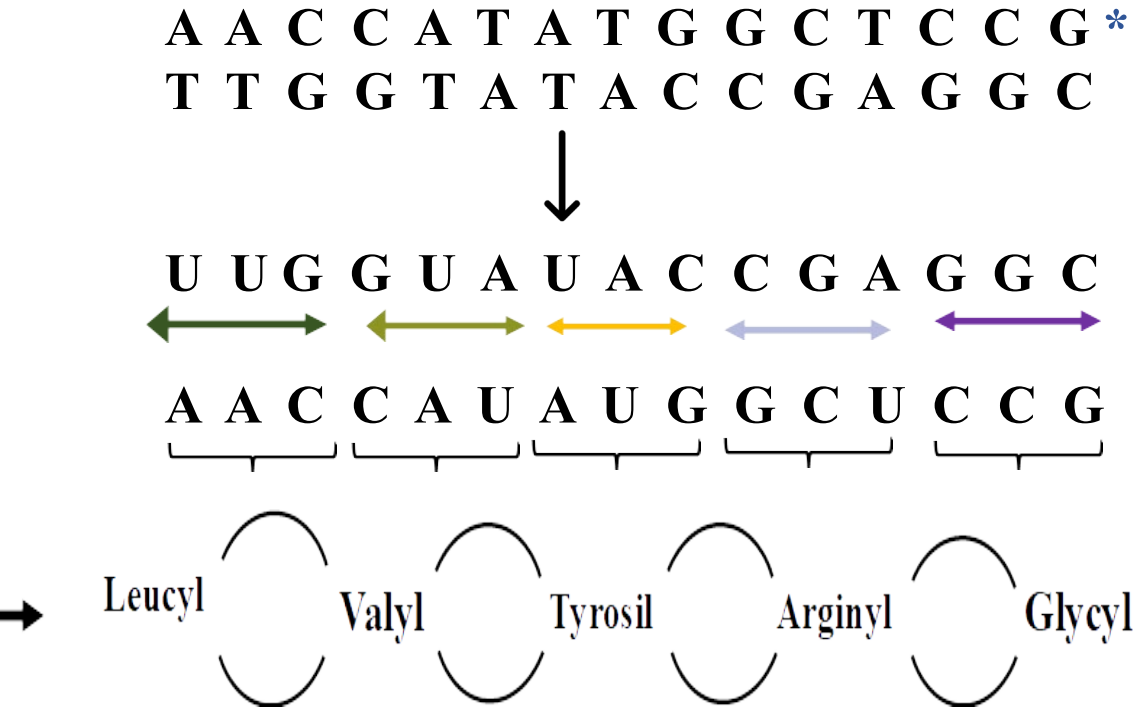
In 1962 Crick, Wathon and Wilkins received Nobel Prize for Physiology or Medicine for their determination of the molecular structure of deoxyribonucleic acid (DNA). R. Franklin had died by cancer (working mostly with X rays). Her contribution honored later and after a lot of scientists and feminist' s reactions.



The two DNA strands are known as polynucleotides as they are composed of simpler monomeric units called nucleotides. Each nucleotide is composed of one of four nitrogen-containing nucleobases (cytosine [C], guanine [G], adenine [A] and thymine [T]), a sugar called deoxyribose, and a phosphate group.

