

Coccidioides Genomic Resources and Comparative Analyses

Jason Stajich

Dept of Microbiology and Plant Pathology

NASEM The Impact and Control of Valley Fever



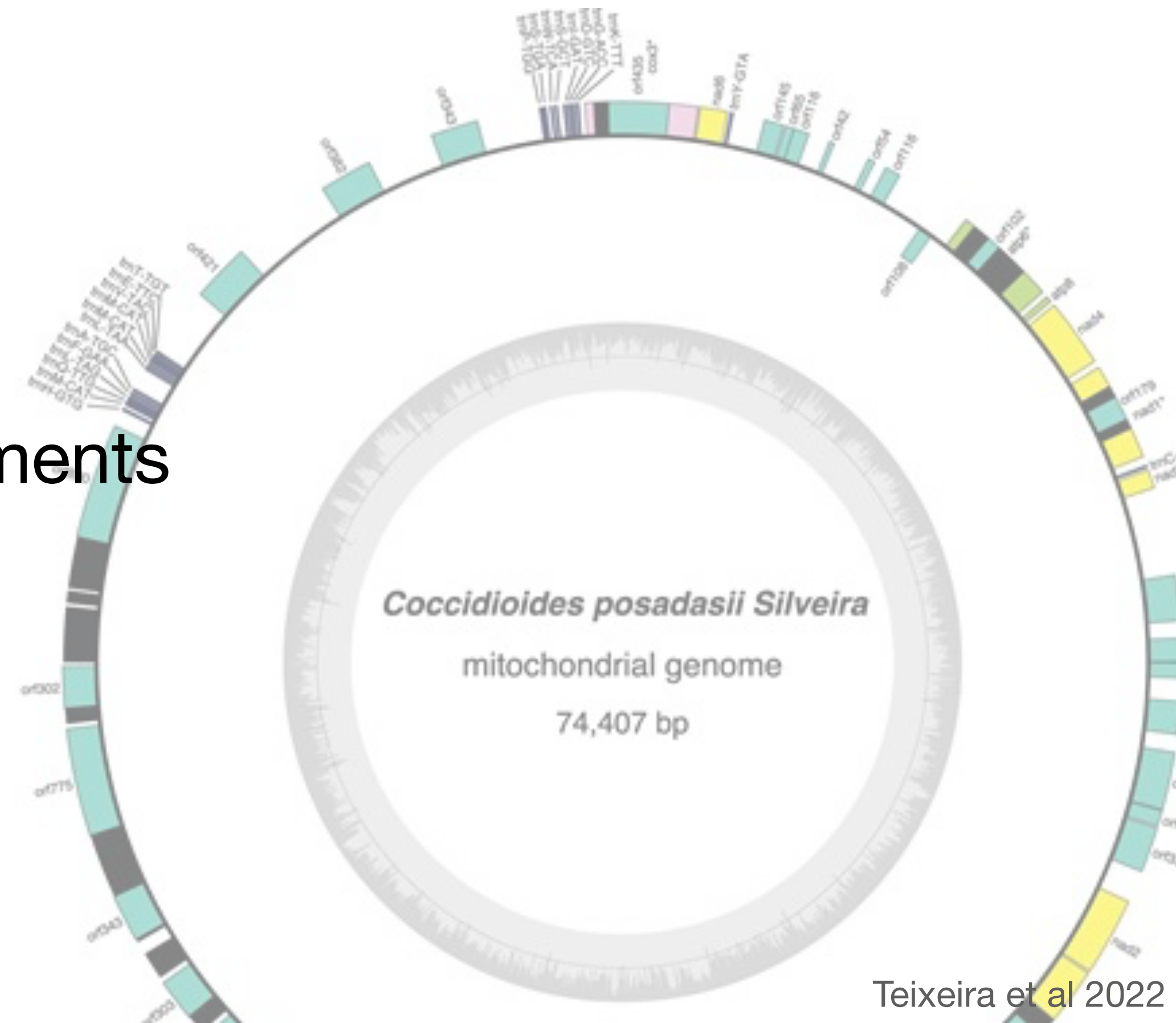
Coccidies Genomics

Building resources to enable investigations of biology and evolution

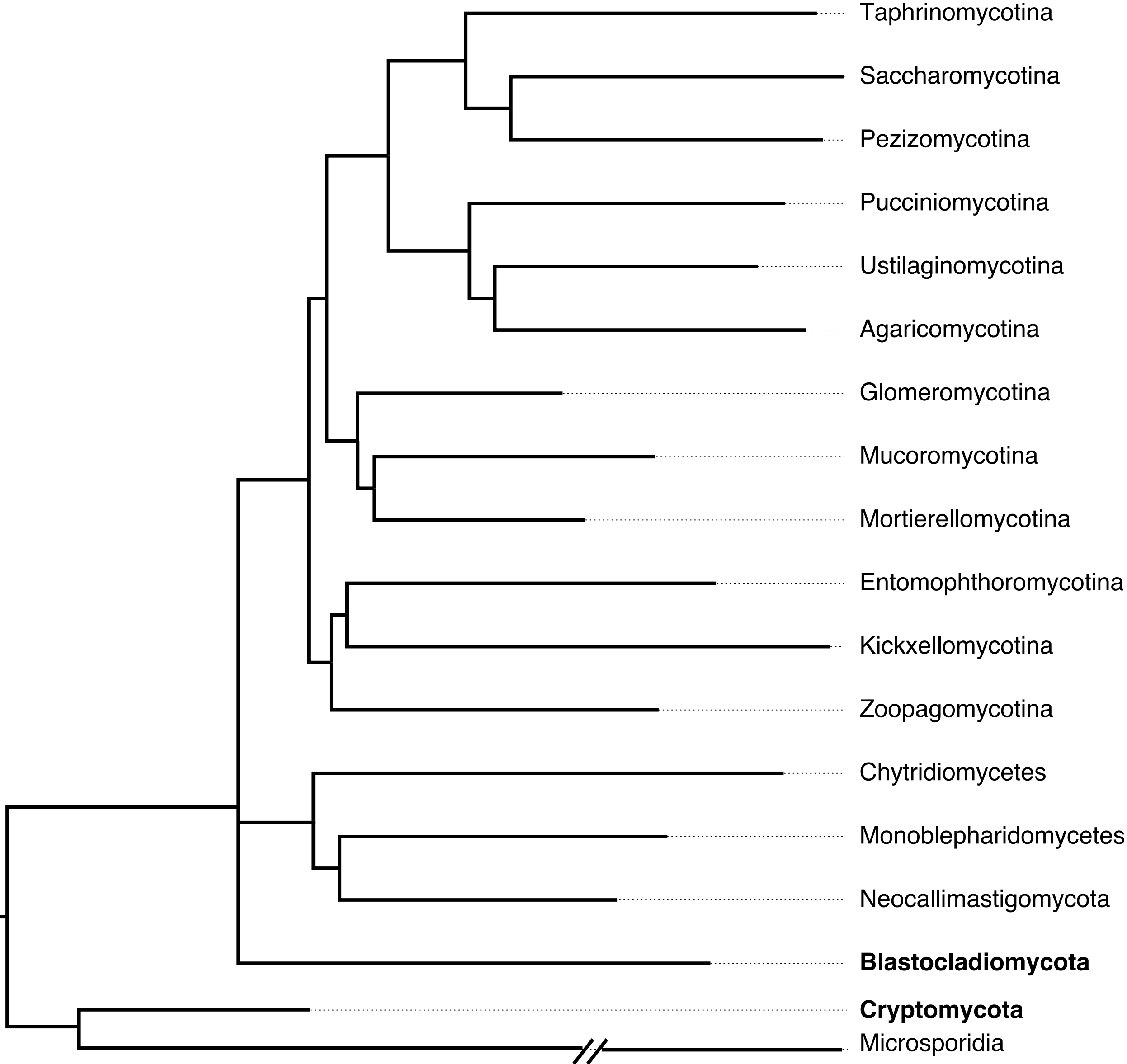
- Investigating the genetic basis of *Coccidioides* unique biology
- Exploring genomic content and stability among isolates and species
- Guiding gene knockouts and genetic manipulation
- Gene expression investigation and developmental stage comparisons
- More understanding of the specific metabolism and anti fungal susceptibility
- Support work towards vaccine production

How are Coccidioides genomes structured?

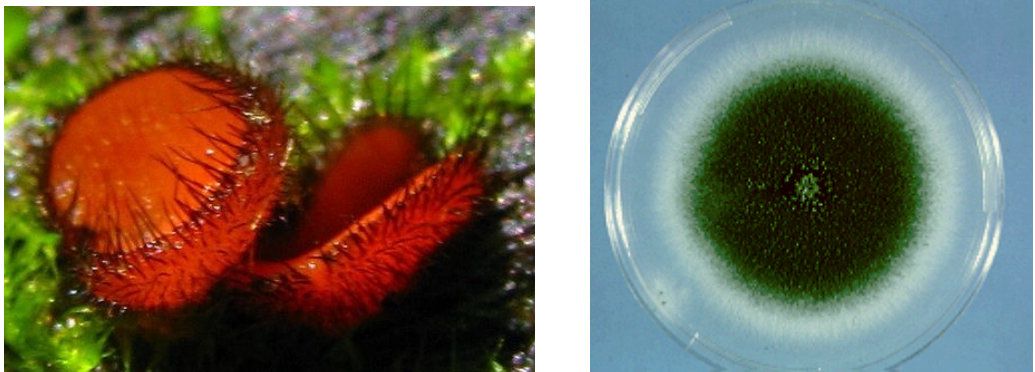
- How large are genomes compared to other fungi and other Onygenales Fungi?
- Gene dense and compact or dispersed?
- Gene families and gene content
- Proliferation and Diversity of Transposable elements



