

# Comparative Analysis of Transcription Factors Families Across Fungal Tree of Life

Asaf Salamov<sup>1</sup>, Igor Grigoriev<sup>1</sup>

<sup>1</sup>DOE Joint Genome Institute, Walnut Creek, CA 94598

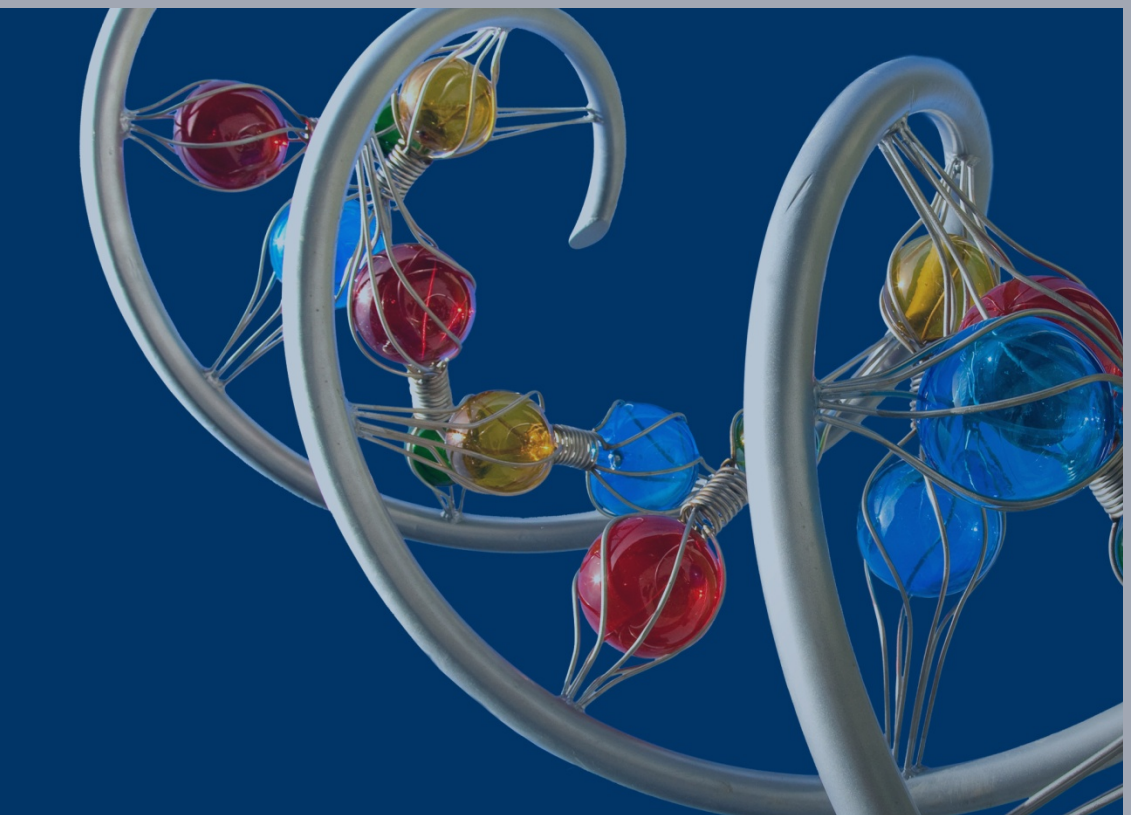
March 2015

The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231

LBNL- 178441

## DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or The Regents of the University of California.



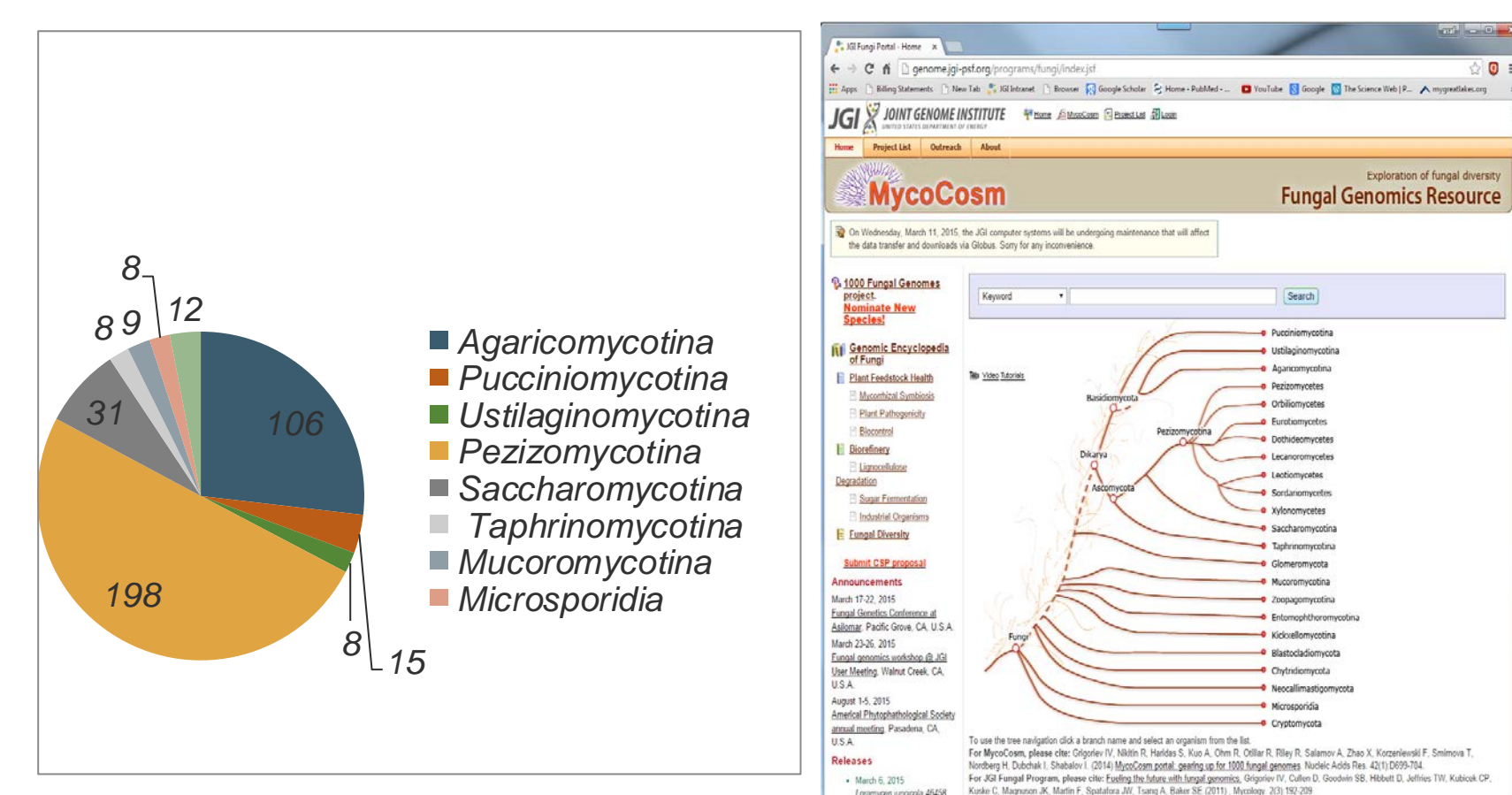