We simply say that the genetic alphabet of DNA is {A, C, G, T}

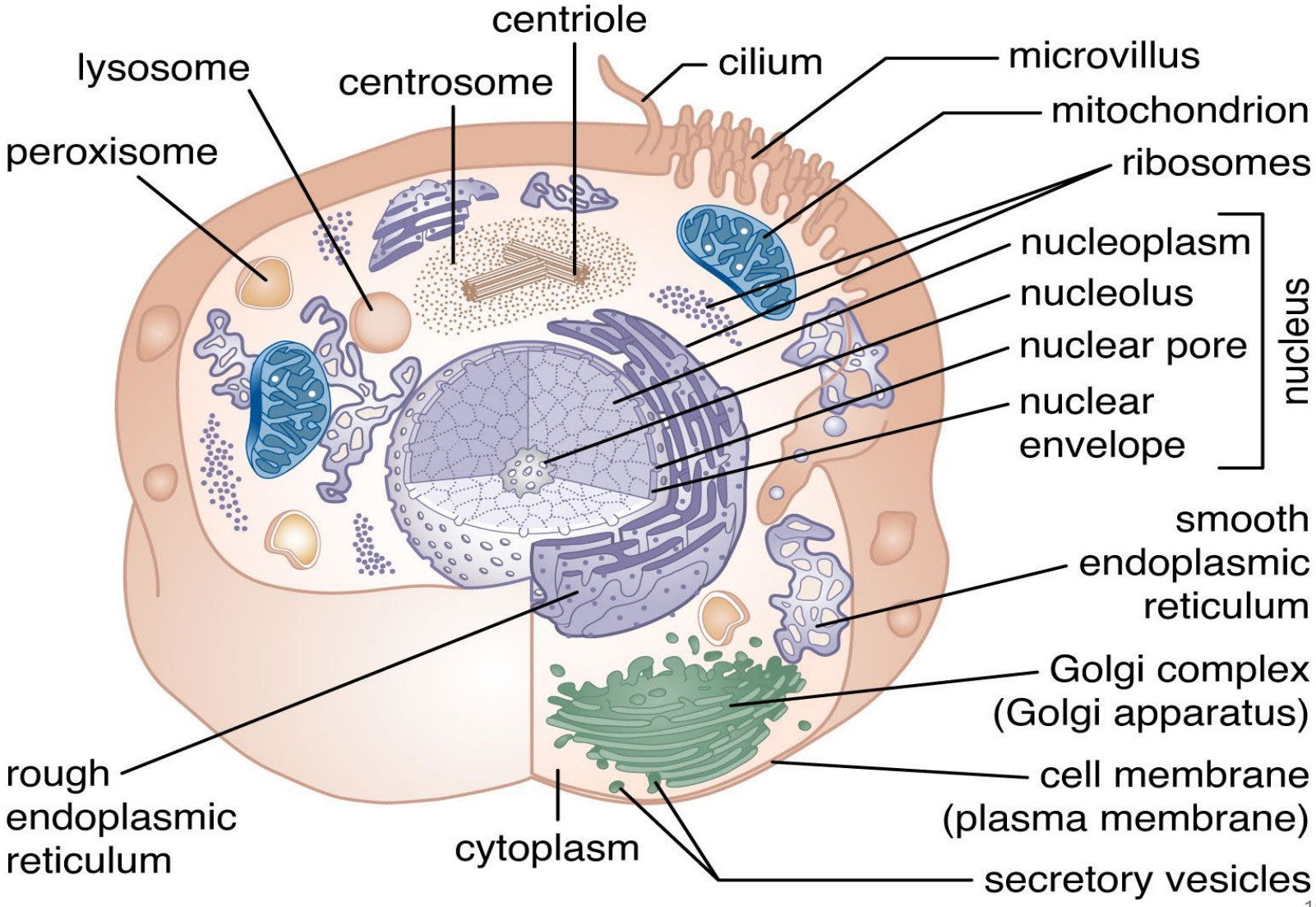

DNA
Μεταγραφή
Εσώνιο
Εξώνιο
Πρόδρομο mRNA
Πόρος πυρηνικής μεμβράνης
Εσώνιο που απομακρύνθηκε από το πρόδρομα mRNA
Ωρίμανση
Πυρηνική μεμβράνη
mRNA
Κυτταρόπλασμα
Μετάφραση
Πολυπεπτιδική αλυσίδα



Polypeptide is a string of amino acids connected together by peptide bonds (The word poly means many, and the word peptide refers to proteins). So, a polypeptide chain is a chain of the building blocks of proteins or amino acids.

The Genetic Code

1st	2nd U	2nd C	2nd A	2nd G	3rd
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	TC	TC	A
	Leu	Ser	TC	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G



An adult has about one quadrillion (10^15) cells !!! And all have the same DNA.

In a few words, DNA makes two important things:

First: It copies itself without stop

Second: It controls the proteins

Cells are not of the same shape or size. It depends on their mission. For example, Egg cell is the biggest and sperm cell is the smallest.

Jacques Monod

In 1965 Jacques Monod (1910-1976) and François Jacob (1920-2013) proved how the genetic information is converted during the formation of proteins by means of a messenger, which proved to the substance we now know as RNA.

