

# Fungal phylogenomics: Getting lost in the moldy forest

---

Jason E Stajich  
Dept of Plant Pathology & Microbiology  
University of California, Riverside

<http://lab.stajich.org>

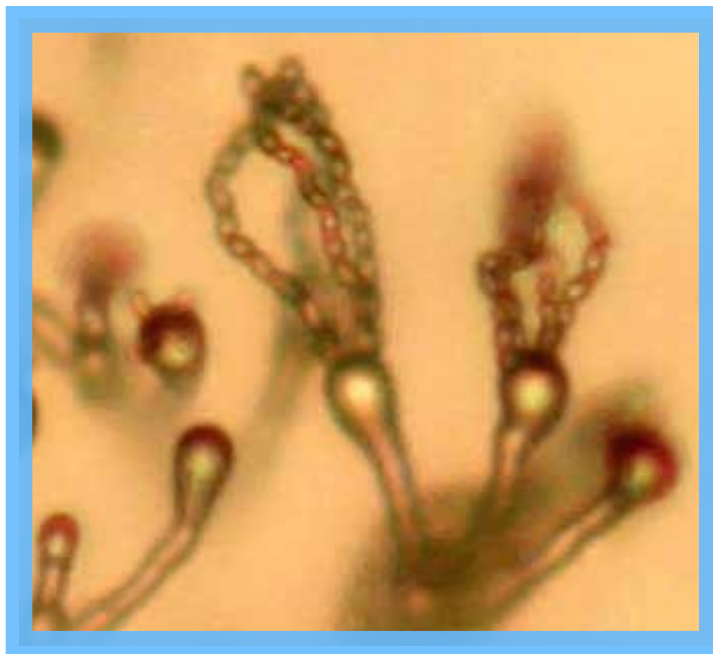
<http://fungalgenomes.org/blog>

<http://fungiDB.org>

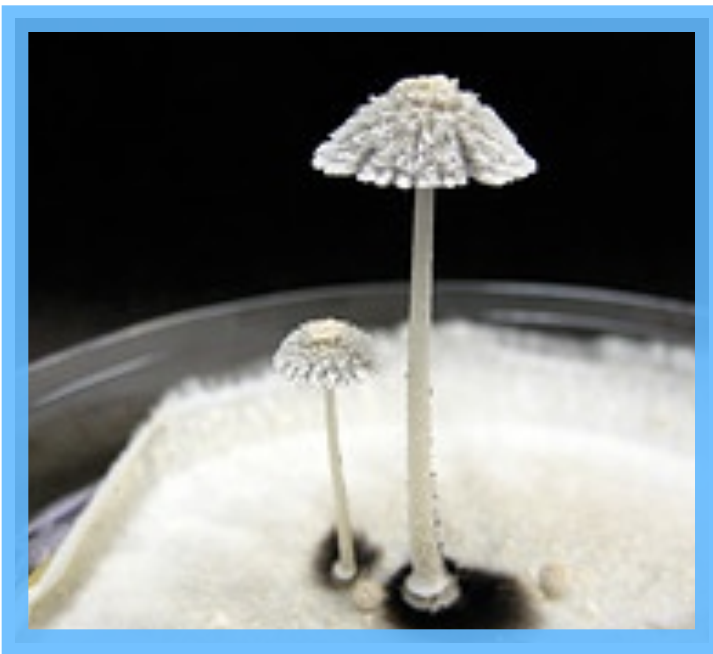
twitter{stajichlab,hyphaltip,fungalgenomes,fungidb}



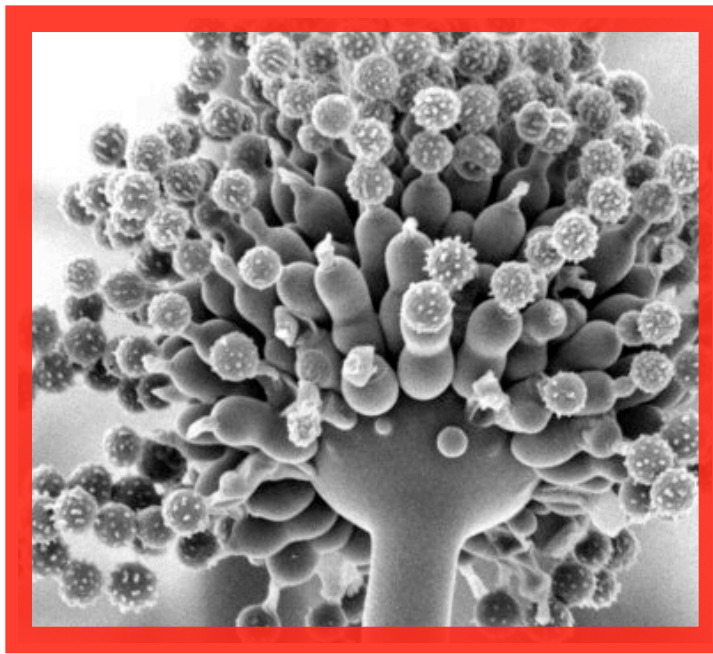
# Fungi have diverse forms, ecology, and associations



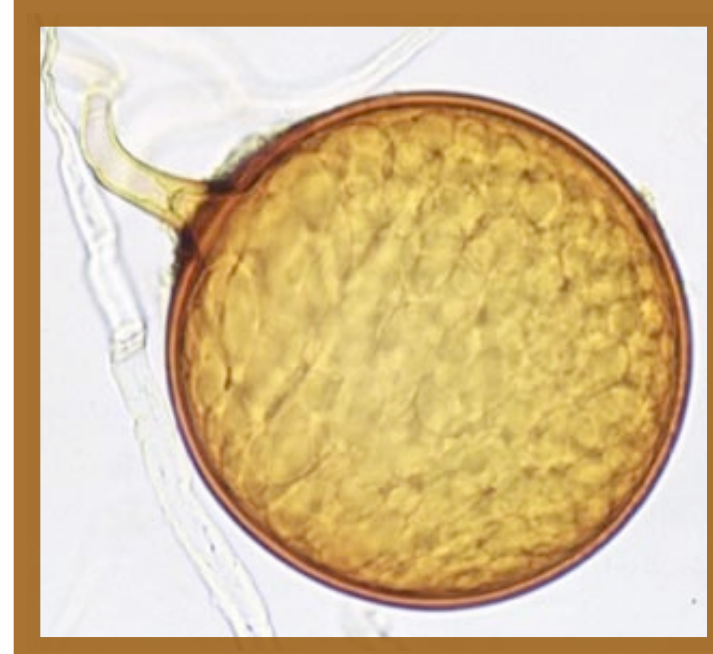
*Cryptococcus neoformans* X. Lin



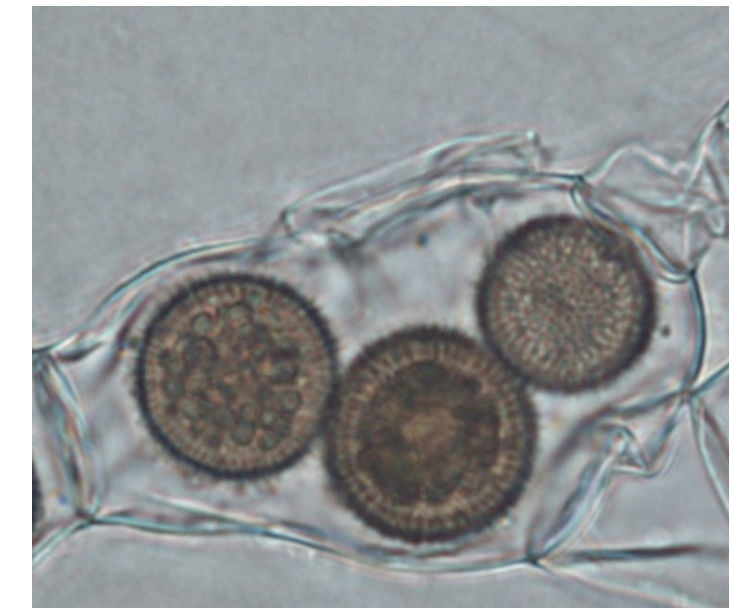
*Coprinopsis cinerea* Ellison & Stajich



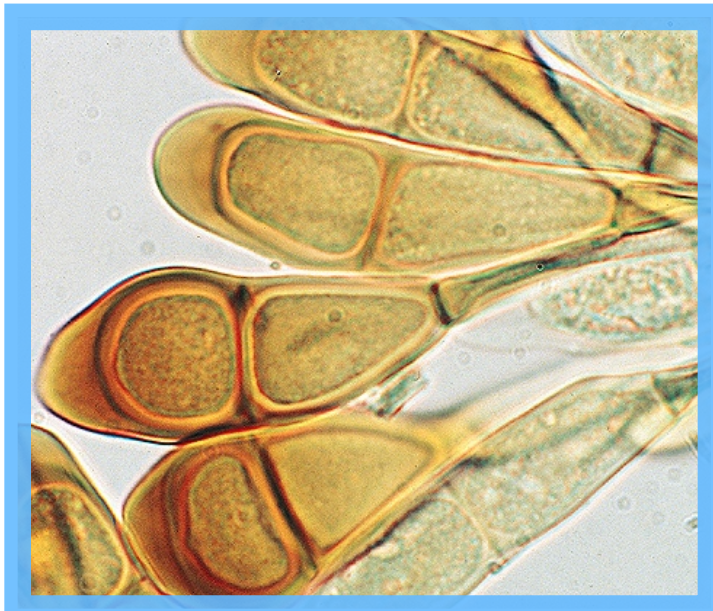
*Aspergillus niger*. N Read



*Glomus* sp. Univ Sydney



*Rozella allomyces*. James et al



*Puccinia graminis* J. F. Hennen



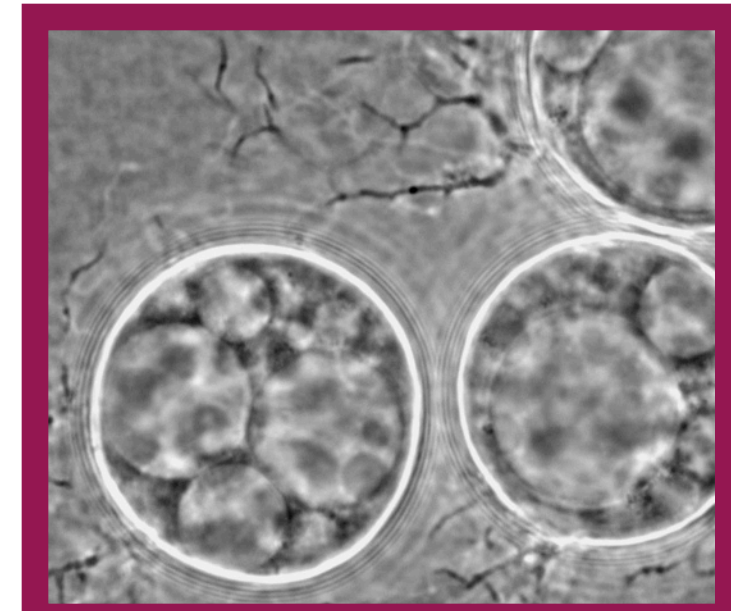
*Laccaria bicolor* Martin et al.



*Neurospora crassa*. Hickey & Reed



*Phycomyces blakesleanus* T. Ootaki



*Batrachochytrium dendrobatidis*  
J. Longcore



*Ustilago maydis* Kai Hirdes



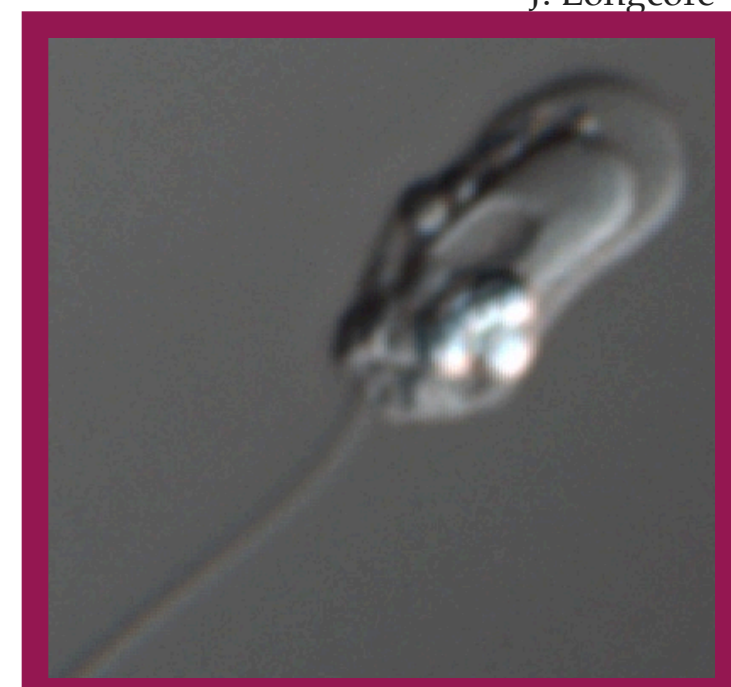
*Amanita phalloides*. M Wood



*Xanthoria elegans*. Botany POTD



*Rhizopus stolonifera*.

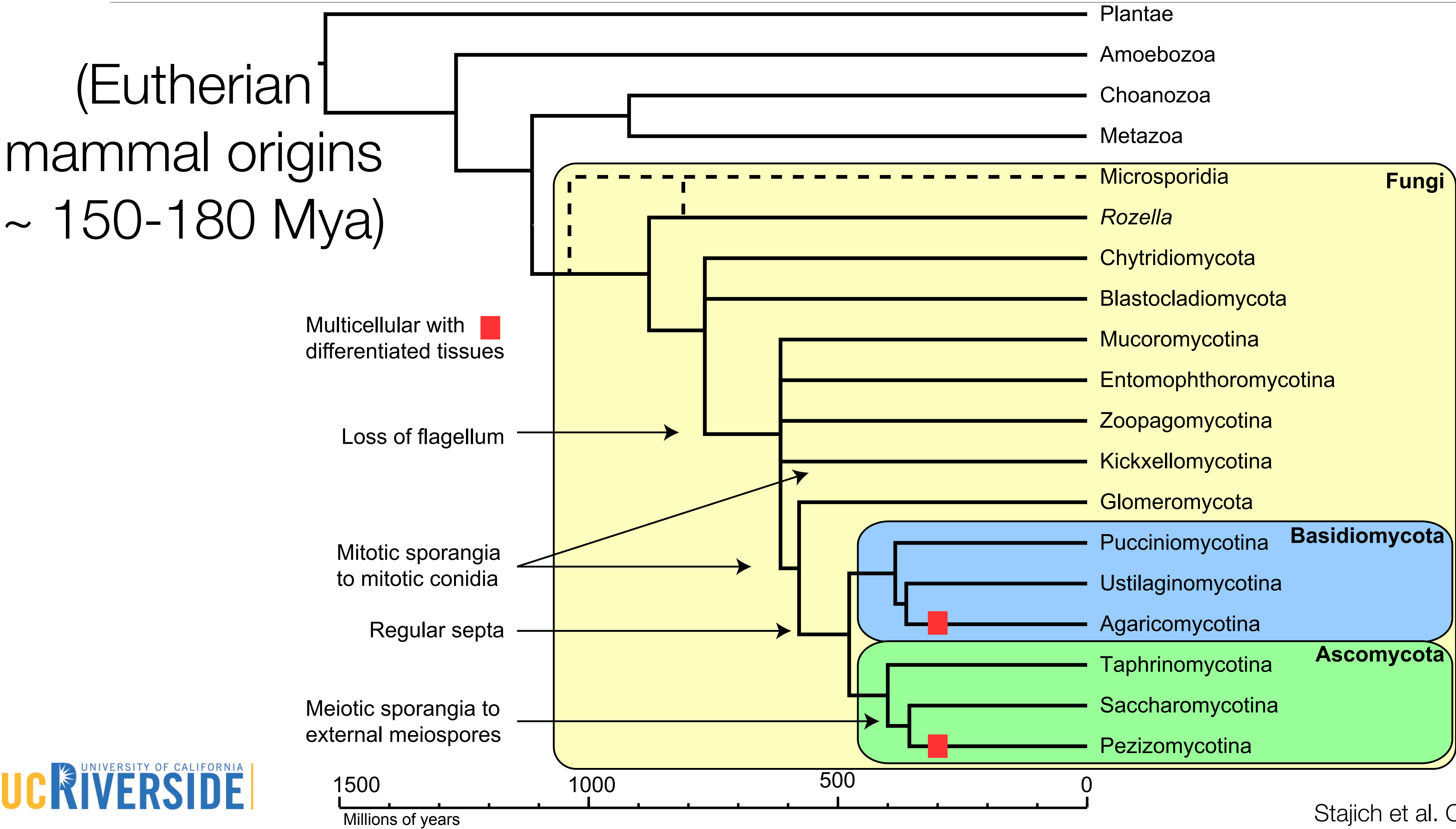


*Blastocadiella simplex* Stajich & Taylor



# Fungi are an old group of organisms

(Eutherian  
mammal origins  
~ 150-180 Mya)



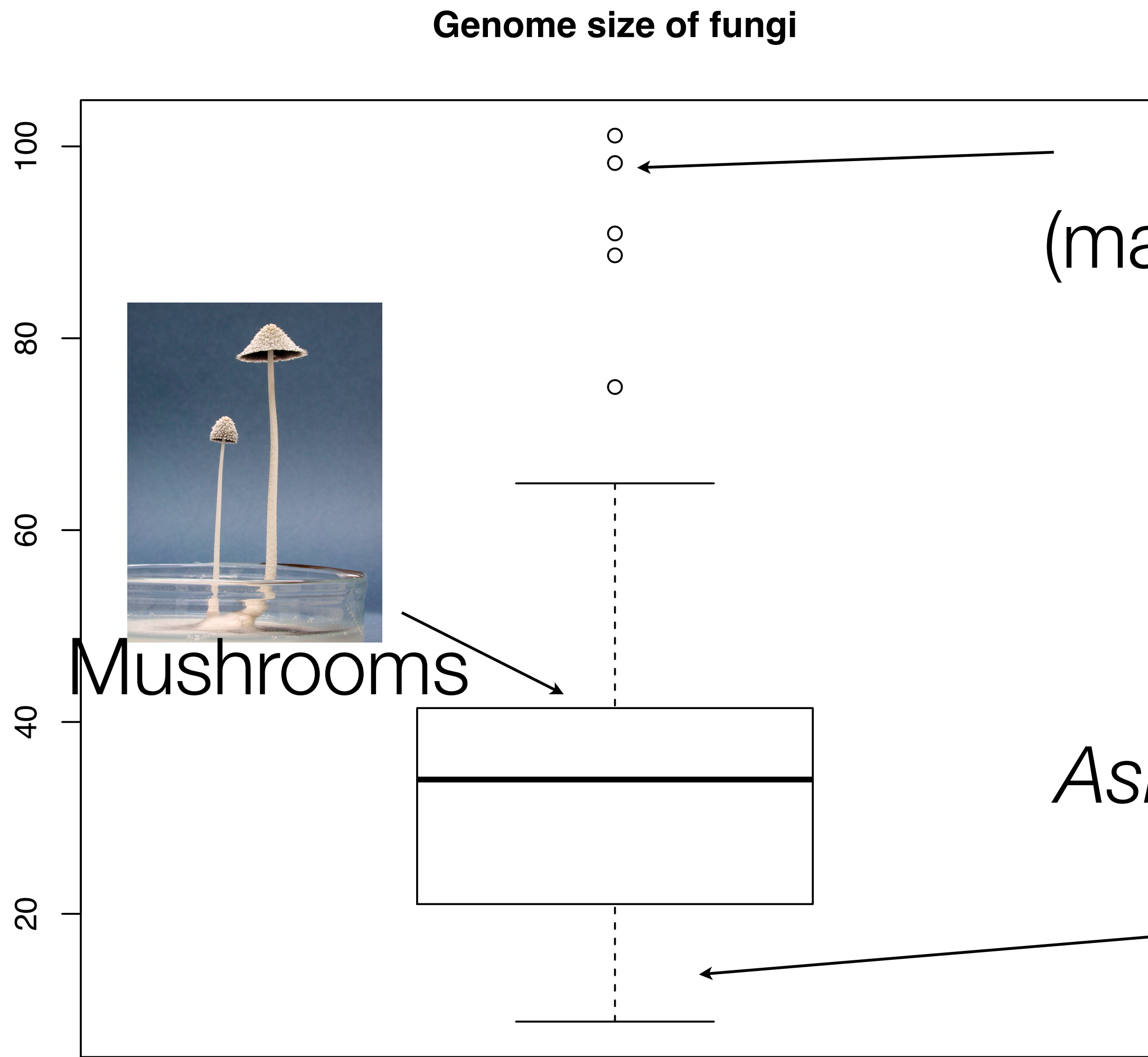
# Awash in Fungal Genomes

---

- Mid-2011: Genomes from 150+ species sequenced and ~400 more in progress/pipeline
- Several multi-strain resequencing projects (*Neurospora* - Ellison PNAS 2011), *Saccharomyces* (Liti Nature 2009), *Coccidoides* (Neafsey Genome Res 2010) and many in progress/proposed.
- Many of sequenced genomes were focused around plant, animal pathogens, and some specific evolutionary questions.
- Now are starting to fill out the tree more to capture the diversity of kingdom and also for studies of molecular evolution among many related species.
- Also efforts targeting specific questions - pathogens and their relatives; wood rotting fungi; flagellated and non-flagellated forms; extremophiles; comparison of growth forms (e.g. yeast forms from phylogenetically very different species)



