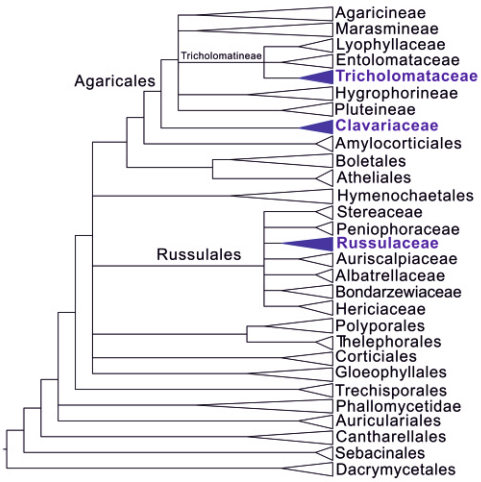
Recovery of species-rank OTUs of agarics (Agaricomycetes, fungi) in metagenomic datasets based on various nrDNA amplicon lengths and positions

Slavomír Adamčík[1], Brian Looney[2], Miroslav Kolařík[3], Marisol Sánchez-García[1], Katarína Adamčíková[5], Miroslav Caboň[1], Gareth W. Griffith[6]

[1]Institute of Botany, Plant Science and Biodiversity Centre, Slovak Academy of Sciences, Dúbravská cesta 9, SK-845 23 Bratislava, Slovakia; [2]Department of Ecology and Evolution Biology, University of Tennessee, 332 Hesler Biology Building, Knoxville, TN 37996-1610, USA; [3]Institute of Microbiology, Czech Academy of Sciences, Vídeňská 1083, CZ-142 20 Praha 4, Czech republic; [4]Biology Department, Clark University, Worcester, Massachusetts 01610, USA; [5]Branch for Woody Plants Biology, Institute of Forest Ecology, Slovak Academy of Sciences Zvolen, Akademická 2, SK-949 01 Nitra, Slovakia; [6]IBERS, Aberystwyth University, Cledwyn Building, Ceredigion, Aberystwyth, Wales SY23 3DD, UK

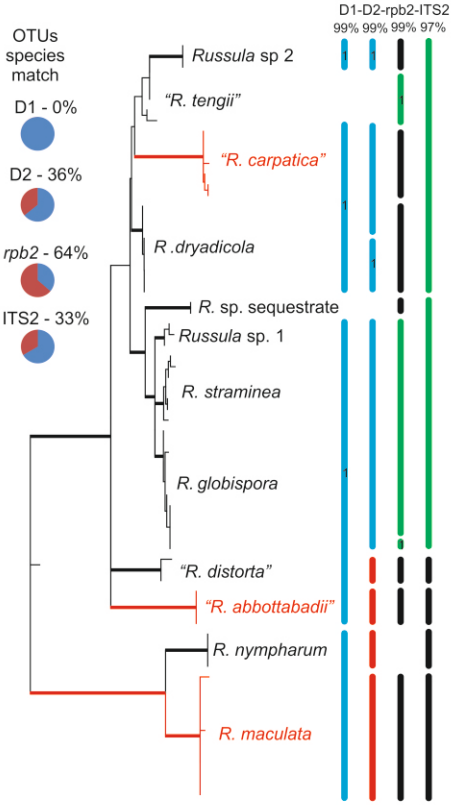


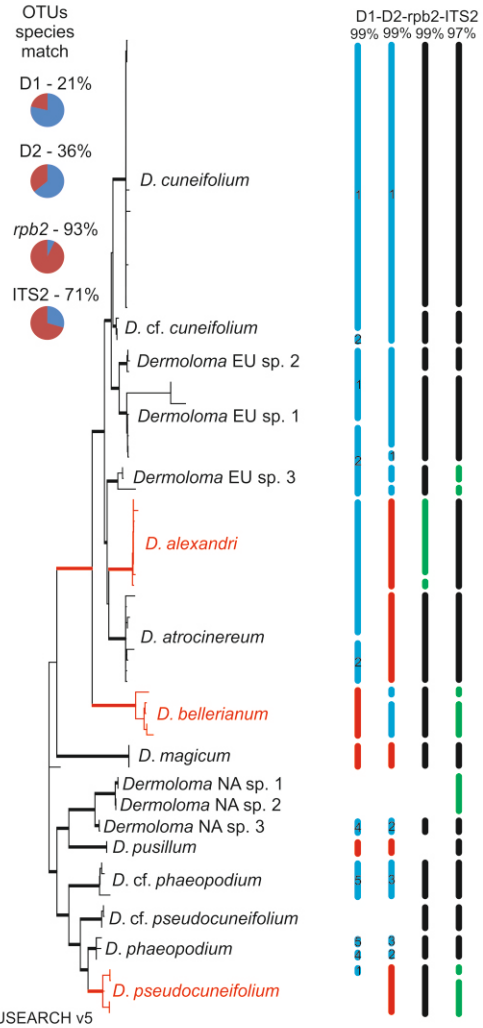
Agarics (Agaricomycetes) are very diverse and abundant fungi in forest and grassland ecosystems. Various next generation sequencing (NSG) techniques allow sequencing of DNA amplicons of different lengths. A variety of loci have been used for metabarcoding of fungi in environmental samples, the most frequently used being the ITS2 region of ribosomal DNA (nrDNA) and reference species sequences of this region are the most widely available in public databases. The ITS2 region has some limitations in the quantification of fungal OTUs and taxon recognition due to high intragenomic variability, unequal sequence length and variable GC contents in some taxa.

