A novel indicator of GC content reflects the evolution of primate species

Marco V. José[1,2,*], Qi Lu[2,], Juan R. Bobadilla[1,2]

[1]*Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México, Apartado Postal 70228, Ciudad Universitaria, CP 04510 México DF, México*
[2]*Centro Internacional de Ciencias A. C, Campus UNAM-UAEM, Col. Chamilpa, CP 62210, Cuernavaca, Morelos, México.*

Abstract

A novel statistics for analyzing the GC content of a given chromosome, which represents the local average distance of GC dinucleotides ($LADGC$), is presented. We found that this statistics is log-linearly inversely related, for a broad range of scales, to the GC content in humans, chimpanzee and rhesus macaque. We also determined the distribution of window sizes for which a given count of GC duplets is found along an entire chromosome. By doing so, we were able to visualize how the GC duplet is distributed along a given chromosome. Interestingly, these distributions followed a quasi-log normal distribution for all chromosomes in the three species and they were fitted to a Fokker-Planck equation. The main difference among these distributions for any given chromosome was that the expected values of the mean distances of GC dinucleotides are slightly higher in humans than in chimpanzees and rhesus macaques. In contrast, the variance has remained approximately constant in the three species. The skewness and kurtosis are slightly different among primate chromosomes. In order to test the direction of evolution, we also randomized the whole nucleotide sequences of all chromosomes and found that these control distributions lie to the right of the actual distributions with a certain degree of overlapping. These observations are consistent with recent studies in which it is found that isochores are vanishing. The reason behind this overlapping of the control distributions is due to the fact that the ratio of GC over the size of each chromosome is practically constant for each species and throughout each of them. One result that stands out is the overlapping of the GC distribution of several primate chromosomes with their controls, indicating that the isochore structure is very difficult to detect. Therefore, these distributions capture simultaneously GC-poor, GC-intermediate, and GC-rich regions.

*Corresponding author:
E-mail address: marcojose@biomedicas.unam.mx;