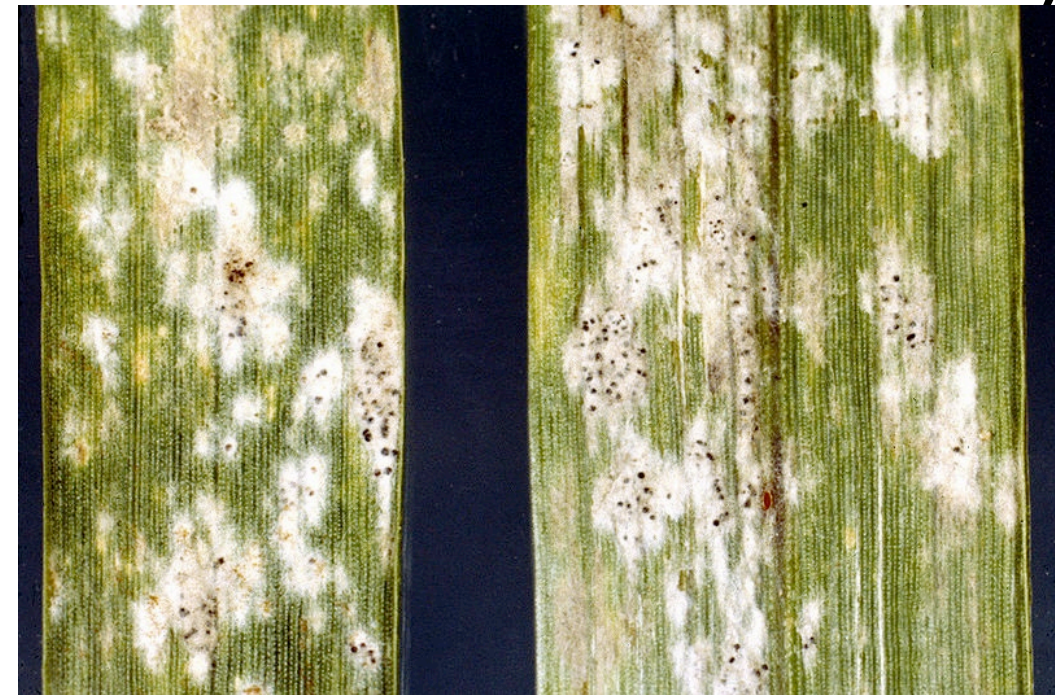
Genome sizes of fungi are 8Mb-100Mb = easy(ish) to sequence



*Melampsora, Blumeria, Puccinia*  
(many Transposable elements)

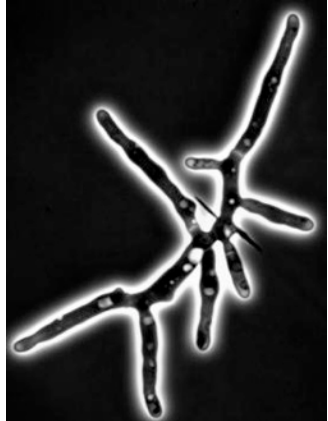


Francis Martin  
[Flickr:fmartin1954/2338045504/](https://www.flickr.com/photos/fmartin1954/2338045504/)



<http://en.wikipedia.org/wiki/File:Barleypowderymildew.jpg>

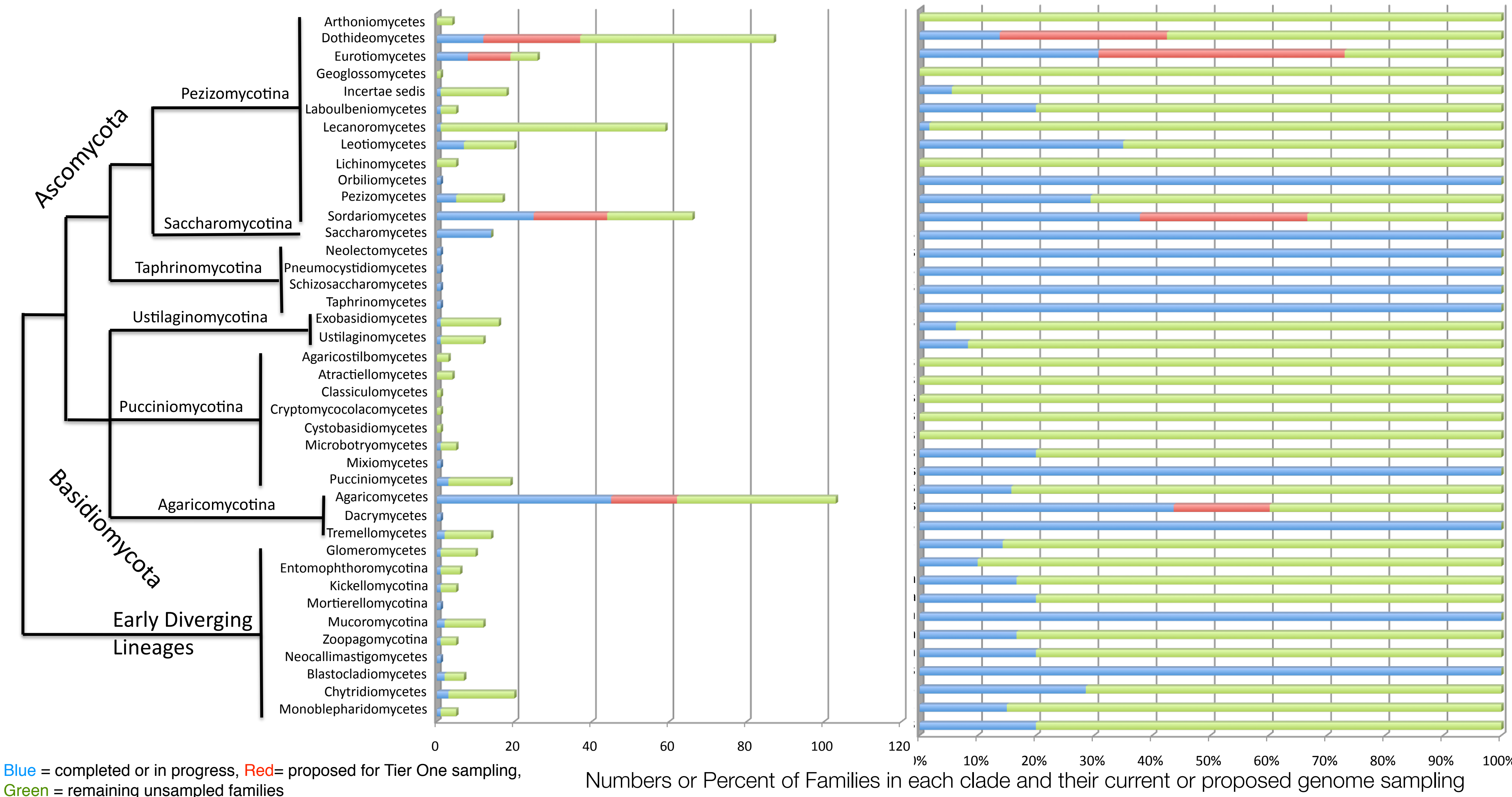
*Ashbya, Malassezia, Candida*  
other yeasts



<http://www.biologie.uni-osnabrueck.de/Genetik/index.php?menuid=12>



# Need to cover more of the phylogenetic diversity: 1000 Fungal genomes project





# Tools to access this data

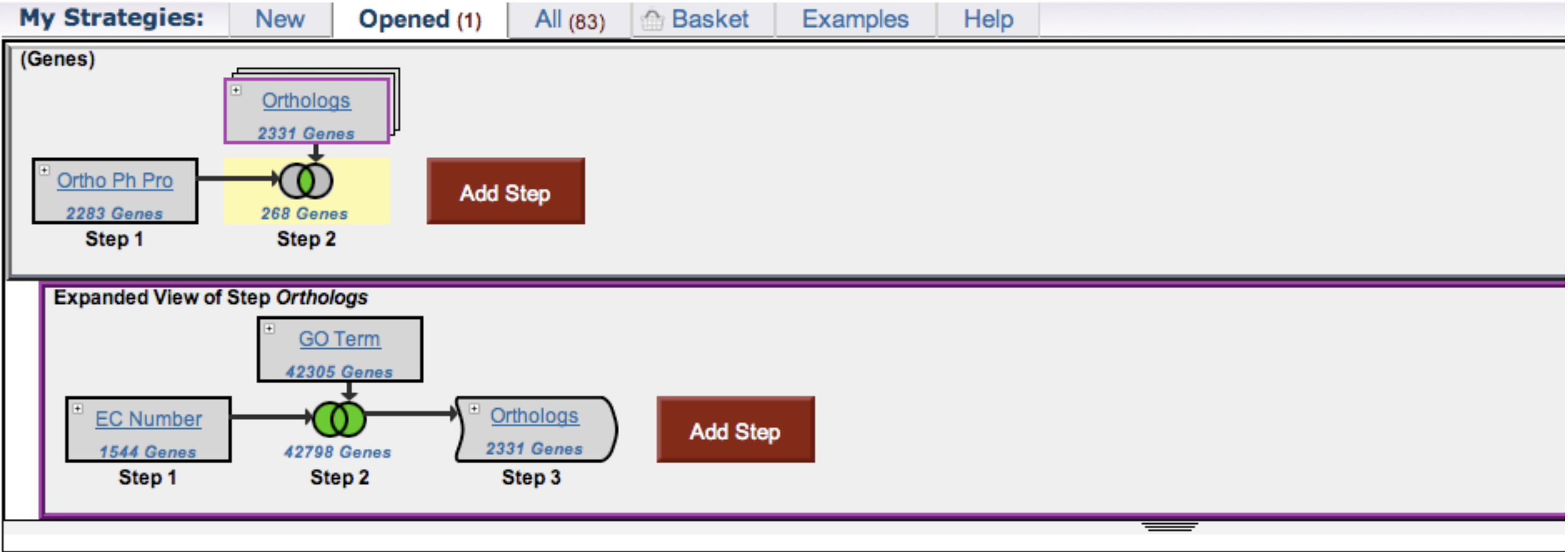
---

- Genome databases at sequencing centers and NCBI
- Comprehensive and integrated systems are not as well developed in Fungi, but some examples of excellent tools highlighted at <http://tools.fungalgenomes.org/>
- Ensembl Fungi, MicrobesOnline, JGI's Mycocosm, Comparative Fungal Genomics, and some targeted to specific clades e.g. *Saccharomyces* (SGD), *Aspergillus* (AspGD), *Candida* (CGD)
- We have launched FungiDB - <http://fungidb.org> to support integrating functional genomics data for data mining as well as standard 'Gene' pages for genes.



# Searching with FungiDB:

## A strategy for drug targets



Filter results by species (results removed by the filter will not be combined into the next step.)

All Results	Ortholog Groups	Eurotiomycetes								Sordariomycetes					Saccharomycotina		Basidiomycota		Rhizopus oryzae
		A.cla	A.fla	A.fum	A.nid	A.nig	A.ter	C.imm H538.4	C.imm RS	F.gra	F.oxy	G.mon	M.ory	N.cra	C.alb	S.cer	C.neo	P.gra	
268	249	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	268	0	0

### Cryptococcus Drug target? - step 2 - 268 Genes

First 1 2 3 4 5 Next Last Advanced Paging

Gene	Genomic Location	Weight	Predicted GO Function	Predicted GO Process
CneoH99_CNAG_07830	CneoH99_Chr 2.4: 1,052,305 - 1,054,670 (-)	20	DNA binding, zinc ion binding	transcription
CneoH99_CNAG_07632	CneoH99_Chr 2.6: 473,070 - 476,242 (+)	20	pyridoxal phosphate binding, transaminase activity	biosynthetic process, cellular amino acid metabolic process
CneoH99_CNAG_07589	CneoH99_Chr 2.11: 160,029 - 161,779 (-)	20	binding, catalytic activity, oxidoreductase activity, zinc ion binding	metabolic process, oxidation reduction
CneoH99_CNAG_07441	CneoH99_Chr 2.5: 1,287,517 - 1,288,271 (-)	20	4-alpha-hydroxytetrahydrobiopterin dehydratase activity	tetrahydrobiopterin biosynthetic process
CneoH99_CNAG_07011	CneoH99_Chr 2.12: 762,209 - 764,308 (+)	20	DNA binding, zinc ion binding	transcription
CneoH99_CNAG_06868	CneoH99_Chr 2.5: 29,970 - 32,039 (-)	20	phosphopyruvate hydratase activity	glycolysis
CneoH99_CNAG_06719	CneoH99_Chr 2.2: 235,880 - 239,543 (+)	20	DNA binding, zinc ion binding	transcription

*C.neoformans* genes with

- no hits in other eukaryotes (excepting fungi), bacteria
- has *S.cerevisiae* orthologs

that have EC terms

OR orthologs in other fungi

with GO metabolic terms associated



# Searching with FungiDB:

A strategy for reannotation

2885 *C.neoformans* genes with desc “conserved hypothetical protein” but 1327 can be assigned name from ortholog in any fungus with EC #, GO terms, or Pfam domains



☐ Filter results by species (results removed by the filter will not be combined into the next step.)

All Results	Ortholog Groups	Eurotiomycetes								Sordariomycetes					Saccharomycotina		Basidiomycota		Rhizopus oryzae
		A.cla	A.fla	A.fum	A.nid	A.nig	A.ter	C.imm H538.4	C.imm RS	F.gra	F.oxy	G.mon	M.ory	N.cra	C.alb	S.cer	C.neo	P.gra	
1327	1229	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1327	0	0

## Annotation(2) - step 4 - 1327 Genes

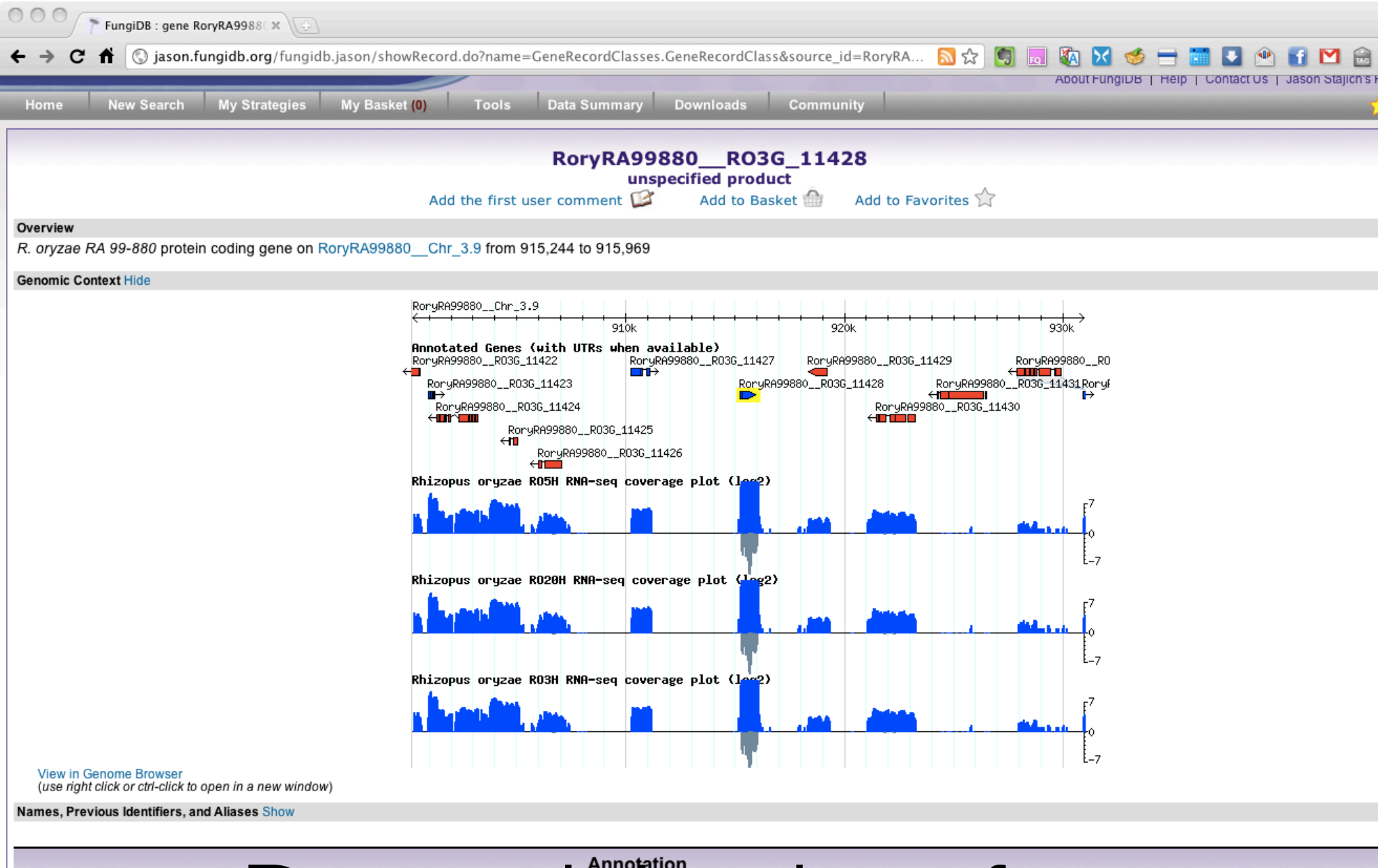
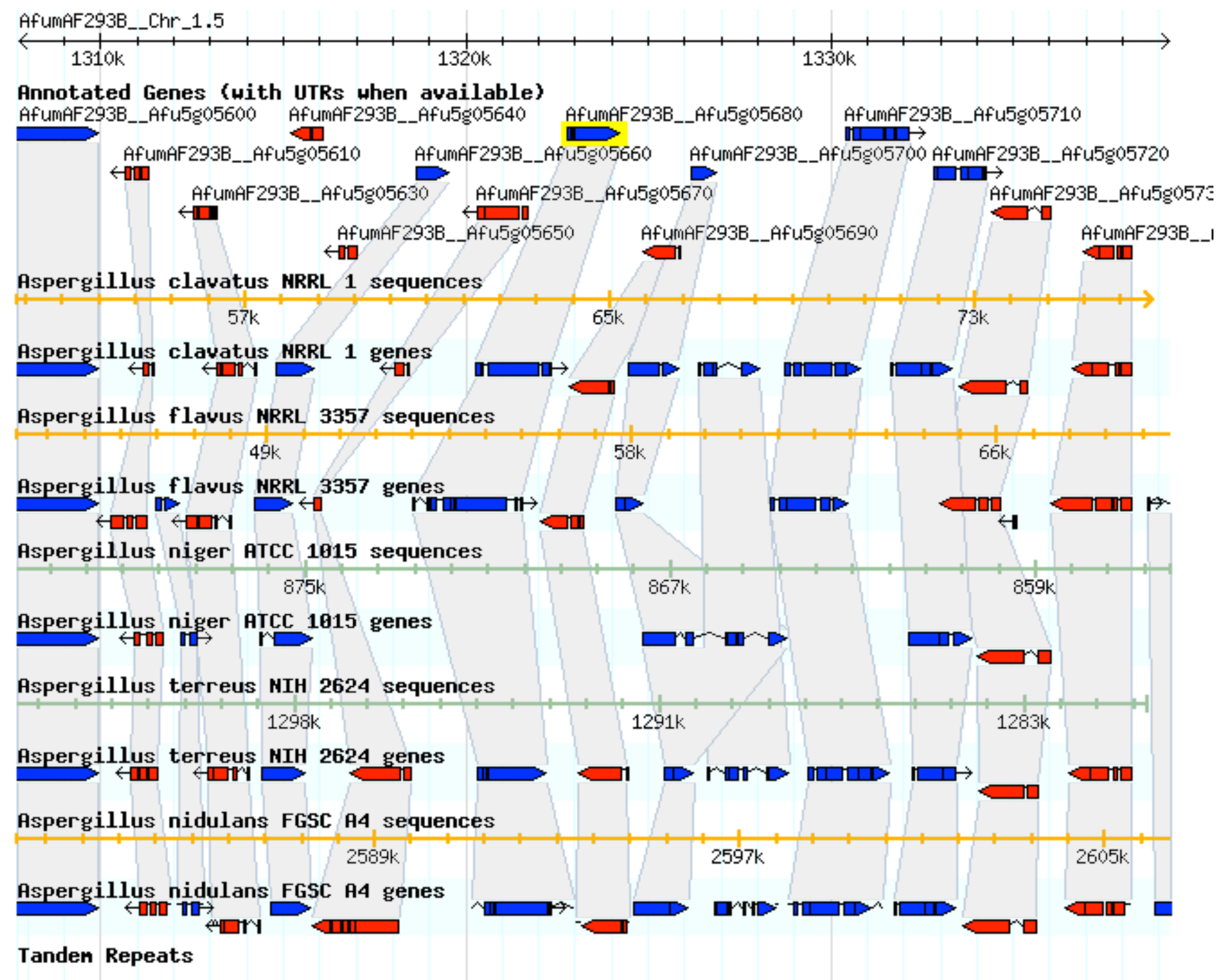
[Add 1327 Genes to Basket](#) | [Download 1327 Genes](#)

First 1 2 3 4 5 Next Last      Advanced Paging      [Select Columns](#)      [Reset Columns](#)

