ExTaxsI: Exploration Taxonomies Information, a tool to build and visualize amplicon database

Chahed A.[1] (lead presenter) a.chahed@campus.unimib.it

Agostinetto G.[1], Di Filippo M.[2], Pescini D.[2], Sandionigi A.[1], Casiraghi M.[1]

[1] Department of Biotechnology and Biosciences, University of Milano-Bicocca, Milan (Italy)

[2] Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Milan (Italy)

Keywords: Metabarcoding, sequence analysis, amplicon database, HTS sequencing

Species' distributions, community composition and the presence of invasive or endangered species, are extremely important information to investigate and understand the ecosystem of interest. This made environmental DNA (eDNA) metabarcoding to spread in many studies, since it is a noninvasive, powerful approach which resolved many challenges of traditional survey methods.

Metabarcoding, as a wide-ranging method, is widely used in various studies (e.g. Comtet, Sandionigi, Viard, & Casiraghi, 2015), emphasising the necessity of reliable results. This approach provides huge amounts of data, due to sequencing being fast and cheap, providing the opportunity to measure and track biodiversity through time and space (e.g. Janzen et al. 2005; Waugh 2007; Borisenko et al. 2008).

However, this massive amount of data has to be processed, analyzed and, through e.g. images or graphs, visualized, which is why the bioinformatic role has become more crucial. Metabarcoding data needs computational pipelines and IT infrastructure to grow over time to match the increasingly data, particularly advanced data management capabilities, that play a key role.

So, to help scientists with and without a strong bioinformatics background, we created ExTaxsI, a tool for taxonomy and data information analysis.

This tool, write totally in Python, presents a variety of functions, like doing queries and getting information from NCBI to create a Taxonomy database or create a world map where are marked each location of the searched records. It is organized in modules which are connected to each other, giving to our tool the flexibility and the possibility to support scientists in any step of taxonomy analysis of DNA sequencing, creation of a customized database and checking the integrity of databases you created from our tool or other sources. Moreover, it also gives the possibilities to visualize assignments considering all principals taxonomic ranks or their coordinates through multiple types of charts making easier to understand more in details your information, even with massive amount of data.