XIII Argentine Congress of Bioinformatics and Computational Biology
XIII International Conference of the Iberoamerican Society of Bioinformatics
III Annual Meeting of the Ibero-American Artificial Intelligence Network for Big BioData

Clade-wide proteome analysis shows widespread non-canonical DCR proteins in Fungi

Lorena Melet[1,2,3], Ivana Orellana[2,3], J. Andrés Rivas-Pardo[2], Nathan Johnson[1,2], Elena A. Vidal[1,2].
[1]Centro de Genómica y Bioinformática, Universidad Mayor.
[2]Millennium Institute for Integrative Biology (iBio).
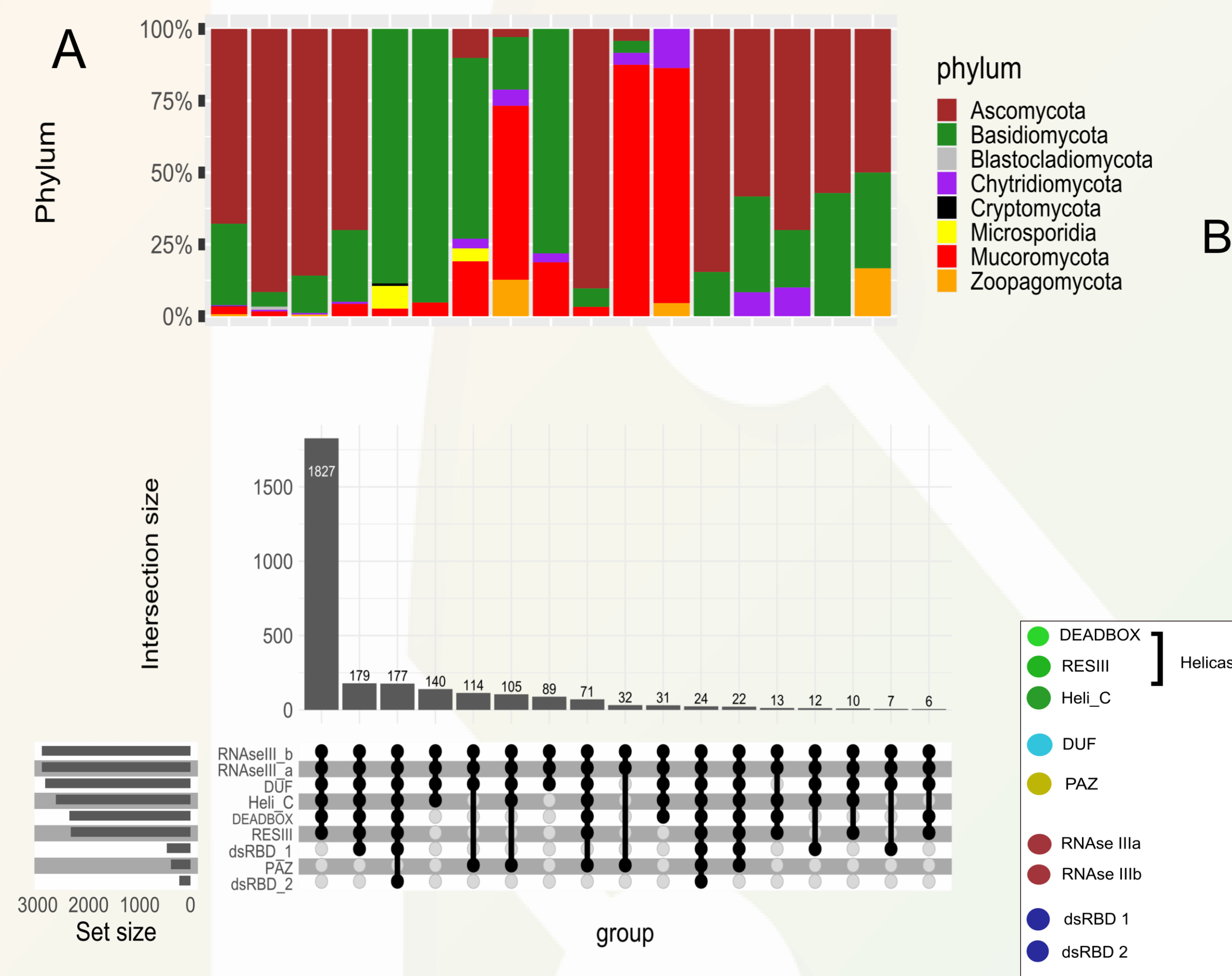[3]Doctorado en Genómica Integrativa, Universidad Mayor.
lorena.melet@mayor.cl