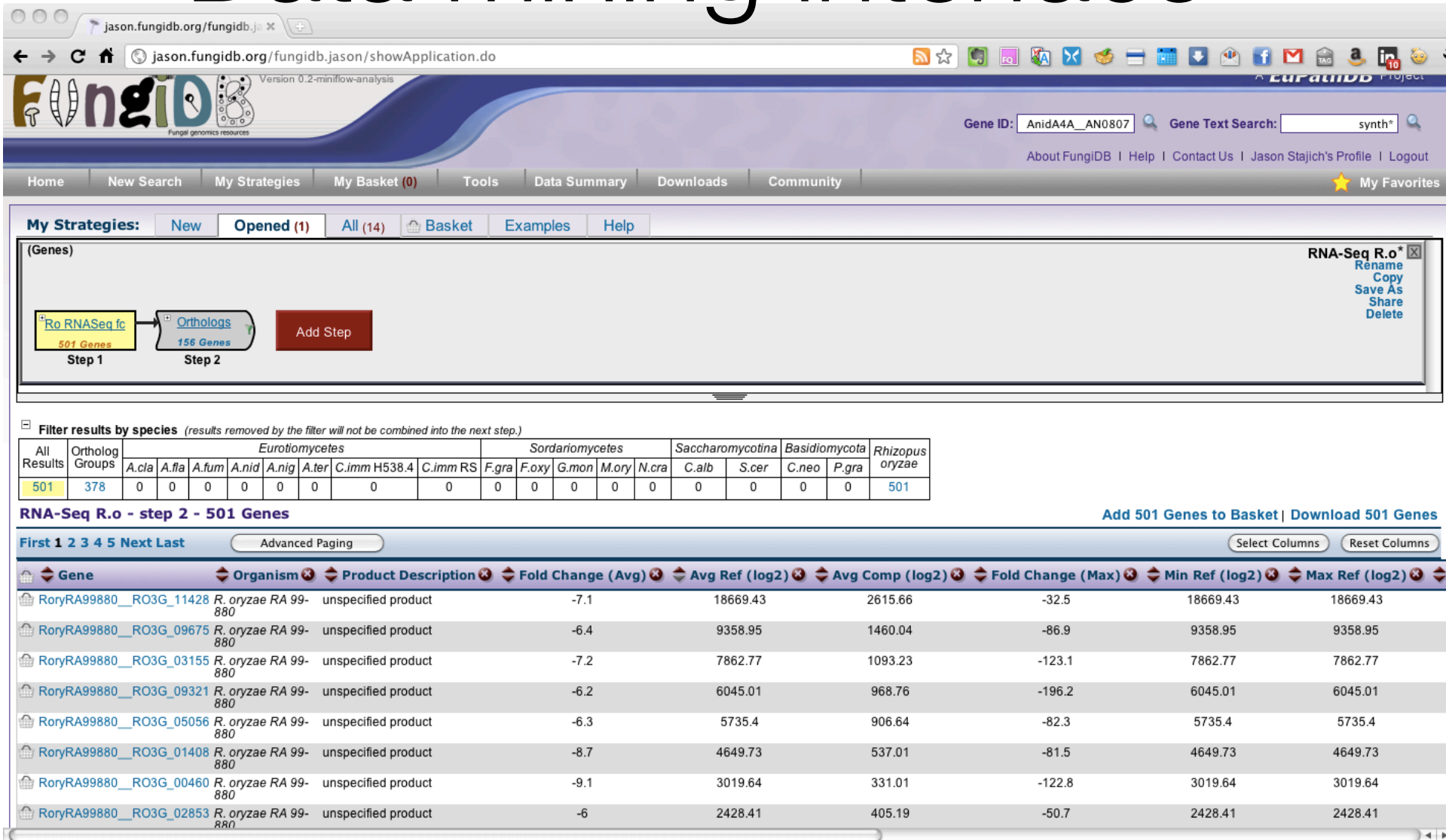
Gene	Genomic Location	Weight	Predicted GO Function	Predicted GO Process	Ortholog Group	SignalP Peptide
CneoH99__CNAG_07913	CneoH99__Chr. 2.12: 637,812 - 638,653 (+)	10	N/A	N/A	OG5_226248	N/A
CneoH99__CNAG_07912	CneoH99__Chr. 2.12: 636,724 - 637,155 (+)	10	N/A	N/A	OG5_226247	N/A
CneoH99__CNAG_07898	CneoH99__Chr. 2.12: 191,256 - 192,798 (+)	10	N/A	N/A	OG5_226246	N/A
CneoH99__CNAG_07890	CneoH99__Chr. 2.14: 796,716 - 797,600 (+)	10	N/A	N/A	OG5_226245	HMM: MSISKSPSLDLCLATLHELLHPSPILSLLL SALDLSAHFQLFAQQLSSDAALVSAL, NN: MSISKSPSLDLCLATLHELLHPSPILSLLL SAL...
CneoH99__CNAG_07883	CneoH99__Chr. 2.14: 482,566 - 484,632 (-)	10	N/A	N/A	OG5_226244	N/A

# Gene page and genome browser

## Synteny Views



## Data mining interface





# Using comparative genomics towards understand pathogen evolution

---

- How do traits like pathogenecity evolve?
- Can comparative genomics indicate meaningful differences that can lead to understanding the basis for pathogenesis?
- Contrasting genomes of pathogens with non-pathogens can suggest recent genomic changes that might be testable in the lab
- Gene duplication is thought to be important source for evolutionary innovation - What role might gene family size change play in adaptation?

# Models for comparing gene family size changes

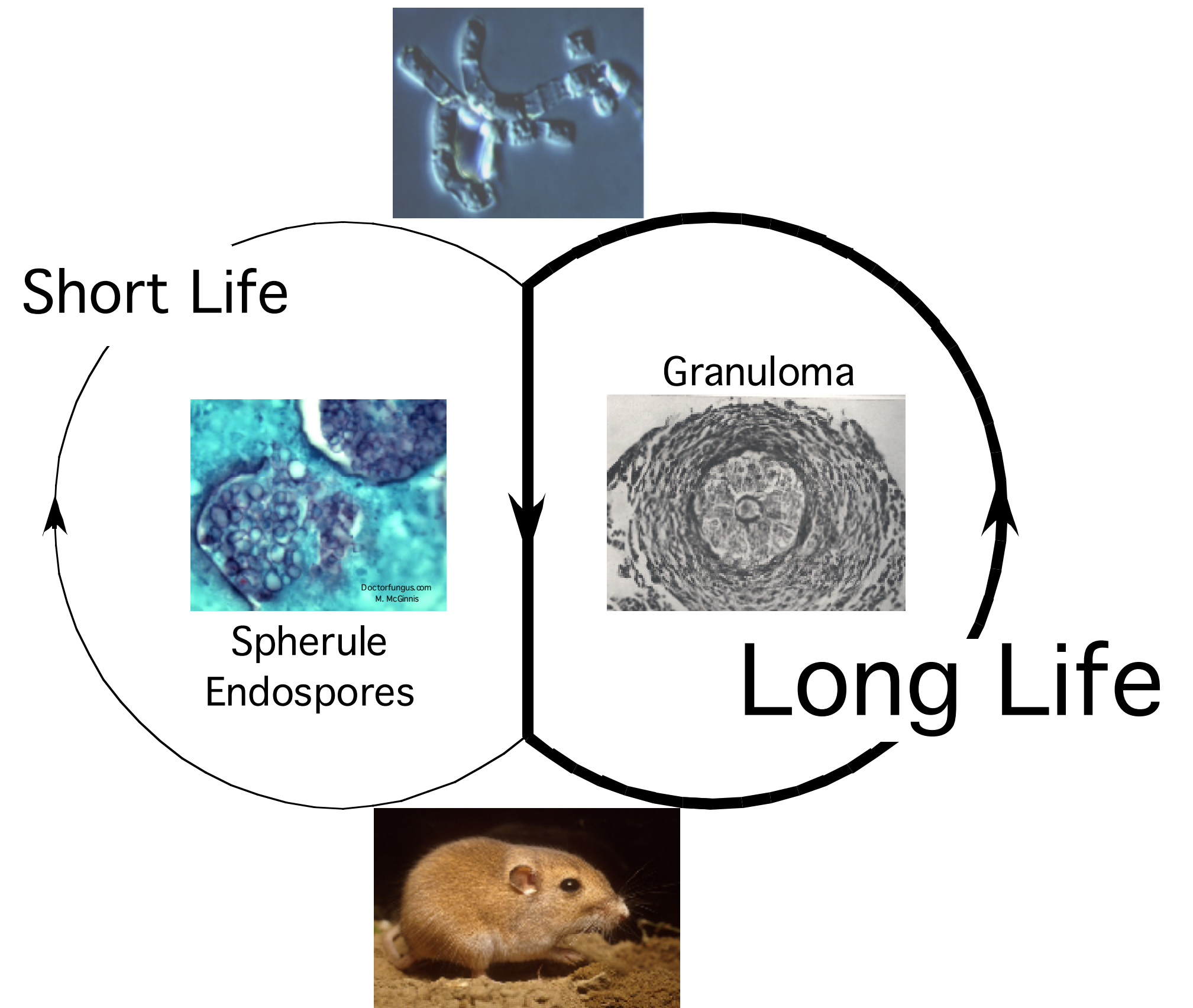
---

- A model for gene family size change that incorporates the birth-death process of gene families which follow a power law distribution (Hahn et al, Genome Res 2005)
- Implemented in a program called CAFE to find unexpectedly large lineage or clade-specific changes in gene family sizes (Hahn Lab @Indiana Univ) (De Bie et al, Bioinformatics 2006)
- Can screen genome family sizes across multiple species to find expectedly large changes (based on counts) which can be verified using gene tree-species tree reconciliation approaches like Notung (Chen, Durand, Farach-Colton, J Comp Bio 2000)



# *Coccidioides*

- Fungal pathogen genomics: Gene families and appetite differences



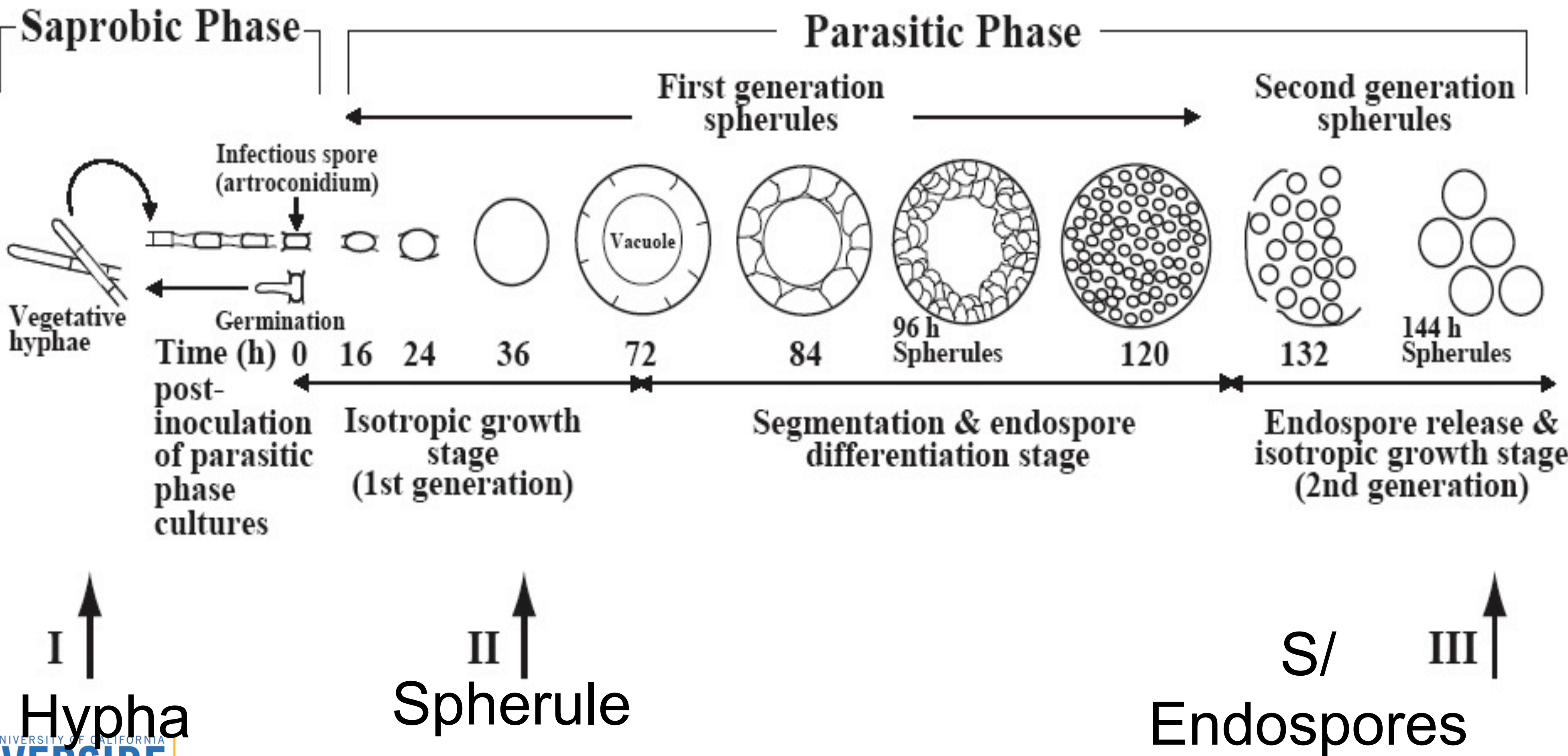
# Human pathogen *Coccidioides*

---

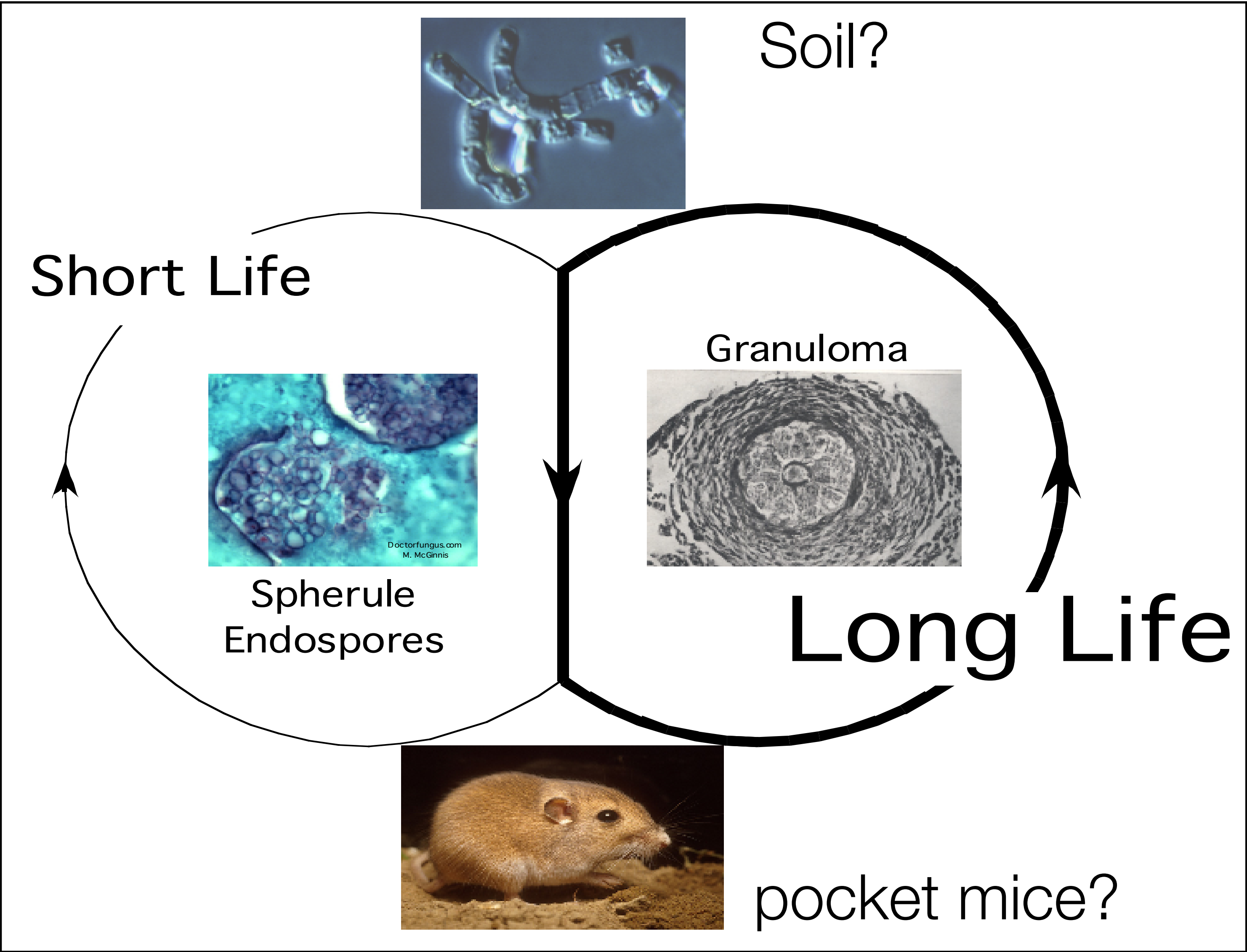
- Coccidioides (Valley fever) - 2 species *C. immitis* and *C. posadasii*
  - Is a primary human pathogen - infects healthy people (most human pathogenic fungi are opportunistic)
  - Endemic in US Southwest, Mexico
  - Requires laboratory BSL3 and is a Select Agent
  - Genomes of 2 species (Sharpton et al Genome Res 2009) and then 18 strains (Neafsey et al Genome Res 2010)
- Comparative analyses of *Coccidioides* spp



# Human pathogen *Coccidioides* Life cycle

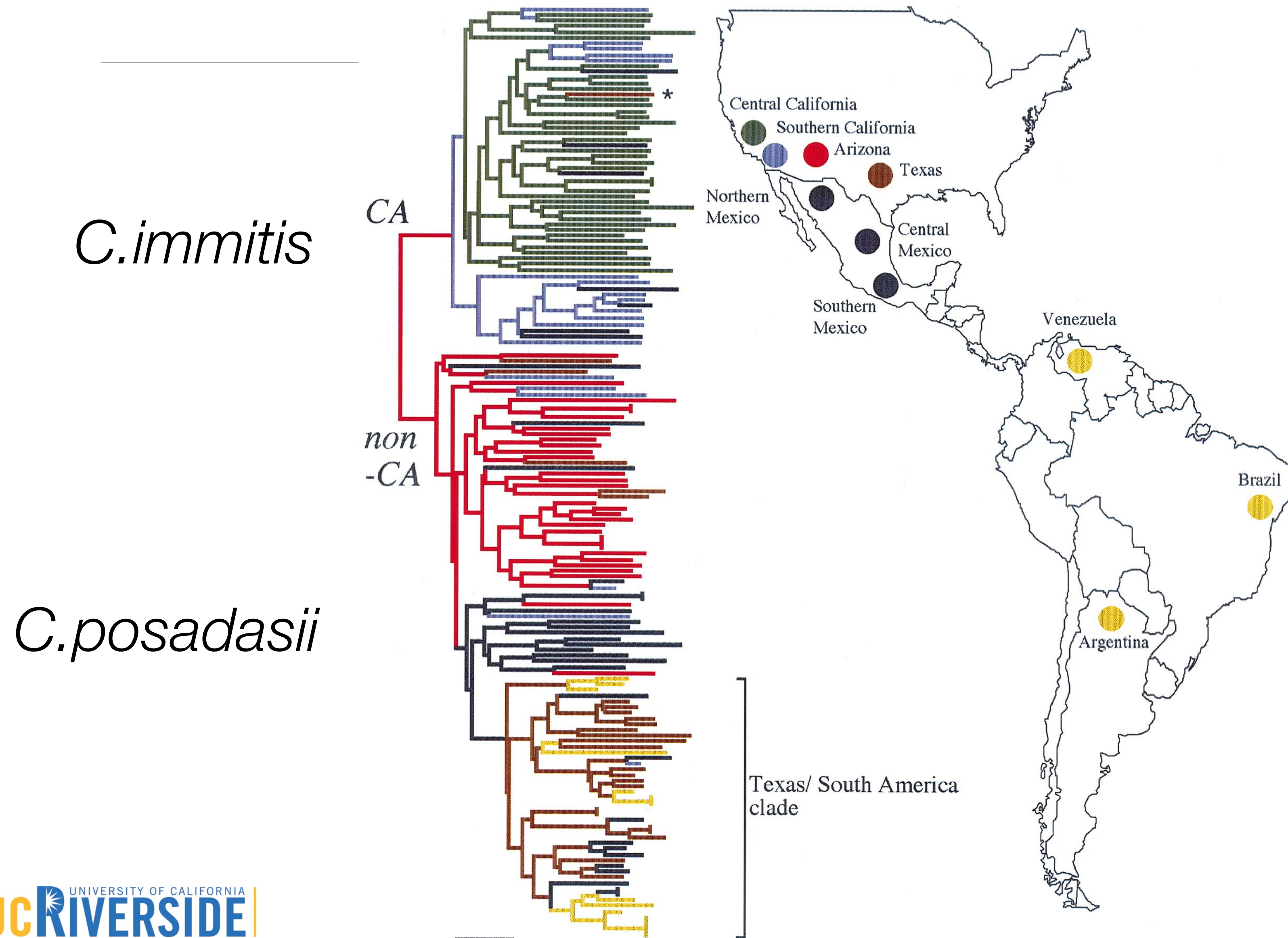


# Coccidioides ecology





# Two species of *Coccidioides* are allopatric



Fisher et al, 2000

# Studying the evolution of a pathogen

---

- Comparing sequences from two *Coccidioides* species, closely related outgroup, and more distant outgroups species:
- Evidence for recent positive selection
- What gene family differences can be identified that distinguish phenotypic groups (mammalian pathogens from non-pathogens)
- Evidence for recent introgression which contains candidate genes for pathogenesis