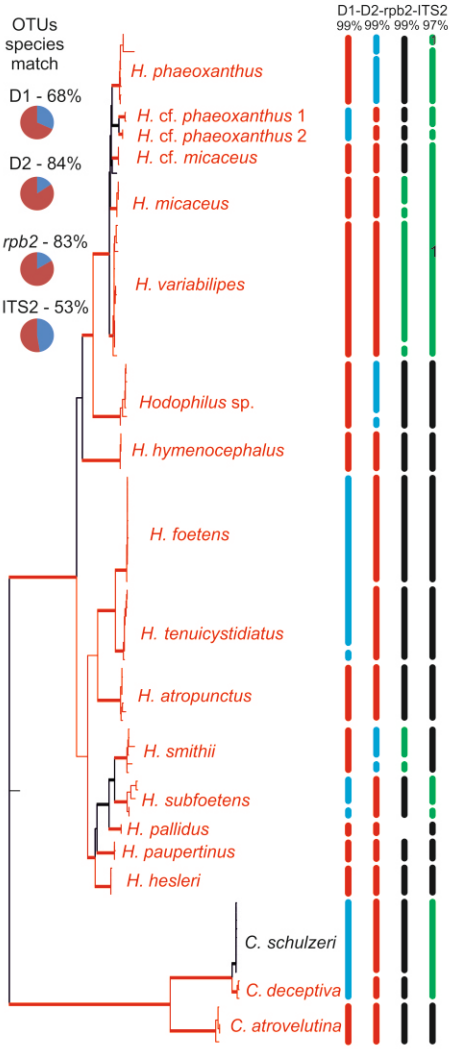
D1 and D2 regions of LSU (large subunit) rRNA gene and a fragment of the second largest subunit of ribosomal polymerase II (*rpb2*) which do not suffer from these problems are tested as an alternatives for metagenomics.

Three evolutionary distant lineages each of closely related agaric species are analysed: *Russula* subsect. *Maculatinae* (Russulaceae), *Dermoloma* (Tricholomataceae) and *Hodophilus* and *Camarophyllopsis* (Clavariaceae).

Multi-locus phylogenies of each group are compared with OTUs defined under 99.5%, 99%, 98%, and 97% similarity thresholds in a simulation of NSG data analyses.







Tree explanation: thick branches - ≥ 95% support in BI analysis
red clades - retrieved by analysis of LSU region
black clades - retrieved by multilocus analysis (ITS-LSU-*rpb2*)
OTUs recognised in NSG datasets using UPARSE algorithm in USEARCH v5

- species rank OTUs in D1/D2 datasets
- species rank OTUs in ITS2 and rpb2 dataset
- OTUs at a non-species rank in D1/D2 dataset
- OTUs at a non-species rank in ITS2 and *rpb2* dataset

Number of OTUs under the defined similarity threshold computed using UPARSE algorithm in Usearch v5.

Dataset	99.5%	99 %	98%	97%	Phylosp.
Dermoloma_D1	14	8	2	2	17
Dermoloma_D2	16	13	11	8	
Dermoloma_ITS2	52	34	27	19	
Dermoloma_*rpb2*		15	13	12	
Hodophilus_D1	25	17	11	8	19
Hodophilus_D2	36	22	16	14	
Hodophilus_ITS2	52	36	25	17	
Hodophilus_*rpb2*		21	14	12	
Russula_D1	2	2	1	1	12
Russula_D2	10	7	5	3	
Russula_ITS2	23	16	11	6	
Russula_*rpb2*		9	6	5	

Specific patterns of each evolutionary lineage:

Russula shows the most closely related species with a weak overlap of OTUs with the multilocus phylogeny
Dermoloma has several OTUs defined by LSU regions with polyphyletic origin
Hodophilus shows good species recovery in all NSG datasets, but with different species combinations

Conclusion:

- *rpb2* datasets have the best species recovery
- to recognise species-rank OTUs, phylogenetic study is needed

Accurate species identification in nsg data is possible by:

- aligning NSG sequences to datasets of existing phylogenetic studies
- oligonucleotic barcoding (hallmarks)

Other risks of identification of species in nsg:

- unrecognised and undescribed species-rank OTUs
- lost of data by different DNA isolation methods, primer bias, PCR bias (chimeric regions)

The project was supported by Slovak national grants APVV-15-210 and Vega 02/0075/14