1 The Dicer proteins: what is known about them in fungal species?.

Dicer is a class-III ribonuclease that plays a fundamental role in the synthesis of small RNAs and the regulation of gene expression. DCRs are ubiquitous in various eukaryotic organisms, including fungi, where they are shown to be indispensable for defense against viruses, the regulation of transposons, and developmental processes. So far, fungal DCRs have primarily been identified in extensively studied phyla such as Ascomycetes, Basidiomycetes, and Chytridiomycetes. However, there is still a significant number of fungal species in which these proteins have not yet been described, and the structural domains present in fungal DCRs remain poorly understood.

In recent years, there has been a notable increase in the availability of fungal genomes and proteomes. More than 1,500 annotated reference genomes with proteome annotations have become accessible. This study undertakes a comprehensive *in silico* analysis, characterizing and identifying fungal proteins on a large scale in over 1,400 species, spanning nine different phyla within the fungal kingdom. This research aims to shed light on the diversity and distribution of DCRs across the fungal kingdom and advance our understanding of their structural attributes and functions.

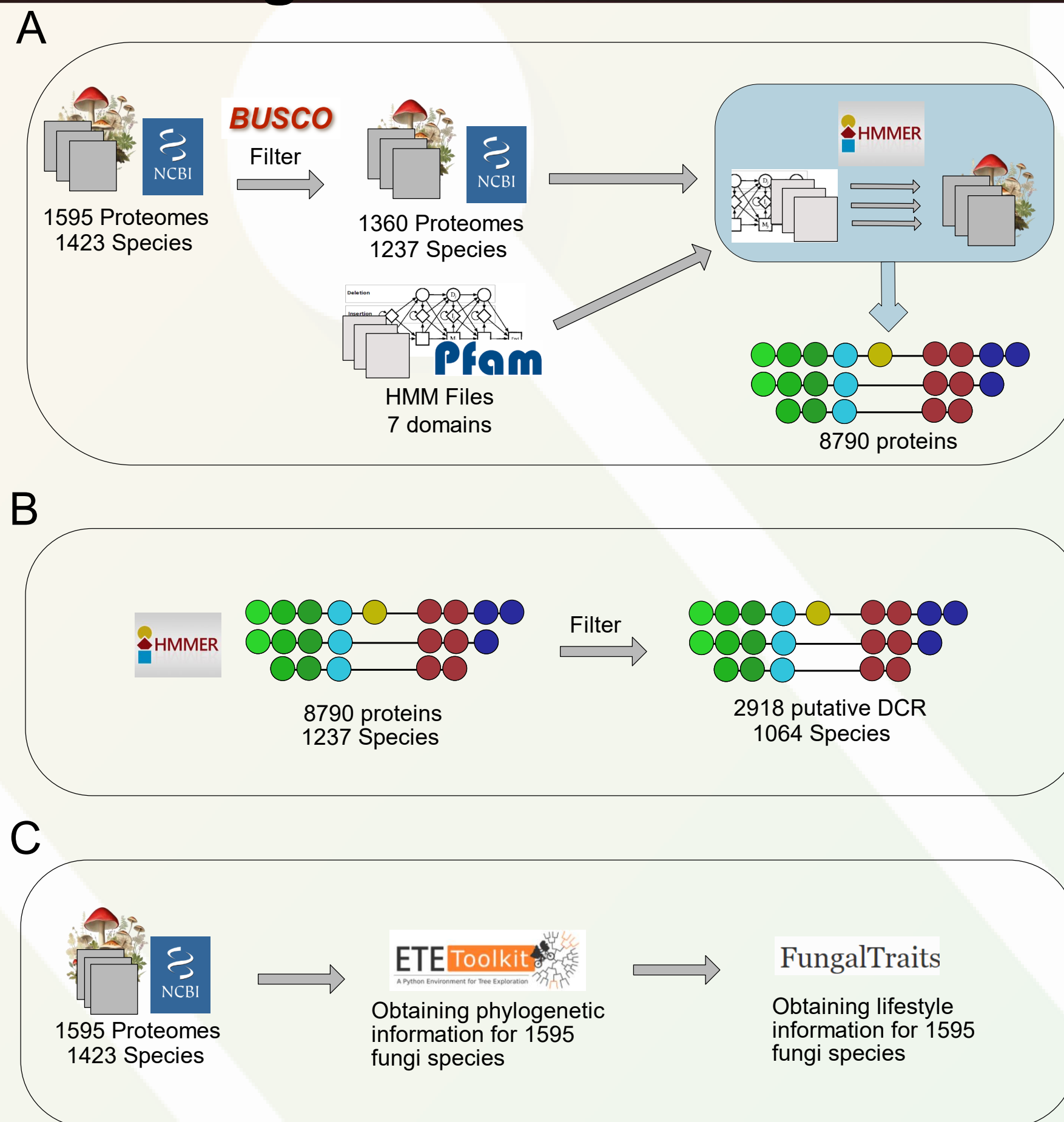## 2 Workflow to characterize Dicer proteins in fungi.