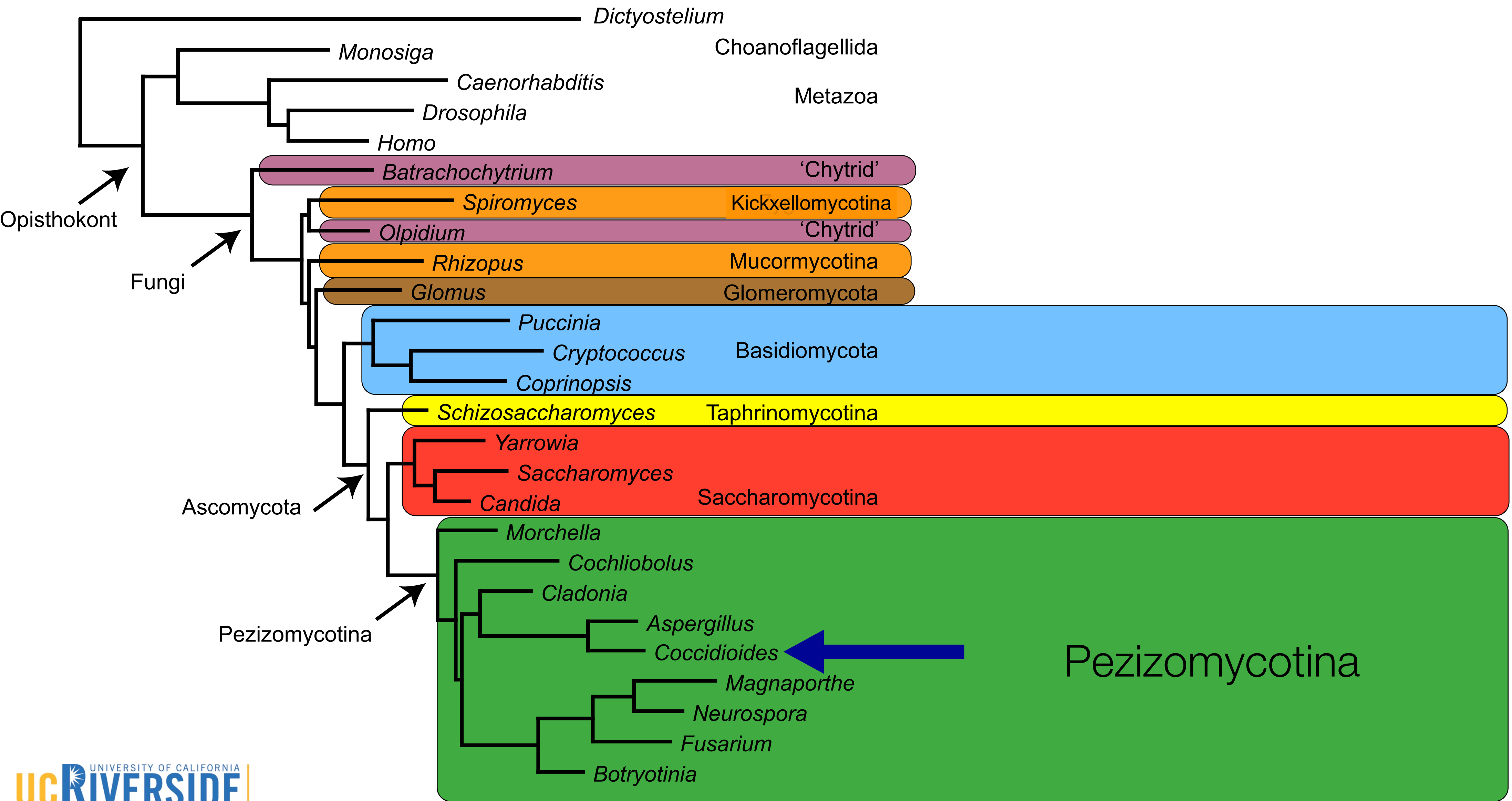


# Gene family changes

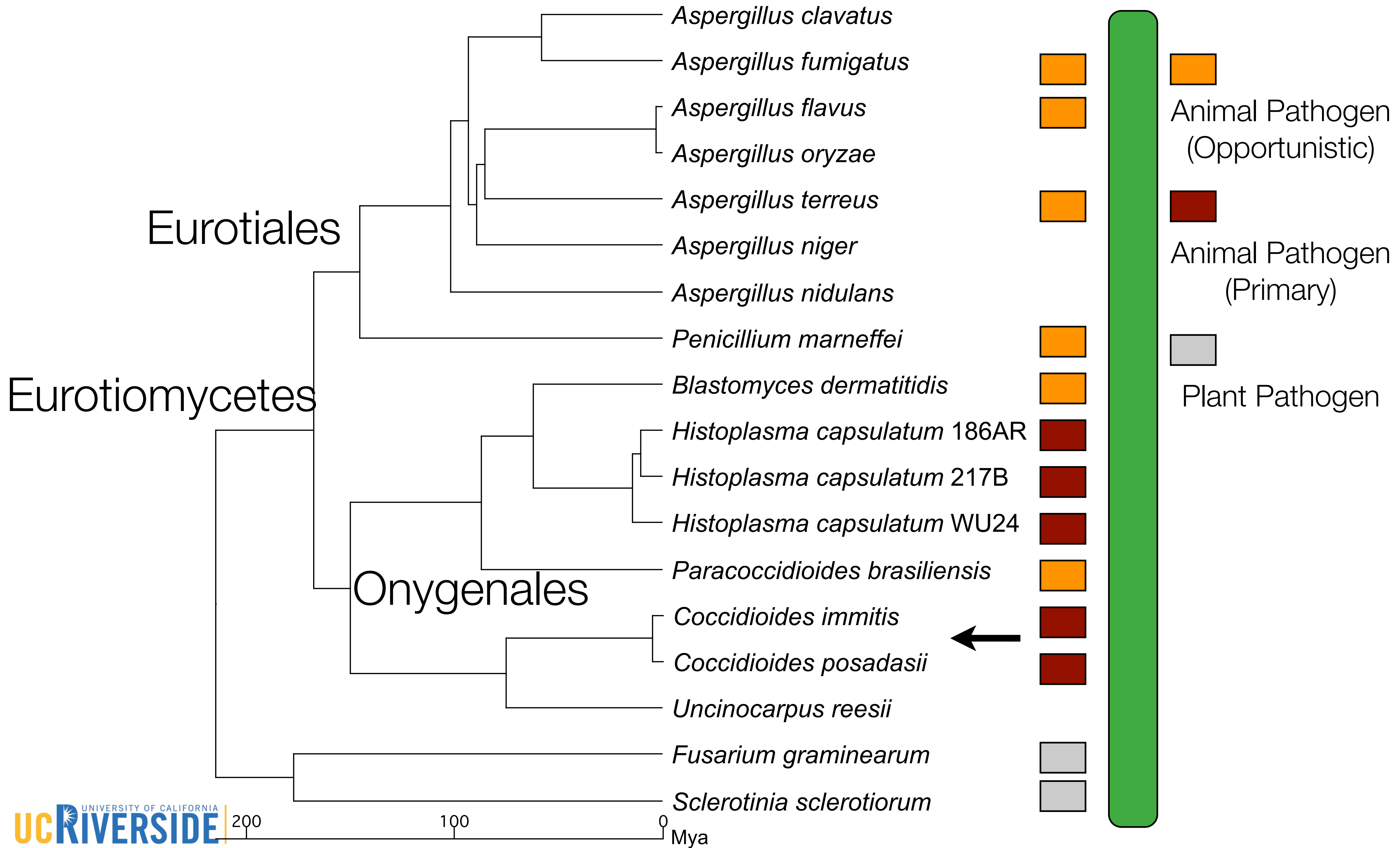
---

- Another mechanism for adaptation may be changes in copy number of a gene family
  - Gene duplication is a source of novelty allowing for changes in the function of one copy if the other maintains original function
  - Expansions of copy number may also be an easy way to get more protein for a particular process
- How important is copy number change in adaptation?

# Genome samples from fungi







# Few protein domains for eating plant material in animal

## pathogens

Animal Pathogen  
(Opportunistic)

Animal Pathogen  
(Primary)

Plant Pathogen

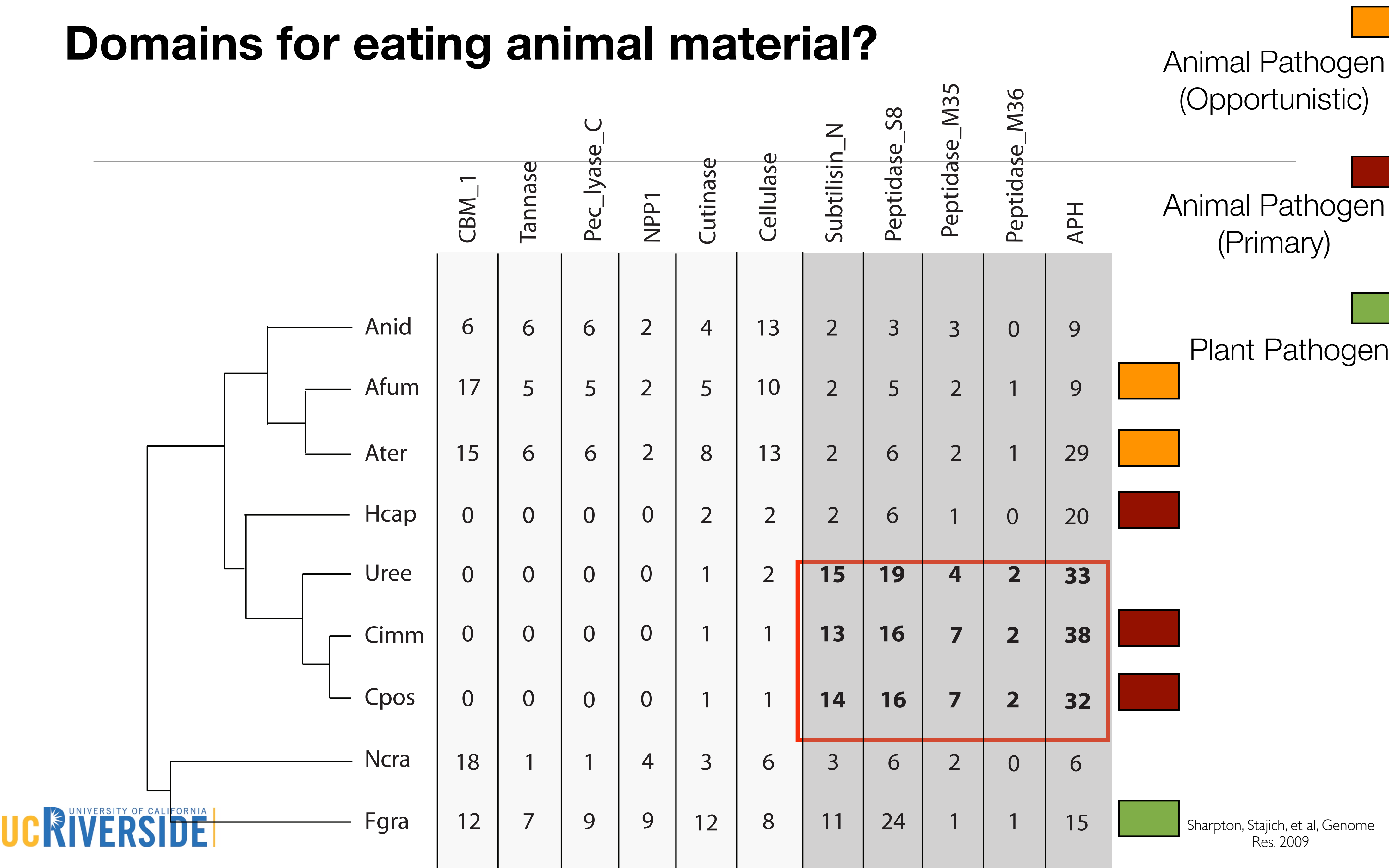
Sharpton, Stajich, et al, Genome  
Res. 2009

Loss of plant  
saprophytic  
enzymes

		CBM_1	Tannase	Pec_lyase_C	NPP1	Cutinase	Cellulase	Subtilisin_N	Peptidase_S8	Peptidase_M35	Peptidase_M36	APH
	Anid	6	6	6	2	4	13	2	3	3	0	9
	Afum	17	5	5	2	5	10	2	5	2	1	9
	Ater	15	6	6	2	8	13	2	6	2	1	29
	Hcap	0	0	0	0	2	2	2	6	1	0	20
	Uree	0	0	0	0	1	2	15	19	4	2	33
	Cimm	0	0	0	0	1	1	13	16	7	2	38
	Cpos	0	0	0	0	1	1	14	16	7	2	32
	Ncra	18	1	1	4	3	6	3	6	2	0	6
	Fgra	12	7	9	9	12	8	11	24	1	1	15



# Domains for eating animal material?



# Keratinases in Onygenales

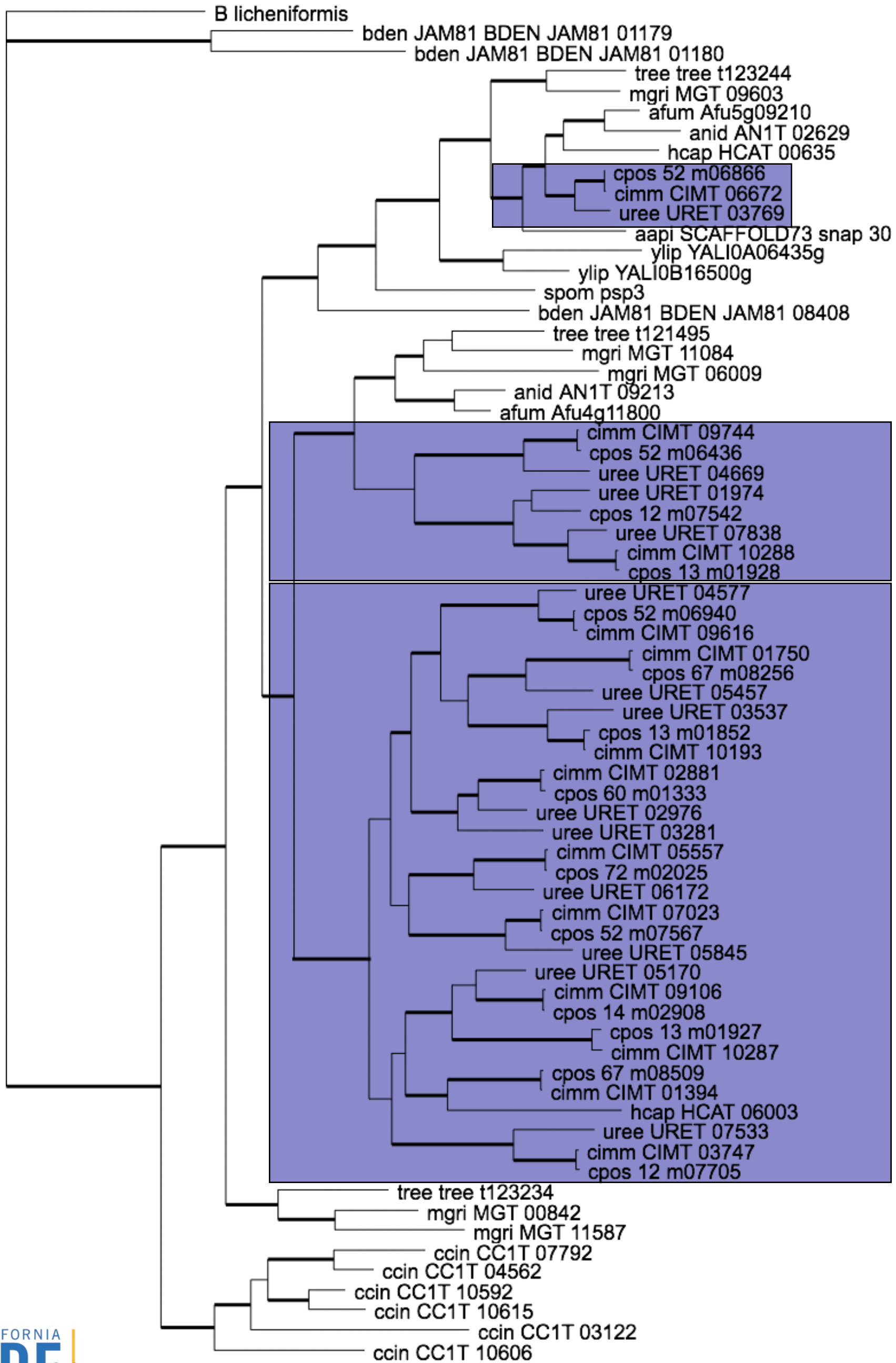


- Onygenales are Keratinophilic
- Domains: Peptidase S8, Subtilisin domains
- Large expansion of putative keratinases in Onygenales



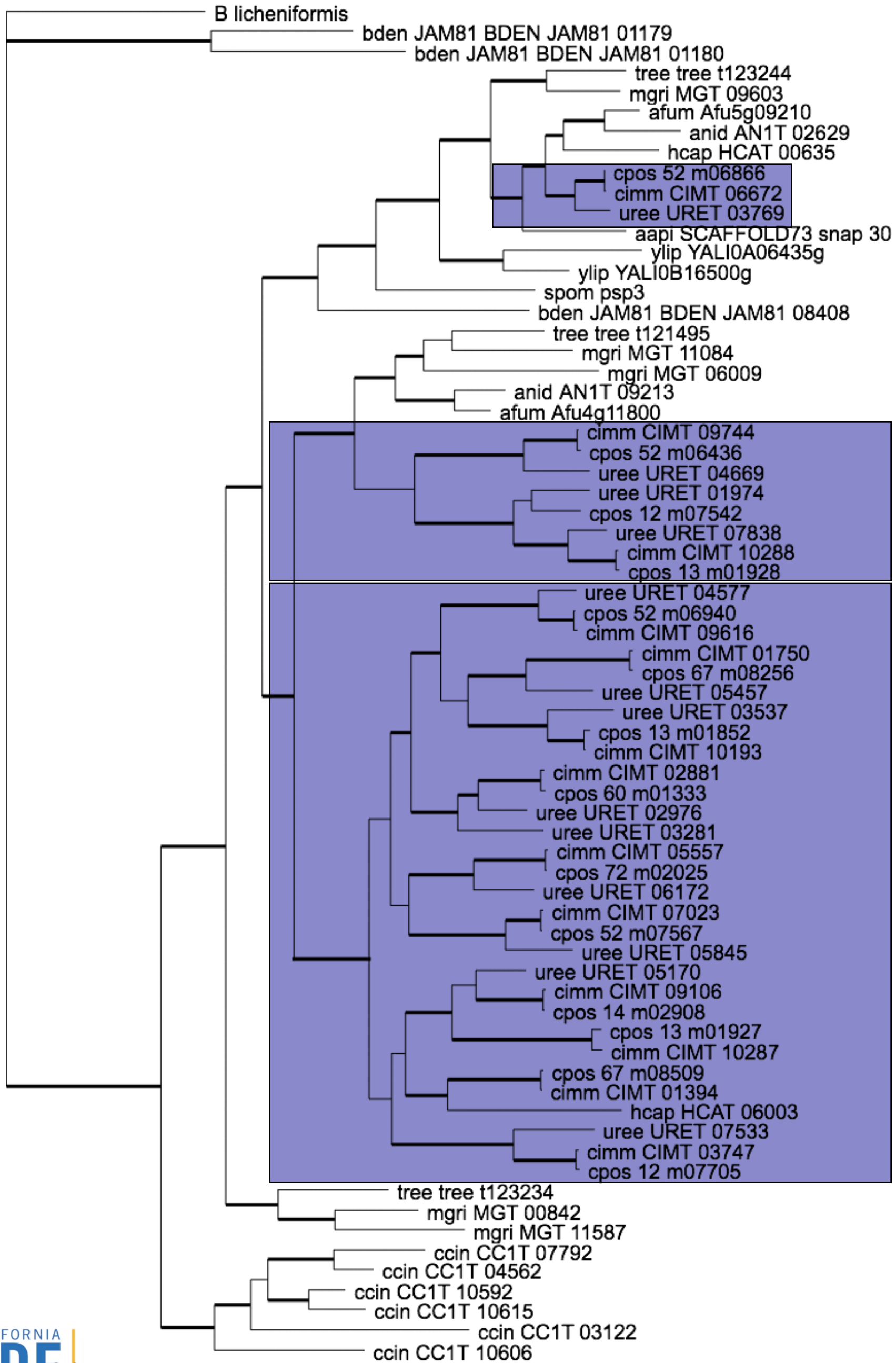
Peptidase S8 expansion  
in Onygenales

14 copies in *Coccidioides*  
1 in *Histoplasma*



# Peptidase S8 expansion in Onygenales

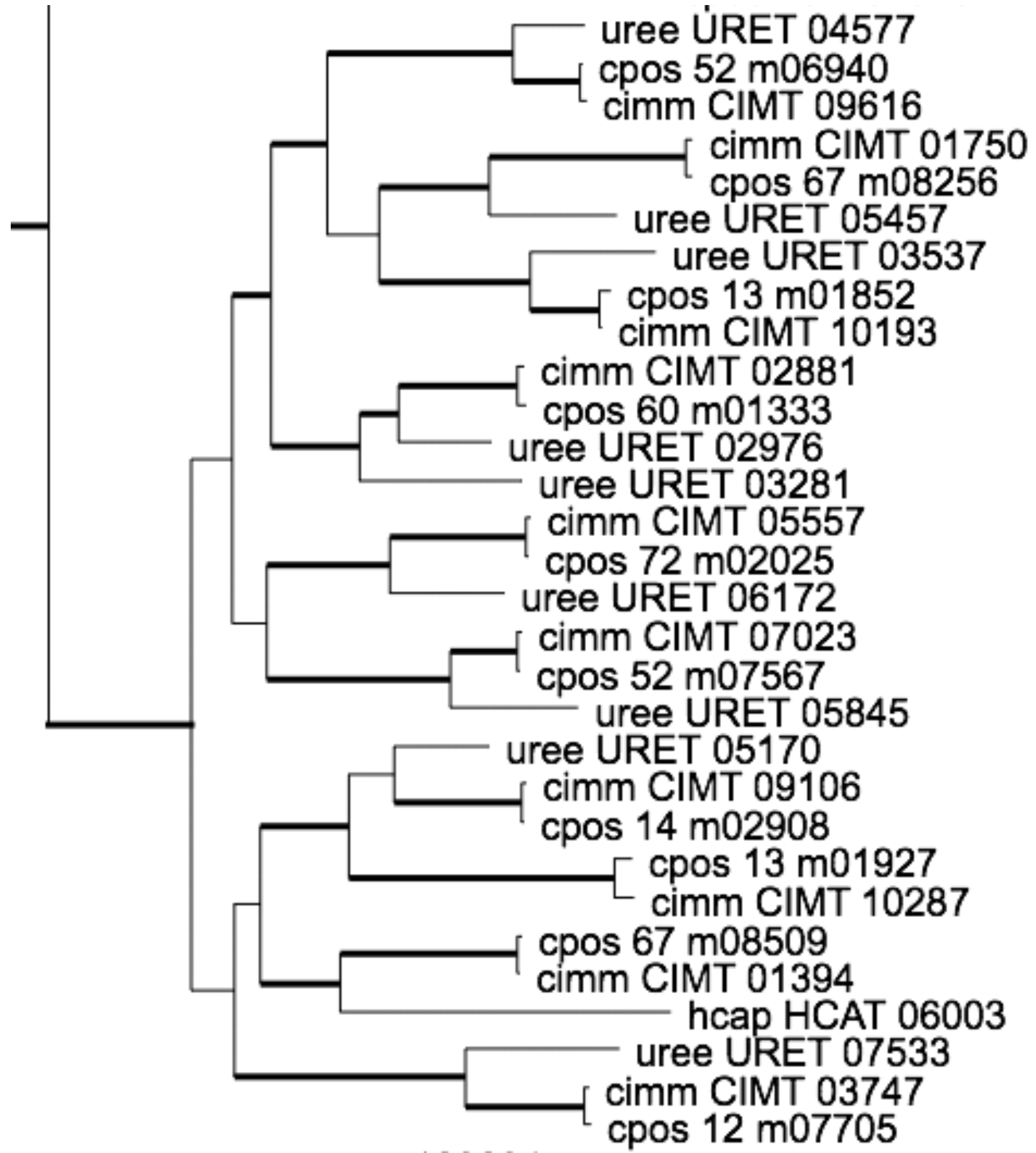
14 copies in *Coccidioides*  
1 in *Histoplasma*



