

ITSoneDB: a specialized ITS1 database for amplicon-based metagenomic characterization of environmental fungal communities

B. Fosso¹✉, M. Santamaria², A. Consiglio³, G. De Caro³, G. Grillo³, F. Licciuli³, S. Liuni³, M. Marzano², G. Pesole¹

¹Dipartimento di Bioscienze, Biotecnologie e Scienze Farmacologiche, Università degli Studi di Bari "A. Moro", Bari, Italy

²Istituto di Biomembrane e Bioenergetica, Consiglio Nazionale delle Ricerche, Bari, Italy

³Istituto di Tecnologie Biomediche, Consiglio Nazionale delle Ricerche, Bari, Italy

Motivations

Metagenomics is experiencing an explosive improvement from the advent of high-throughput next-generation sequencing (NGS) technologies which allows an unprecedented large-scale identification of microorganisms living in almost every environment. In particular, the use of amplicon-based metagenomic approach to explore the diversity of fungal environmental communities is increasingly expanding. At the species level, a number of studies have used the non-conserved internal transcribed spacers (ITS) 1 and 2 of the ribosomal RNA genes cluster as genetic markers to explore the fungal taxonomic diversity. Particularly, ITS1 is gaining an increasing popularity as better discriminating species marker in Fungi because of its higher variability compared to ITS2. Starting from the total DNA extracted from any environmental sample, this locus can be easily amplified with taxonomically universal primers and sequenced by means of high-throughput next generation platforms. Reference databases and robust supporting taxonomies are crucial in assigning phylogenetic affiliation to the huge amount of produced sequences. Even if a large number of ITS1 sequences are collected in public databases, a specialized resource focused particularly on this region, where sequences identity, boundaries and taxonomic assignment are validated, is still needed at present. In this work we present ITSoneDB, a new comprehensive collection of ITS1 sequences belonging to Fungi Kingdom.

Methods

ITSoneDB has been generated and populated using a multi-step Python workflow. In the first step the ribosomal RNA gene cluster sequences of Fungi including the target ITS1 region were retrieved from Genbank. Then, ITS1 start and end boundaries were extracted from the Features Tables annotations, if available. In order to infer, validate and, eventually, redesign the ITS1 location, Hidden Markov Model (HMM) profiles of flanking genes for 18S and 5.8S ribosomal RNA, generated from their reference alignments stored in RFAM database, were mapped on the entire collection of retrieved nucleotide sequences, by means of the *hmmsearch* tool from HMMER 3.0 package.

Results

At present, ITSoneDB includes 405,433 taxonomically arranged sequence entries provided with ITS1 both start and end positions defined by GenBank annotations and/or HMM based method. ITSoneDB front-end is a JAVA platform-based website for data browsing and downloading. The database can be queried by species or taxon name, GenBank accession ID or by "expanding" the target rank on a detailed fungal taxonomical tree. The complete ITS1 sequences dataset collected in ITSoneDB is available in Fasta format and the users can extract and locally save all or selected queried ITS1 sequences for further analysis.

Availability

<http://itsonedb.ba.itb.cnr.it/>