## Introduction

Transcription factors (TFs) are proteins that regulate the transcription of genes, by binding to specific DNA sequences.

Based on literature (Shelest, 2008; Weirauch and Hughes, 2011) collected and manually curated list of DBD Pfam domains (in total 62 DBD domains)

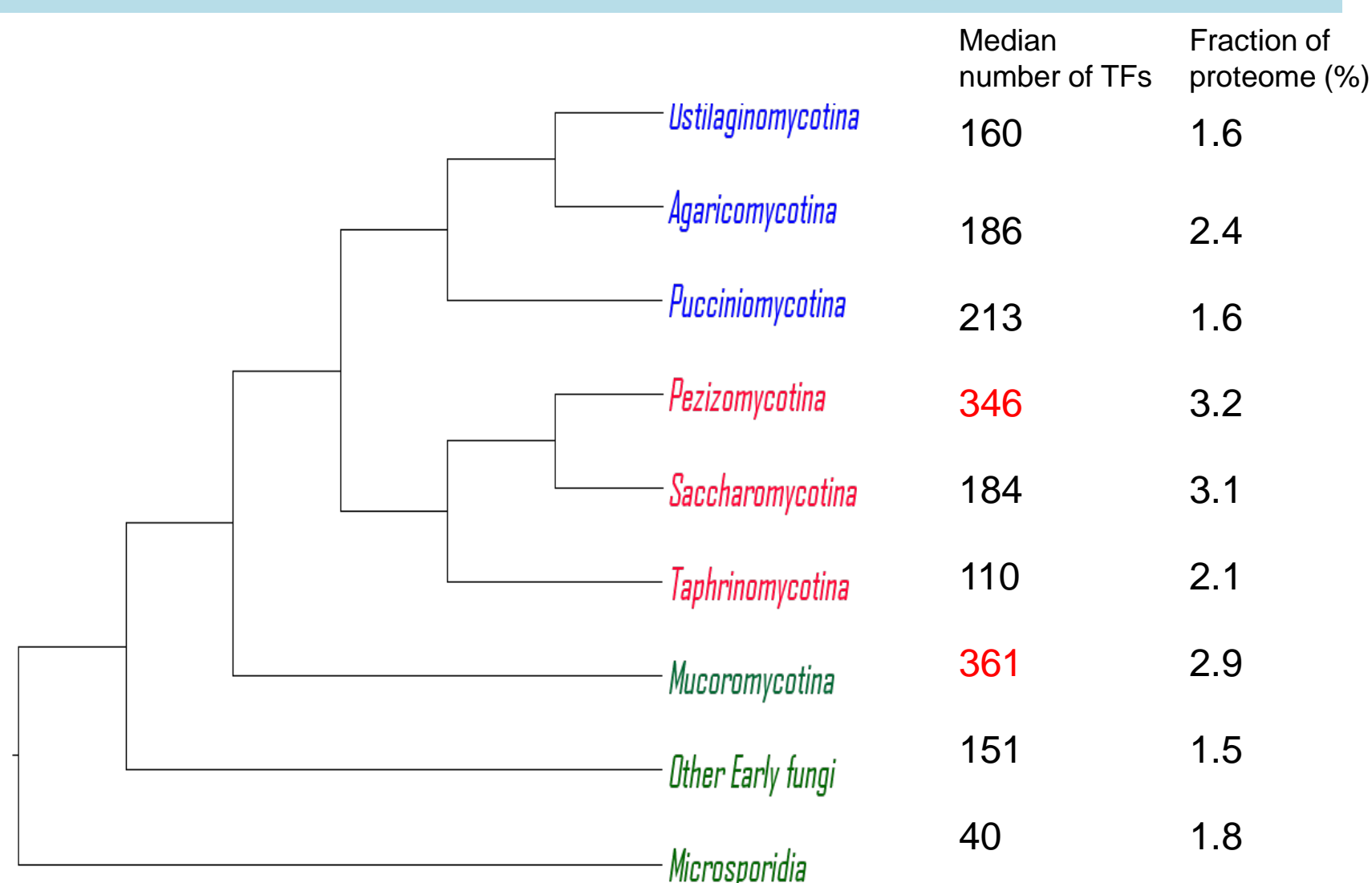
We looked for distribution of TFs in 395 fungal genomes plus additionally in plant genomes (Phytozome), prokaryotes (IMG), some animals/metazoans and protists genomes