In the same year Jacques Monod, Francois Jacob and Andre Lwoff received the Nobel Prize in Physiology or Medicine.

In 1972 J. Monod published a book with the title "Chance and Necessity" in Paris and very soon it was translated to many languages as well as in Greek. The book is an essay on the natural philosophy.

.

J. Monod used as the title of his book the opinion of ancient Greek philosopher Democritus about the world .
Democritus was born about 460BC in Abdera of Thrace (N.G) and he traveled widely in the East.

Moreover, J. Monod explained his thoughts about the probable problems which arise after an important discovery. He believed that scientists must continue their research for finding "Truth" resembling their struggle to that of Sisyphus. Thus, he refers a specific extract of the book of Albert Came (Nobel Prize for Literature 1957) "The myth of Sisyphus".

Professors Fotis Kafatos (1940-2017) and Thanasis Fokas (1952-) two Greek internationally distinguished scientists (Biology, Mathematics, Medicine etc) gave an interesting interview to Tziotzios, a graduate student of Cambridge University, expressing their thoughts about the relationship of Mathematics and Biology.

www.damtp.cam.ac.uk/user/tf227/tziotzios.pdf

Fotis Kafatos emphasized that, because the struggle for 'Truth" is endless, a scientist must be modest.

In 1959 Eugene Paul Wigner (1902-1995) theoretical physicist and mathematician, who was honored with Nobel Prize in Physics of 1963, wrote an article with title:

«The Unreasonable Effectiveness of Mathematics in Natural Sciences».

The article impressed scientists and it has used for many special topics. If one person writes the first two words of the title in a searching machine, he will find a lot of articles concerning specific sciences includind Biology too.

In 2012 Joel E. Cohen (1944-) mathematician and biologist, who works in Rockefeller University in New York and in the Earth Institute of Columbia University, published an article with the ingenious title:

«Mathematics Is Biology's Next Microscope, Only Better; Biology Is Mathematics' Next Physics, Only Better» .

Cohen highlighted that the following topics are necessary to be developed for helping certain problems of Biology.

Understand computation

Find better ways to model multi-level systems

Understand data mining, simultaneous inference and Statistical de identification

Understand probability, risk and uncertainty

The following article is written by Arthur M. Lesk (Ptinceton, Cambridge, one of the founders of the programma «Biocompurting Programme at the European Molecular Biology Laboratory in Heidelberg» and Prof. of Biochemistry of the University Pensylvana State of USA).

“The Unreasonable Effectiveness of Mathematics in Molecular Biology”, Mathematical Intelligencer; Spring 2000, Vol. 22 Issue 2, p28 [9].

According to A. Lesk **there are doubts** for the effectiveness of mathematics in molecular Biology because the experiments on living organism depend on combination of:

Laws of Physics and Chemistry (initial conditions)
The mechanisms of evolution
Historical accidents

In the following we present the references of the above presentation including some relative articles written by members mostly of our department (*with modesty*).

We give an elementary example to understand the basic assumptions.
We consider the DNA alphabet {A, C, G, T} and an arbitrary word or a piece of DNA strand of interest

AAGCGGCAAAGAAG

If we observe a sequence like the following

AAGCGGCAAAGAAGCGGCAAAGAAG

how many appearances of the word shall we count? One or two?

Basic hypotheses concern the appearances of letters:
They **are independent** or have a **Markovian dependence**,
or **Discrete time semi-Markov**. Under specific assumptions we derived limit theorems and certain bounds using Chen-Stein method on total variation distance, as well as combinatorial arguments. Also, the last two decades researchers started to use **Hidden Markov chains .**

In conclusion, we present the branches of Mathematics which play an important role in Biological problems according to the reference [10]:

Statistics and Stochastic Processes
Dynamical Systems Theory
Nonlinear Partial Differential and Functional Equations
Classical Analysis
Topology and Geometry

References

1. www.damtp.cam.ac.uk/user/tf227/tziotzios.pdf_

2. Watson J.D. and Crick F.H.C., «A Structure for Deoxyribose Nucleic Acid», *Nature* **171**, 737-738, 1953.

3. Ζακ Μονό, « Η Τύχη και η Αναγκαιότητα», εκδόσεις Ράππα, 1971.

