

An indirect evidence of the Fibonacci Strings model for DNA Repetitive Sequences growth

Michel E. Beleza Yamagishi

Laboratório de Bioinformática Aplicada
Av. André Tosello, 209 Campinas – SP 13083-970

Paula R. Kuser Falcão

Laboratório de Bioinformática Aplicada
Av. André Tosello, 209 Campinas – SP 13083-970

Fábio Danilo Vieira

Laboratório de Bioinformática Aplicada
Av. André Tosello, 209 Campinas – SP 13083-970

Roberto Hirochi Herai

Falcudades Network
Av. Ampélio Gazeta s/n Nova Odessa – SP

Abstract

Almost half of the mammalian genome is composed of repetitive elements. Repetitive elements, or simply repeats, can be subdivided into those that are tandemly arrayed and those that are interspersed. Examples of the first class are microsatellites, minisatellites and telomeres; examples of the second class are the transposable elements or transposons. Such considerable amount of repeats should play some role in the living organisms. Actually, the influence of transposons present in the human germ line on gene expression can be envisaged by the fact that roughly one quarter of all analyzed human promoters regions harbor sequences derived from these elements. Nevertheless, the most part of the repetitive elements lacks any recognizable function, and seems to be part of the famous “junk” DNA. It is interesting to notice that in terms of those eukaryotes that have their genomes completed there is an approximately linear correlation between genome size and the total number of DNA repetitive elements present, although the contribution is more significant in larger genomes. It is possible to propose mathematical models to the tandem repeats sequence growth. Possibly the simplest is based on Fibonacci strings. In this model, the words are considered over the genetic alphabet and start with two arbitrary strings. These strings or fragments of them may be excised and become candidates for replication and concatenation, and this growth process can go on iteratively. Of course, some error may be introduced in the concatenation step what turns the problem of recognition of this process really difficult. However, if this process really takes place, then some kind of signature should be left behind. In this work, we present some results that are indirect evidences of Fibonacci String model for DNA repetitive sequences growth.

References

- [1] Yamagishi, M. E. B. & Shimabukuro, A. I., Nucleotide Frequencies in Human Genome and Fibonacci Numbers (2007), Bulletin of Mathematical Biology, in press.
- [2] Bannert, N. and Kurth, R. (2004); Retroelements and the human genome: new perspectives on an old relation, Proceedings of the National Academy of Sciences.101:14572-79
- [3] Chargaff, E. (1951);Structure and function of nucleic acids as cell constituents, Fed. Proc., 10, 654-659.
- [4] Dress, A., Giegerich, R., Grunewald, S. and Wagner, H. (2003); Fibonacci-Cayley Numbers and Repetition Patterns in Genomic DNA, Annals of Combinatorics,7, 259-279
- [5] Do, J. H. and Choi D-K (2006); Computational Approaches to Gene Prediction, Journal of Microbiology, 44, 137-144.
- [6] Fibonacci, L. and Singler, L. E (Translator) (2002);Fibonacci's Liber Abaci,New York, Ed. Springer-Verlag.
- [7] Forsdyke, D. R. and Bell, S. J. (2004);A discussion of the application of elementary principles to early chemical observations, Applied Bioinformatics,3,3-8.
- [8] Gregory, T. R. (2005);The C-value enigma in plants and animals: a review of parallels and an appeal for partnership, Annals of Botany, 95: 133-146
- [9] Jordan, I. K., Rogozin, I. B., Glazko, G. V. and Koonin, E. V. (2003);Origin of a substantial fraction of human regulatory sequences from transposable elements, Trends Genet, 19, 68-72
- [10] Mitchell, D. and Bridge, R. (2006);A test of Chargaff's second rule, BBRC, 340, 90-94.
- [11] Nocedal, J. and Wright, S. J. (2000);Numerical Optimization, Springer Series in Operations Research, New York, Ed. Springer-Verlag.
- [12] Salzberg, S. L. and Yorke J. A. (2005);Beware of misassembled genomes, Bioinformatics, 21: 4320-21
- [13] Watson, J. D. and Crick, F. H. C. (1953); Molecular Structure of Nucleic Acids, Nature, 4356, 737.