I

II

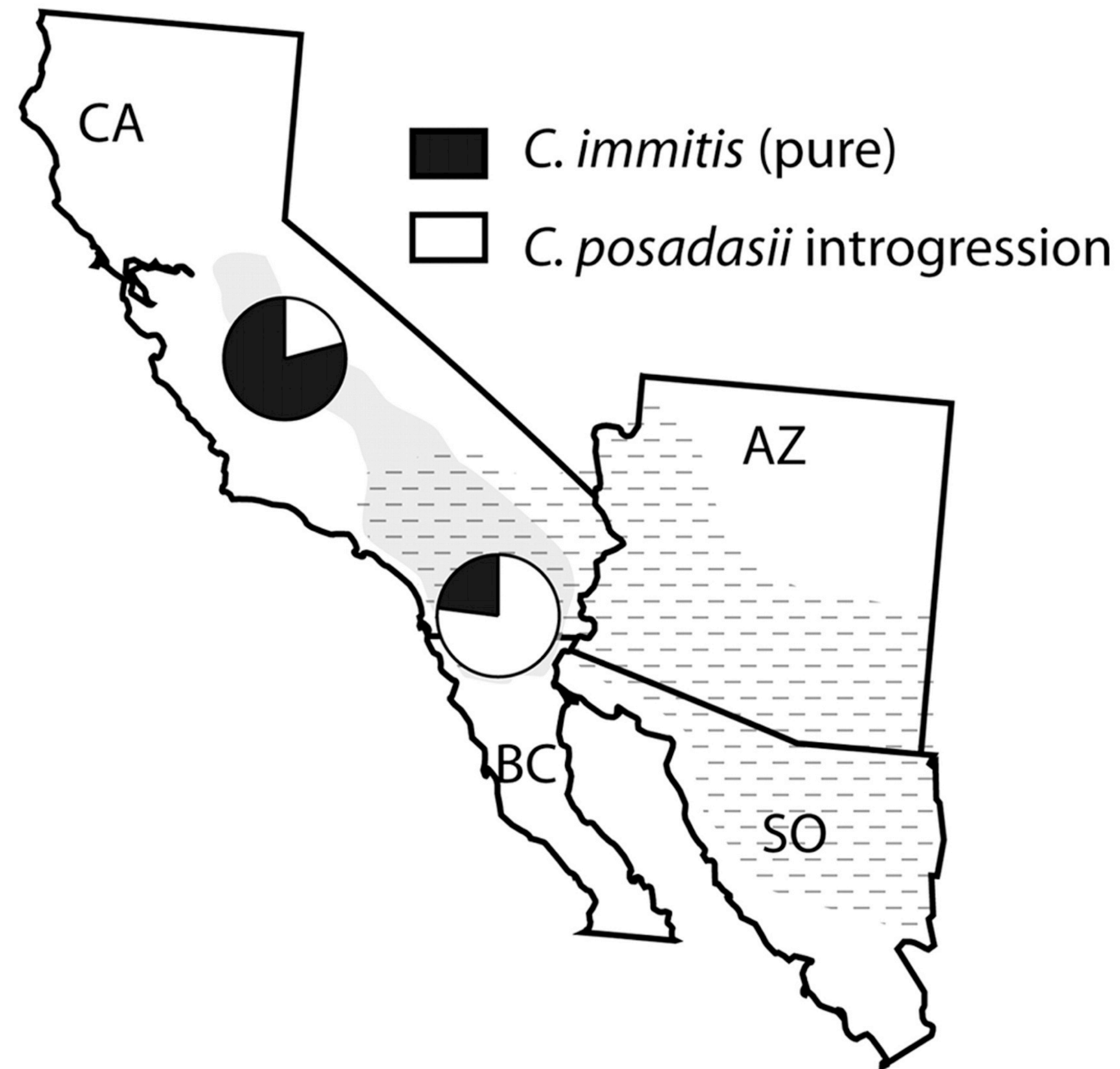
III





# Population genomics

- Revealed no loci with evidence for balancing selection - so little evidence for long standing host-pathogen battle as in *Plasmodium*.
- Sliding window  $F_{ST}$  analysis revealed some regions of reduced divergence and followup revealed a regions of recent intergression. Directionality looks to be mostly from Cp and into Ci.
- One gene implicated in pathogenesis, MEP4, is found in introgressed region



# Towards identifying genes underlying adaptation

---

# Towards identifying genes underlying adaptation

---

- Coccidioides is found in desert soil and associated with animals - perhaps has a long term animal association without inducing response from the host.



# Towards identifying genes underlying adaptation

---

- *Coccidioides* is found in desert soil and associated with animals - perhaps has a long term animal association without inducing response from the host.
- Loss of genes involved in plant product metabolism suggests nutritional shift in Onygenales from relatives in Eurotiales

# Towards identifying genes underlying adaptation

---

- *Coccidioides* is found in desert soil and associated with animals - perhaps has a long term animal association without inducing response from the host.
- Loss of genes involved in plant product metabolism suggests nutritional shift in Onygenales from relatives in Eurotiales
- Expansion of only a few gene families, seem to be involved in metabolism - none are *Coccidioides* specific though.

# Towards identifying genes underlying adaptation

---

- *Coccidioides* is found in desert soil and associated with animals - perhaps has a long term animal association without inducing response from the host.
- Loss of genes involved in plant product metabolism suggests nutritional shift in Onygenales from relatives in Eurotiales
- Expansion of only a few gene families, seem to be involved in metabolism - none are *Coccidioides* specific though.
- From population genetics analysis - failed to find regions under balancing selection suggesting there is not a long-term arms race between host and pathogen.



# Towards identifying genes underlying adaptation

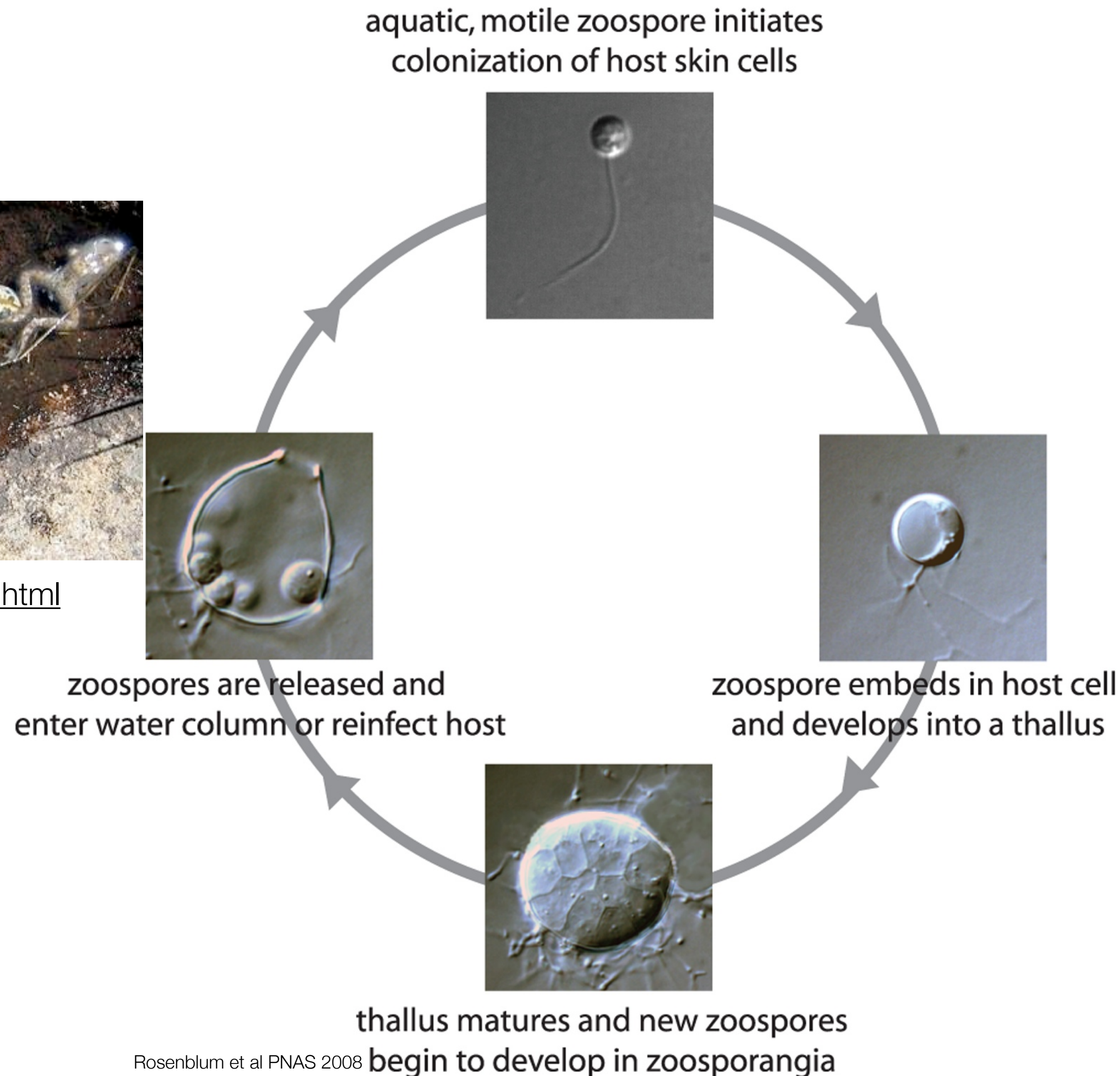
---

- *Coccidioides* is found in desert soil and associated with animals - perhaps has a long term animal association without inducing response from the host.
- Loss of genes involved in plant product metabolism suggests nutritional shift in Onygenales from relatives in Eurotiales
- Expansion of only a few gene families, seem to be involved in metabolism - none are *Coccidioides* specific though.
- From population genetics analysis - failed to find regions under balancing selection suggesting there is not a long-term arms race between host and pathogen.
- Evidence for introgression between the species and perhaps imported novel alleles that are important for adaptation in both species to their animal hosts.

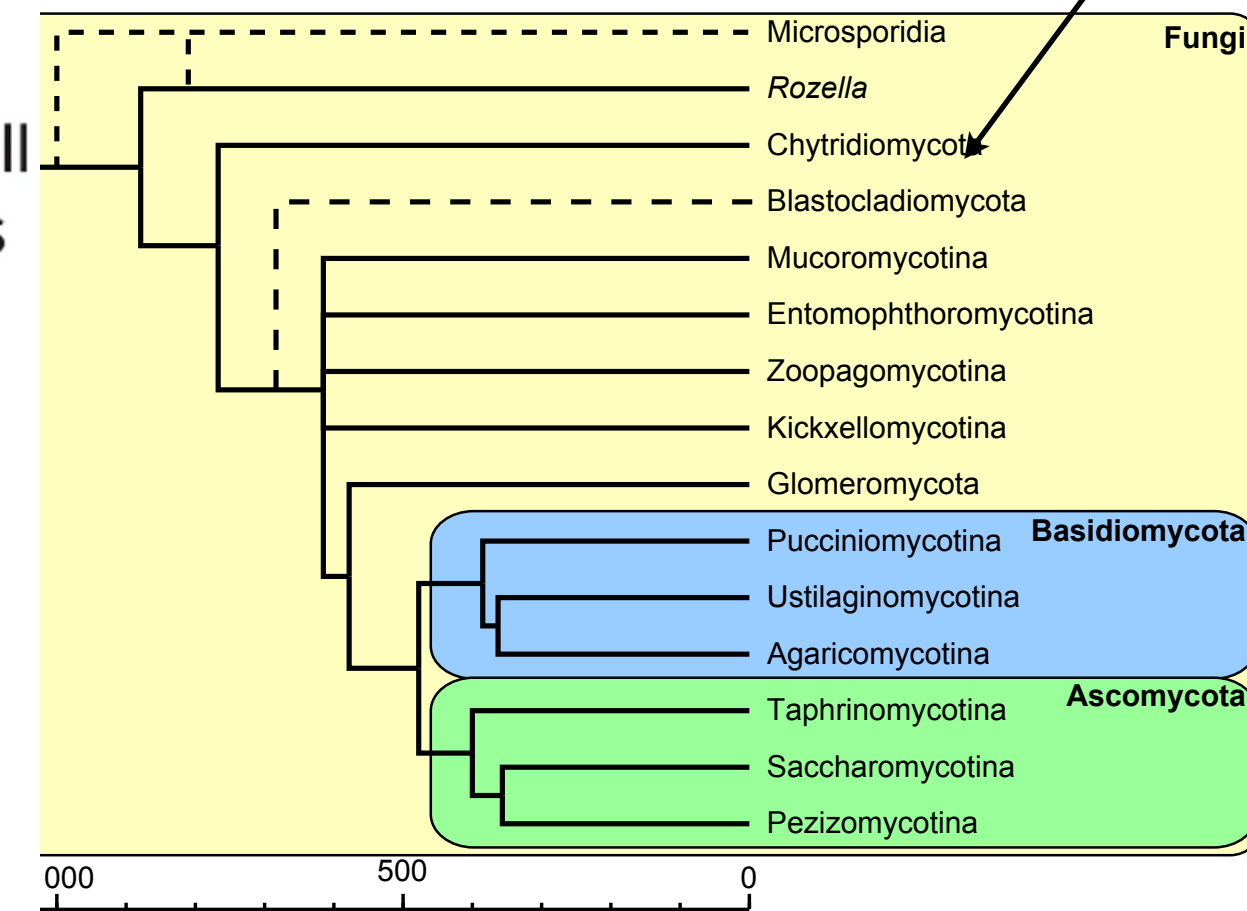
# *Batrachochytrium dendrobatidis* (Bd)- a major cause of amphibian decline



[http://civr.ucr.edu/chytrid\\_fungus.html](http://civr.ucr.edu/chytrid_fungus.html)



*Bd* is a Chytrid fungus



Rosenblum et al PNAS 2008



# Bd genome sequence projects

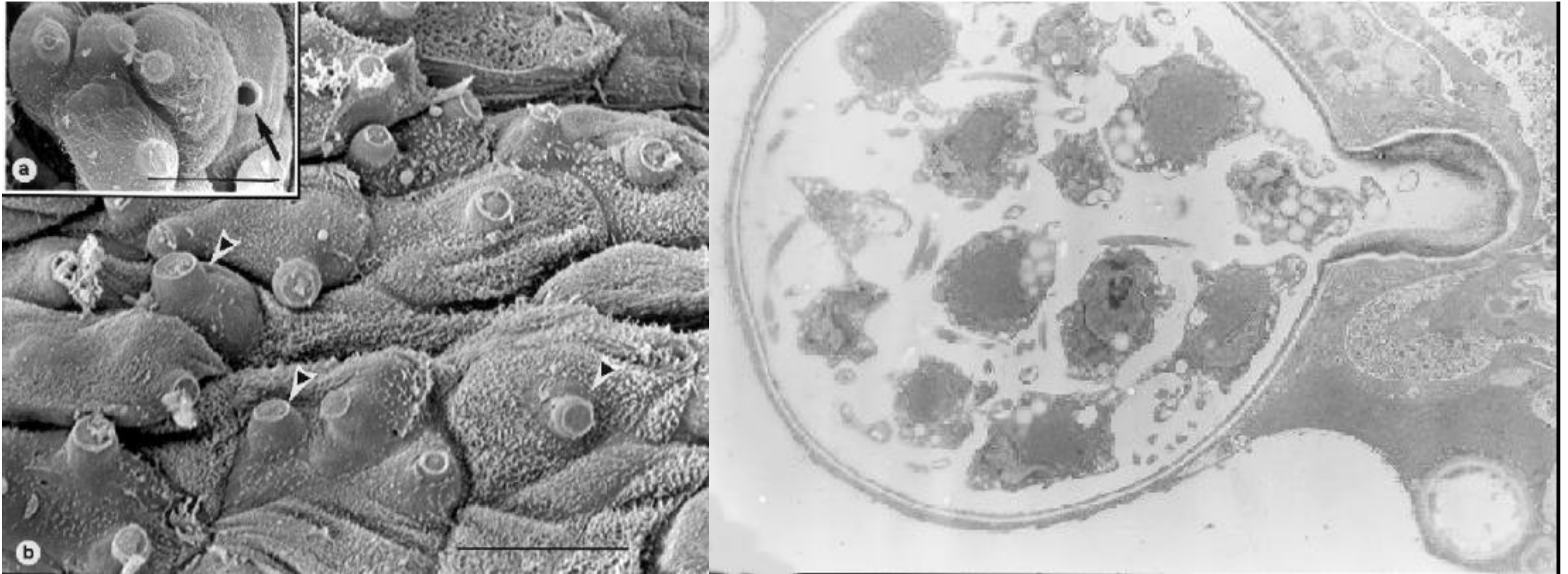
---

- 2 strains sequenced at JGI and Broad; allow whole genome comparison between strains
- Found that genome is diploid, but with large regions with loss of heterozygosity (LOH)
- Identified gene family changes that might suggest mechanisms of pathogenicity
- Greater understanding of what the early fungus was like
- With collaborators we are sequencing 20 more strains for population genomics to better understand population dynamics, trace origin of diversity, and understand the LOH as independent or shared events.