4. Wigner E. P., «The Unreasonable Effectiveness of Mathematics in the Natural Sciences», Communications on pure and applied Mathematics, Vol. XIII, 001-14, 1960.

5. Joel E. Cohen, «Mathematics is Biology's Next Microscope, Only Better; Biology is Mathematics'Next Physics, Only Better», *PLoS Biology*. **2** (12), 2012.

6. Lothaire, M.(the pseudonym of a group of mathematicians, many of whom were students of Marcel-Paul Schützenberger), Applied Combinatorics on Words, Encyclopedia of Mathematics and its Applications, **105**, Cambridge University Press, 2005.

7. Chen Louis, “Poisson approximation for dependent trials”. Ann. Probab. 3, 534-545, 1975.
8. Charls M. Stein, «A bound for the error in the normal approximation to the distribution of a sum of dependent random variables», Sixth Berkeley Stanford Symposium, pages 583-602, 1972.
9. Arthur M. Lesk, “The Unreasonable Effectiveness of Mathematics in Molecular Biology”, Mathematical Intelligencer; Spring 2000, Vol. 22 Issue 2, p28.
10. Levin S, editor. Mathematics and biology: The interface. Challenges and opportunities. Lawrence Berkeley Laboratory Pub-701. Berkeley (California): University of California; 1992. Available:http://www.bio.vu.nl/nvtb/Contents.html via the Internet. Accessed 20 October 2004.

11. Chryssaphinou, O. , S. Papastavridis. “A limit theorem for the number of non-overlapping occurrences of a pattern in a sequence of independent trials", Journal of Applied Probability 25, 428-431, 1988.
12. Chryssaphinou, O. , S. Papastavridis. “A limit theorem on the number of overlapping appearances of a pattern in a sequence of independent trials" , Probability Theory and Related Fields 79, 129-143, 1988.

13. Chryssaphinou, O. , S. Papastavridis “The occurrence of sequence patterns in repeated dependent experiments", Probability Theory and Its Applications 167-173, 1990.

14. Chryssaphinou, O. , S. Papastavridis, E. Vaggelatou, “On the Number of Appearances of a Word in a Sequence of I.I.D. Trials, Methodology and Computing in Applied Mathematics, 1:3, 329-348, 1999.

15. Chryssaphinou, O. , S. Papastavridis, E. Vaggelatou, Applications of Compound Poisson Approximation“, Probability and Statistical Models with Applications (eds. N. Balakrishnan, M. V. Koutras and C. Charalambides), 41-62, Chapman and Hall/CRC, 2000.

16. Chryssaphinou, O. , S. Papastavridis, E. Vaggelatou, “ Poisson Approximation for the non-overlapping appearances of several words in Markov chains“, Combinatorics, Probability and Computing Comb., 10, 293-308, (2001).

17. A.D. Barbour, O. Chryssaphinou, " Compound Poisson Appr
The Annals of Applied Probability, 2001, Vol. 11, No 3, 964-1002**.**

18. Chryssaphinou, O., M. Karaliopoulou and N. Limnios ,
"On Discrete Time semi-Markov chains and applications in words occurrences",
Communications in Statistics: Theory and Methods, 37,(8): 1306-1322, 2008.

19. Karaliopoulou, M, On the number of word occurrences in a semi-Markov sequence of letters, ESAIM: Probability and Statistics, Tome 13 (2009), pp. 328-342

20. Garcia-Maya, B. I., Karaliopoulou, M., & Limnios, N. (2022). Asymptotic properties of words in semi-Markov sequences. International Journal of Biomathematics. DOI: https://doi.org/10.1142/s1793524522500139

National Museum of Denmark in Copenhagen.

We would like to end presenting The **Seikilos epitaph**, which is an Ancient Greek inscription (200-100 BC), that preserves the oldest surviving complete musical composition, including musical notation. The Epitaph was discovered in 1883 by Sir W. M. Ramsay in Aydin and close to Smyrna (Ionian area).. The poem is:

"ὅσον ζῆς, φαίνου/ μηδὲν ὅλως σὺ λυποῦ/
Πρὸς ὀλίγον ἔστι τὸ ζῆν / τὸ τέλος ὁ χρόνος ἀπαιτεῖ".

"As long as you're alive, shine, don't be sad at all; life is short, time asks for its due".

Thank you
for your presence and attention