

Unsupervised Classification of DNA Barcodes Species Using Multi-Library Wavelet Networks

Authors : Abdesselem Dakhli, Wajdi Bellil, Chokri Ben Amar

Abstract : DNA Barcode, a short mitochondrial DNA fragment, made up of three subunits; a phosphate group, sugar and nucleic bases (A, T, C, and G). They provide good sources of information needed to classify living species. Such intuition has been confirmed by many experimental results. Species classification with DNA Barcode sequences has been studied by several researchers. The classification problem assigns unknown species to known ones by analyzing their Barcode. This task has to be supported with reliable methods and algorithms. To analyze species regions or entire genomes, it becomes necessary to use similarity sequence methods. A large set of sequences can be simultaneously compared using Multiple Sequence Alignment which is known to be NP-complete. To make this type of analysis feasible, heuristics, like progressive alignment, have been developed. Another tool for similarity search against a database of sequences is BLAST, which outputs shorter regions of high similarity between a query sequence and matched sequences in the database. However, all these methods are still computationally very expensive and require significant computational infrastructure. Our goal is to build predictive models that are highly accurate and interpretable. This method permits to avoid the complex problem of form and structure in different classes of organisms. On empirical data and their classification performances are compared with other methods. Our system consists of three phases. The first is called transformation, which is composed of three steps; Electron-Ion Interaction Pseudopotential (EIIP) for the codification of DNA Barcodes, Fourier Transform and Power Spectrum Signal Processing. The second is called approximation, which is empowered by the use of Multi Library Wavelet Neural Networks (MLWNN). The third is called the classification of DNA Barcodes, which is realized by applying the algorithm of hierarchical classification.

Keywords : DNA barcode, electron-ion interaction pseudopotential, Multi Library Wavelet Neural Networks (MLWNN)

Conference Title : ICCIT 2015 : International Conference on Computer and Information Technology

Conference Location : Paris, France

Conference Dates : April 27-28, 2015