KINGDOM FUNGI



Ascomycota
Onygenales



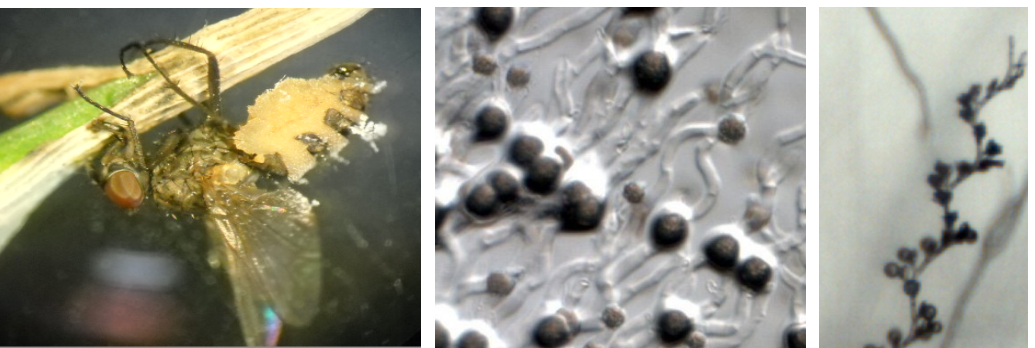
Basidiomycota



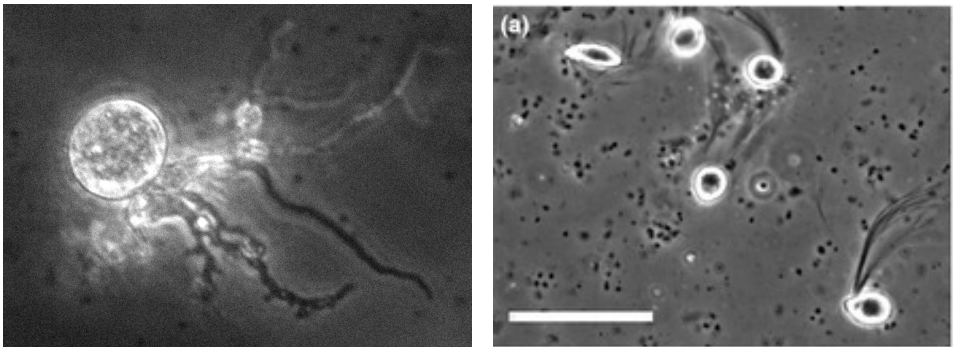
Mucoromycota



Zoopagomycota

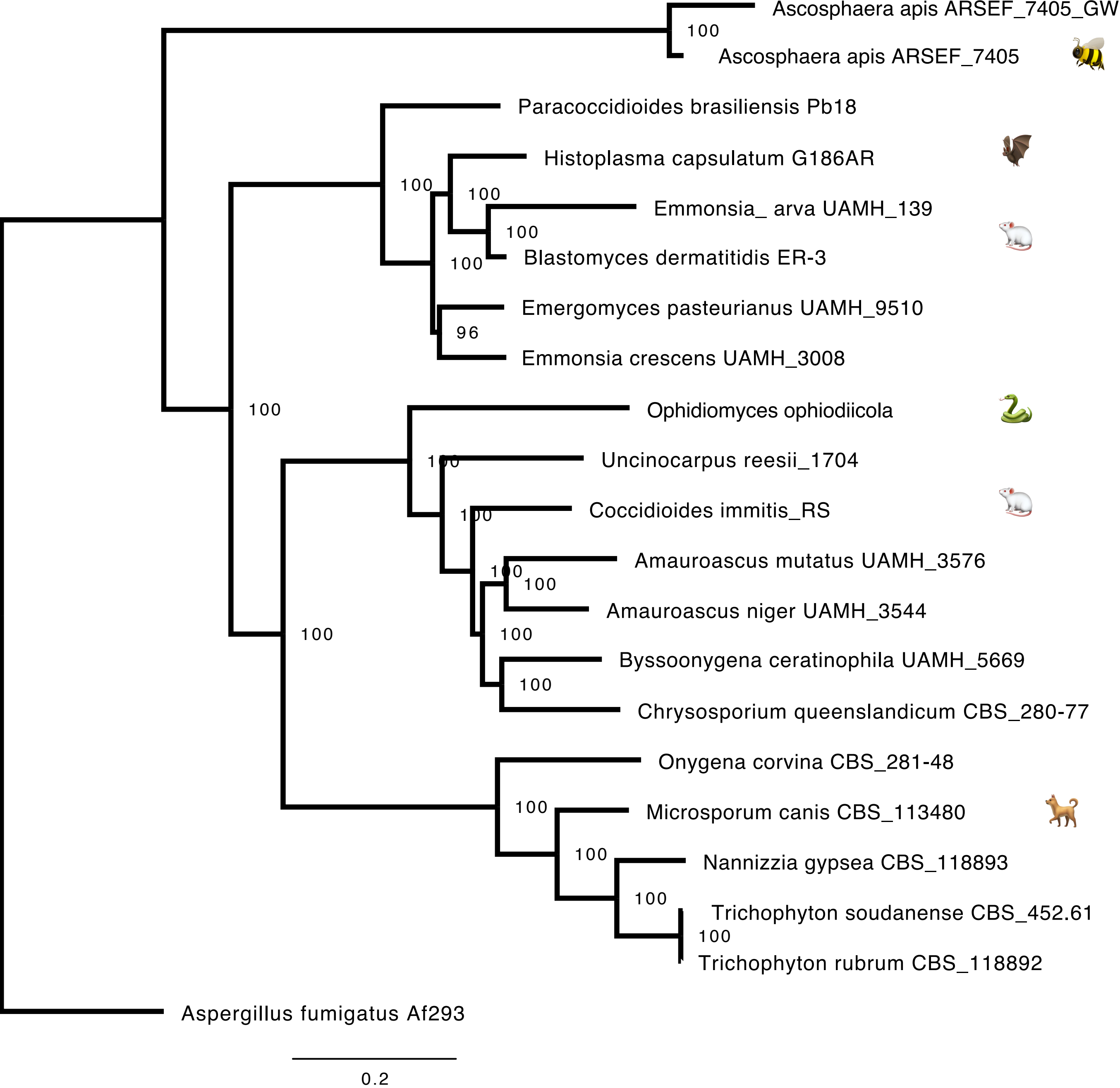


Chytridiomycota

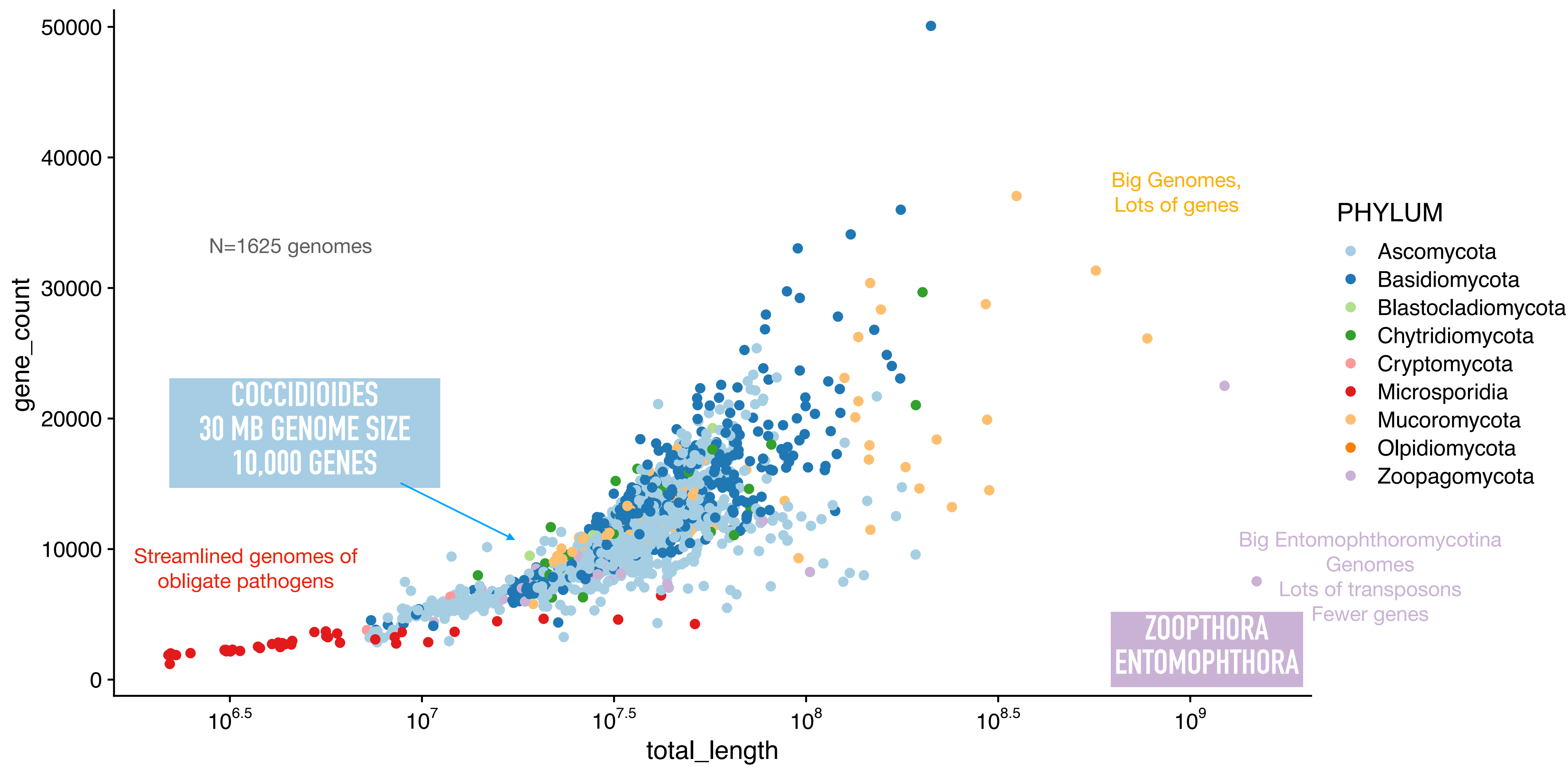


Phylogeny of the Onygenales fungi

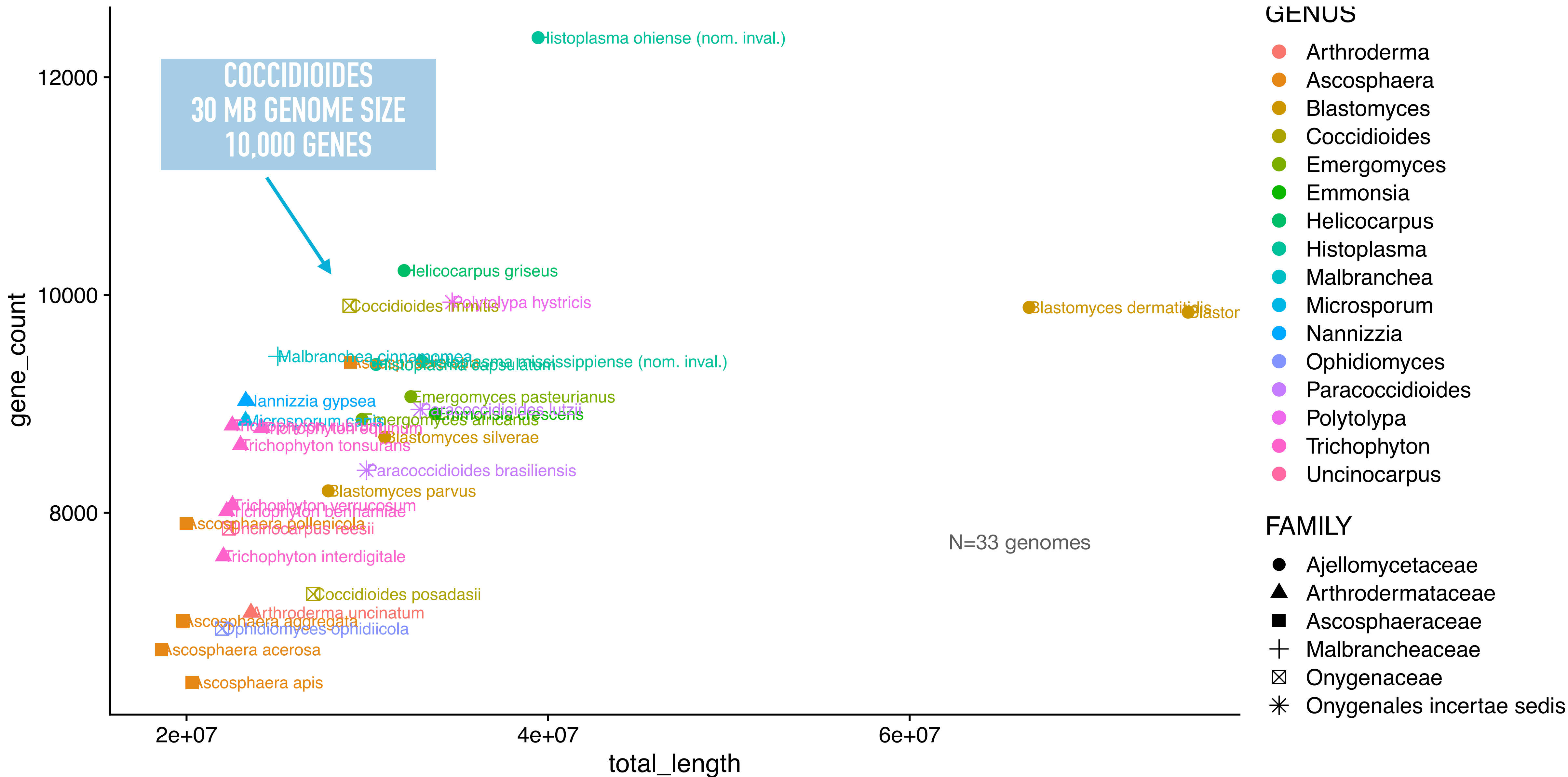
Tree of species with sampled genomes



How does genome size vary across the Fungal Kingdom?