## 395 fungal genomes from MycoCosm



<https://jgi.doe.gov/fungi>

## TF distribution across fungal clades

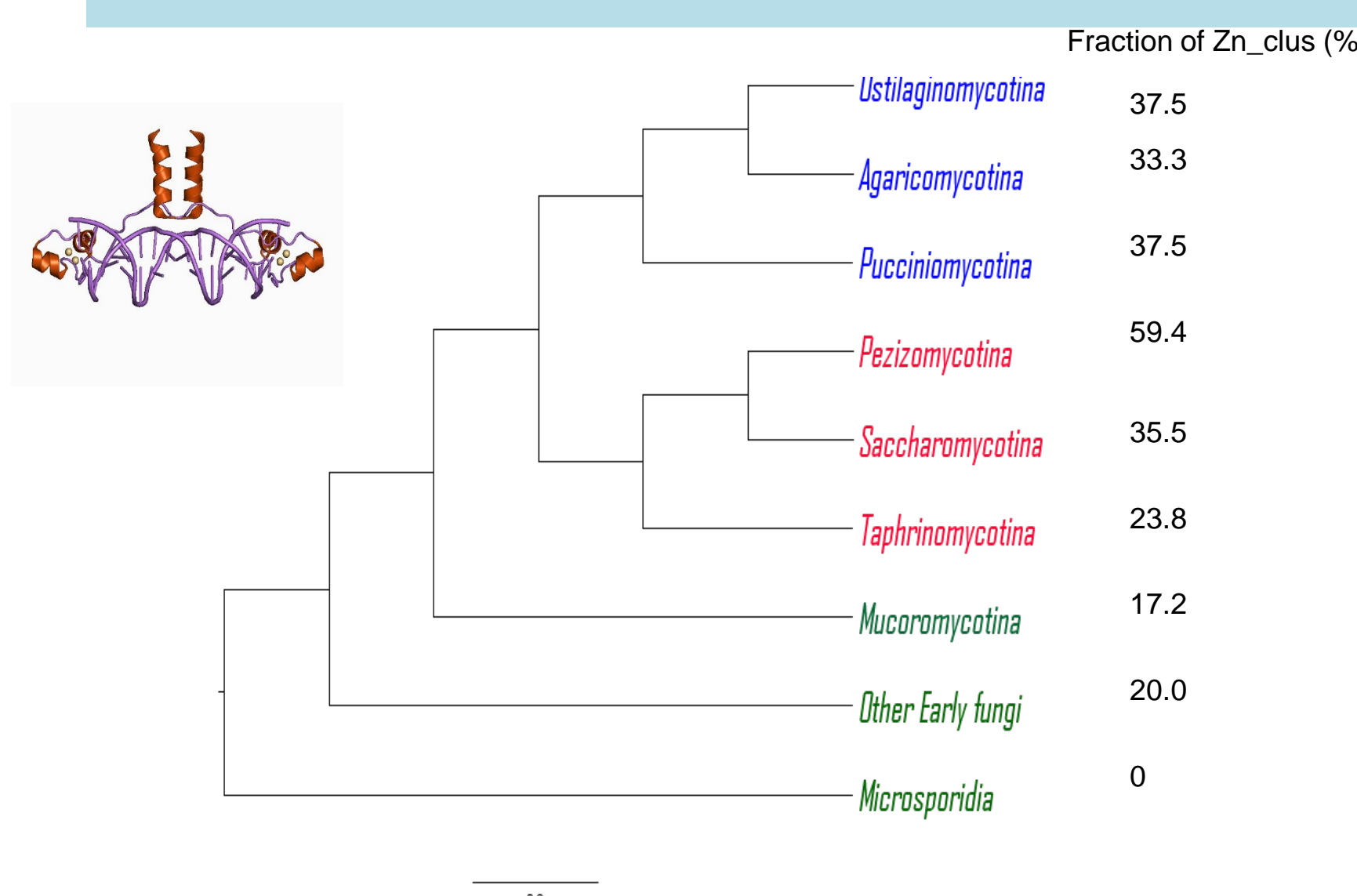


Median number of TFs per genome ~245, 0.5-6% per proteome

## Top fungal transcription factor families

DBD-domain	TF family
Zn_clus, Fungal_trans, Fungal_trans_2	Fungal-specific Zn cluster
HLH	Basic helix-loop-helix
Homeobox	Homeodomain
CBFD_NFYB_HMF	Histone-like TFs
bZIP_1, bZIP_2	Leucine zipper
zf-C2H2	Zinc finger
GATA	GATA zinc finger
HMG_box	High-mobility group
HSF_DNA_bind	Heat-shock factor
Fork-head	Fork-head protein
SRF-TF	SRF-type TF

## Fungal-specific binuclear Zn-cluster superfamily – the largest fungal TF family accounting for ~50% of all TFs



Correlation of numbers per genome of Zn\_clus TFs with numbers of transporters and SM backbone genes

	Zn_clus	Fungal_trans	Fungal_trans_2	av.corr	Other 58 TFs range
MFS_1	0.89	0.92	0.87	0.08	-0.25 - 0.64
Sugar_tr	0.90	0.92	0.89	0.08	-0.22 - 0.65
PKS	0.68	0.70	0.68	0.02	-0.28 - 0.52
NRPS	0.45	0.49	0.46	-0.02	-0.027 - 0.42
Total SM	0.62	0.67	0.59	0.03	-0.28 - 0.58
Total number of genes	0.23	0.24	1	0.21	-0.31 - 0.63

## Colocalization of Zn\_clus TFs with transporters and SM backbone genes

In most genomes Zn\_clus transcription factors statistically significantly ( $P < 0.05$ ) located adjacent to transporter genes (MFS, Sugar\_tr)

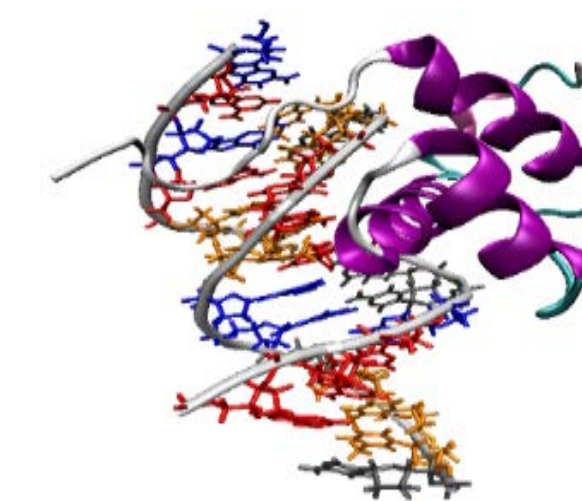
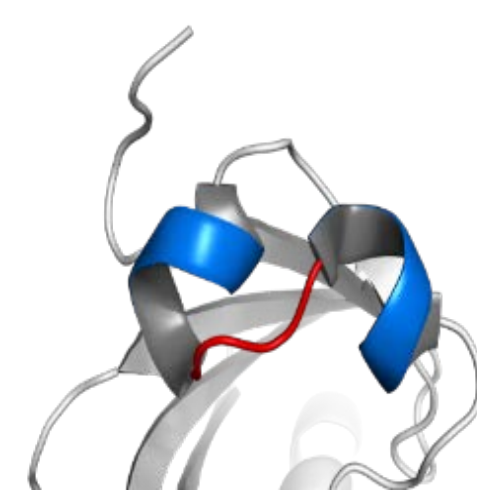
Clade	Number of genomes (fraction)
Pezizomycotina	161 (81.3%)
Saccharomycotina	19 (61.3%)
Agaricomycotina	6 (5.7%)
Ustilaginomycotina	4 (50.0%)
Pucciniomycotina	3 (20%)
Taphrinomycotina	3 (37.5%)

In many Pezizomycotina genomes Zn\_clus transcription factors statistically significantly ( $P < 0.05$ ) located within 20 genes to PKS/NRPS genes

