

Author's Accepted Manuscript

Circular code motifs in genomes of eukaryotes

Karim El Soufi, Christian J. Michel



PII: S0022-5193(16)30205-3
DOI: <http://dx.doi.org/10.1016/j.jtbi.2016.07.022>
Reference: YJTBI8747

To appear in: *Journal of Theoretical Biology*

Received date: 26 February 2016
Revised date: 16 June 2016
Accepted date: 16 July 2016

Cite this article as: Karim El Soufi and Christian J. Michel, Circular code motif in genomes of eukaryotes, *Journal of Theoretical Biology* <http://dx.doi.org/10.1016/j.jtbi.2016.07.022>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Circular code motifs in genomes of eukaryotes

Karim El Soufi, Christian J. Michel*

Theoretical bioinformatics

ICube

University of Strasbourg, CNRS

300 Boulevard Sébastien Brant

67400 Illkirch, France

Email: kelsoufi@unistra.fr, c.michel@unistra.fr

* Corresponding author

Keywords: circular code motifs, circular code, bijective transformation circular code, permuted circular code, genomes of eukaryotes.

Abstract

A set X of 20 trinucleotides was identified in genes of bacteria, eukaryotes, plasmids and viruses, which has in average the highest occurrence in reading frame compared to its two shifted frames (Michel, 2015; Arquès and Michel, 1996). This set X has an interesting mathematical property as X is a circular code (Arquès and Michel, 1996). Thus, the motifs from this circular code X , called X motifs, have the property to always retrieve, synchronize and maintain the reading frame in genes. In this paper, we develop several statistical analyses of X motifs in 138 available complete genomes of eukaryotes in which genes as well as non-gene regions are examined. Large X motifs (with lengths of at least 15 consecutive trinucleotides of X and compositions of at least 10 different trinucleotides of X among 20) have the highest occurrence in genomes of eukaryotes compared to its 23 large bijective motifs, its two large permuted motifs and large random motifs. The largest X motifs identified in eukaryotic genomes are presented, e.g. an X motif in a non-gene region of the genome *Solanum pennellii* with a length of 155 trinucleotides (465 nucleotides) and an expectation $\mathbb{E} = 10^{-71}$. In the human genome, the largest X motif occurs in a non-gene region of the chromosome 13 with a length of 36 trinucleotides and an expectation $\mathbb{E} = 10^{-11}$. X motifs in non-gene regions of genomes could be evolutionary relics of primitive genes using the circular code for translation. However, the proportion of X motifs (with lengths of at least 10 consecutive trinucleotides of X and compositions of at least 5 different trinucleotides of X among 20) in genes/non-genes of the 138 complete eukaryotic genomes is about 8. Thus, the X motifs occur preferentially in genes, as expected from the previous works of 20 years.

Download English Version:

<https://daneshyari.com/en/article/6368915>

Download Persian Version:

<https://daneshyari.com/article/6368915>

[Daneshyari.com](https://daneshyari.com)