

SEARCH OF PHASE SHIFTS OF DIFFERENT PERIOD LENGTHS IN THE GENOMES OF C.ELEGANS, D.MELANOGASTER AND S.CEREVISIAE

Pugacheva V.M., Frenkel F.E., Korotkov E.V.

Prospekt 60-letiya Oktyabrya d.7 k.1, Moscow, Russian Federation 117312

We describe a new mathematical method for finding very diverged short tandem repeats containing a single indel. The method involves comparison of two frequency matrices: a first matrix for a subsequence before shift and a second one for a subsequence after it. A measure of comparison is based on matrix similarity. The approach developed was applied to analysis of the genomes of *C.elegans*, *D.melanogaster* and *S.cerevisiae*. They were investigated regarding the presence of tandem repeats having repeat length equal to 2 and 4-11 nucleotides. A number of phase shift regions for these genomes was approximately 2.2×10^4 , 1.5×10^4 and 1.7×10^2 , respectively. Type I error was less than 5%. The mean length of fuzzy periodicity and phase shift regions was about 220 nucleotides.

The regions of fuzzy periodicity having single insertion or deletion occupy substantial parts of the genomes: 5%, 3% and 0.3%, respectively. Only less than 10% of these regions have been detected previously. That is, the number of such regions in the genomes of *C.elegans*, *D.melanogaster* and *S.cerevisiae* is dramatically higher than it has been revealed by any known methods. We suppose that some found regions of fuzzy periodicity could be the regions for protein binding.