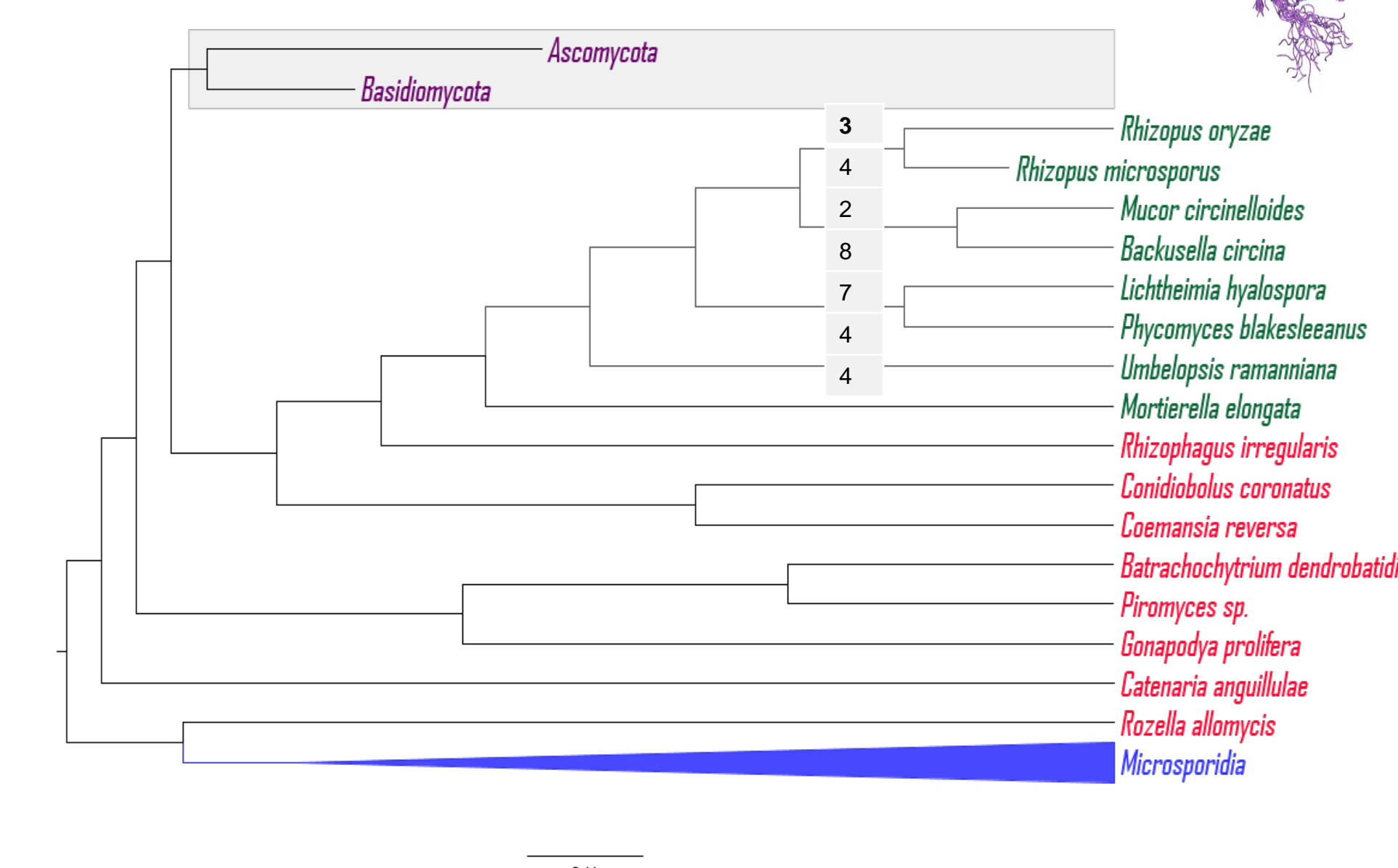
Clade	Number of genomes (fraction)
Pezizomycotina	53 (26.8%)
Agaricomycotina	9 (8.5%)
Neocallimastigo	1
Taphrinomycotina	1 (12.5%)

