ITSoneDB: a reliable resource for the investigation of the human Mycobiome
Category Health

Giuseppe Defazio (1), Bruno Fosso (2), Monica Santamaria (2) and Graziano Pesole (1,2).
(1)Department of Biosciences, Biotechnology and Biopharmaceutics, University of Bari "A. Moro", via E. Orabona 4, Bari 70125, Italy
(2) Institute of Biomembranes, Bioenergetics and Molecular Biotechnologies, Consiglio Nazionale delle Ricerche, via Amendola 122/O, Bari 70126, Italy

Aim: In the last ten years the human microbiome investigation has raised a constantly increasing interest both in the scientific community and in public opinion. The gained evidences have clearly defined how it deeply influences both the physiological and pathological processes. Even if most of the attention was focused on prokaryotes, more recently the survey of fungal communities, the so called mycobiome [1], gained momentum. In this framework, the advent of Next Generation Sequencing (NGS) technologies and the DNA metabarcoding (exploits marker genes to taxonomically classify species) has allowed to rapidly and accurately investigate complex microbiomes. Regarding the mycobiome, the ITS1 (Internal Transcribed Spacer 1) of the eukaryotic ribosomal gene cluster is the most promising marker. The taxonomic annotation of microbial species relies on well-curated databases. In this respect we carried out a benchmark assessment of ITSoneDB [2] and UNITE [3] as reference databases for the investigation of fungal communities.

Method: In order to obtain an in-silico fungal mock community we have randomly drawn 1,000 fungal genomes from the Genome NCBI database. The ITS1 regions were in-silico amplified by using *primersearch* and an *in-house* developed Python script. Then, Illumina 2x250 paired end ITS1 sequence amplicons have been generated in silico by using the Art-Illumina software. The obtained mock community includes 109 Families, 163 Genera and 73 Species. The sequences have been analysed by using QIIME2 by using both ITSoneDB and UNITE as reference databases.

Results: The databases performances in correctly profiling the mycobiome relative abundances have been assessed at Family, Genus and Species ranks by using the Pearson correlation (R^2). Both ITSoneDB and UNITE have shown a $R^2=1$ at Family and Genera level. At the Species level ITSoneDB outperformed UNITE by obtaining more accurate results both in terms of classified sequences (Table 1) and relative abundance correlation (Figure 1).

Table 1: Percentage of assigned reads at each taxonomic level and relative abundance correlation (R2).

		Family	Genus	Species
ITSoneDB	**Assigned**	98.9%	98.8%	81.7%
	R^2	1.00	1.00	1.00
UNITE	**Assigned**	98.6%	96.9%	1.9%
	R^2	1.00	1.00	0.01
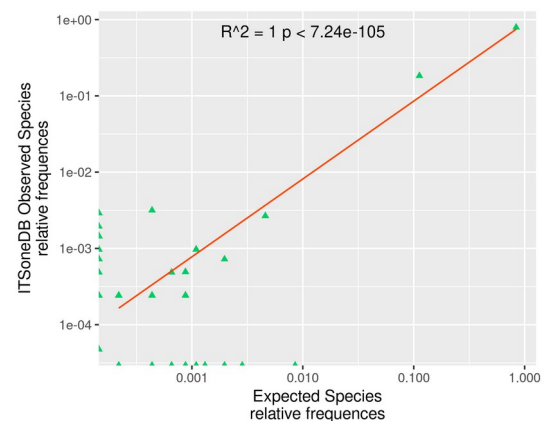
Figure1. Scatter-plot representing the relationship between the observed and expected species relative abundances.

Conclusion: UNITE taxonomic assignments were very consistent at Family and Genus levels but not at species rank. Remarkably, as ITSoneDB results are consistent at every taxonomic level it represents a state of the art resource for the metabarcoding-based investigation of fungal microbial communities.

[1] S. Raimondi *et al.*, «Longitudinal Survey of Fungi in the Human Gut: ITS Profiling, Phenotyping, and Colonization», *Front. Microbiol.*, vol. 10, lug. 2019.
[2] M. Santamaria *et al.*, «ITSoneDB: a comprehensive collection of eukaryotic ribosomal RNA Internal Transcribed Spacer 1 (ITS1) sequences», *Nucleic Acids Res.*, vol. 46, n. D1, pagg. D127–D132, gen. 2018.
[3] U. Kõljalg *et al.*, «UNITE: a database providing web-based methods for the molecular identification of ectomycorrhizal fungi: Methods», *New Phytol.*, vol. 166, n. 3, pagg. 1063–1068, mar. 2005.