

## REVISION OF MISASSIGNED BARCODING SEQUENCES IN PUBLIC DATABASES, THE CASE OF INTERNAL TRANSCRIBED SPACER IN WHITISH TRUFFLES

**Beatrice Belfiori<sup>1</sup>, Claudia Riccioni<sup>1</sup>, Maurizio Cenci<sup>1</sup>, María Belén Pildain<sup>234</sup>, Andrea Rubini<sup>1</sup>**

<sup>1</sup>CNR-IBBR: Institute of Biosciences and BioResources, Perugia Division, Italy

<sup>2</sup>Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)

<sup>3</sup>Centro de Investigación y Extensión Forestal Andino Patagónico (CIEFAP), Esquel, Argentina

<sup>4</sup>Universidad Nacional de la Patagonia SJB, Esquel, Argentina

E-mail: [beatrice.belfiori@ibbr.cnr.it](mailto:beatrice.belfiori@ibbr.cnr.it)

**Key words:** *Tuber borchii*, BLASTn, misidentification, molecular taxonomy, morphology

Taxonomy of *Tuber* has advanced significantly in recent years, mainly thanks to the massive use of barcoding sequences such as the Internal Transcribed Spacer (ITS) of rDNA. The number of ITS sequences of *Tuber* specimens deposited in public databases such as GenBank is about 4200 entries. Among *Tuber* spp., some groups such as the “whitish truffles species complex” are rich in species that share high morphological similarity and are difficult to identify. This has generated the accumulation in the databases of sequences from specimens not accurately identified. The main reason that led to this situation is the absence of a taxonomic review procedure, given that the reliability of GenBank taxonomic information depends only on the validity of data provided by the researchers. The presence of barcode sequences from wrongly classified specimens in GenBank limited the utility of powerful tools for quick species identification such as BLASTn search and increased the risk of propagating taxonomic errors to newly deposited sequences. With the aim of contributing to mitigate such problems (i.e. by proposing corrections for wrongly identified specimens), we performed an inventory of all ITS sequences from *Tuber* specimens classified in the whitish species complex. We retrieved approximately 1400 sequences from GenBank, belonging to the main clades “puberulum”, “maculatum”, “gibbosum” and “latisporum”. Based on phylogenetic analyses we identified terminal nodes containing specimens with contrasting species names, as probable cases of incorrect species attribution (approx. 15%). The main criteria we followed to identify reference specimens with reliable species attribution were sequences cited in valuable taxonomic studies and sequences belonging to type specimens. Further, we are doing ITS sequencing and morphological examination of specimens collected by ourselves, to obtain additional reference samples.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 101007623.