## Most of universal eukaryotic TFs are expanded in Mucoromycotina

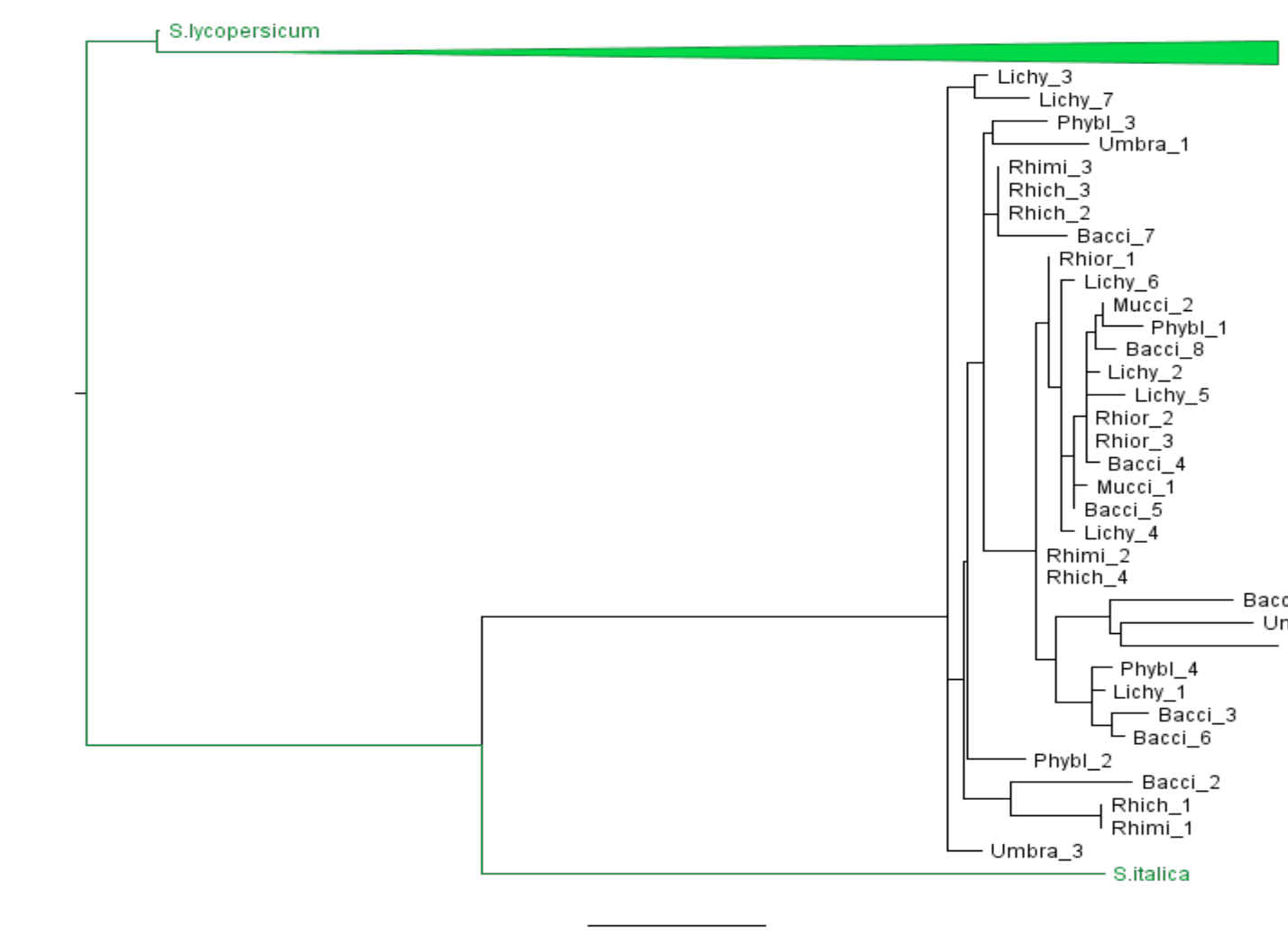
	Fungi	Mucor.	Metazoan	Plants	Protists
HLH	8	37	42	145	0
Homeobox	5	26	96	102	7
bZIP_1	5	16	15	87	10
bZIP_2	4	16	19	83	7
GATA	4	21	33	0	0
