

Qualitative comparison of four DNA metabarcoding markers for species recovery sensitivity based on *Hodophilus* fungi amplicon sequence dataset

Vasilii Shapkin¹, Slavomir Adamčík¹, Miroslav Caboň¹

¹Institute of Botany, Plant Science and Biodiversity Centre, Slovak Academy of Sciences, Dúbravská cesta 9, SK-845 23 Bratislava, Slovakia

Introduction:

Nowadays species-level taxonomic resolution for prokaryotes using 16S amplicon sequencing is common. However there is a big challenge to achieve the same quality of results for fungal ITS amplicon sequencing due to the high variabilities in the genome structure of fungi or the absence of distinction in ITS sequences between different species.

Thus, in order to achieve better results there is a trend on the side stepping from classic (ITS2) to less popular but promising (D1, D2, Rpb2) metabarcoding markers.

Methods:

The total of 476 amplicon sequences in 29 *Hodophilus* fungi species were collected from GenBank. From them, 163 - ITS, 113 - D1, 113 - D2 and 87 - Rpb2 marker sequences. The RAxML program was used for phylogenetic analysis and the VSEARCH algorithm for clustering of the sequences (Fig. 1).

To evaluate the ability of the DNA metabarcoding markers to recover the *Hodophilus* fungi species the following parameters have been selected (Fig. 2):

MOTU performance (MP) - the ability of the DNA metabarcoding marker to correctly recover only relevant members of the *Hodophilus* fungi species.

Species performance (SP) - the ability of the DNA metabarcoding marker to correctly recover relevant members of the *Hodophilus* fungi species.

Species and MOTU performance defined as the true positive rate and the positive predictive value $TP/(TP+FP)$, respectively, where TP is the number of MOTUs agreed with phylogenetic species, FN is the number of phylogenetic species not disagreed with MOTUs, and FP is the number of MOTUs disagreed with phylogenetic species. For the species performance if several MOTUs were assigned to the same *Hodophilus* fungi species, both were considered to be one true positive.

The F1 score - defined as the harmonic mean of the Species and MOTU performance - $2 \times (MP \times SP) / (MP + SP)$. Species/MOTU performance and F1 score criterias theoretically range between 0 and 1. They equal 1 when the DNA metabarcoding marker is perfect and successfully identifies all members of the *Hodophilus* fungi species.

Results:

The best species performance showed by ITS2 (100%) and Rpb2 (95%) marker genes. The best MOTU performance showed by D2 (77%) and Rpb2 (77%) markers.

The best overall result achieved using Rpb2 marker gene sequences clustered with 99% similarity threshold - F1 score 0.85.

Contacts:

E-mail: vasilii.shapkin@savba.sk, Vasilii Shapkin, PhD student

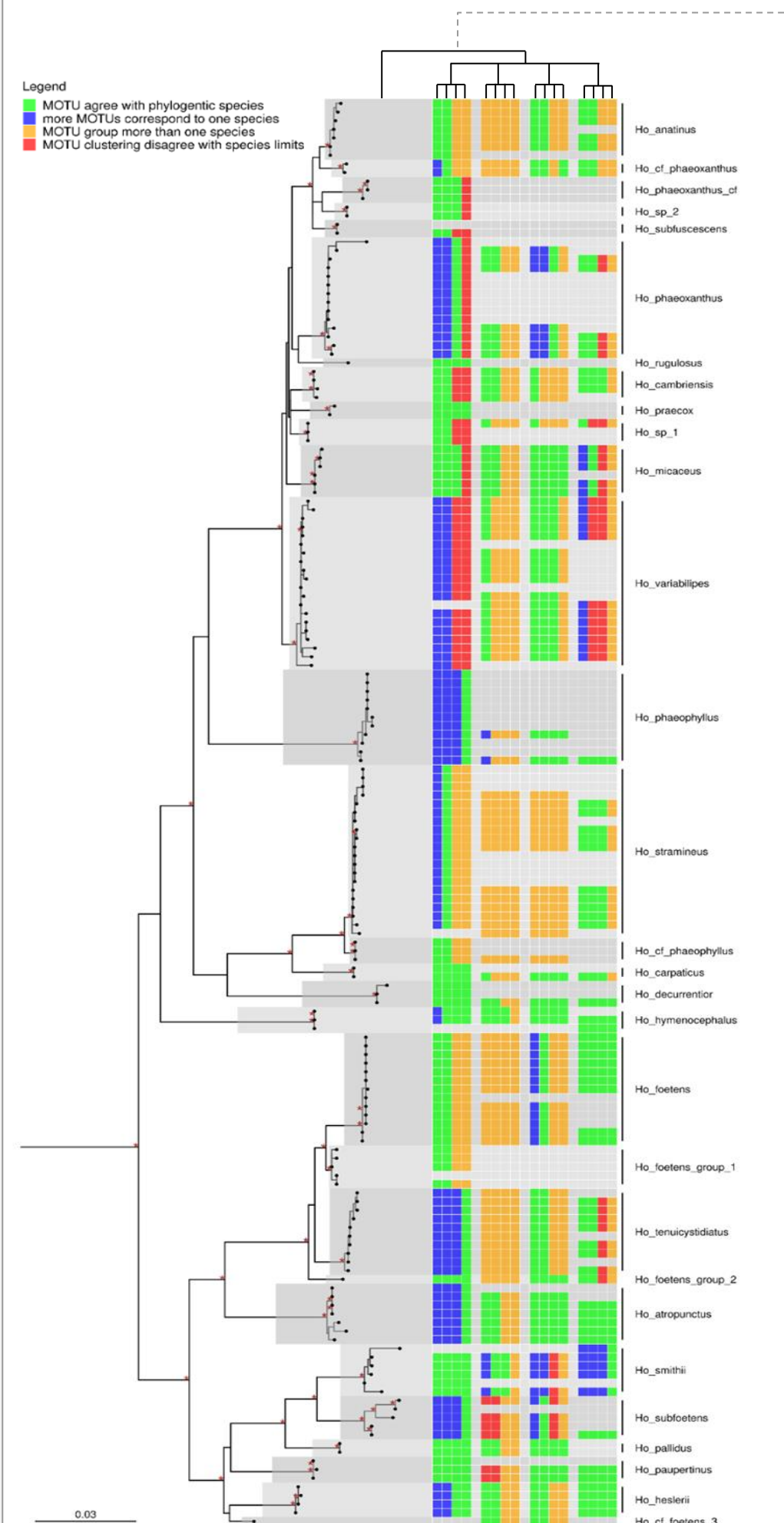


Figure 2. Qualitative comparison of the DNA metabarcoding markers performances

	ITS2				D1				D2				Rpb2			
	99.5	99	98	97	99.5	99	98	97	99.5	99	98	97	99.5	99	98	97
MP	0.37	0.51	0.52	0.78	0.57	0.59	0.2	N/P	0.62	0.77	0.69	0.71	0.65	0.77	0.71	0.75
SP	1.00	1.00	0.62	0.48	0.61	0.43	0.09	N/P	0.91	0.83	0.48	0.43	1.00	0.95	0.65	0.45
F1	0.54	0.68	0.57	0.60	0.59	0.50	0.12	N/P	0.74	0.80	0.56	0.54	0.79	0.85	0.68	0.56

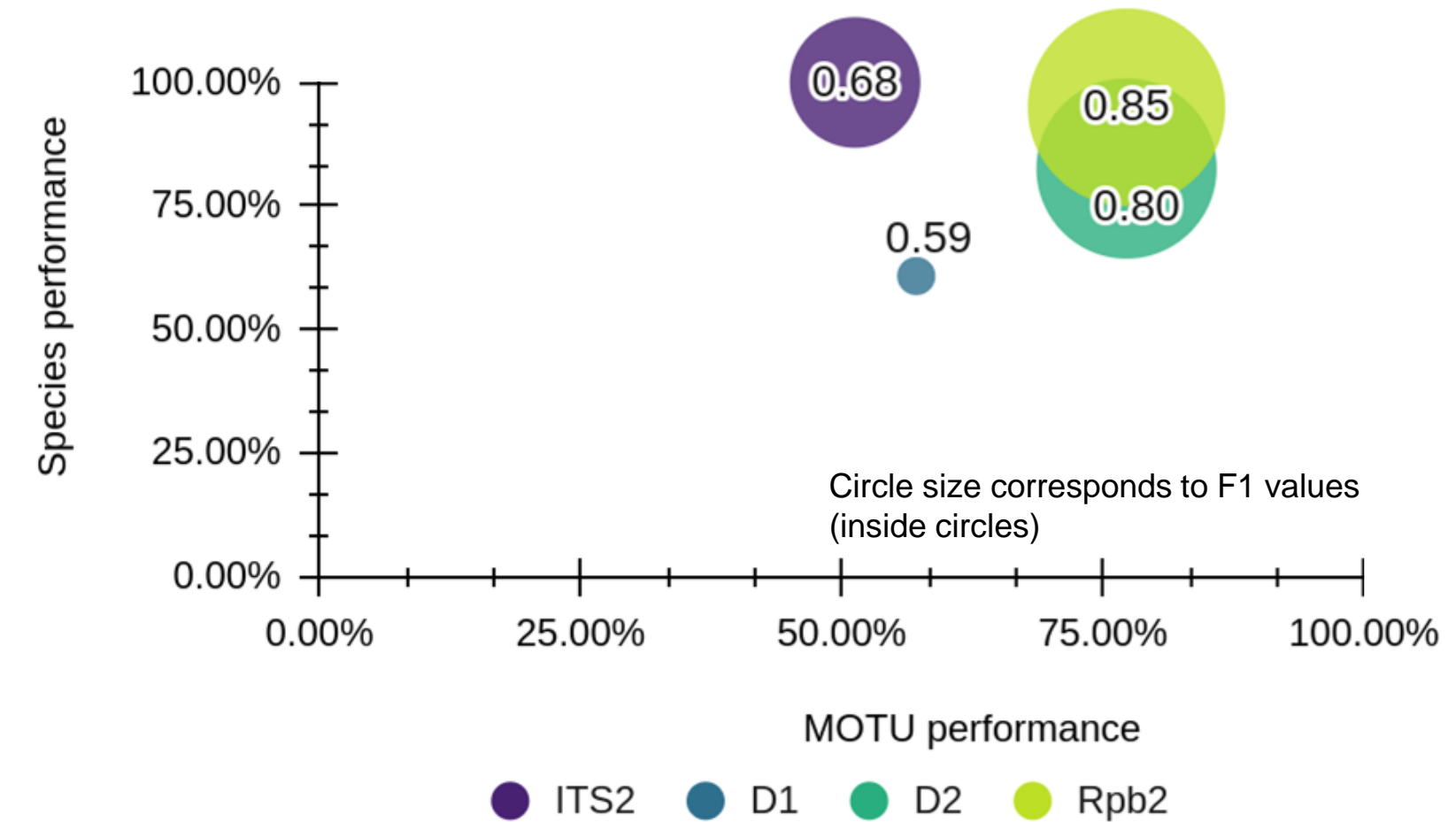


Figure 1. Phylogenetic tree of *Hodophilus* fungi species with clustering performance heatmap