How does genome size vary across the Onygenales



Genome and Transcriptome Resources

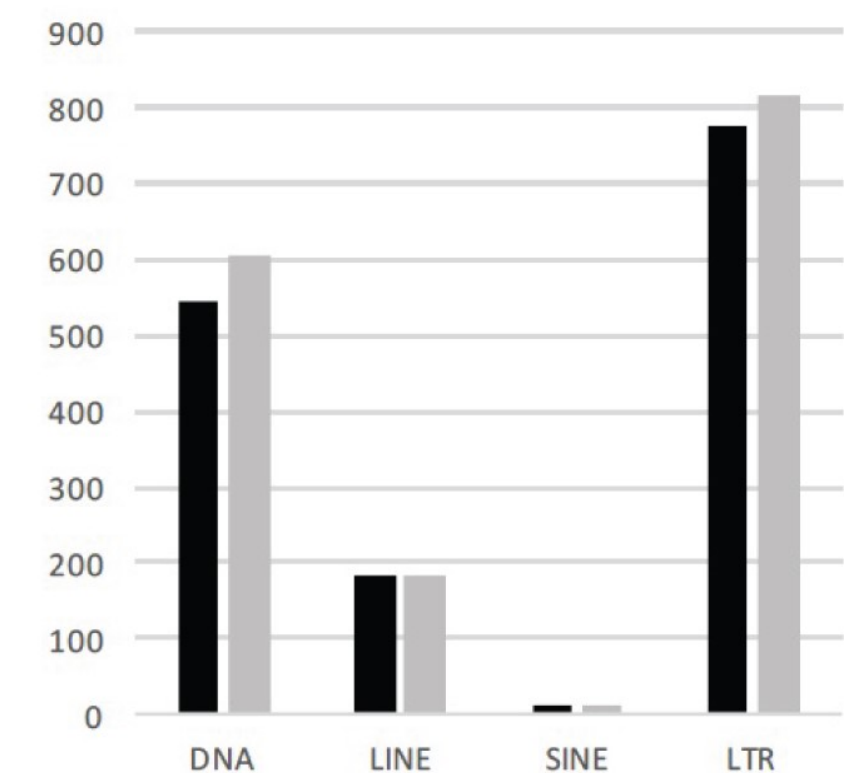
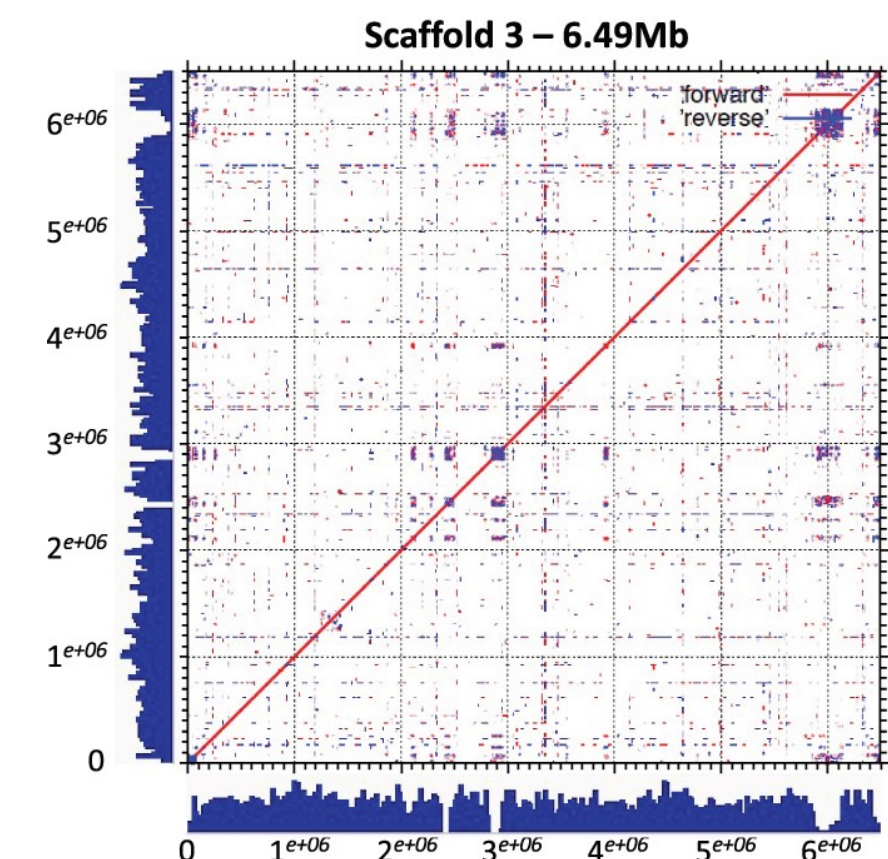
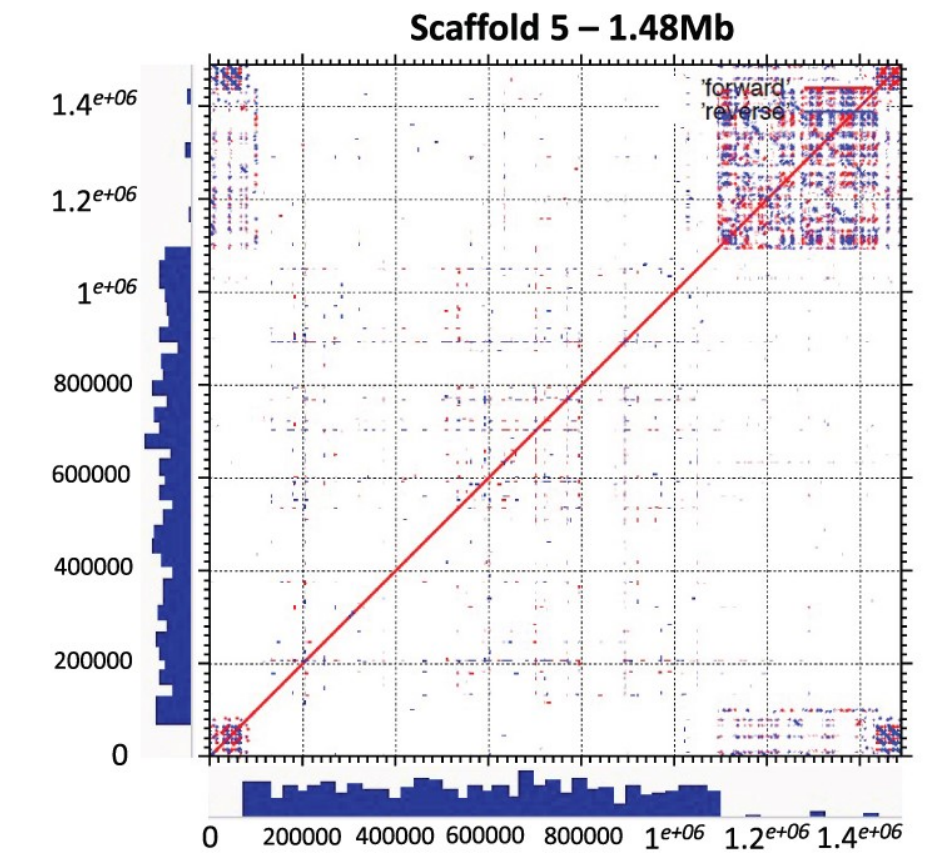
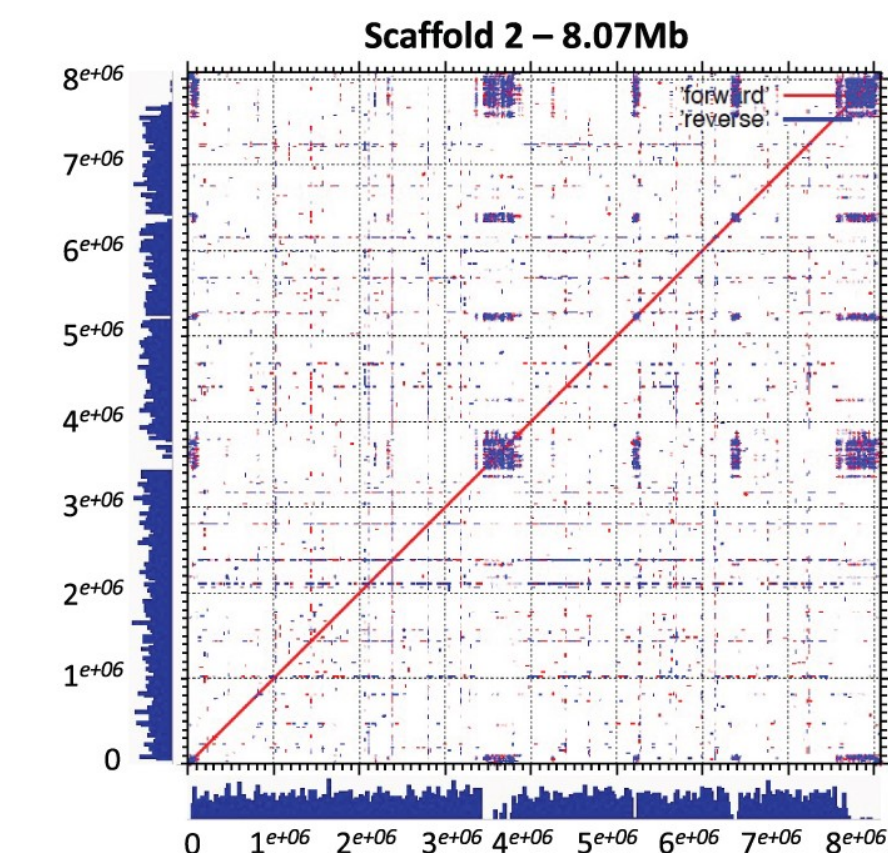
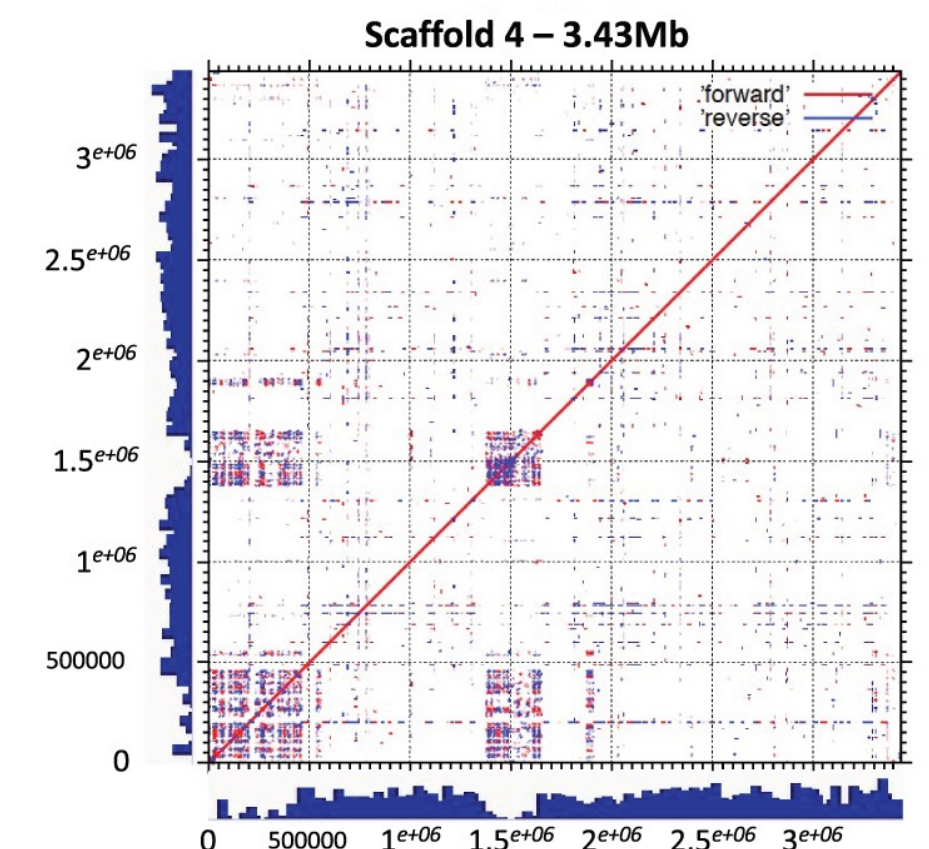
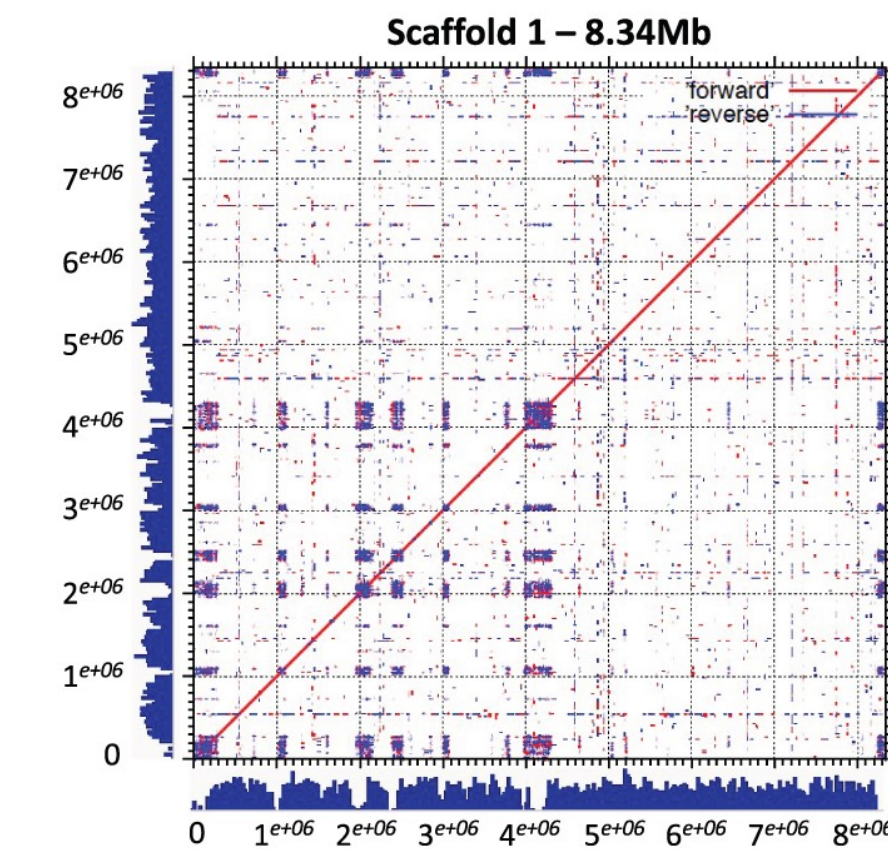
Some of the available resources