# *Bd* grows intercellularly

Formally described by Joyce Longcore (U Maine) (Mycologia 1999)



Rick Speare, Lee Berger, Alax Haytt  
James Cook University, Townsville, Australia



# Phylogenomic profiling

---

- For each gene in the the target (*Bd*) genome, look to see which have homologs in other fungi, animals, plants
- Classify genes by its profile as to when it must have arisen based on identified homologs
- Also compare well-studied organisms (*S. cerevisiae*, *N. crassa*) to see which genes were missing
- For *Bd*/Chytrids several trends appeared
  - Missing: some cell wall genes, spindle-pole body
  - Present in *Bd* but not other non-chytrid fungi: Flagella, some signaling pathways, effector like proteins
  - Some expansions of gene families

# Fungal cell wall evolution- a view from earliest branches

Ergosterol

Plasma membrane

Yeast Cell Wall

GPI- anchored  
protein

mannoproteins

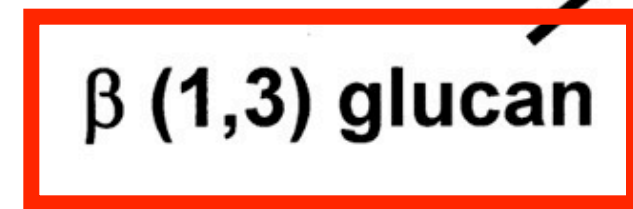
What is the fungal cell wall  
made of? Sugar polymers

$\beta$  (1,6) glucan

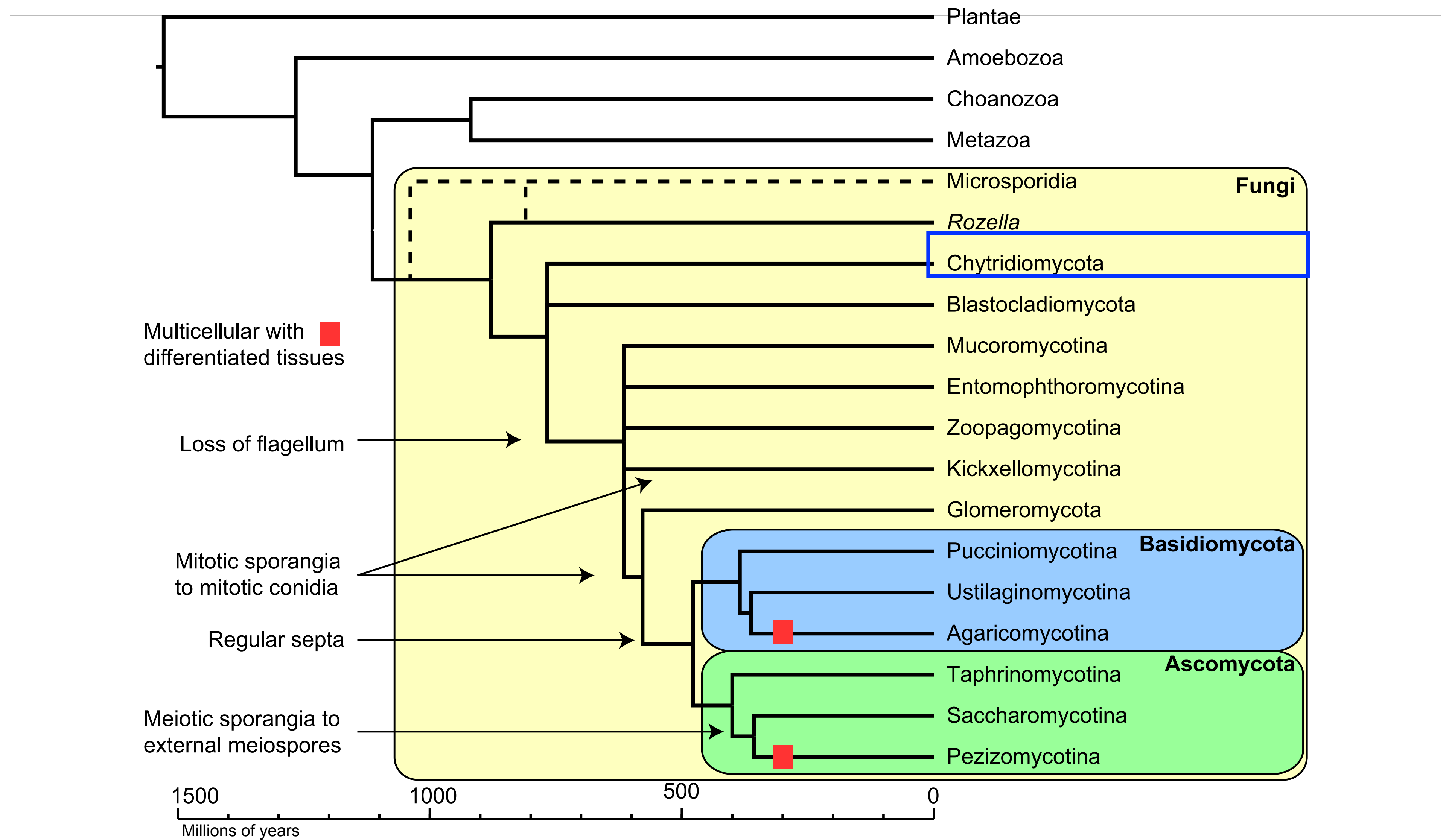
Chitin

$\beta$  (1,3) glucan

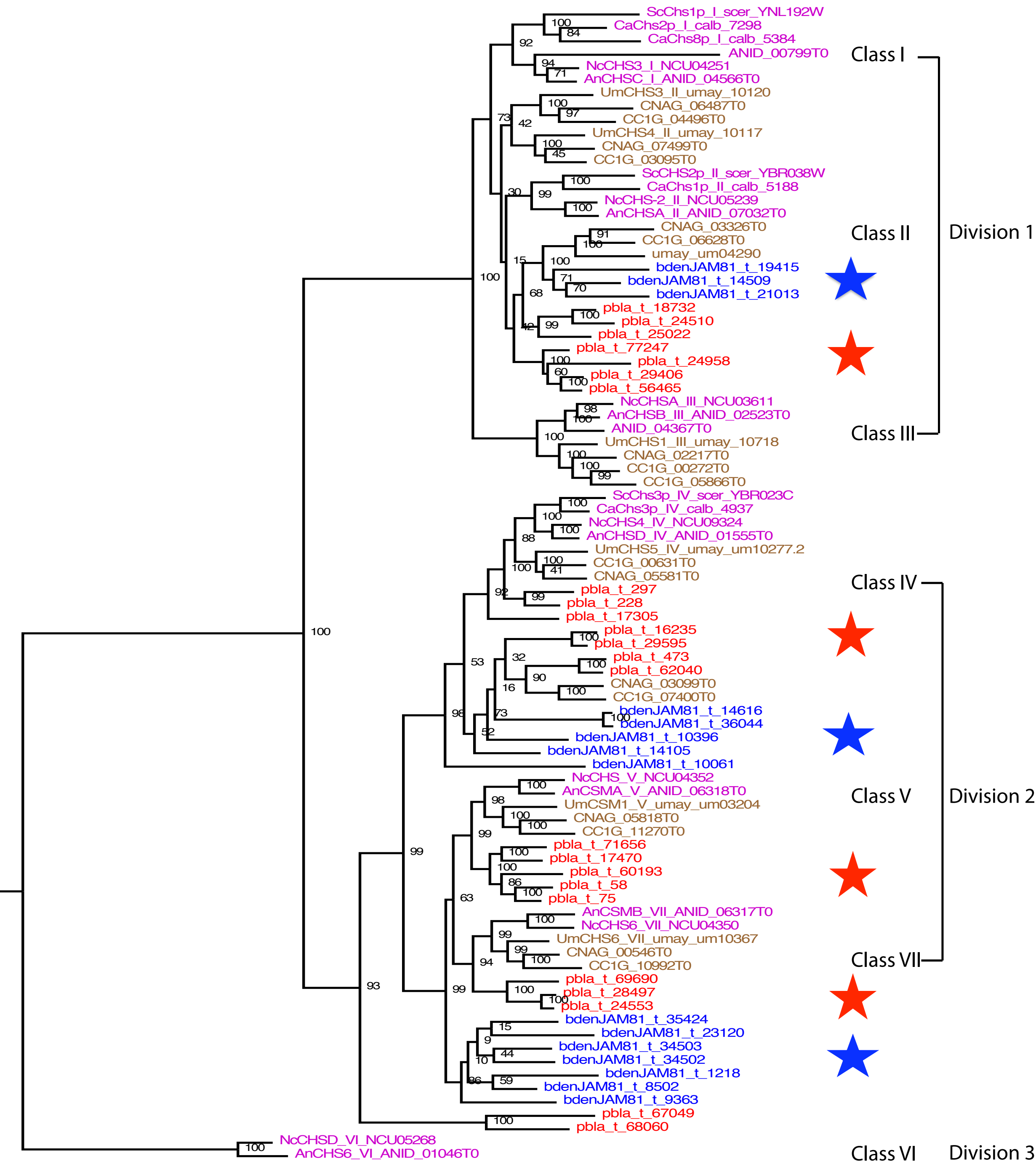
Some fungi also have  
 $\alpha$  (1,3) glucan







# Chitin Synthase Gene Evolution

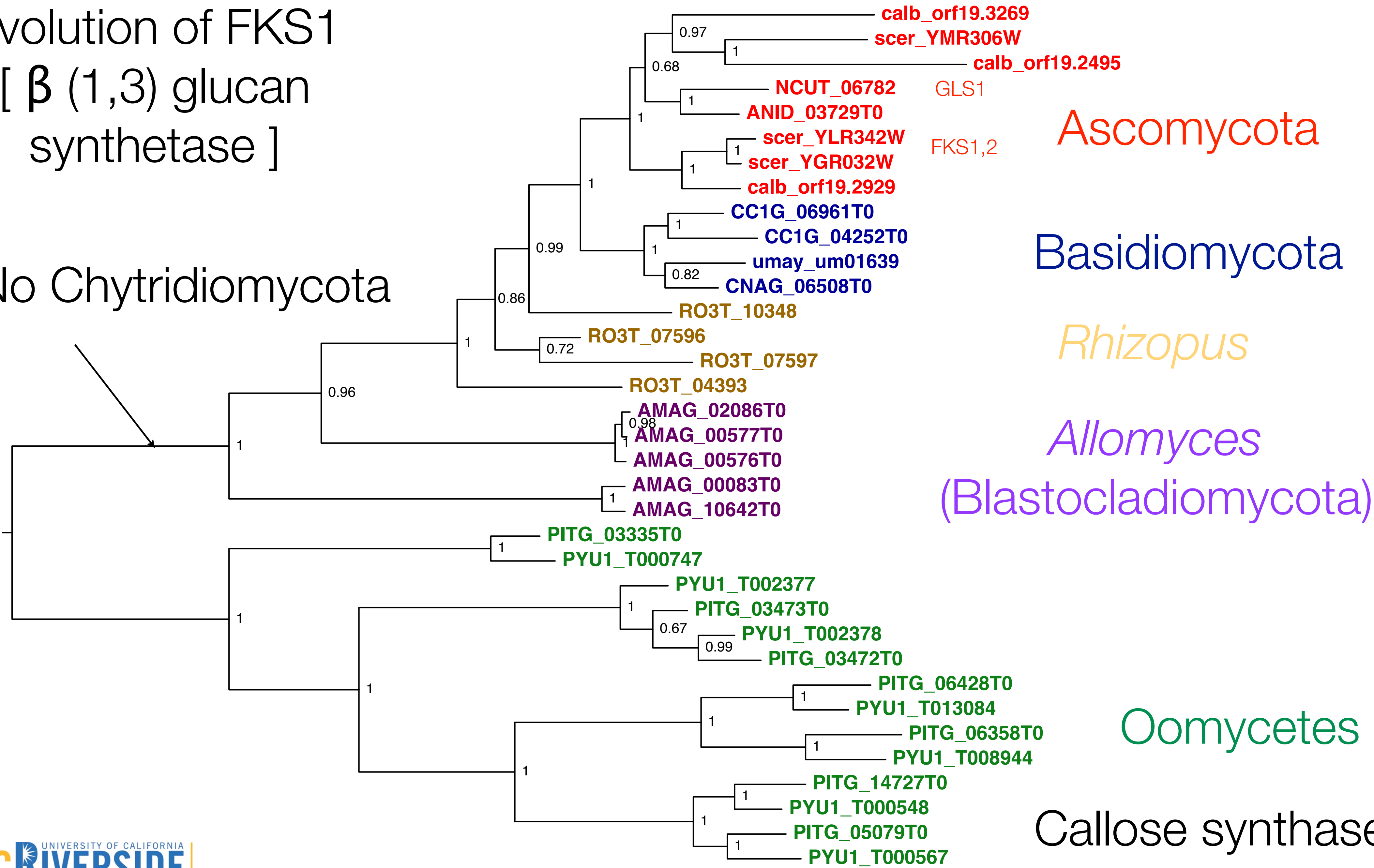


# Evolution of $\beta$ (1,3) glucan synthase



# Evolution of FKS1 [ $\beta$ (1,3) glucan synthetase ]

No Chytridiomycota



No 1,3 Beta-glucan?

*B. dendrobatidis* cell wall biochemical analysis

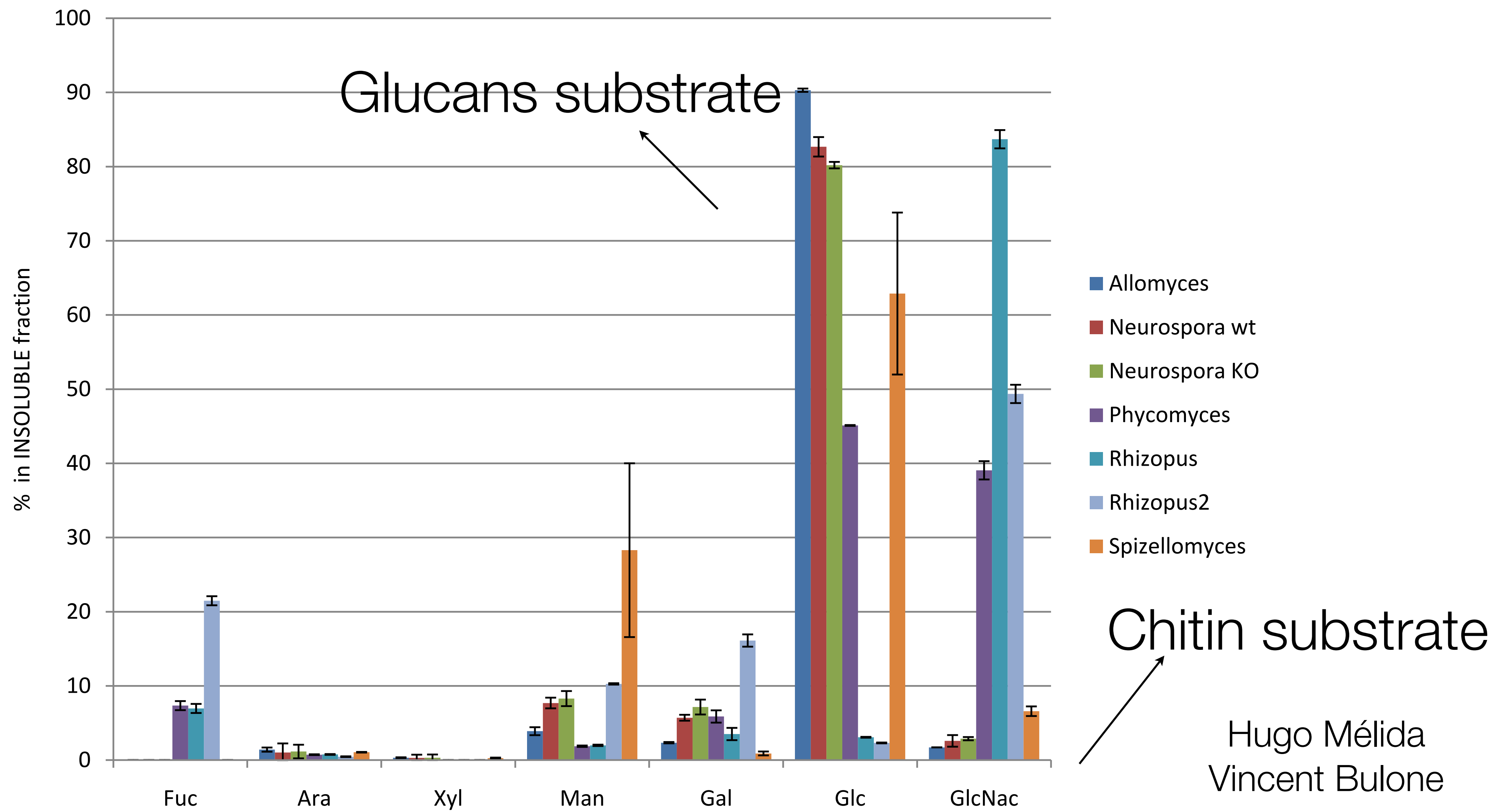
$\beta$ (1,3)-glucan	$\beta$ (1,6)-glucan	$\alpha$ (1,3)-glucan	Cellulose $\beta$ (1,4)-glucan	Chitin
<b>X</b>	<b>X</b>	<b>X</b>	✓	✓

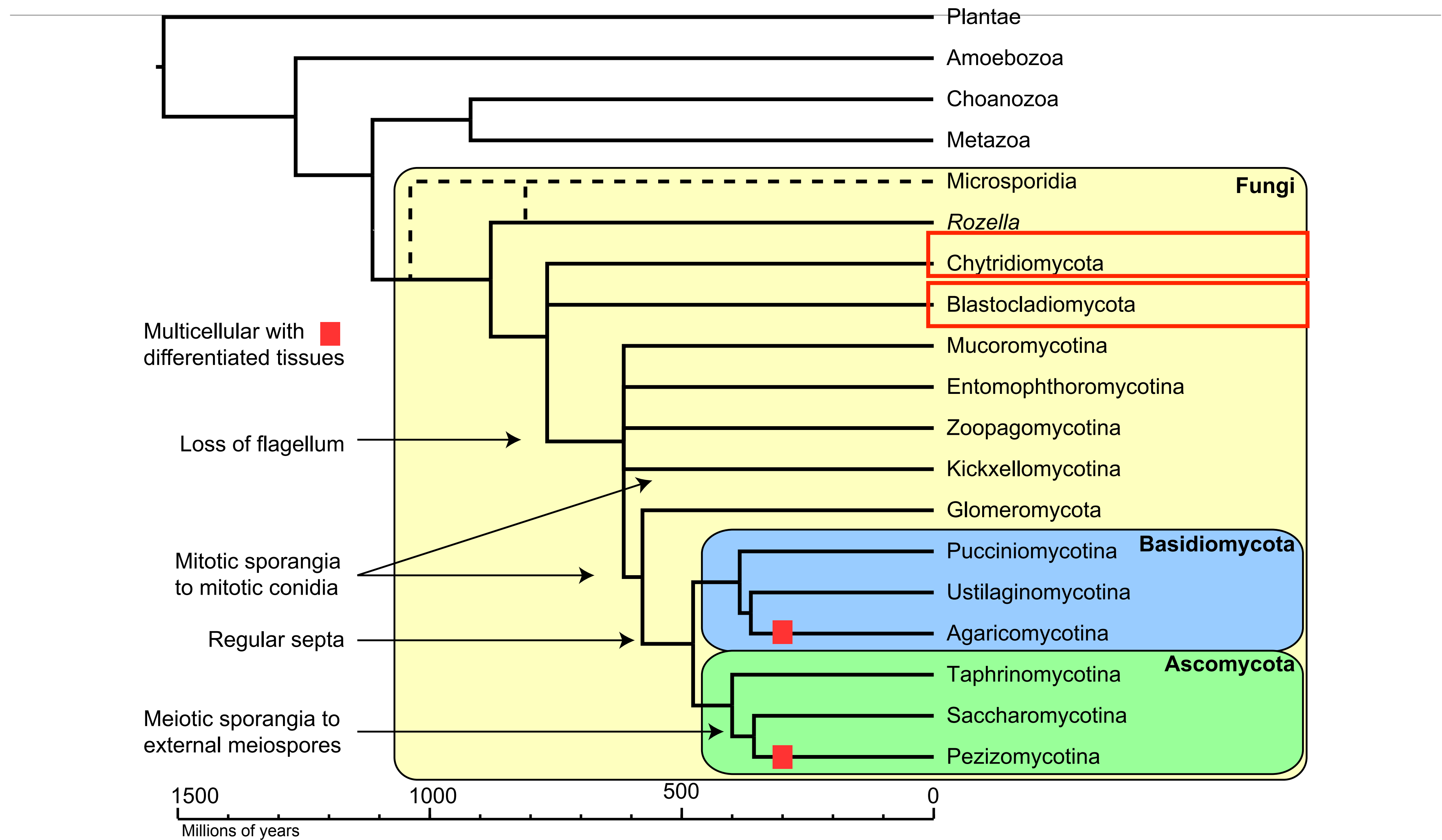
A putative cellulose synthase gene can be found in genome of *Bd*  
but also found in *N. crassa* and other Fungi