HSF	2	16	2	25	15
Fork-head	2	7	18	0	0
SRF-TF	2	9	2	89	1
NDT80_Phog	1	6	1	0	0



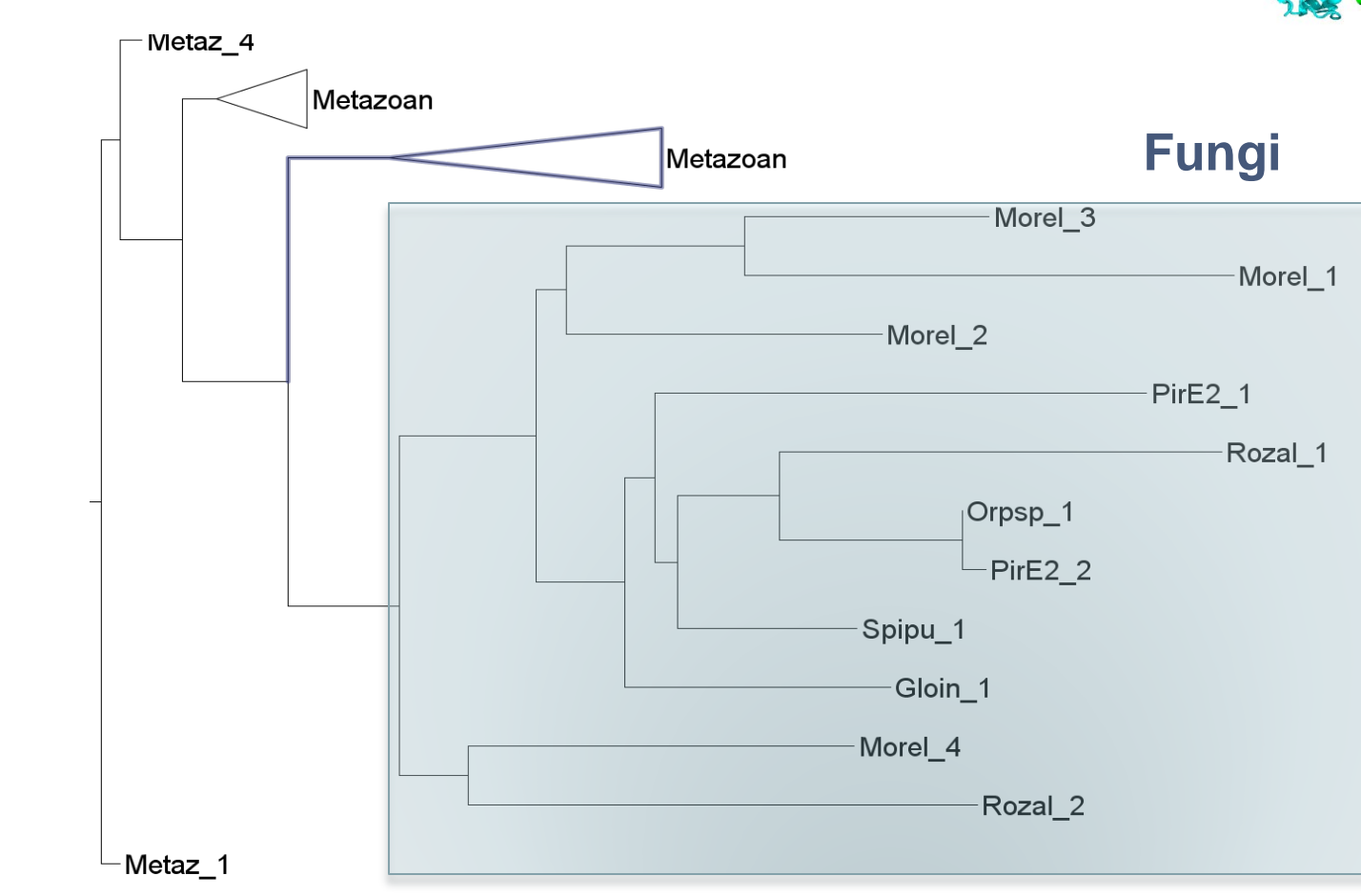
## Plant-specific WRKY transcription factors are present in 2-8 copies in Mucomycotina genomes but absent in other fungal clades