- *Coccidioides immitis* and *C. posadasii* genome resources - Sanger sequencing
 - Broad Institute - Fungal Genome Initiative, TIGR - JCVI (2004-2006)
- Population sequencing (Illumina) - U Arizona, CDC, TGen and Northern Arizona University
- RNA-seq profiling and Microarray work - UC Berkeley, UT Southwest, UCSD, ...
- Reference genomes with long read genome sequencing (PacBio)— NAU and U Arizona

New Genome Resources

Long read genome assemblies to improve contiguity and gene recovery

- Goals to improve assembly to examine full gene content
- *C. posadasii* Silveira PacBio reference genome (Barker Lab)
- New PacBio generated genomes (Orbach Lab)
- 5 chromosomes



Assembly & Annotation Statistics

Telomere to Telomere assembly recovery

PacBio HiFi enables near complete genomes

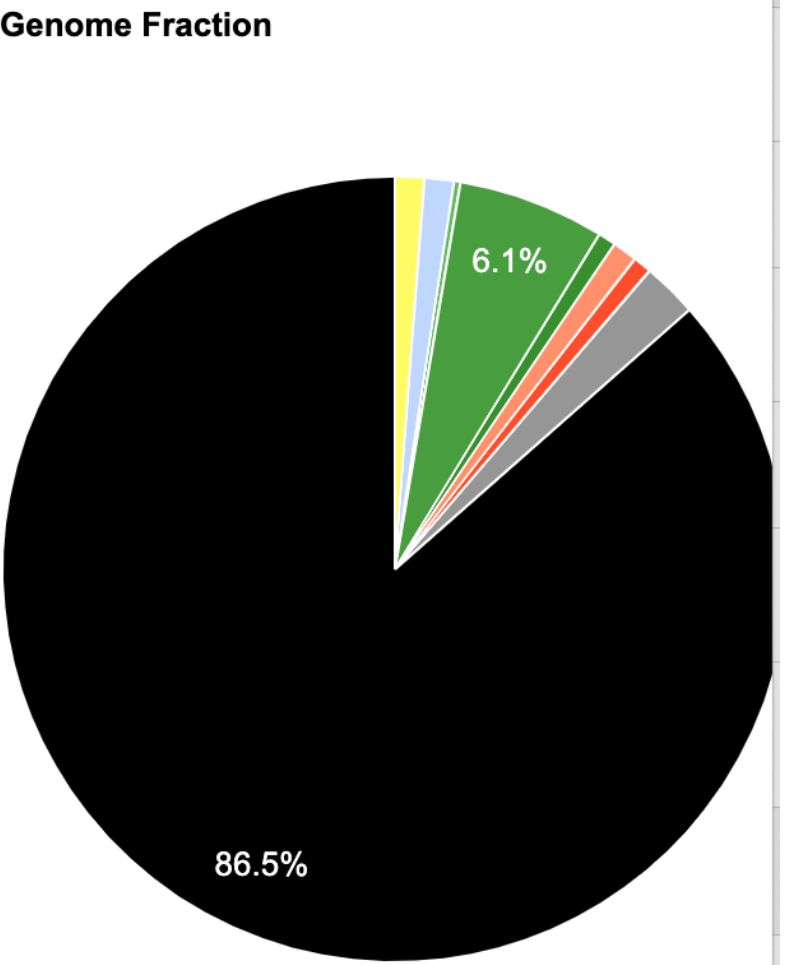
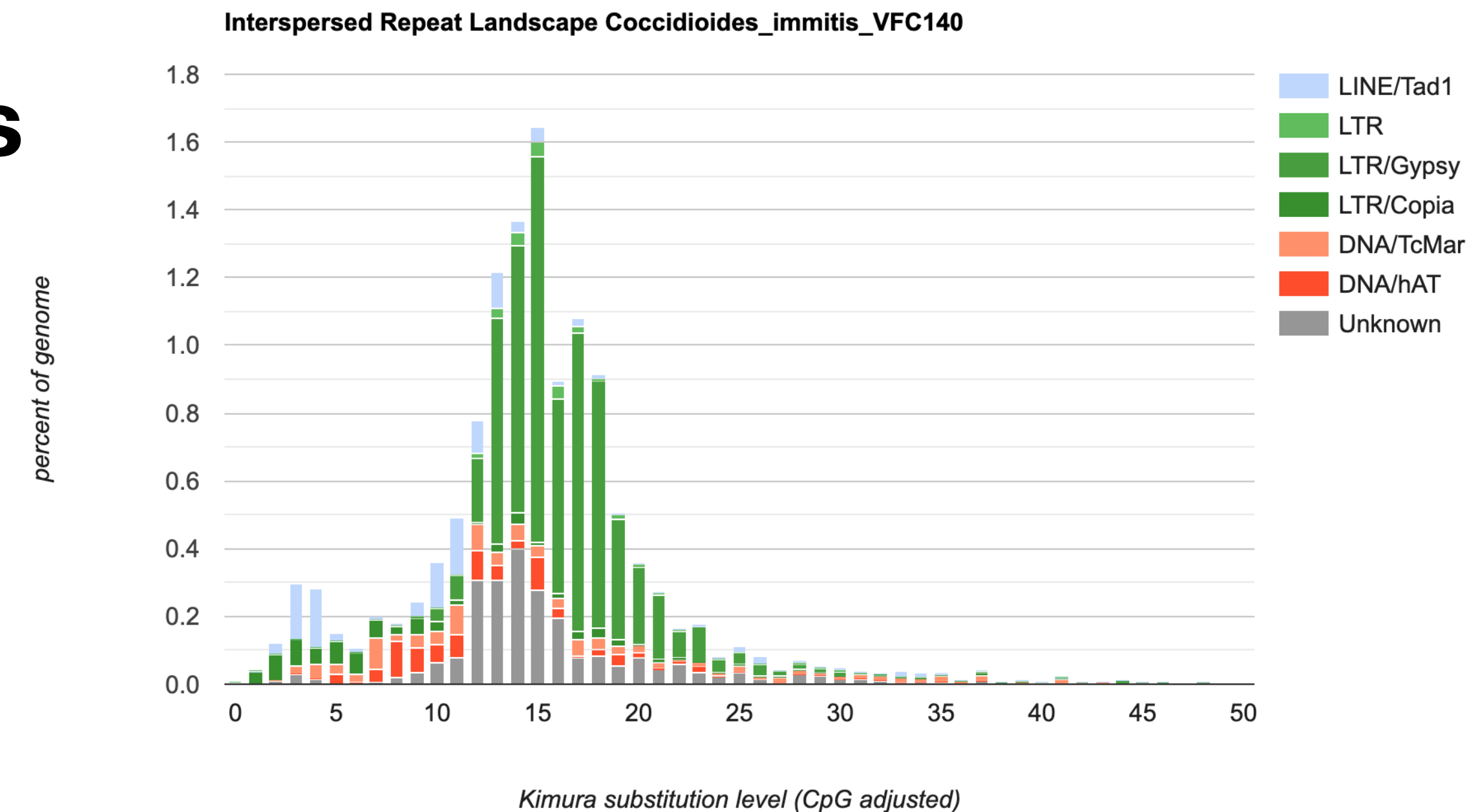
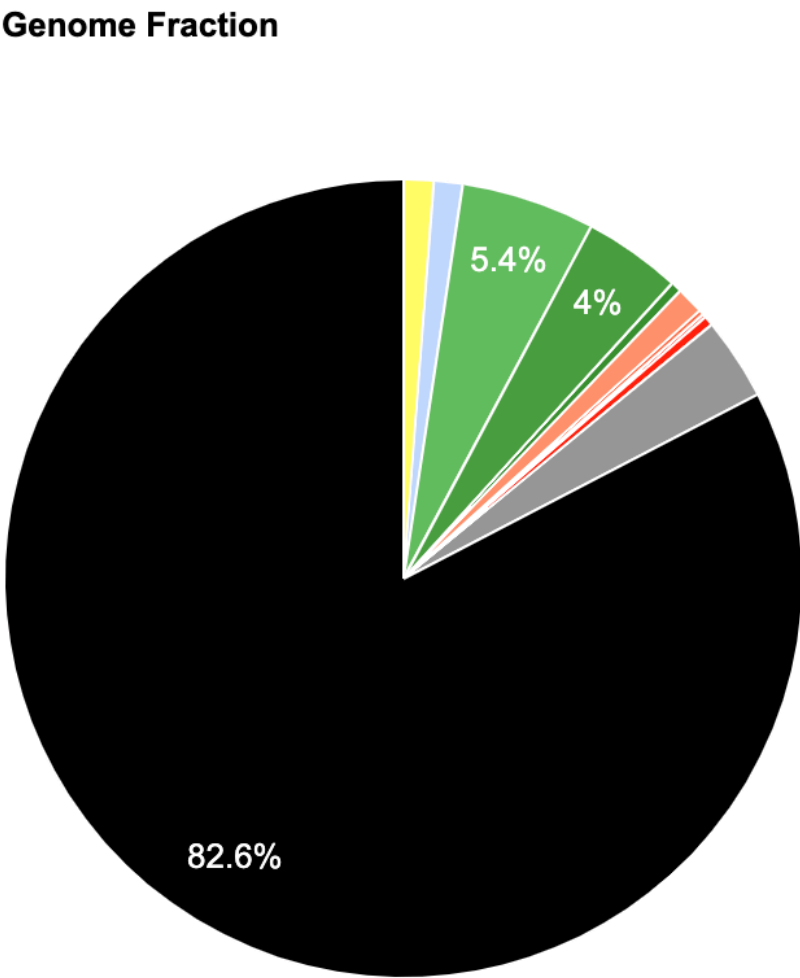
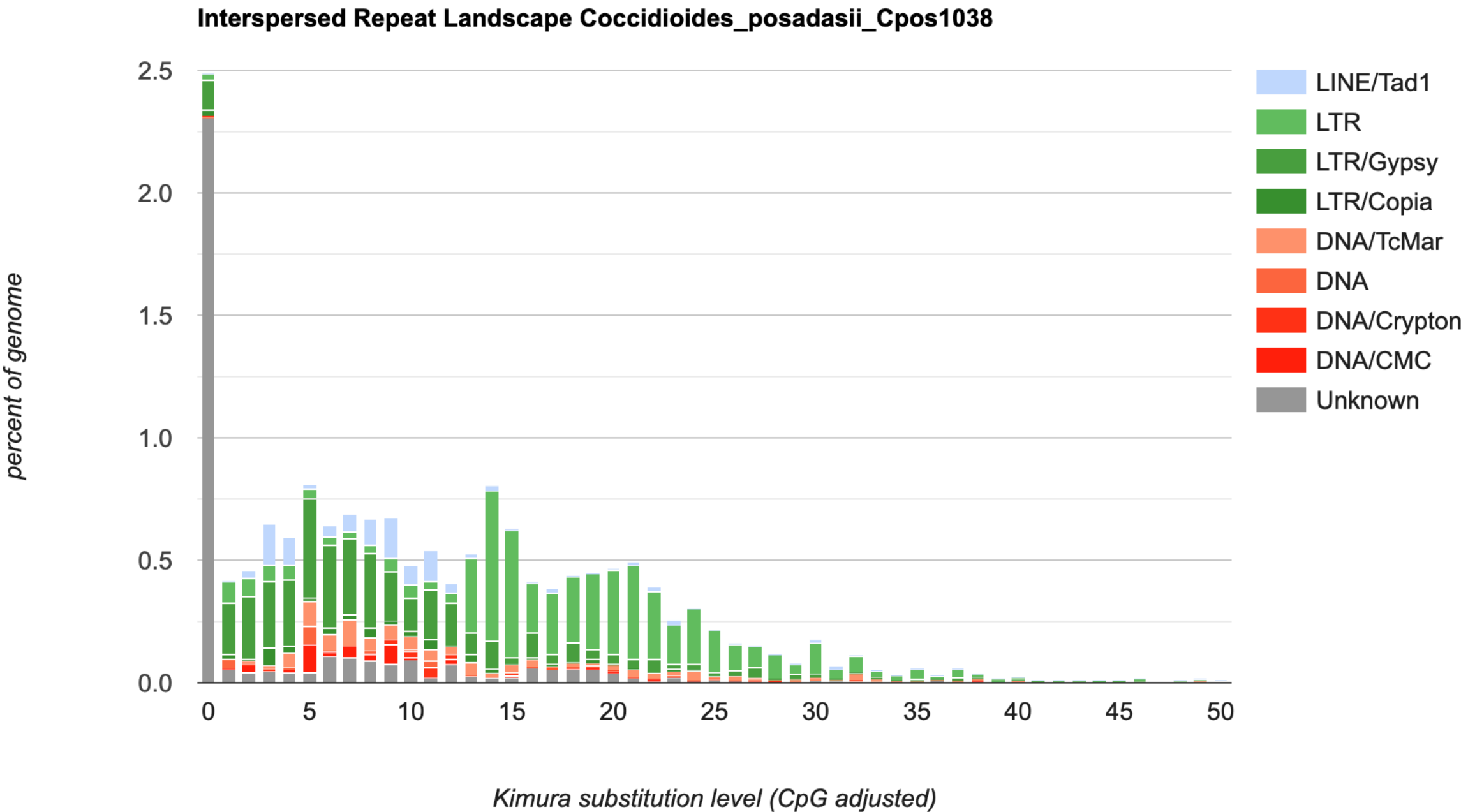
Predicted gene content can vary due to accuracy of prediction models and available RNA-Seq evidence.

Also may reflect genomic differences among isolates

	Size (Mb)	Scaffolds	Telomere containing scaffolds	Complete T2T	Predicted genes	Repeat Content
<i>C. immitis</i> CiB10637	28.0	14	5	4	8,493	12.98%
<i>C. immitis</i> CiB10992	28.4	37	5	4	8,564	12.83%
<i>C. immitis</i> VFC140	31.4	81	5	4	9,964	13.51%
<i>C. posadasii</i> Cpos1038	30.3	90	5	3	8,803	17.82%
<i>C. posadasii</i> Silveira *	28.1	8	5	-	10,228	16.91%

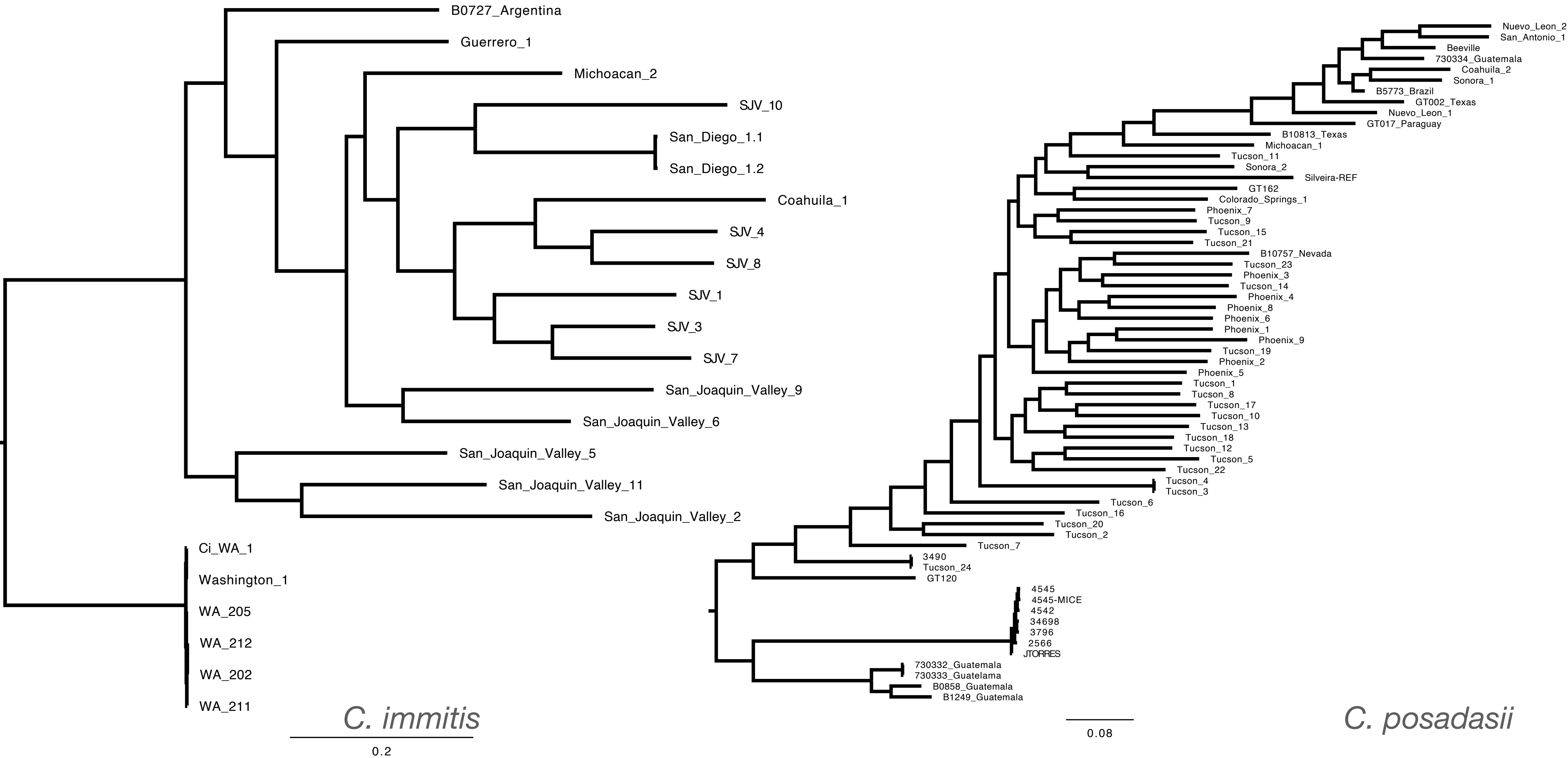
Age and timing of repetitive element activity