with JP Latgé, M. Fisher

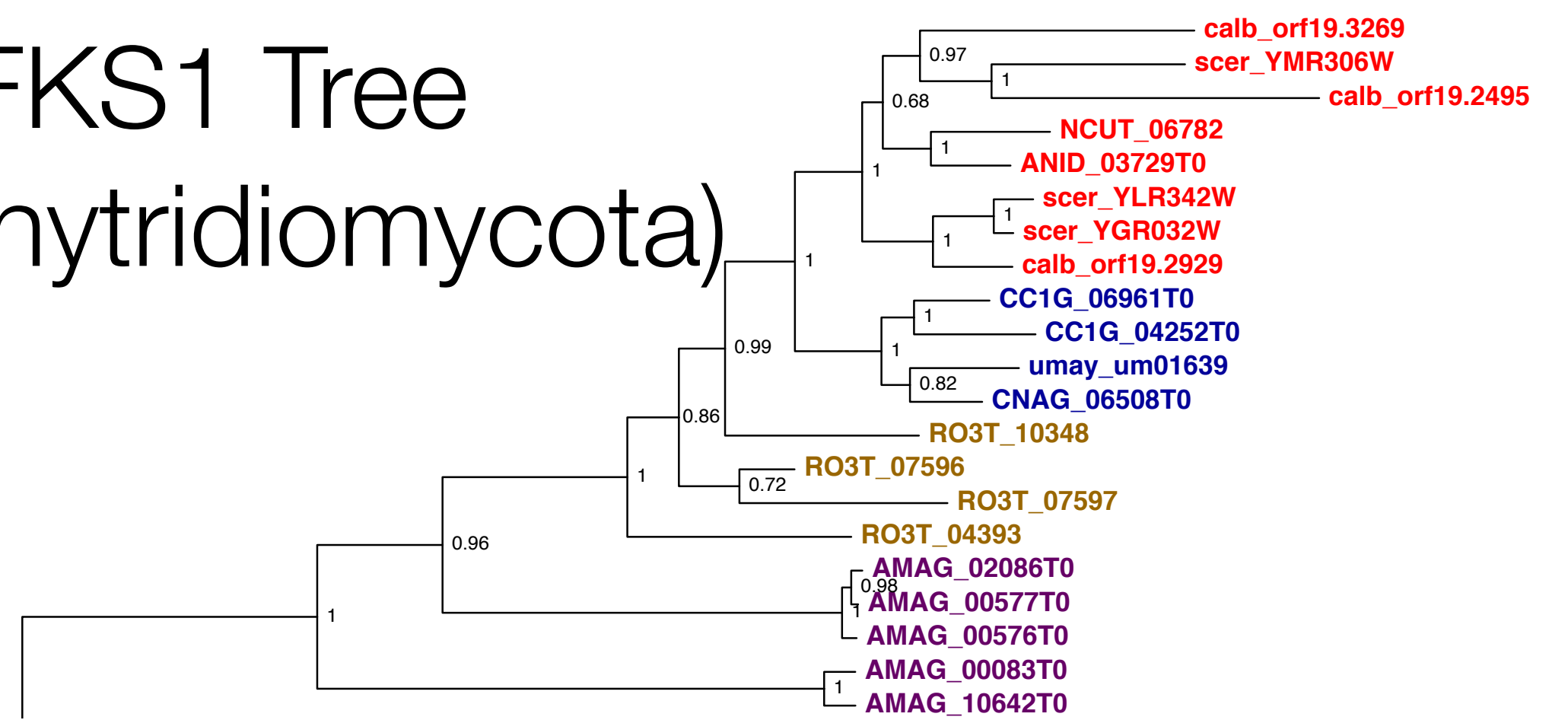


# Monosaccharide breakdown of several fungal cell walls

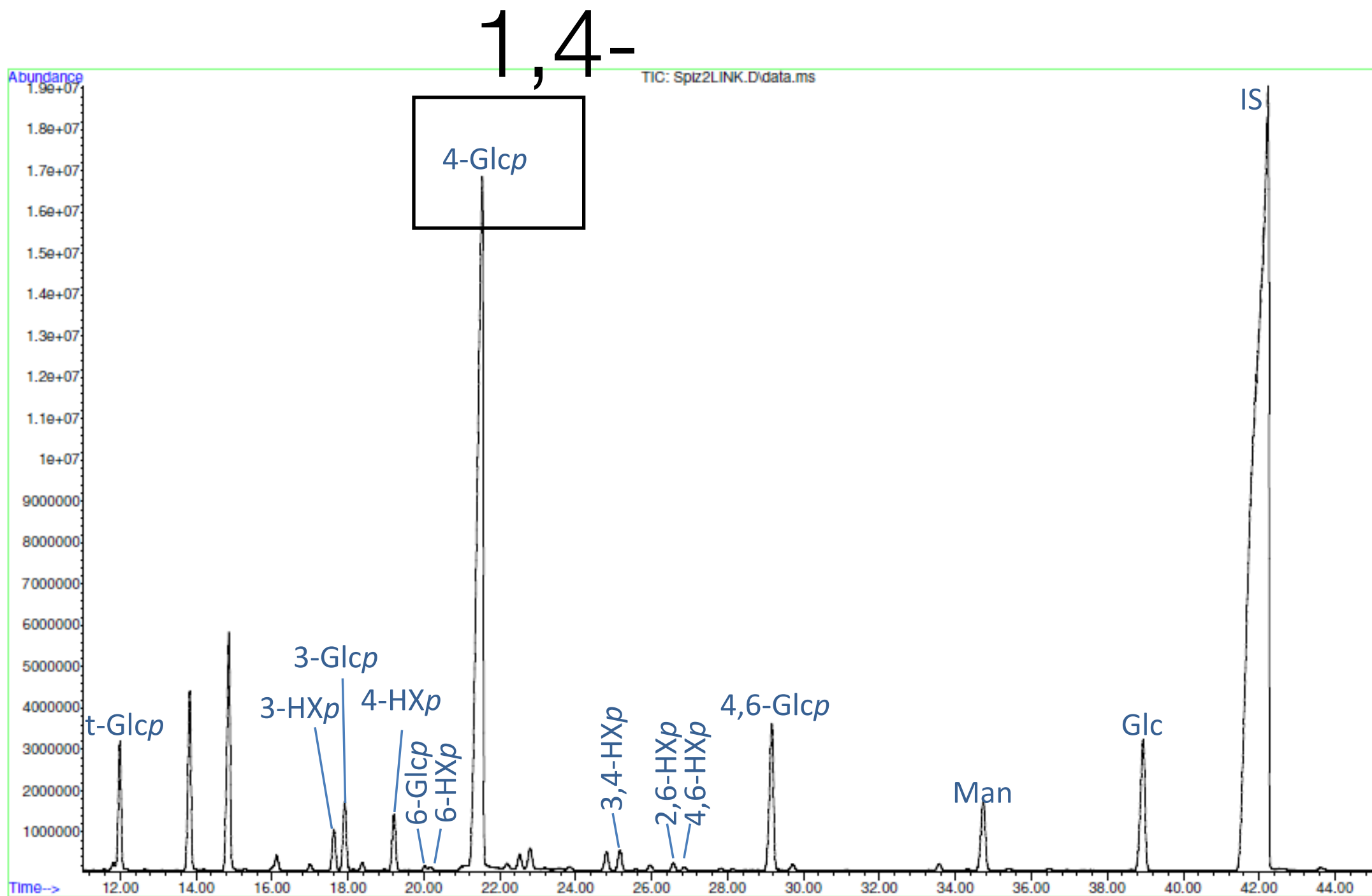




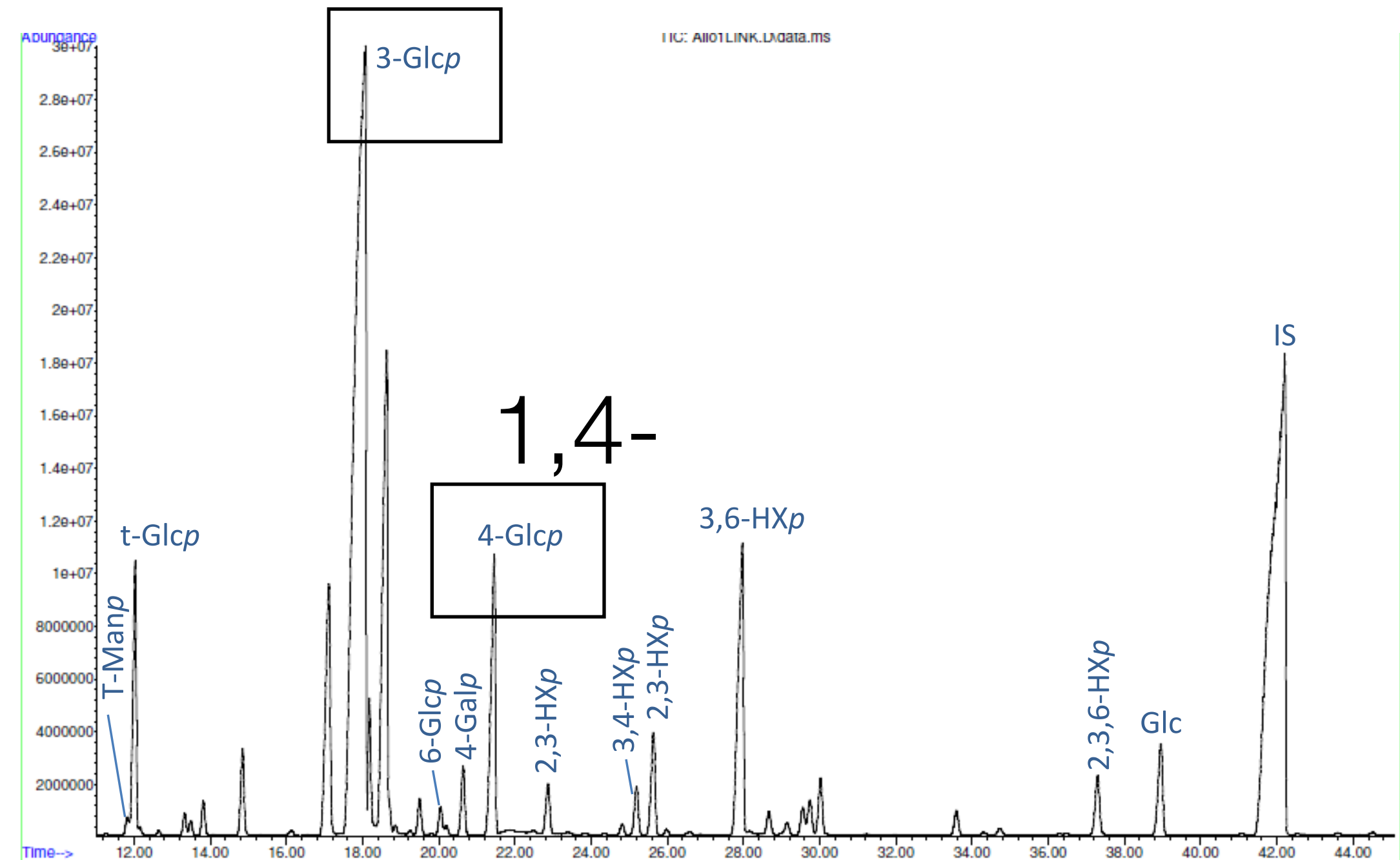
# FKS1 Tree (No Chytridiomycota)



1,3-



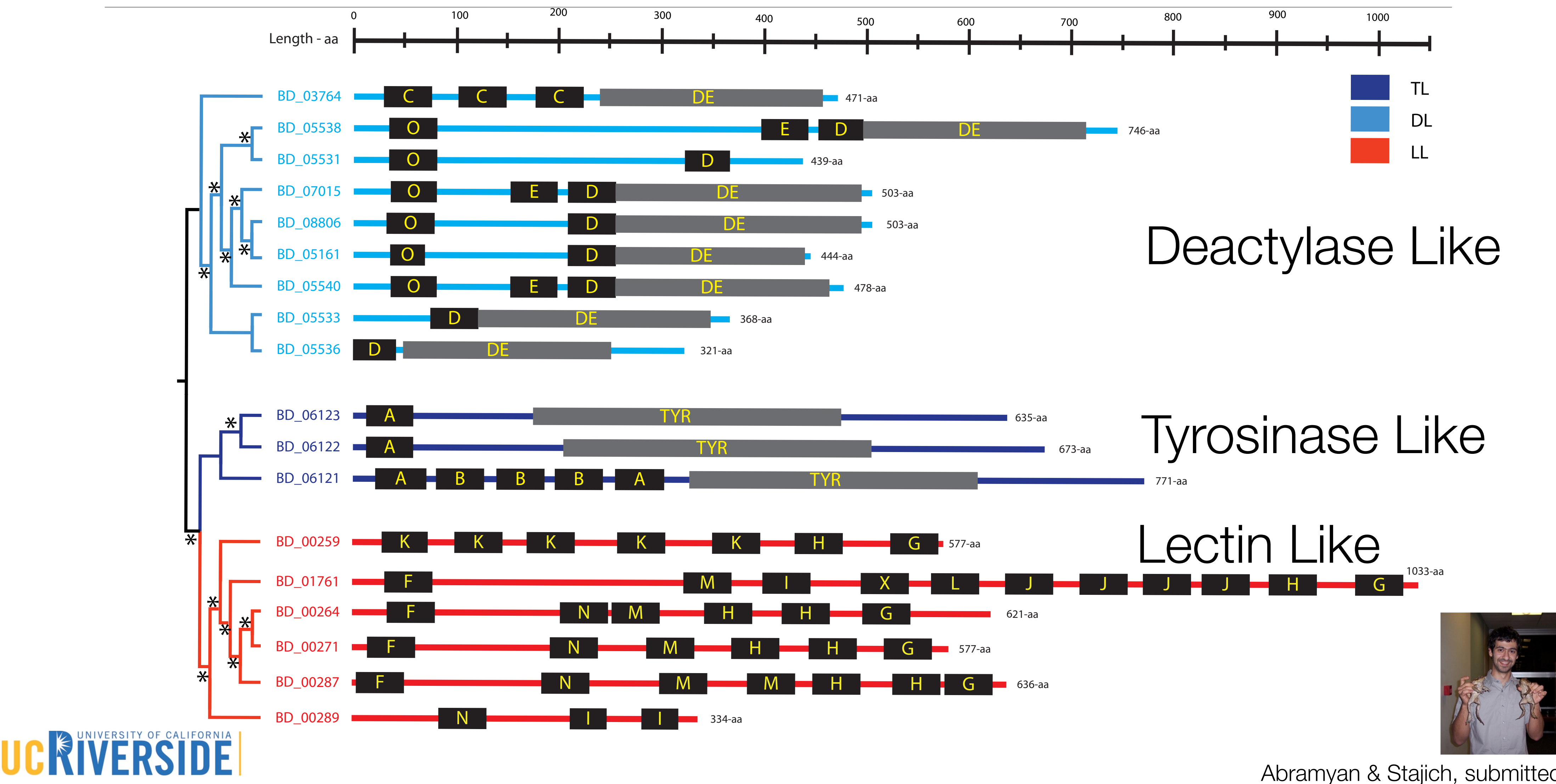
Spizellomyces



Allomyces



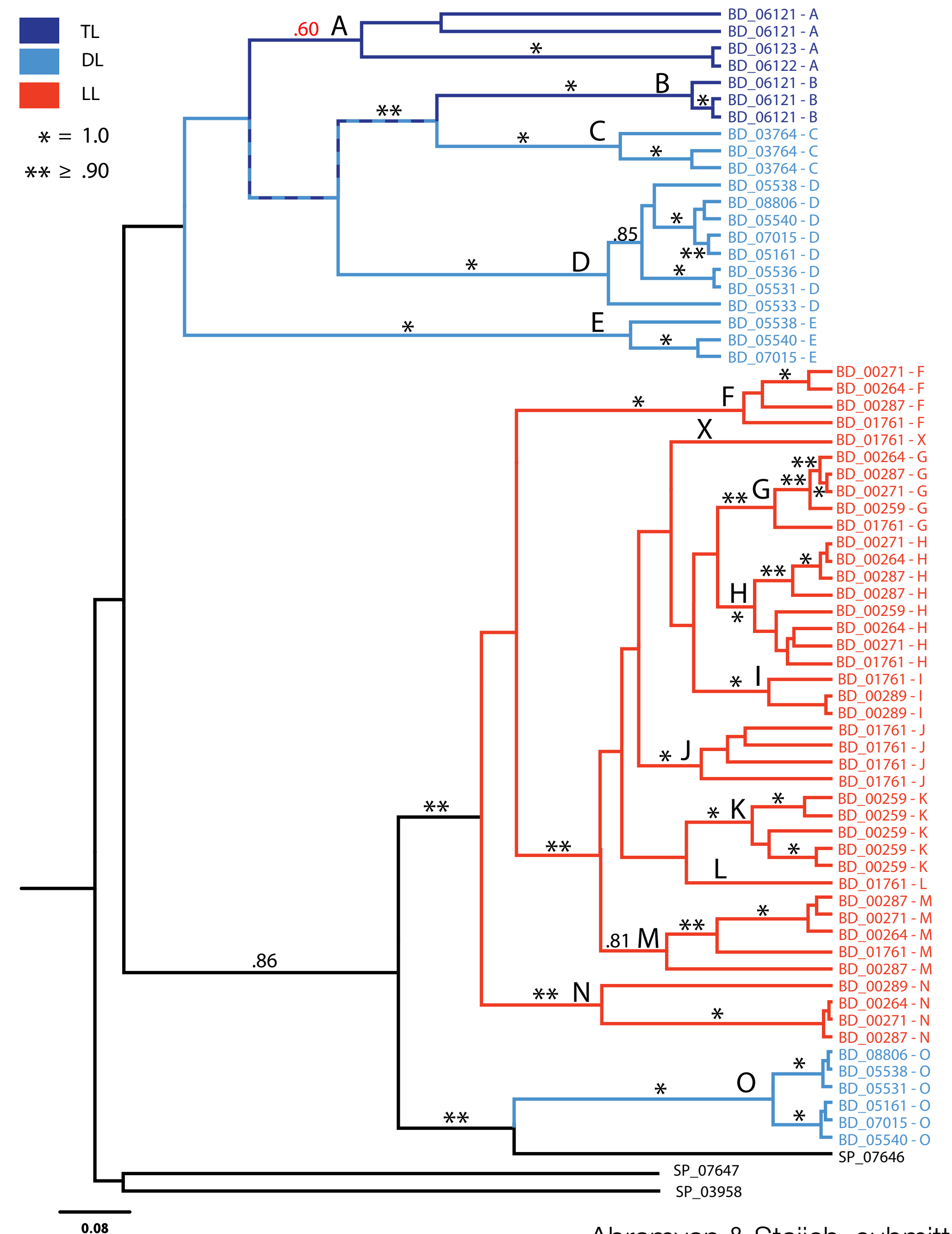
# Expansion of CBM18 domain - Chitin binding gene family



Abramyan & Stajich, submitted

# Domain tree shows clade-specific grouping

Domains evolved from mostly tandem  
duplications with some intergene  
duplications.



# CBM18 expansion

---

- Largest copy number of CBM18 in all fungi, and most number of domains (11) in a single locus
- Evidence for positive selection among copies of the domains based on codon analyses
- CBM18 thought to bind chitin, could be involved in binding its own chitin to cloak the cells from the host immunity
- Could also bind chitin-related molecules in animals attach more firmly to the animal cells