Fig. 1. Identification of DCR proteins in fungi.

A) **Recovery of proteomes.** 1595 annotated and reference proteomes up to November 2022 were retrieved from NCBI, which correspond to 1423 unique species. Evaluation of the quality of proteomes was performed with BUSCO v.5.3.2. Dicer protein identification was carried out with HMMER, using highly conserved Dicer domains within eukaryotic species: PF00270/IPR011545 (DEAD/DEAH box), PF04851/IPR006935 (ResIII), and PF00271/001650 (Helicase C terminal), PF00636/IPR000999 (RNAse III), PF02170/IPR003100 (PAZ), PF03368/IPR005034 (DUF/Dicer dimerization), and PF00035/IPR014720 (Double-stranded RNA binding motif).

B) **Putative Dicer protein filter and selection.** An E-value of 10^{-3} was used as a cutoff for a putative domain hit. All the proteins having at least two RNAse III domain were conserved, and the sequence of these proteins were extracted from their respective proteomes.

C) **Species classification.** Using the taxid for the different species we obtained the phylogenetic information with the ETE 3 toolkit. Additionally, with FungalTraits[2] we added information related to the lifestyle of those species.

3 The fungal species in this study exhibit diverse architectures in their DCR proteins.

Fig. 2. Frequency of composition domains and representation of the putative DCR proteins.

A) **Domain composition of putative DCRs found in fungal proteomes.** The upper graph displays the percentage of species within each phylum, corresponding to the intersections found in the lower graph. In the lower graph, set size bars represent the frequency of each domain, intersection size bars indicate the number of proteins found within those domains out of the total, and black dots illustrate the domains present in each intersection.

B) **Putative DCR classification.** Each domain is color-coded. The spatial arrangement of domains is to scale and reflects their relative distribution. Domains marked with an asterisk (*) are larger than usual. The identified protein architectures are grouped into (1) Canonical DCRs, (2) Non-canonical DCRs A and B, and (3) non-DCRs. The numbers on the left represent the number of putative proteins found for each type using HMMER.