Diversity of TEs within Coccidioides genomes



**How plastic are the *Coccidioides*
genomes?**

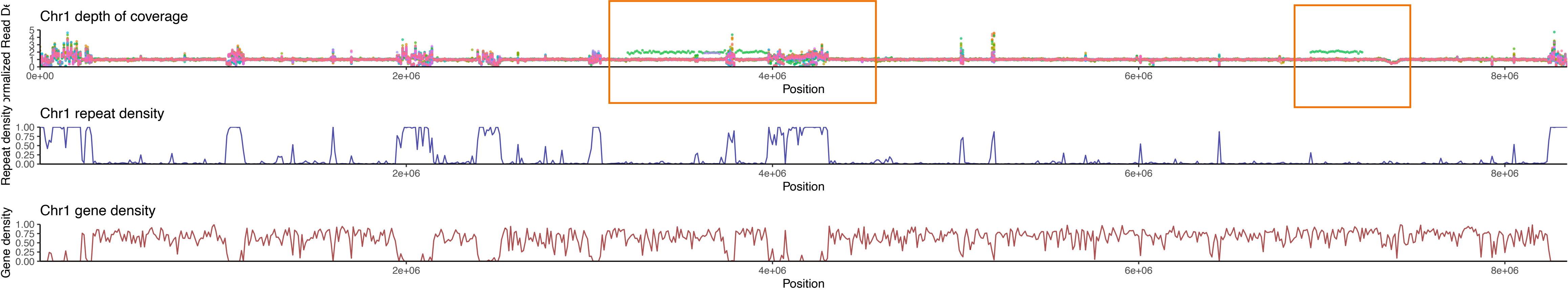
Population diversity sampling of two species of *Coccidioides*



Comparison of genomic variability across populations

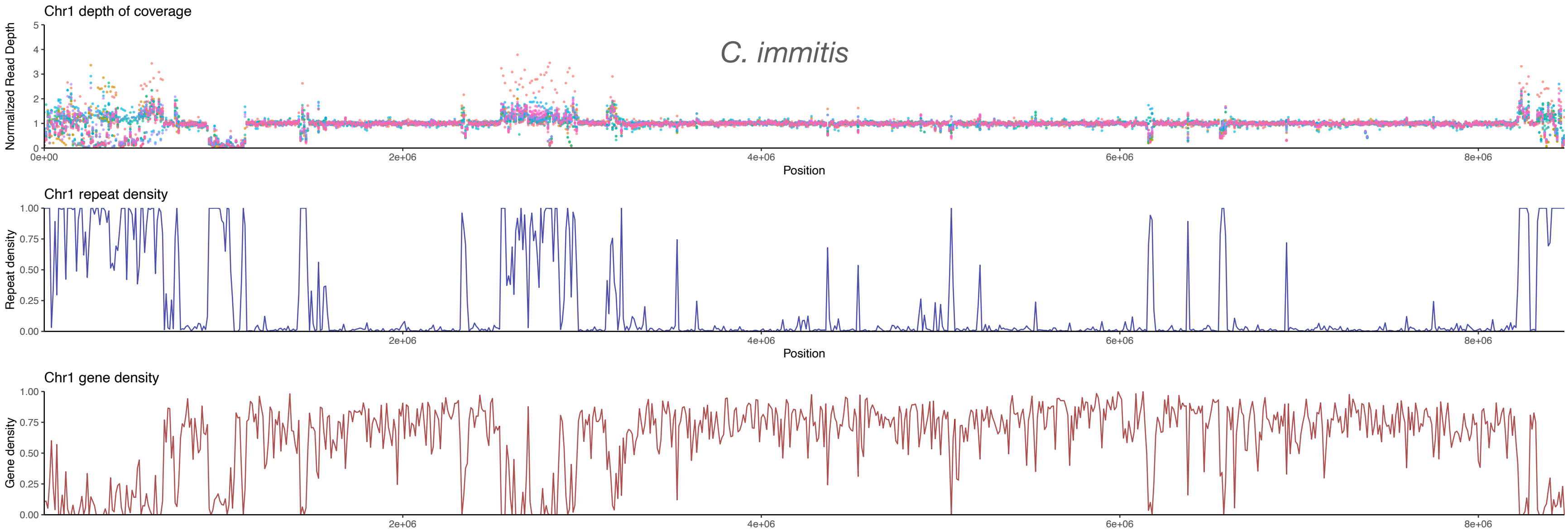
Sliding window plots of coverage (Illumina reads mapped to genome) and TE and Gene content

C. posadasii



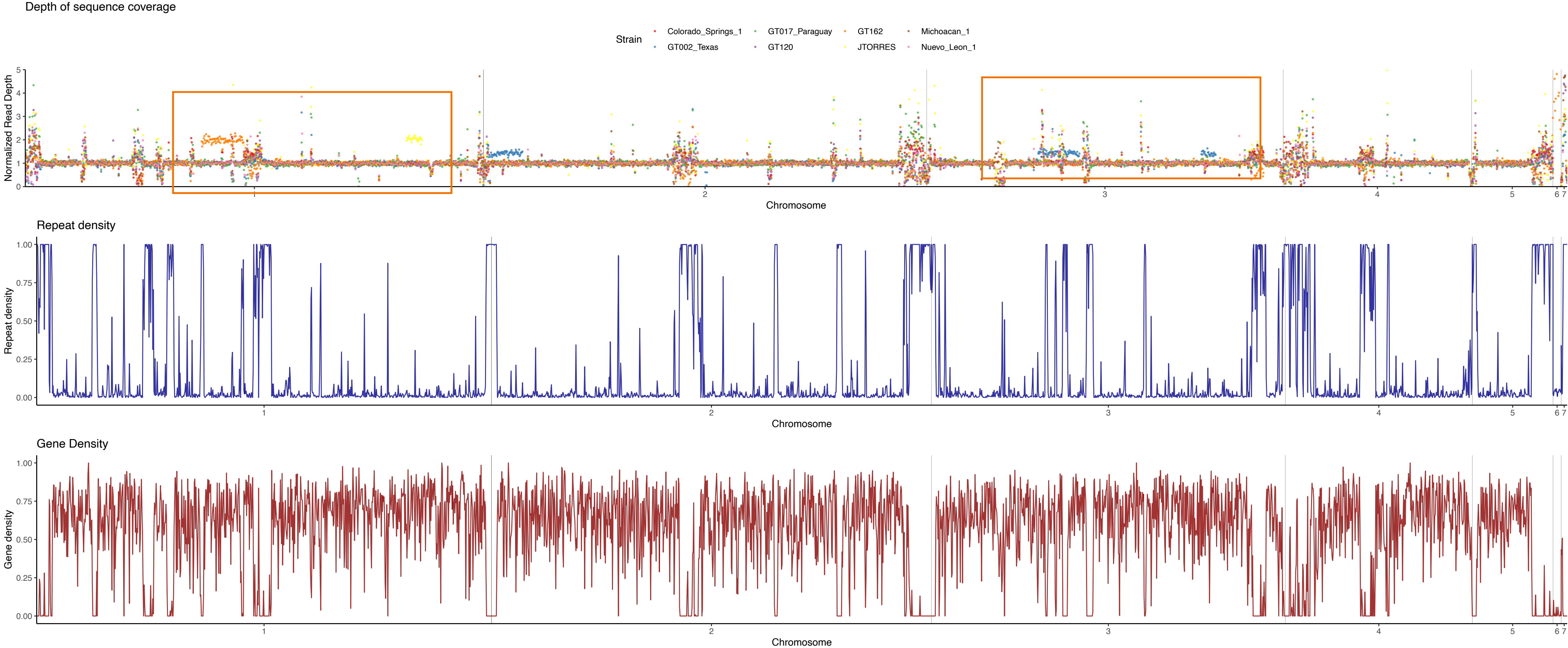
Colors are different strains mapped to genome

C. immitis



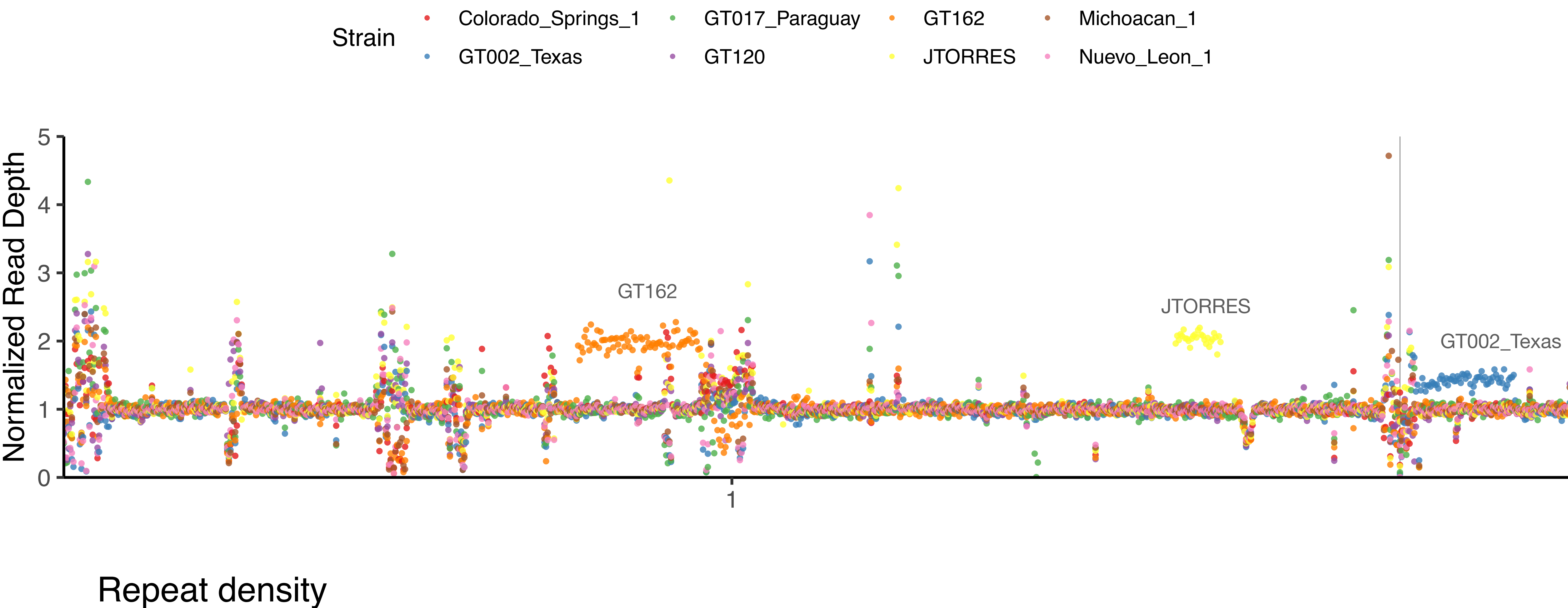
C. posadasii Copy Number Variation regions

Unique CNVs to different strains (different colors), some relate to genes



C. posadasii Copy Number Variation regions

Unique CNVs to different strains (different colors), some relate to genes



Search for...

expand all | collapse all

Filter the searches below...?