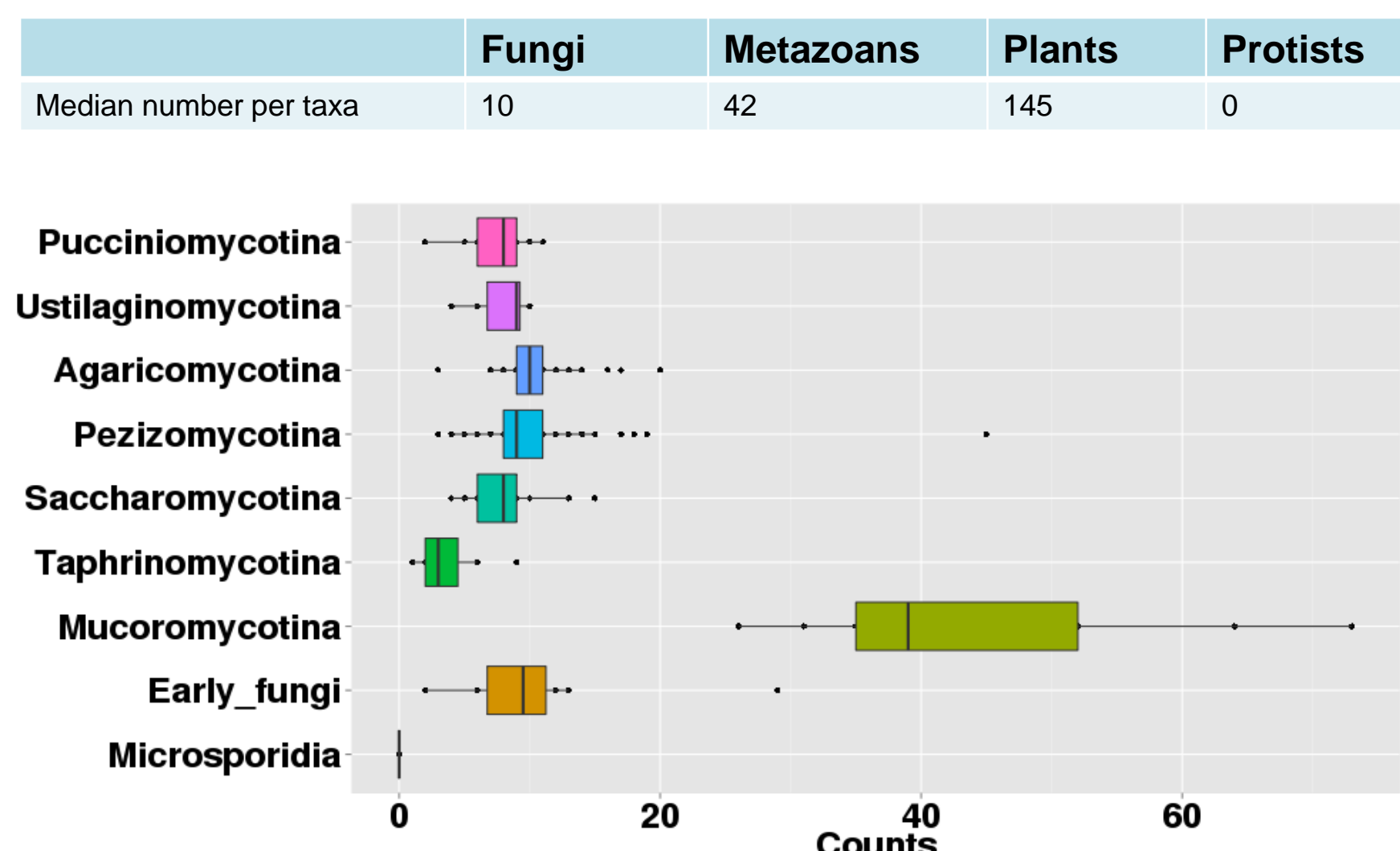
WRKY gene tree with closest plant sequences (horizontal transfer)



