

The Westerdijk fungal data resources for fungal identification

Dr Duong Vu¹, Dr Marizeth Groenewald¹, Gerard Verkley¹

¹*Westerdijk Fungal Biodiversity Institute*

At the Westerdijk Fungal Biodiversity Institute, we maintain a world-renowned microbial genetic resource, the CBS collection with more than 100K living strains of micro-organisms, representing a large percentage of the species in the fungal kingdom that have been cultured to date. In the WI-DNA barcoding project, we have generated more than 200,000 DNA barcode sequences (ITS and LSU), for fungal identification. Two large barcode datasets were released to GenBank in 2016 and 2019 as reference sequences for yeast and filamentous fungal identification respectively, an unprecedented data release event in global fungal barcoding efforts to date (Vu et al 2016, 2019). Both datasets are globally used for describing new fungal species, for taxonomic reclassification, and for the identification of mycobiota from environmental samples (eDNA). They were integrated into ARISE, an Authoritative and Rapid Identification System for Essential biodiversity information, for species identification and for monitoring biodiversity in the Netherlands. The datasets were also used for identifying mycobiota from the gardens around the Netherlands in Citizen Science projects. The filamentous fungi dataset was awarded the Dutch Data Prize 2022.

Based on our barcode datasets, we were able to develop a number of bioinformatics tools to improve accuracy, precision, and speed of fungal identification (Vu et al. 2012, 2014, 2018, 2020). In most ecological studies, only one single similarity cut-off such as 97% is used for species identification. As more and more barcodes were generated, it gradually became clear that the use of single, static threshold for identification is problematic. We demonstrated in our recent study (Vu et al. 2022) that metabarcoding loses resolution and scientific explanatory power by relying on a single similarity cut-off for taxonomic assignment. We introduced dnabarcoder, a software tool to compute dynamic similarity cut-offs for different clades for fungal identification. It was shown that dynamic similarity cut-offs assigned fewer sequences than the traditional similarity cut-offs, but the accuracy and precision were significantly improved.

References

- Vu, D., Groenewald, M., Szöke, S., et al. (2016). *Studies in Mycology* 85, 91-105.
- Vu, D., Georgievskaya, S., Szöke, S., et al. (2018). *Bioinformatics*, 1-3.
- Vu, D., Groenewald, M., de Vries, M., et al. (2019). *Studies in Mycology* 92, 135-154.
- Vu, D., Groenewald, M. & Verkley, G. (2020). *Scientific reports* 10, 12628.
- Vu, D., Nilsson, H.N. & Verkley, G. (2022). *Molecular Ecology Resources*, <https://doi.org/10.1111/1755-0998.13651>.