Fuzzy logic as a tool for taxonomic identification of protozoa from the Kinetoplastea class

MARIA AUGUSTA DARIO[1], ANTONIO REVAIL ALVES PEREIRA[2], WANESSA DE SOUSA MEDEIROS[1,3], FLÁVIO LUIS DE MELLO[4], MARCELLO GOULART TEIXEIRA[3], ANA MARIA JANSEN[1], MÁRCIO GALVÃO PAVAN[5], SAMANTA CRISTINA DAS CHAGAS XAVIER[1,3]

[1]LABORATÓRIO DE BIOLOGIA DE TRIPANOSOMATÍDEOS, INSTITUTO OSWALDO CRUZ, FIOCRUZ, RIO DE JANEIRO, RJ, BRASIL.
[2]INSTITUTO DE COMPUTAÇÃO, UNIVERSIDADE FEDERAL DO RIO DE JANEIRO, RJ, BRASIL.
[3]PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA COMPUTACIONAL E SISTEMAS, INSTITUTO OSWALDO CRUZ, FIOCRUZ, RIO DE JANEIRO, RJ, BRASIL
[4]DEPARTAMENTO DE ENGENHARIA ELETRÔNICA E COMPUTAÇÃO, UNIVERSIDADE FEDERAL DO RIO DE JANEIRO, RJ, BRASIL.
[5]LABORATÓRIO DE MOSQUITOS TRANSMISSORES DE HEMATOZOÁRIOS, INSTITUTO OSWALDO CRUZ, FIOCRUZ, RIO DE JANEIRO, RJ, BRASIL.

Kinetoplastea class exhibits remarkable diversity, including free-living protozoa and parasites capable of infecting vertebrates, invertebrates, and plants. The organisms' taxonomic classification is currently being extensively revised due to new features obtained from DNA sequencing and phylogenetic studies. Parasitic species are in constant evolution, resulting in individuals from closely related species having in common different proportions of genetic material. In this study a mixed clustering approach was adopted in that Fuzzy C-Means, which captures uncertainties with greater flexibility compared to traditional methods, was combined with Hierarchical Clustering, which provides a hierarchical view of data relationships to classify kinetoplastids. Unaligned kinetoplastid sequences were obtained through next-generation sequencing (18S rDNA gene, Illumina platform) from different hosts and Brazilian biomes to form clusters, and then, validated the groups based on total inertia to classify Kinetoplastea species. Sequences from species such as *Trypanosoma cruzi* (DTUs TcI, TcII, TcIV, and Tcbat), *T. cruzi marinkellei*, *T. dionisii*, *T. rangeli*, *T. janseni*, *T. sp.* Neobat, *T. lewisi*, *T. cascavelli*, *T. minasense*, *Crithidia mellificae*, *Leishmania infantum*, *Neobodo*, *Parabodo*, and representatives of the Bodonidae family were analyzed. The results suggested 4 to 6 clusters. *T. cascavelli*, *T. cruzi*, and *T. rangeli* formed unique groups, with *T. cascavelli* being the most homogeneous. Free-living protozoa formed a separated group only as the number of clusters increased. Other species exhibited heterogeneous classifications, reinforcing the potential of fuzzy logic in evolutionary analysis. The combination of Fuzzy C-Means and Hierarchical Clustering to analyze Kinetoplastea revealed homogeneity in *T. cascavelli* and heterogeneity in free-living species. Fuzzy logic proved effective in capturing genetic uncertainties, advancing the taxonomic and evolutionary understanding of these organisms.

Supported by: Fiocruz, Ufrj, CAPES, Faperj
Keywords: Kinetoplastid, taxonomy, fuzzy logic