Genes

Organisms

Popset Isolate Sequences

Genomic Sequences

Genomic Segments

SNPs

ESTs

Metabolic Pathways

Overview of Resources and Tools

- 

Take a Tour



Getting Started



Search Strategies



Genome Browser



Transcriptomic Resources



Phenotypic Data



Analyze My Data



Downloads



How to Submit Data



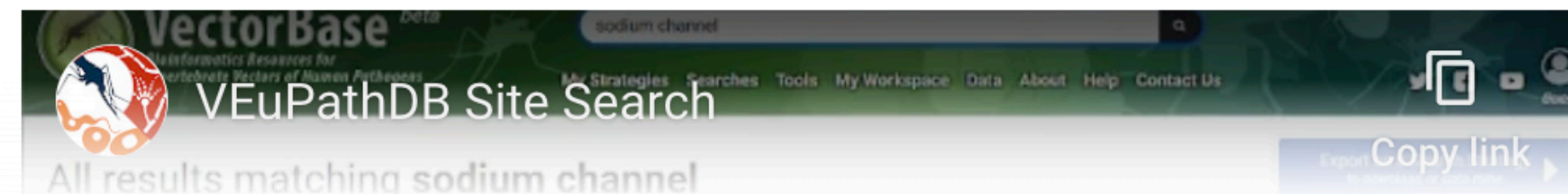
Curation and Annotation

Getting Started

VEuPathDB is packed with data, tools and visualizations that can help answer your research questions. We gather data from many sources, analyze according to standard workflows, and present the results for you to mine in a point and click interface. Here's how to get started:

SITE SEARCH: Explore the site; find what you need

Enter a term or ID in the site search box at the top of any page. The site search finds documents and records that contain your term and returns a summary of categorized matches. Its easy to find genes, pathways, searches, data sets and more with the site search.

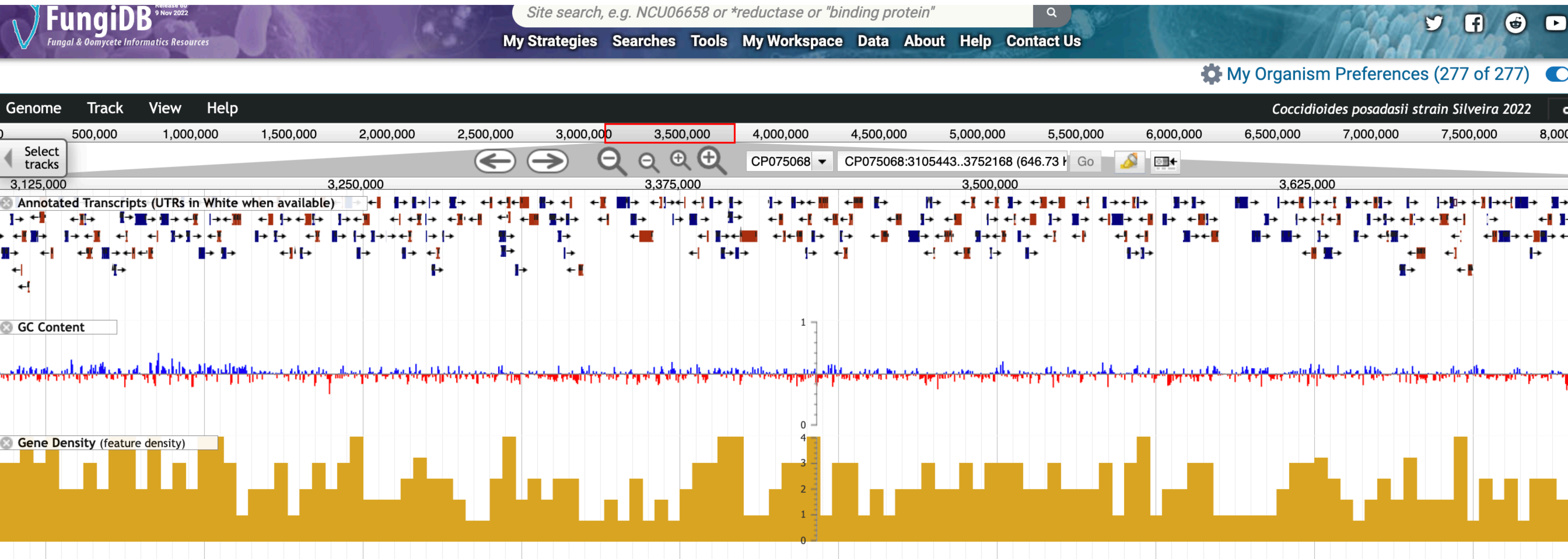


FungiDB / VEuPathDB

NIH supported integrated database and analysis of fungal genomes and resources

VuEuPathDB/FungiDB resources for fungal genomics

Genome Browser View of Features



VuEuPathDB/FungiDB resources for fungal genomics



Release 60
9 Nov 2022

Fungal & Oomycete Informatics Resources

Site search, e.g. NCU06658 or *reductase or "binding protein"

🔍

My Strategies

Searches

Tools

My Workspace

Data

About

Help

Contact Us

🐦

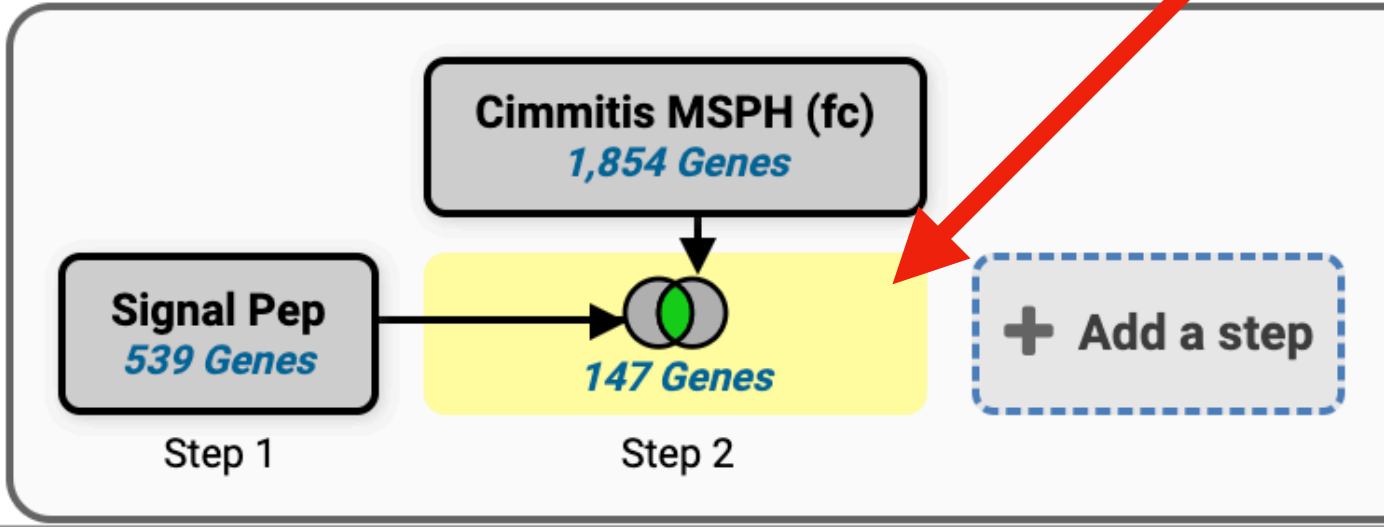
📘

📺

📺

📺

Secreted Spherule Genes *



147 Genes (141 ortholog groups)

For the Experiment

☒ Transcriptional analysis of Coccidioides immitis mycelia and spherules - Sense

☐ Transcriptional analysis of Coccidioides immitis mycelia and spherules - Antisense

return Genes

that are

with a Fold change >=

between each gene's expression value (or a Floor of)

in the following Reference Samples

☐ Mycelia

☐ Young spherules

☒ Mature spherules

and its expression value (or the Floor selected above)

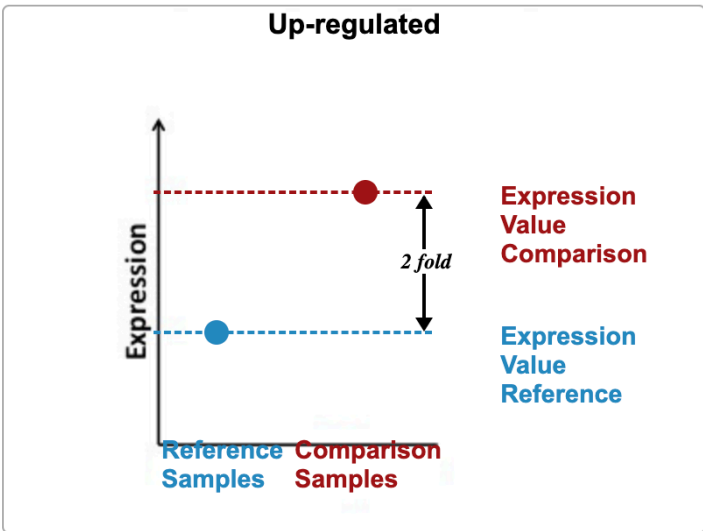
in the following Comparison Samples

☒ Mycelia

☐ Young spherules

☐ Mature spherules

Example showing one gene that would meet search criteria (Dots represent this gene's expression values for selected samples)



For each gene, the search calculates:

$$\text{fold change} = \frac{\text{comparison expression value}}{\text{reference expression value}}$$

and returns genes when fold change >= 2.

You are searching for genes that are up-regulated between one reference sample and one comparison sample.

My Organism Preferences (277 of 277)

enable



Gene Results Genome View Analyze Results

Genes: 147 Transcripts: 148 ☐ Show Only One Transcript Per Gene

◀ 1 2 3 ... 8 ▶ Rows per page: 20

Download

Send to...