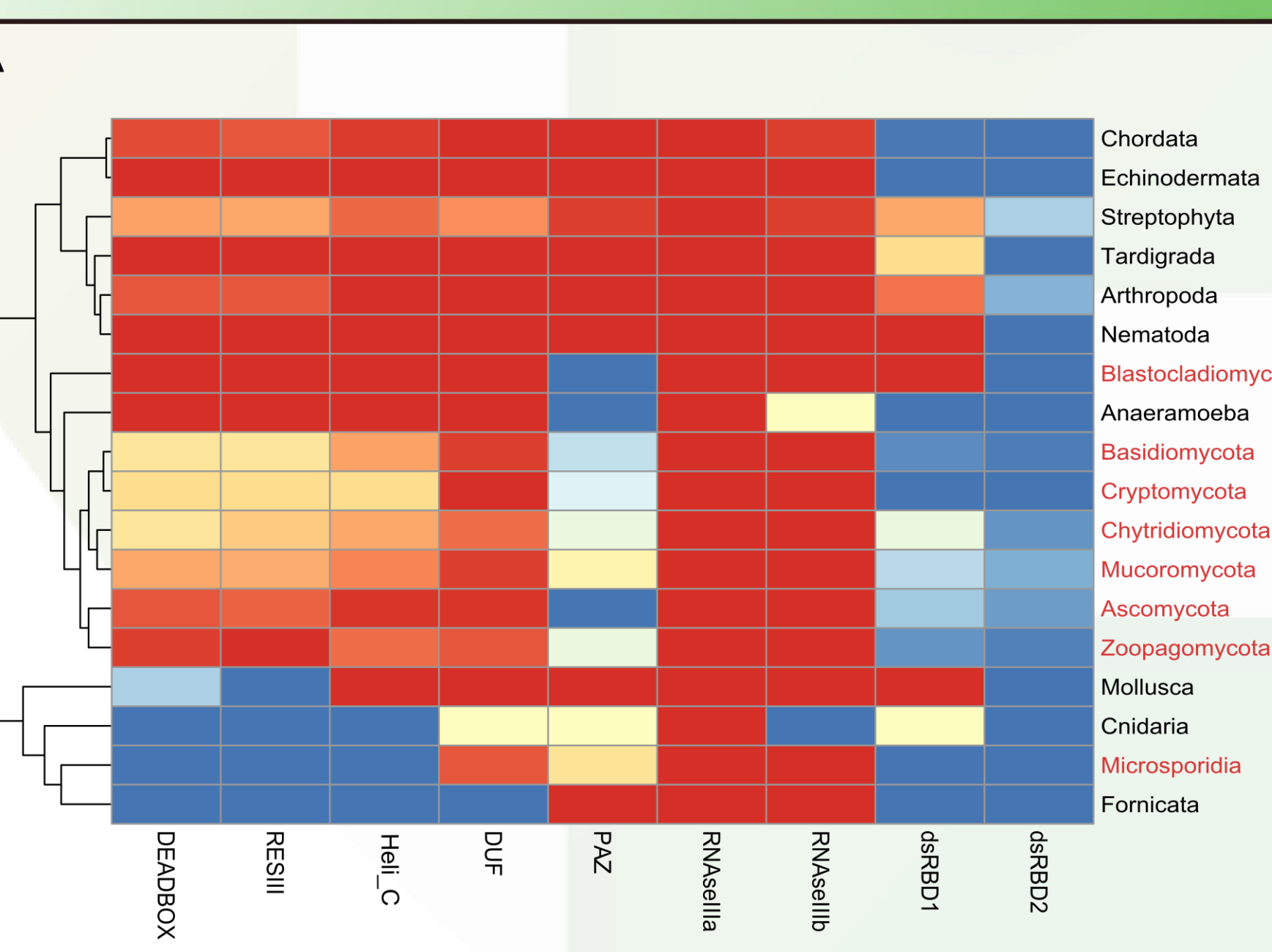## 4 Fungal lifestyle correlates with DCR protein architecture.

Fig. 3. Clustered heatmap comparing the presence of canonical domains among putative fungal DCR proteins and annotated DCRs from NCBI.

A) **Domains conserved between Phyla.** The Heatmap shows in red color letters the different Phyla for the fungal species and black colors letters correspond to DCR from NCBI. The canonical DCR protein domains are on the x-axis and species grouped by Phylum on the y-axis.

B) **Lifestyle as a relevant factor for putative fungal infections.** Fungal species were grouped by lifestyle in red color letters and the DCR from NCBI species was grouped by species type in an arbitrarily manner.

