ALAIN ARNEODO, Université de Lyon, Laboratoire Joliot Curie and Laboratoire de Physique, Ecole Normale Supérieure de Lyon, France

Low frequency rhythms in human DNA sequences: from genome-wide sequence analysis to the detection of replication origins in higher eukaryotes

We explore the large-scale compositional heterogeneity of human autosomal chromosomes through the optics of the wavelet transform (WT) microscope. We show that the GC content displays relaxational nonlinear oscillations with two main frequencies corresponding to 100 kb and 400 kb which are well recognized characteristic sizes of chromatin loops and loop domains involved in the hierarchical folding of the chromatin fiber. These frequencies are also remarkably similar to the size of mammalian replicons. When further investigating deviations from intrastrand equimolarities between A and T and between G and C, we corroborate the existence of these two fundamental frequencies as the footprints of the replication and/or transcription mutation bias and we show that the observed nonlinear oscillations enlighten a remarkable cooperative organization of gene location and orientation. When further investigating the intergenic and transcribed regions flanking experimentally identified human replication origins and the corresponding mouse and dog homologous regions, we reveal that for 7 of 9 of these known origins, the (TA + GC) skew displays rather sharp upward jumps, with a linear decreasing profile in between two successive jumps. We present a model of replication with well positioned replication origins and random terminations that accounts for the observed characteristic serrated skew profiles. We further use the singularity tracking ability of the WT to develop a methodology to detect the origins of replication. We report the discovery of 1024 putative origins of replications in the human genome. The statistical analysis of the distribution of sense and anti-sense genes around these origins strongly suggests that the origins of replication play a fundamental role in the organization of mammmalian genomes. Taken together, these analyses show that replication and gene expression are likely to be regulated by the structure and dynamics of the chromatin fiber.