Nonextensive Analysis of Human DNA

Thomas Oikonomou

June 27, 2006

We apply nonextensive statistics to four human chromosomes (chr.: 6, 16, 17, 19), to study possible differences in the statistics of their coding and noncoding segments. From the Genbank data base all coding and noncoding segments of the above mentioned chromosomes are obtained and their size distributions are calculated. For all distributions (coding and noncoding) we use two exponents r, q to describe the short and long scale behavior respectively. For the coding size distributions the exponents vary as $5.2 \leq r \leq 5.7$ and $1.45 \leq q \leq 1.50$ while for the noncoding size distributions they vary as $2.4 \leq r \leq 3.2$ and $1.50 \leq q \leq 1.72$. There is thus an obvious discrepancy in the statistics governing the coding and noncoding size distributions which reflects their different functional roles in the genome.