Animal-specific T-box TFs, involved in limb and heart development are present at some early-diverging fungi

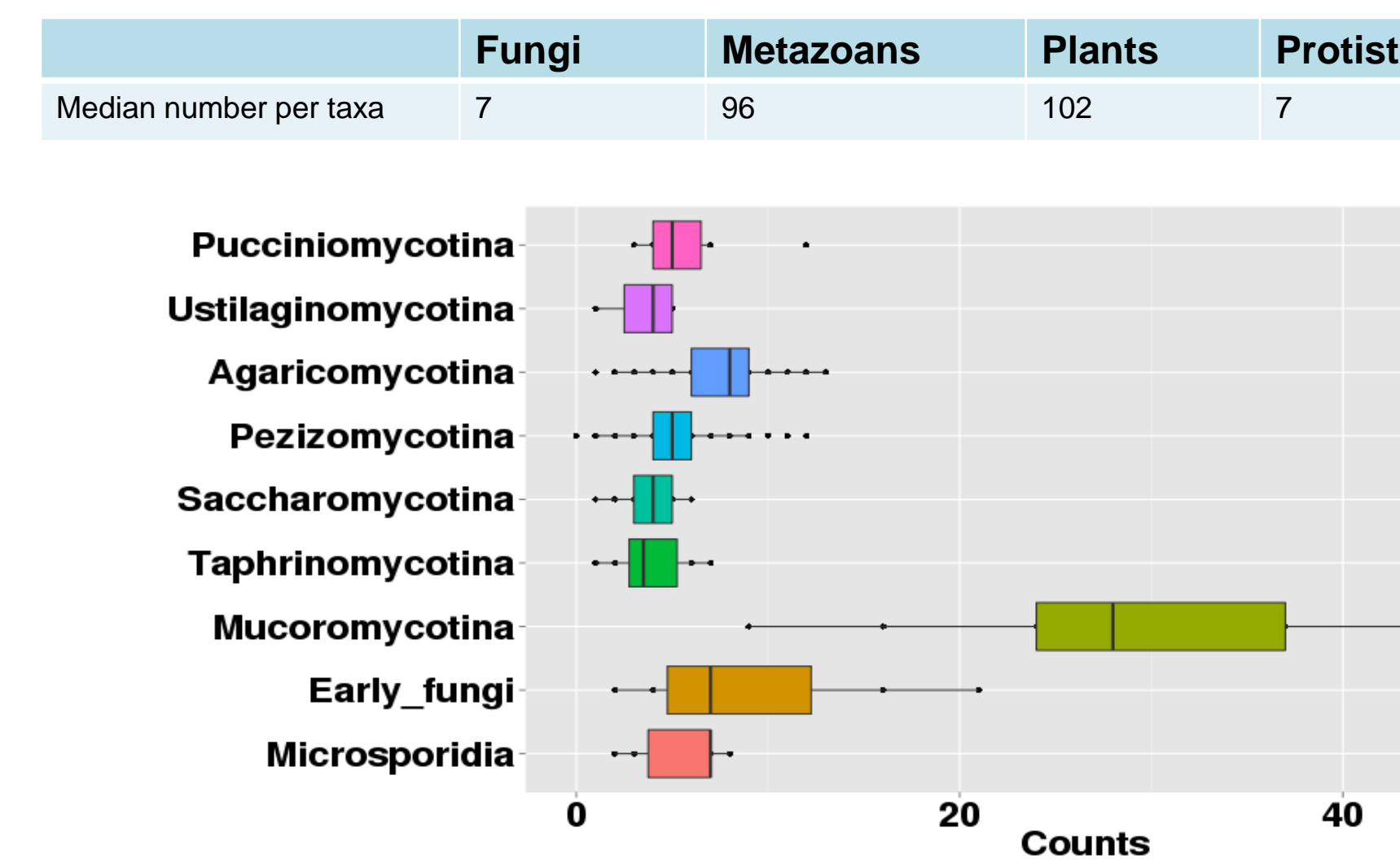


## Eukaryotic Basic helix-loop-helix (HLH) TF family



Mucoromycotina clade gained +8 genes ( $P = 0.001$ , CAFÉ)

## Homeobox (TFs with homeodomain fold, found for example in animal Hox genes) family



Mucoromycotina clade gained +5 genes ( $P = 0.004$ , CAFÉ)

## SUMMARY

- Fungal-specific Binuclear zinc finger family (Zn\_clus) constitutes the largest transcription factor family in the fungal kingdom, accounting for ~60% of all TFs in Pezizomycotina

- At least some fraction of Zn\_clus TFs seem to be expanded in complex with other genes in SM clusters predominantly in Pezizomycotina clade

- Most of universal eukaryotic TF families are expanded in Mucoromycotina

- Early divergent fungi have plant and animal specific TFs absent in Dikarya