Abstract

Tandem repeats occur frequently in eukaryotic and prokaryotic genomic sequences. They are associated with several inherited human diseases, DNA fingerprinting, evolution and regulatory processes. In spite of their importance, detection of tandem repeats is still not resolved in the sense that the current existing detection tools do not give the same results for a given input sequence. This is mainly due to the differences in the methods adopted by the search algorithms and the different parameter settings needed when they are executed. This paper proposes an efficient method to identify all exact and approximate tandem repeats within a given DNA sequence and also identifies the presence of any changes brought about by mutation. The method first identifies all potential tandem repeats by clustering using K-means method, followed by biclustering to filter out the actual repeats along with the position of occurrence of approximate tandem repeats. The results obtained by this method are consistent with that of existing methods.

References


Index Terms

Computer Science Biomedical

Keywords

Tandem Repeat, DNA Sequence, Micro Satellites, Mini Satellites, Clustering