Abstract

Accurate classification of organisms into taxonomical hierarchies based on genomic sequences is currently an open challenge, because majority of the traditional techniques have been found wanting. In this study, we employed mitochondrial DNA (mtDNA) genomic sequences and Digital Signal Processing (DSP) for accurate classification of Eukaryotic organisms. The mtDNA sequences of the selected organisms were first encoded using three popular genomic numerical representation methods in the literature, which are Atomic Number (AN), Molecular Mass (MM) and Electron-Ion Interaction Pseudopotential (EIIP). The numerically encoded sequences were further processed with a DSP based cepstral analysis to obtain three sets of Genomic Cepstral Coefficients (GCC), which serve as the genomic descriptors in this study. The three genomic descriptors are named AN-GCC, MM-GCC and EIIP-GCC. The experimental results using the genomic descriptors, backpropagation and radial basis function neural networks gave better classification accuracies than a comparable descriptor in the literature. The results further show that the accuracy of the proposed genomic descriptors in this study are not dependent on the numerical encoding methods.

Keywords

Atomic number Cepstral EIIP Genomic Molecular mass mtDNA