

# dnabarcoder: an open-source software package for analyzing and predicting DNA sequence similarity cut-offs for fungal sequence identification

Duong Vu<sup>1</sup>, Henrik Nilsson<sup>2</sup>, and Gerard Verkley<sup>1</sup>

<sup>1</sup>Westerdijk Fungal Biodiversity Institute

<sup>2</sup>University of Gothenburg

January 12, 2022

## Abstract

The accuracy and precision of fungal molecular identification and classification are challenging, particularly in environmental metabarcoding approaches as these often trade accuracy for efficiency given the large data volumes at hand. In most ecological studies, only a single similarity cut-off value is used for sequence identification. This is not sufficient since the most commonly used DNA markers are known to vary widely in terms of inter- and intra-specific variability. We address this problem by presenting a new tool, dnabarcoder, to analyze and predict different local similarity cut-offs for sequence identification for different clades of fungi. For each similarity cut-off in a clade, a confidence measure is computed to evaluate the resolving power of the genetic marker in that clade. Experimental results showed that when analyzing a recently released filamentous fungal ITS DNA barcode dataset of CBS strains from the Westerdijk Fungal Biodiversity Institute, the predicted local similarity cut-offs varied immensely between the clades of the dataset. In addition, most of them had a higher confidence measure than the global similarity cut-off predicted for the whole dataset. When classifying a large public fungal ITS dataset – the UNITE database – against the barcode dataset, the local similarity cut-offs assigned fewer sequences than the traditional cut-offs used in metabarcoding studies. However, the obtained accuracy and precision were significantly improved.

## Hosted file

dnabarcoder\_DV\_HN\_GV.docx available at <https://authorea.com/users/455216/articles/552575-dnabarcoder-an-open-source-software-package-for-analyzing-and-predicting-dna-sequence-similarity-cut-offs-for-fungal-sequence-identification>