

Symmetry properties of pentamer usage in non-coding DNA - (session: Other)

Emanuele Bultrini, Elisabetta Pizzi, Paolo Del Giudice, Clara Frontali

Istituto Superiore di Sanità, Roma

A set of pentanucleotides characterising non-coding regions of a specific genome can be extracted from introns using Principal Component Analysis; this set of words discriminates between introns, their randomised counterparts and exons [1]. The procedure was applied to sequences from different species as *C. elegans* and *D. melanogaster*. A genome-wide analysis revealed that the introns' vocabulary usage is typical also of many intergenic portions, constituting a sort of "background" of each genome, and makes it possible to segment intergenic sequences into intron-like and non-intron-like regions, the latter of which possibly contain functional elements.

The vocabulary is consistently characterised by a symmetry property: it is almost entirely composed of reverse complementary oligos, and the level of symmetry for intron-like sequences is greater than what would be expected from Chargaff's second parity rule; in fact symmetry of DNA sequences is usually observed on large scale, mainly because of statistical reasons [2], while on smaller scale very different results are obtained for introns, that show much higher levels of symmetry than exons. It has been shown that randomised sequences are more symmetrical than the real ones, when a suited symmetry measure is adopted [3], as we also checked in our data sets. However, we show in the present work that intron and intron-like sequences have high levels of symmetry that don't arise equally from all pairs of reverse-complementary pentamers, but are mainly due to the small set of vocabulary words. SOLO POSTER

- [1] Bultrini, E., Pizzi, E., Del Giudice, P., Frontali, C., 2003. Pentamer vocabularies characterizing intron and intron-like intergenic tracts from *Caenorhabditis elegans* and *Drosophila melanogaster*. *Gene* 304, 183-192.
- [2] Prabhu, V. V., 1993. Symmetry observations in long nucleotide sequences. *Nucleic Acids Res.* 21, 2797-2800.
- [3] Baisnée, P.-F., Hampson, S., Baldi, P., 2002. Why are complementary DNA strands symmetric? *Bioinformatics* 18,1021-1033.