
Padova, 18 aprile 2020

DNA SEQUENCE SYMMETRIES FROM MAXIMUM ENTROPY: THE ORIGIN OF THE CHARGAFF'S SECOND PARITY RULE

“The leading force shaping the DNA sequence in the genomes is the entropy and that the major cause of all these symmetries is the randomness”

DNA is the molecule of life on this planet, and most living organisms rely on double-stranded DNA to perpetuate their genetic code. This biological information is the main target of evolution. However, the symmetries found within the genomes can emerge from the physical peculiarities of the double helix DNA molecule itself and the maximum entropy principle alone, rather than from biological or evolutionary pressure.

Professor Piero Fariselli of the University of Turin in collaboration with professors Cristian Taccioli, Luca Pagani and Amos Maritan of the University of Padua (Italy) created a mathematical model able to explain the origin of the so-called "Chargaff's second parity rule". Furthermore, their new theory, named Generalized Chargaff Theory, formulates precise genomic predictions that have been validated on bacteria, archaea and mammals, including humans. The study is published in the prestigious journal "Briefings in Bioinformatics" in the volume of April 2020.

Chargaff's second parity rule has been a mystery for more than 50 years. In 1968 the Austrian biochemist Erwin Chargaff discovered that in the single strand of a double-stranded DNA molecule, the adenines number was almost identical to the number of thymines and, in the same way, the number of cytosines was almost equal to the number of guanines. To date, there was no explanation for such symmetry. In recent years, many research groups have tried to clarify the reason for this biological oddity, but nobody has ever been able to obtain valid predictions.



Prima Regola di Chargaff:

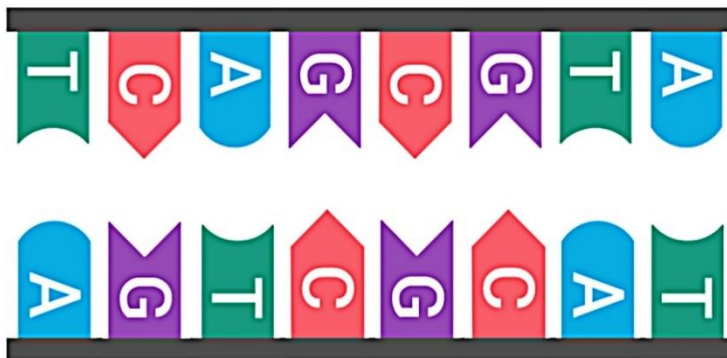
Considerando la coppia di filamenti

Numero A = Numero T

Numero C = Numero G

Validità: Sempre

Motivo: Accoppiamento selettivo delle basi della doppia elica



Seconda Regola di Chargaff:

Per singolo filamento

Numero A = Numero T

Numero C = Numero G

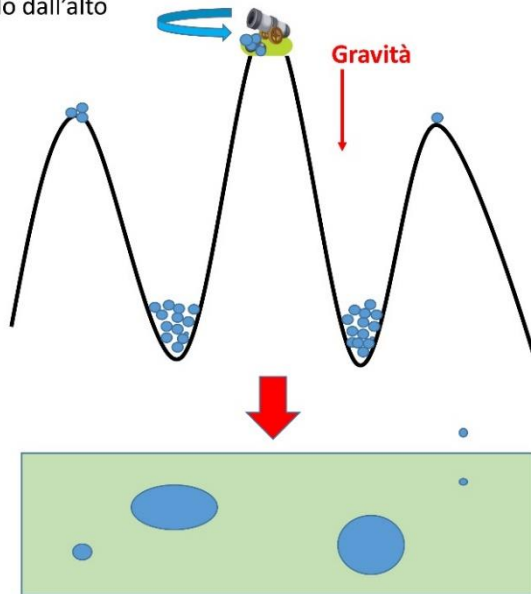
Validità: approssimata, tanto più precisa quanto il DNA è lungo

Motivo: Dimostrato che dato il vincolo della doppia elica è statisticamente la distribuzione più probabile.

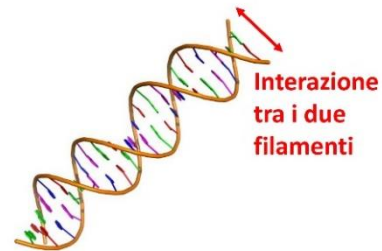
«By means of a multidisciplinary collaboration that includes biologists, anthropologists and theoretical physicists, we have succeeded in explaining the origin of Chargaff's second parity rule, through the hypothesis that this genomic symmetry emerges from the DNA double helix constraint and the principle of maximum randomness (entropy) – **says Cristian Taccioli, Professor of molecular biology and bioinformatics at the M.A.P.S Department of the University of Padova** -. Moreover, we described the evolution of the genome in terms of energy rather than focusing only on natural selection. Our idea is that DNA, like all the other systems in the universe, follows the stream of entropy, trying to reach equilibrium and maximizing its stability. Our results show that processes that increase the entropy of a double-stranded DNA molecule are favoured during the evolution of living organisms. »

«We hypothesize that the leading force shaping the DNA sequence in the genomes is the entropy and that the major cause of all these symmetries is the randomness. However, randomness does not imply uniformity and equality. As an example, the 'random' process of blindly throwing stones in a rugged landscape generates a higher probability of finding the stones in the valleys than on top of the hills.»

Lancio casuale di palle di cannone in tutte le direzioni con l'effetto della forza di gravità e la superficie montagnosa genera distribuzioni inaspettate guardando dall'alto



L'interazione tra i due filamenti di DNA può generare simmetrie all'interno delle sequenze che possono essere dovute semplicemente alla massima casualità durante il processo evolutivo.



Numero **A** = Numero **T**
 Numero **C** = Numero **G**
 ...
 Numero **ACGTT...** = Numero **AACGT...**
 ...

«We introduced a new paradigm stating that the free energy of DNA is the primary target of evolutionary forces that shape the structure of DNA. Metaphorically, the genome is often defined as a book, in which the ink represents the biological information it encodes. With our work, we also include the paper (entropy and physical characteristics of the double helix) of which the book is made – **say Luca Pagani and Amos Maritan professors of the Departments of Physics and Biology of the University of Padova** -. The exceptions to this trend could also offer further opportunities to measure the energy content of biological information embedded in DNA sequences modeled by natural selection.»

This work might also be crucial in the next future for its application in biotechnology. Deciphering the mechanisms that favor the long-term survival of one random DNA sequence over another at energetic level, can provide crucial insights into genome structure, and the design of synthetic DNA constructs.

Link: <https://doi.org/10.1093/bib/bbaa041>