For both heat maps the colors range from 0 to 100%, with Red being 100% and Blue being 0%.
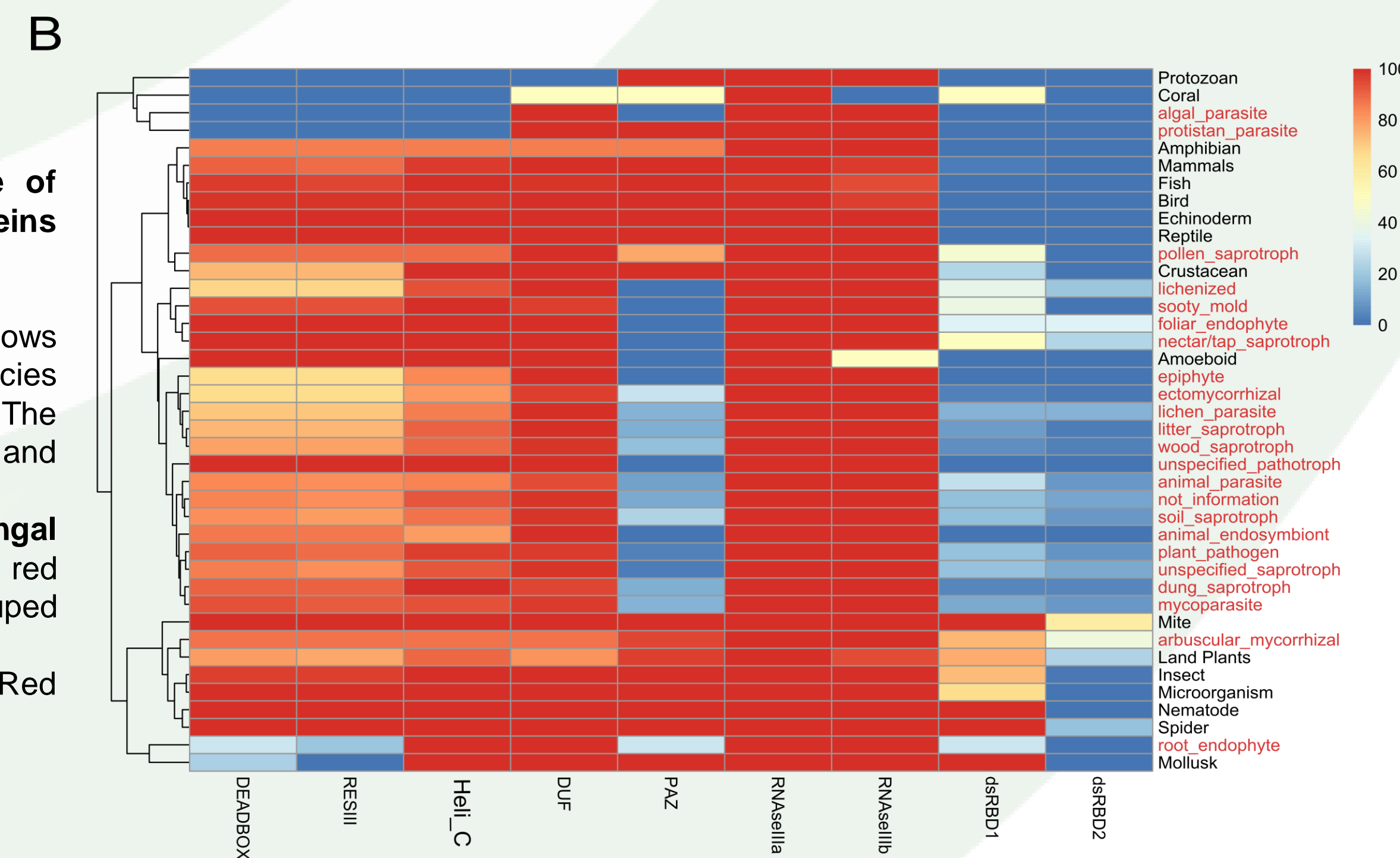
Conclusion

- We were able to find putative Dicer proteins in at least 1237 species of the 1423 species with proteomes in NCBI.
- As in previous studies, we also found that the canonical PAZ domain is less frequent in fungal DCR proteins.
- At least for fungi with an arbuscular mycorrhizal lifestyle it seems that they show a domain composition like land plants.
- This could indicate that DCRs from arbuscular mycorrhizal fungi could generate a repertoire similar to small RNAs from plants.
- The next phase of our research involves conducting an evolutionary analysis of these Dicer proteins and establishing a correlation between their structural characteristics and the small RNA profiles within these species

References

1. Lax, C. et al.,. 2020. The evolutionary significance of RNAi in the fungal kingdom. International Journal of Molecular Sciences, 21(24), 9348.
2. Põlme, S. et al., 2020. FungalTraits: a user-friendly traits database of fungi and fungus-like stramenopiles. Fungal diversity, 105, 1-16.