Add Columns

	Gene ID	Transcript ID	Genomic Location (Gene)	Product Description	C. immitis Saprobic vs Parasitic Growth - Saprobic Hyphae - unique (C.i. Saprobic/Parasitic RNA-Seq)	C. immitis Saprobic vs Parasitic Growth - Parasitic Spherules - unique (C.i. Saprobic/Parasitic RNA-Seq)
🛒	CIMG_09696	CIMG_09696-t26...	GG704912:498,208..499,596(+)	proline-rich antigen	14474.59	4103.31
🛒	CIMG_00925	CIMG_00925-t26...	GG704911:6,043,708..6,044,849(-)	hypothetical protein	11355.69	53.43
🛒	CIMG_06973	CIMG_06973-t26...	GG704912:3,988,013..3,989,128(+)	hypothetical protein	1928.17	2929.23
🛒	CIMG_03870	CIMG_03870-t26...	GG704916:2,403,434..2,405,017(-)	hypothetical protein	1927.25	95.37

CIMG_00925

expand all | collapse all

Search section names...

▶ 1 Gene models

▶ 2 Annotation, curation and identifiers

▶ 3 Link outs

▶ 4 Genomic Location

▶ 5 Literature

▶ 6 Taxonomy

● ▼ 7 Orthology and synteny

▶ 8 Genetic variation

▶ 9 Transcriptomics

▶ 10 Sequence analysis

▶ 11 Sequences

▶ 12 Structure analysis

▼ 13 Protein features and properties

Ortholog Group

Orthologs and Paralogs within FungiDB

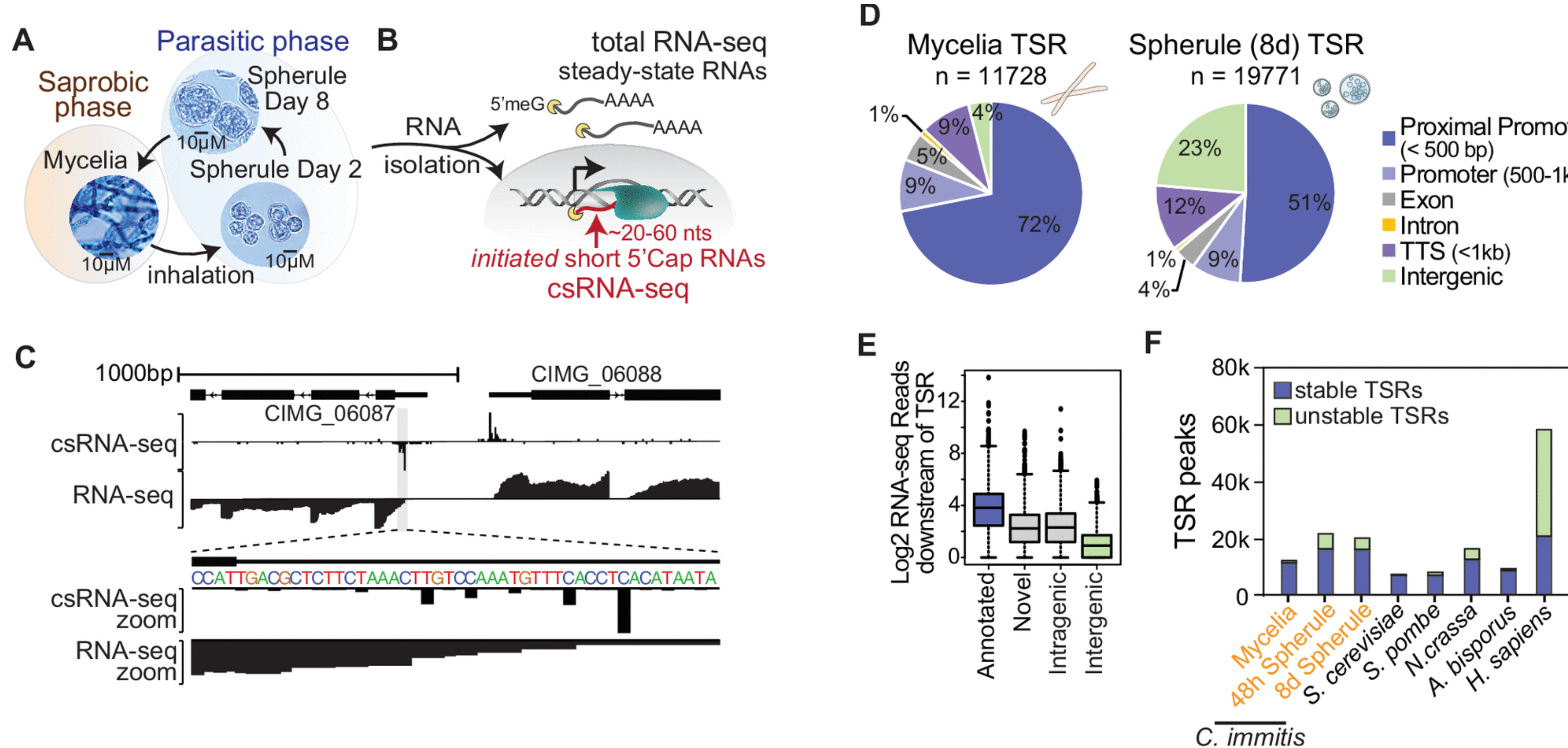
Synteny

Attributes and Protein Browser

☐	PV08_11231	unspecified product	Exophiala spinifera CBS 89968
☐	Z517_10248	hypothetical protein	Fonsecaea pedrosoi CBS 271.37
☐	FCIRC_2371	putative SED1-abundant cell surface glycoprotein	Fusarium circinatum NRRL 25331
☐	FCIRC_5584	sporozoite surface 2 precursor	Fusarium circinatum NRRL 25331
☐	FCIRC_9224	clock-controlled 6 (CCG-6)	Fusarium circinatum NRRL 25331
☐	FFUJ_04919	related to sporozoite surface protein 2 precursor	Fusarium fujikuroi IMI 58289
☐	FFUJ_07326	probable clock-controlled protein 6 (CCG-6)	Fusarium fujikuroi IMI 58289
☐	FFUJ_08434	probable SED1-abundant cell surface glycoprotein	Fusarium fujikuroi IMI 58289
☐	FGRAMPH1_01G07345	clock-controlled 6	Fusarium graminearum PH-1

↕ Name	↕ Interpro ID	↕ Primary ID	↕ Secondary ID	↕ Description	↕ Start Min	↕ End Min	↕ E-value	↕ Transcript ID(s)
InterProScan	N/A	IPR038843	Sed1/Spi1	Cell wall protein Sed1/Spi1	4	115	N/A	CIMG_00925-t26_1

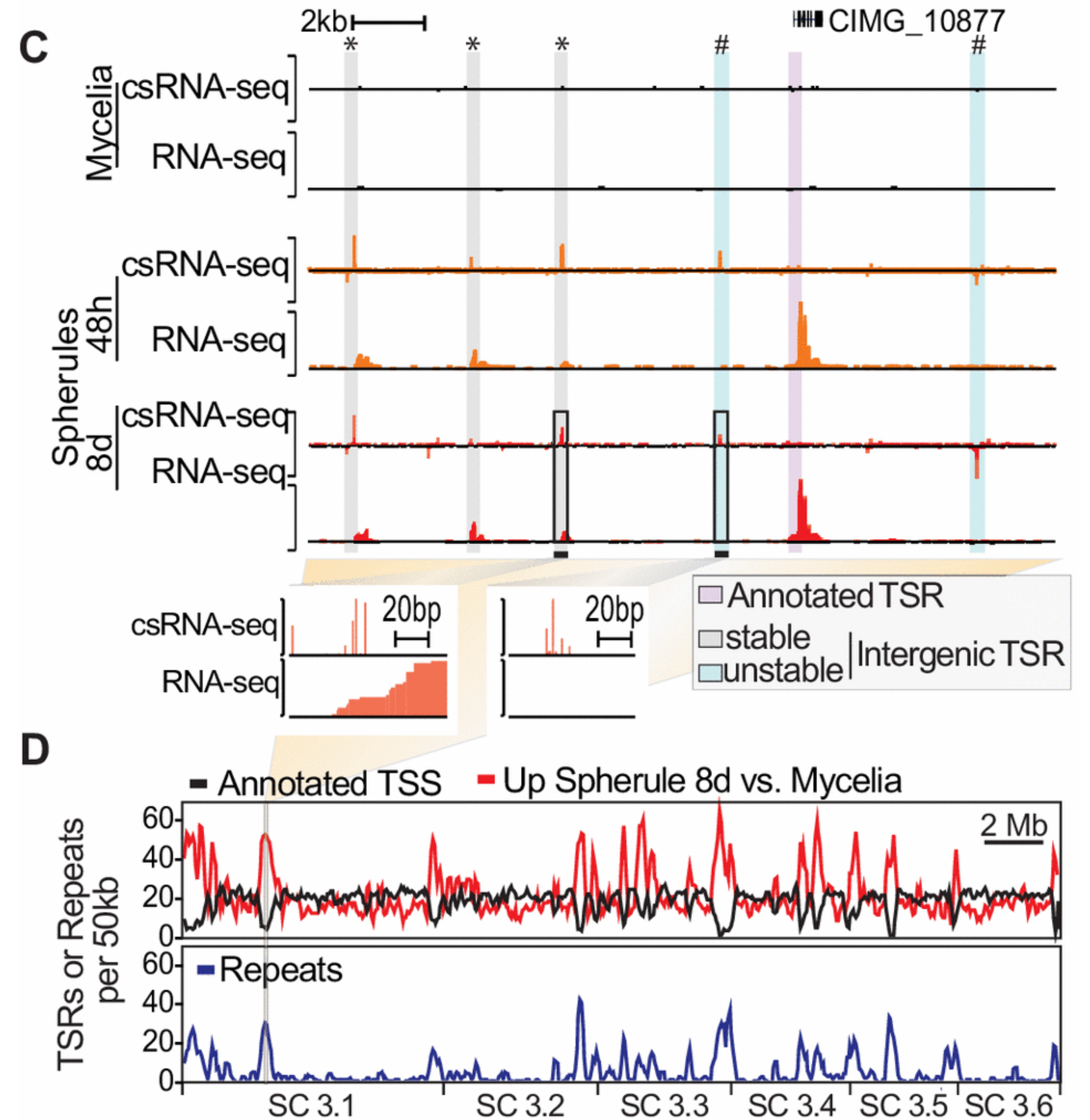
Capped small RNA sequencing (csRNAseq) for capturing actively initiated transcripts from genes in *C. immitis*



csRNAseq

Improve annotation and defining transcriptionally active genome regions

- Defining of promotor regions and regulatory transcribed start regions (TSR) such as enhancers
- 11,000 TSRs in mycelia, 2x as many in Spherules
- Most of these found within 500 bp of 5' end of genes and there were often multiple of upstream of a gene initiating transcription
- Alternative regulation beyond *cis*-regulation in mycelia-spherule transition



Acknowledgements

Kelsey Aadland - UC Riverside

Bridget Barker, Marcus Teixeira - Northern Arizona University

Marc Orbach - University of Arizona

Sasha Duttke, Aaron Carlin, Theo Kirkland, Josh Fierer - UC San Diego

Anita Sil and UC Valley Fever Consortium

Funding Support for this work



Other Funding Support

