Poster H-32

Characterization of homogeneous and heterogeneous regions in the human genome and compositional features of individual chromosomes

Authors:
Victor Sabbia (*Laboratorio de Organización y Evolución del Genoma*)
Rosina Piovani (*Laboratorio de Organización y Evolución del Genoma*)
Hugo Naya (*Insituto Pasteur Montevideo*)
Hector Musto (*Laboratorio de Organización y Evolución del Genoma*)

Short Abstract: In the present work we A) describe the homogeneous vs heterogeneous categorization of long fragments along the human genome and the uneven distribution of genes between them. B) We show the results of a study of compositional correlations of individual chromosomes, their genes, introns and surrounding isochores.

Long Abstract:
The human genome is compositionally heterogeneous, in other words, high-molecular weight GC-rich and GC-poor regions are interspersed along the chromosomes. This organization was discovered in the bovine genome more than 30 years ago by using density gradient centrifugation. Further research with the same approach showed that it is characteristic of all vertebrates, and was confirmed by the analysis of genome sequences, mainly from human and mouse. What could be considered as the “traditional” view, developed mainly by Bernadi and co-workers states that “vertebrate genomes are mosaics of isochores – namely, of long, compositionally homogeneous DNA segments that can be subdivided into a small number of families characterized by different GC levels”. The isochore length was proposed to be > 300 kb and the homogeneity in base composition above 3 Kb. In the case of the human genome, which shows typical features for the genomes of most mammals, exist two “light” (GC poor) isochore families called L1 and L2, which represents 62% of the genome, and three “heavy” (GC rich) families called H1, H2 and H3, which amount to 22, 9 and 3-4% of the genome, respectively. Interestingly, GC-rich isochores contains 20 times more genes than GC-poor ones.

Although throughout the past decades lot of work has been done in terms of compositional analysis of genomes as a whole, there is still missing a comprehensive study at the chromosome level, especially in the light of actual gene discovery data.

In the present work we will show two different results.
First we shall introduce our own algorithm (derived from the originally described by Nekrutenko and Lee) aimed to categorize the human genome into compositionally homogeneous and heterogeneous regions. We show 1) the magnitude of the influence of the parameters used for obtaining the two kind of regions, 2) the change of the ratio homogeneous/heterogeneous caused by small changes in the parameters and finally we chose a set of parameters based on the same authors which classifies the 40% of the genome as homogeneous. Using this parameter set we show 3) the pattern of distribution of

the different regions, and 4) the strong unevenness of the distribution of genes according to the heterogeneity level since 80% of the genes are located in the 60% heterogeneous regions. 5) Finally, we propose a “homogeneous-heterogeneous” banding of each chromosome.

Second, we have made an analysis of the compositional properties of genes, their introns and their landscape isochores for the whole human genome and particularly for each chromosome. We investigated the correlations among the genes and with the whole chromosome data for more than non redundant 14,000 human genes whose sequence and chromosome localization have been confirmed. Unexpectedly, we found that several chromosomes display patterns that are clearly different from the whole genome. Finally, we confirm that the orientation of the coding sequences along each chromosome is not random. We postulate that these features could be related with the level of transcription of each region and with the position of each chromosome within the nucleus.