# Tyrosinase expansions in *Bd* and other fungi

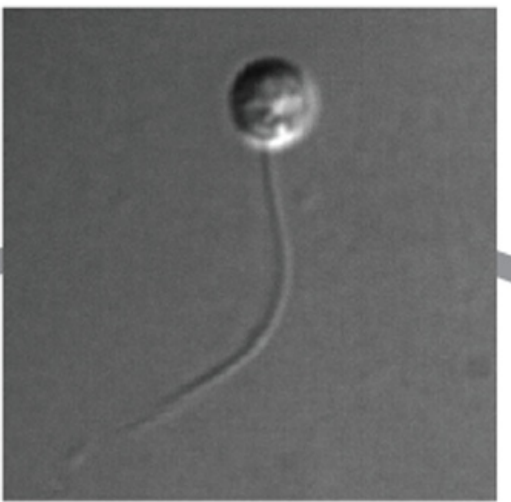
## Allomyces



H.Whistler via Liu et al BMC Evol Biol 2006

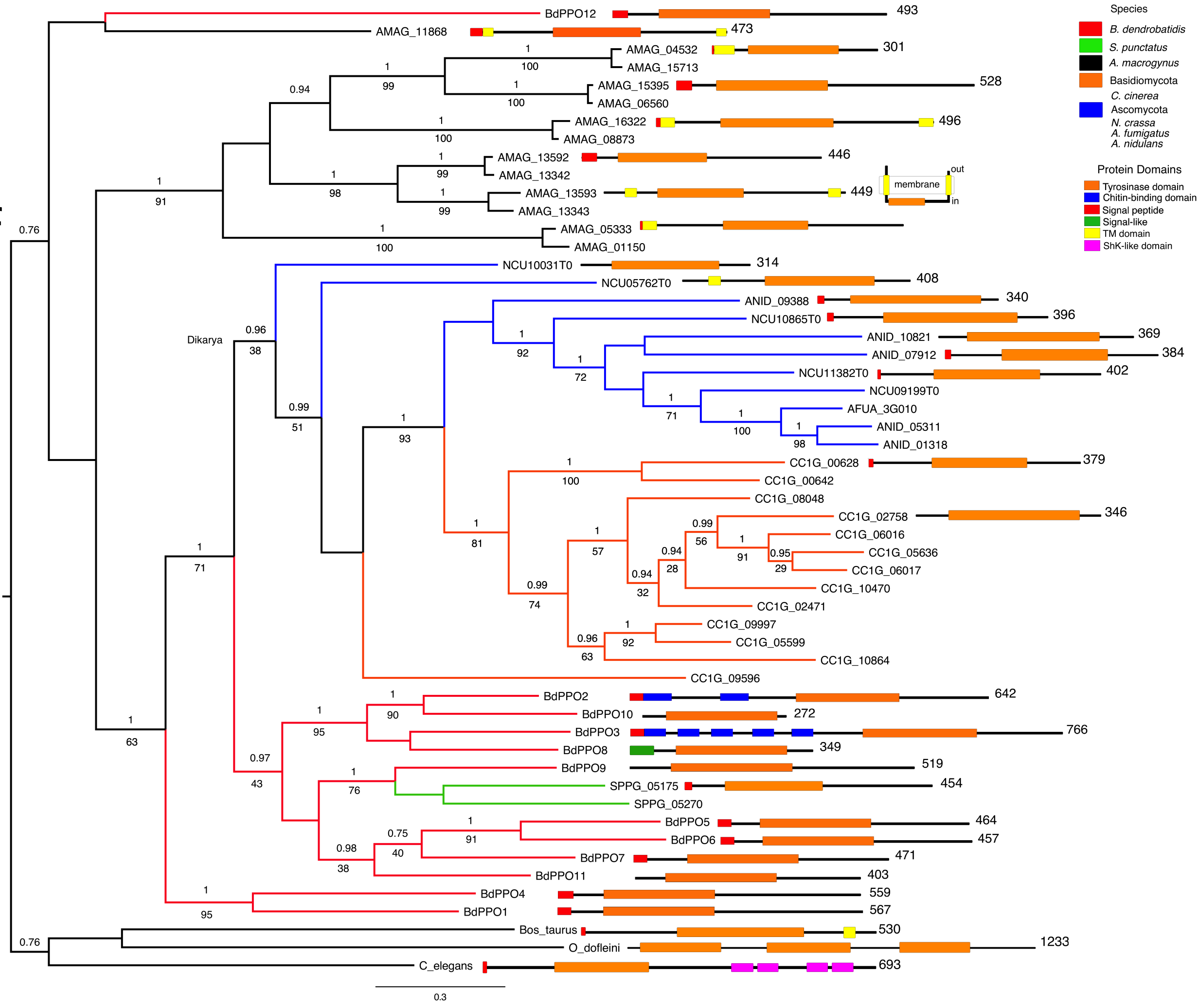


Hajime Muraguchi via Stajich et al PNAS 2010



E.Medina

Synthesis of melanin, pigments and other polyphenols



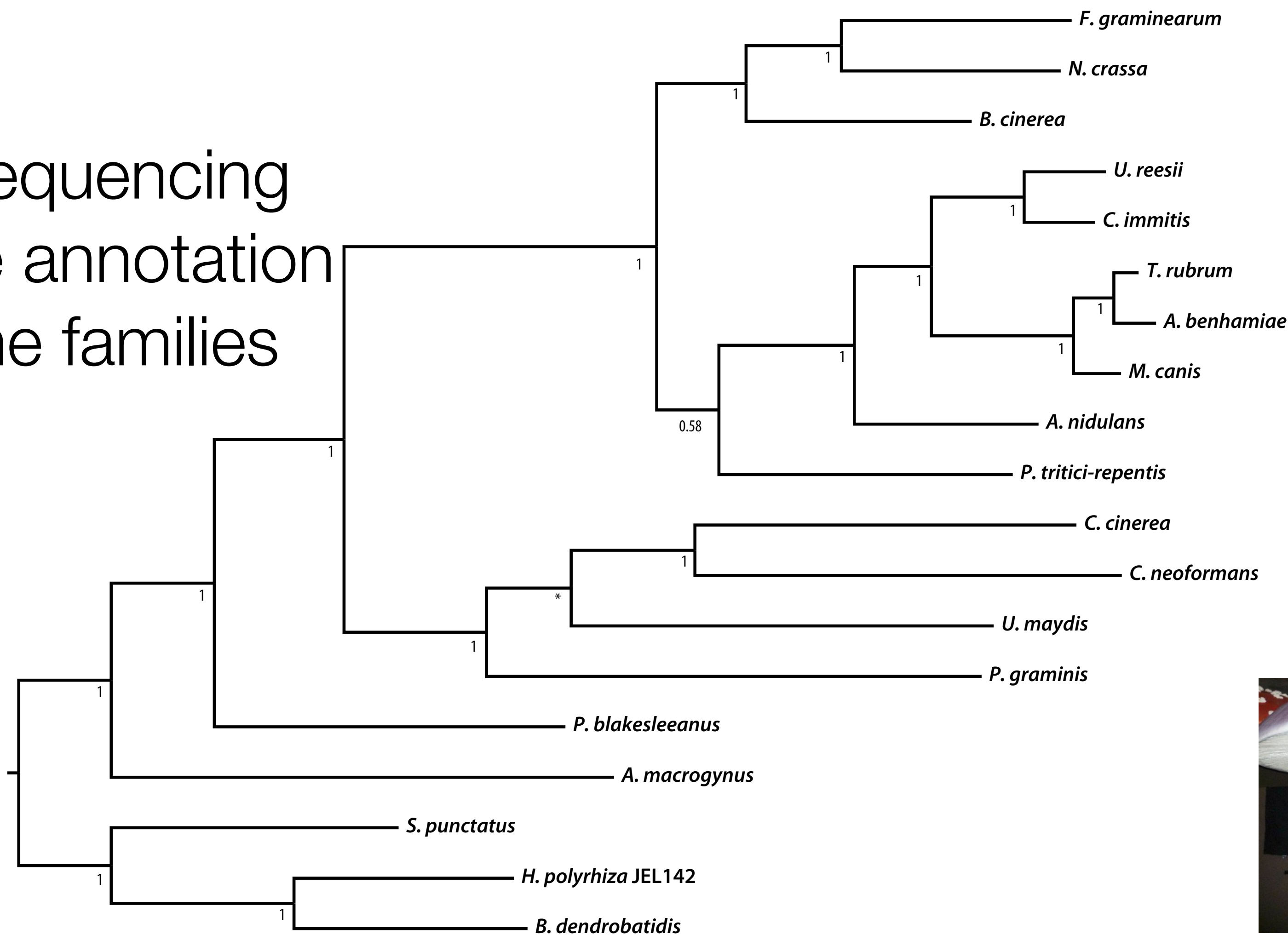
# But are these changes important?

---

- Just observing big families in a genome is nice, but does it mean that changes are really related to recent adaptation?
  - The branches on some gene trees are short, indicating duplications are recent
- A better is to polarize the changes to the *Bd* branch by having a closer species than 50-100 Mya...
- So is there a species closer to *Bd* we can use? ...

# *Homolaphylictis polyrhiza* JEL142 is a close(ish) relative

454 genome sequencing  
MAKER genome annotation  
OrthoMCL gene families



0.09



Joneson et al, Submitted



# Contrasting some family sizes among the Chytrids

	M36	S41	Tyrosinase	CBM18
<i>Allomyces</i>	31	0	16	3
<i>Spizellomyces</i>	3	3	3	3
Hp	3-5	3	15	5
<b>Bd</b>	<b>38</b>	<b>32</b>	12	<b>14-16</b>

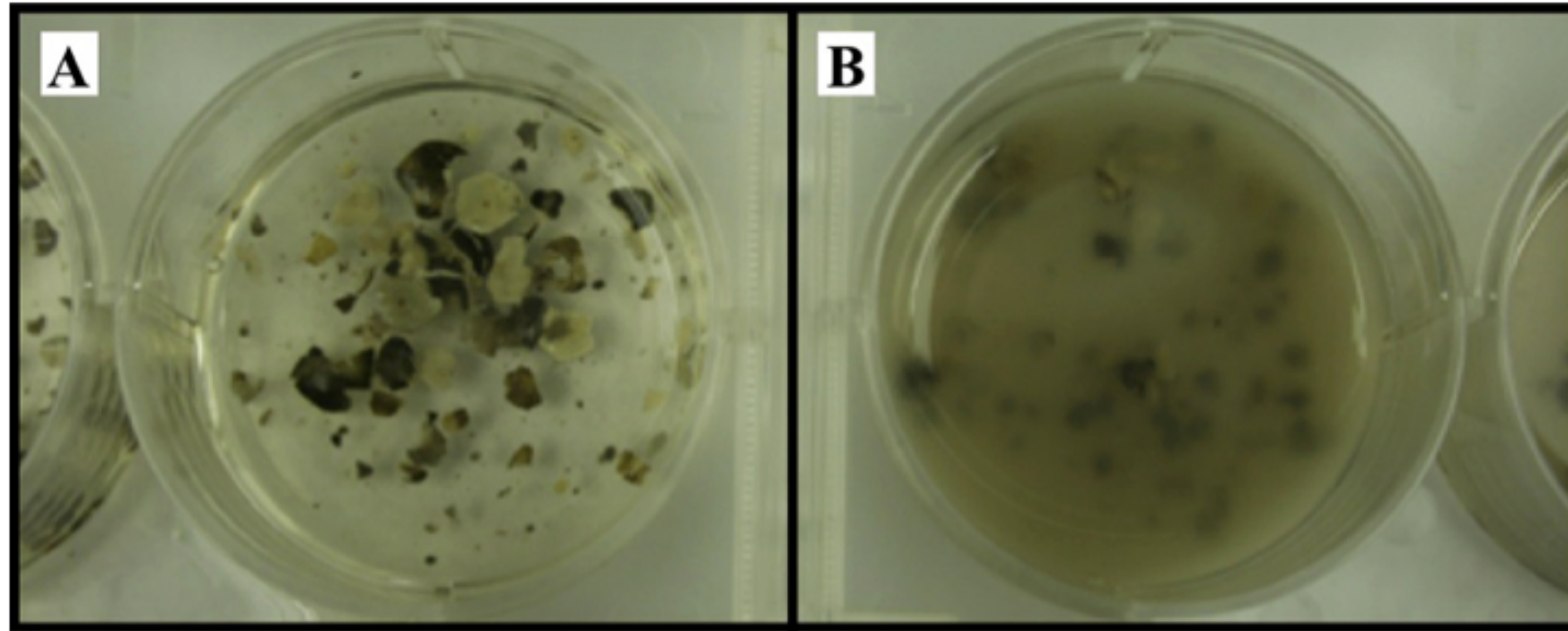
Fungalysin      Serine protease

Fungalysins - thought to be keratinases (break down keratin in an amphibian skin...)

Serine protease expansions in *Coccidoides* and relatives maybe related to breaking down on animal matter

*H. polyrhiza* won't grow on frog skin; doesn't cause mortality in frogs

---

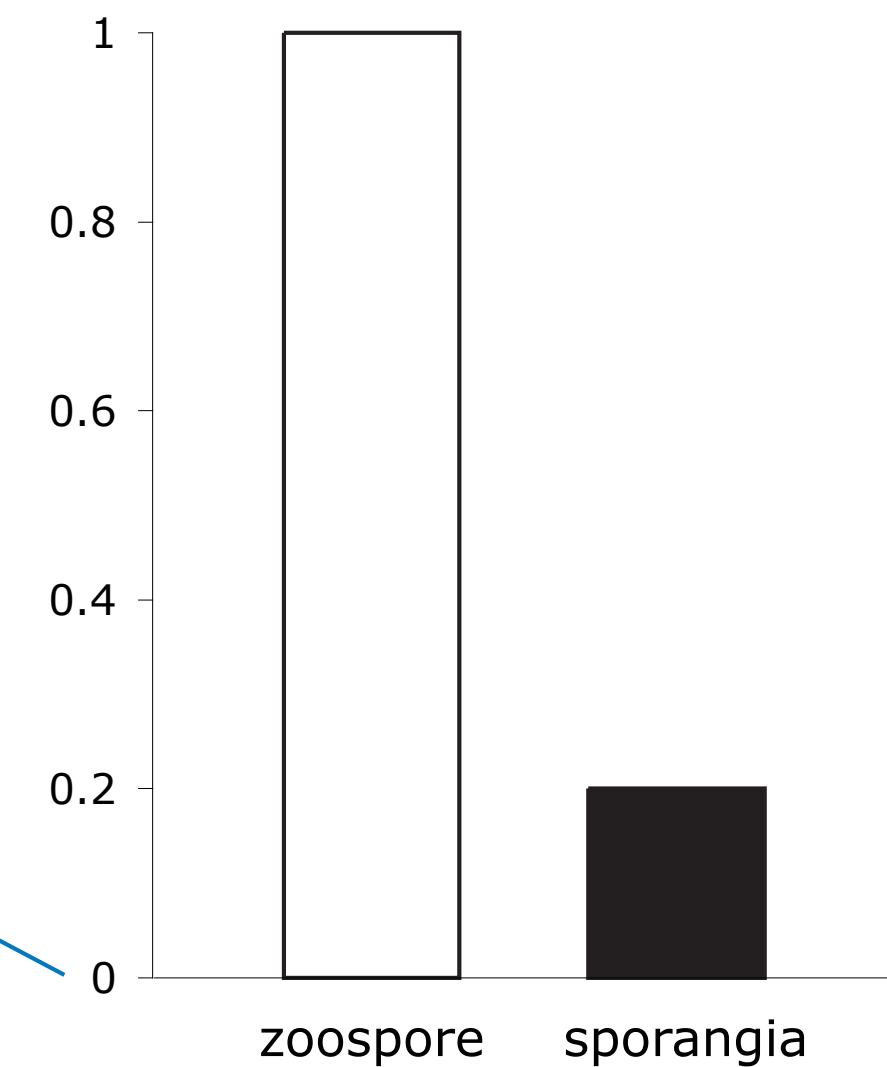
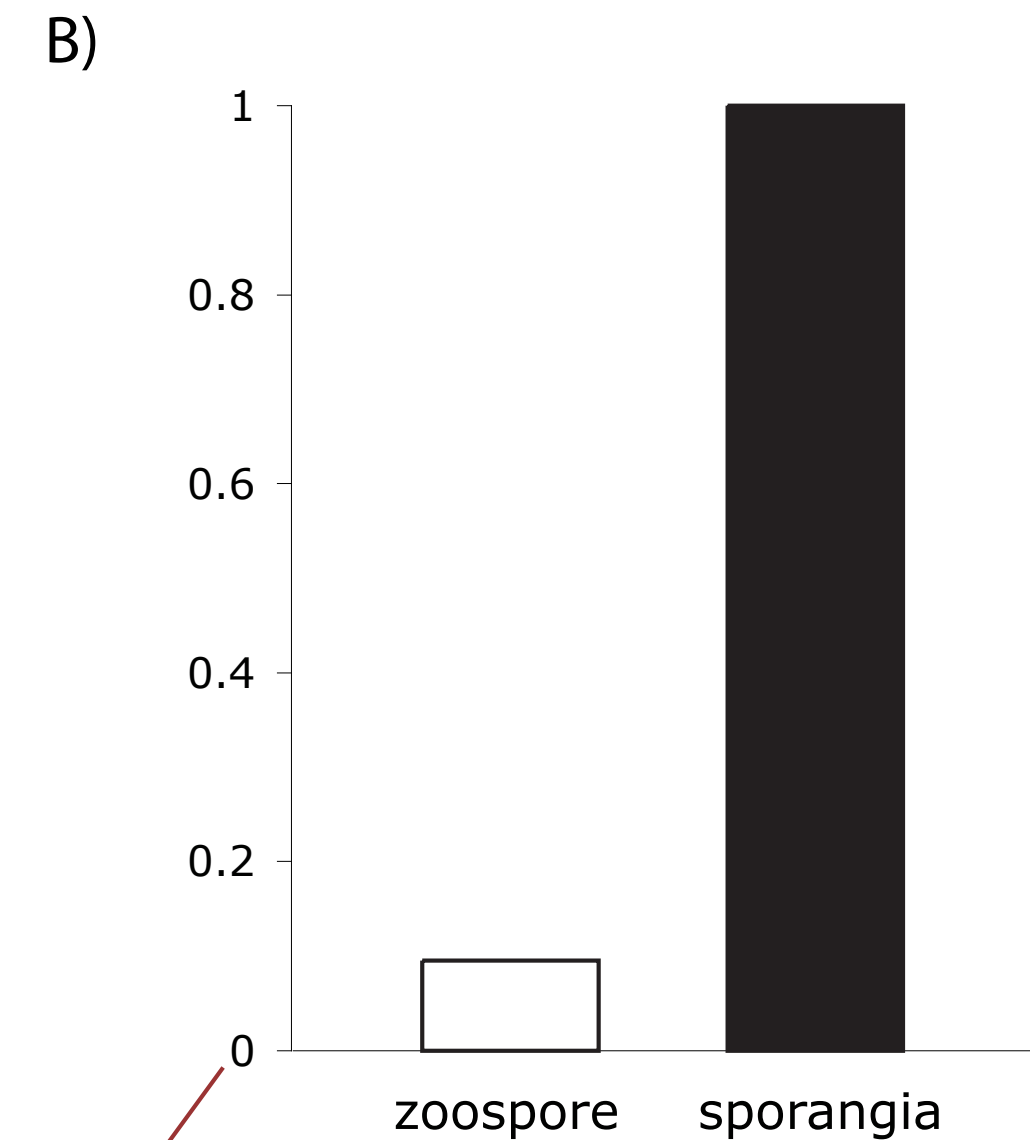
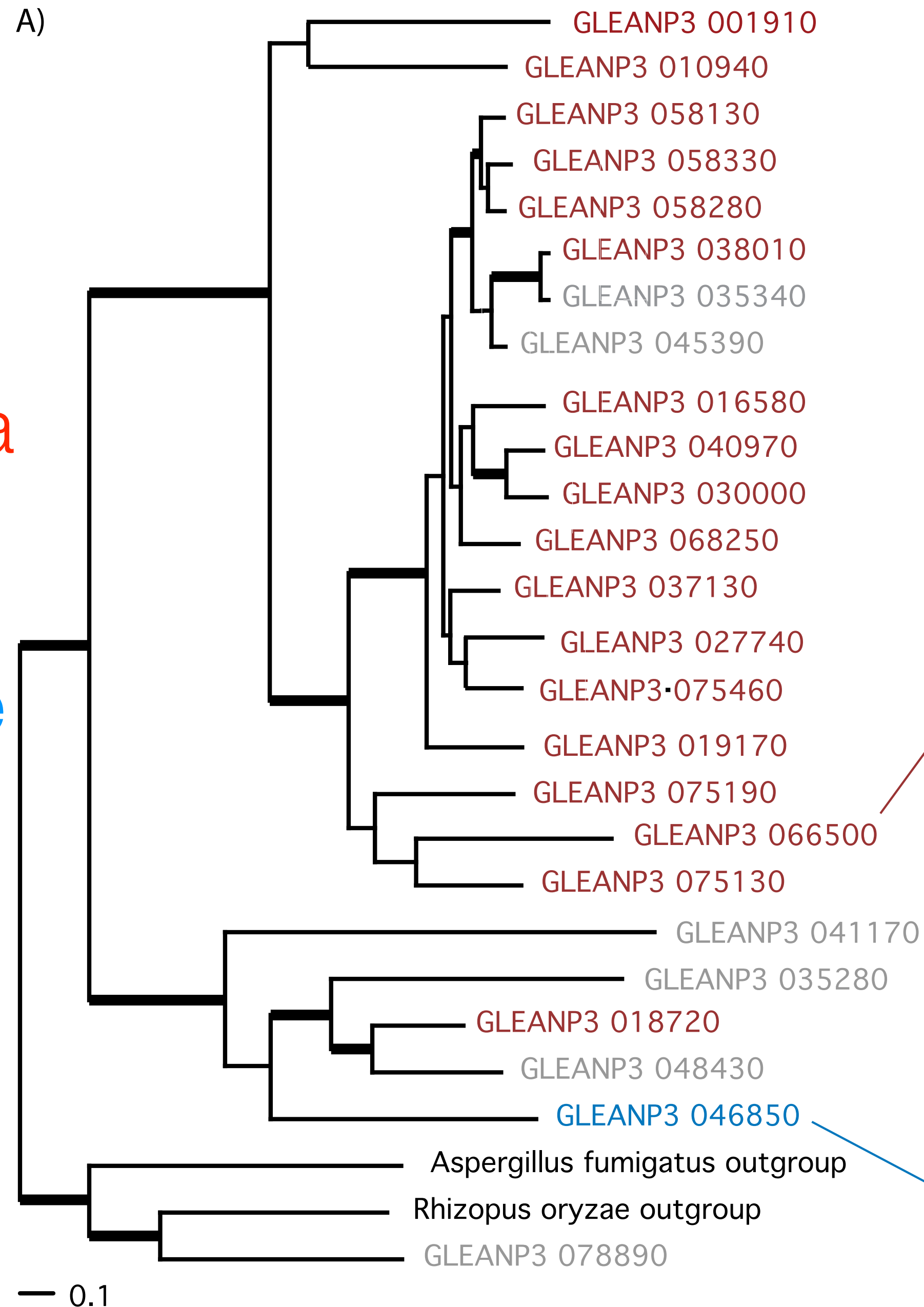


*Hp*- frog skin intact

*Bd*- frog skin degraded

# Fungalysin M36

up in sporangia  
no expression  
diff  
up in zoospore



Rosenblum et al PNAS 2008



# Bd summary

---

- Comparisons to close species can polarize some differences in lineage specific expansions
  - Findings of M36, S41 expansions seem to be recent to *Bd* and could be a functional link to pathogenesis
  - Expansion of Tyrosinase from counts is less compelling, but gene tree analyses suggests recent expansions on *Bd-Hp* branch
  - Lack of FKS1 in early Chytridiomycota lineages suggests a more recent origin of this gene than origin of the Fungi and predicts timing of some changes in fungal cell wall composition
  - CBM18 expansion may be related to adhesion, future experiments to test this
- Future: We are employing a population genomic approach, resequencing 24 strains of *Bd* and hope to understand more about origins and variation in the genome from these data.

# Summary

---

- Unraveling the evolution of pathogens can be tricky, more so when mode of pathogenesis is not obvious (e.g. not just a toxin gene)
- Comparative genomics at the scale of gene and gene families can suggest changes that may be important in adaptation of a species.
- Connecting these molecular changes to pathogenecity is still needed to understand the role these expansions may play - but provides rich experimental fodder for the laboratory.



# Acknowledgements

## Coccidioides

Tom Sharpton {UCSF} John Taylor {UC Berkeley}  
Dan Neafsey {Broad Institute}  
Bridget Barker, Marc Orbach, Steve Rounsley {U  
Arizona}

## Cell Wall profiling

Hugo Mérida, Vincent Bulone {KTH, Stockholm}

## Stajich lab @UCR lab

Steven Ahrendt  
Divya Sain  
Yizhou Wang  
Yi Zhou  
Jessica De Anda  
Annie Nguyen

W.M. Keck  
Foundation



BURROUGHS  
WELLCOME  
FUND 



ALFRED P. SLOAN  
FOUNDATION

## B. dendrobatidis

John Abramyan {UCR}  
Edgar Medina {Universidad de Los Andes,  
Colombia}  
Christina Cuomo {Broad Institute}  
Tim James {U Michigan}  
Suzanne Joneson, Tom Poorten, Erica Rosenblum  
{U Idaho}  
Igor Grigoriev {JGI}

## H. polyrhiza

Suzanne Joneson, Erica Rosenblum {U Idaho}  
Shin-Han shiu {Michigan State}  
Joyce Longcore {U Maine}  
Tim James {U Michigan}

<http://lab.stajich.org>

<http://mobedac.org>

<http://fungidb.org>