

Organização e apresentação de arquivos e resultados

Desirrê Petters-Vandresen

Módulo I – Genômica no Estudo de Microrganismos

Princípios gerais

- Uma boa organização e apresentação das informações, anotações, predições, resultados e conclusões obtidos no estudo:
 - Compreensível para os leitores
 - Tem relação com as perguntas, hipóteses e objetivos do estudo
 - Facilita a percepção e observação de aspectos e tendências gerais
 - Destaca eventuais particularidades de interesse ao estudo
 - Permite que os dados (ou métodos) desenvolvidos possam ser utilizados em estudos futuros

Arquivos brutos ou de resultados (output de softwares)

- Devem ser disponibilizados sempre que possível
- Garantem a reproduzibilidade do estudo e servem de base para vários estudos futuros
- Exemplos:
 - Reads de sequenciamento de genoma e transcriptoma
 - Montagens de genomas e transcriptoma
 - Anotações de genes, TEs e anotações funcionais



Alguns exemplos

SRA SRA Advanced

Full ▾

SRX9601541: *Phyllosticta citricarpa* CBS 111.20 Standard Draft
1 PACBIO_SMRT (PacBio RS II) run: 6.1M spots, 53.4G bases, 12.6Gb downloads

External Id: JGI-SRA-181144

Submitted by: DOE Joint Genome Institute (JGI)

Study: *Phyllosticta citricarpa* CBS 111.20 Standard Draft genome sequencing
[PRJNA677129](#) • [SRP294733](#) • [All experiments](#) • [All runs](#)
[show Abstract](#)

Sample: *Phyllosticta citricarpa* CBS 111.20 Standard Draft
[SAMN16773214](#) • [SRS7805318](#) • [All experiments](#) • [All runs](#)
Organism: *Phyllosticta citricarpa*

Library:

Name: GUZGO
Instrument: PacBio RS II
Strategy: WGS
Source: GENOMIC
Selection: RANDOM
Layout: SINGLE
Construction protocol: Regular (DNA)

Runs: 1 run, 6.1M spots, 53.4G bases, 12.6Gb

Run	# of Spots	# of Bases	Size	Published
SRR13162647	6,063,335	53.4G	12.6Gb	2020-11-29

ID: 12544978

ASM698474v1

Organism name: [Phyllosticta citricarpa \(ascomycetes\)](#)

Infraspecific name: Strain: LGMF06

BioSample: [SAMN09869335](#)

BioProject: [PRJNA486917](#)

Submitter: Instituto Agronomico

Date: 2019/07/15

Assembly level: Scaffold

Genome representation: full

GenBank assembly accession: GCA_006984745.1 (latest)

RefSeq assembly accession: n/a

RefSeq assembly and GenBank assembly identical: n/a

WGS Project: [QXPX01](#)

Assembly method: CLC Genomics Workbench v. 6.5.1

Expected final version: no

Genome coverage: 850x

Sequencing technology: Illumina HiSeq

IDs: 3822611 [UID] 12078858 [GenBank]

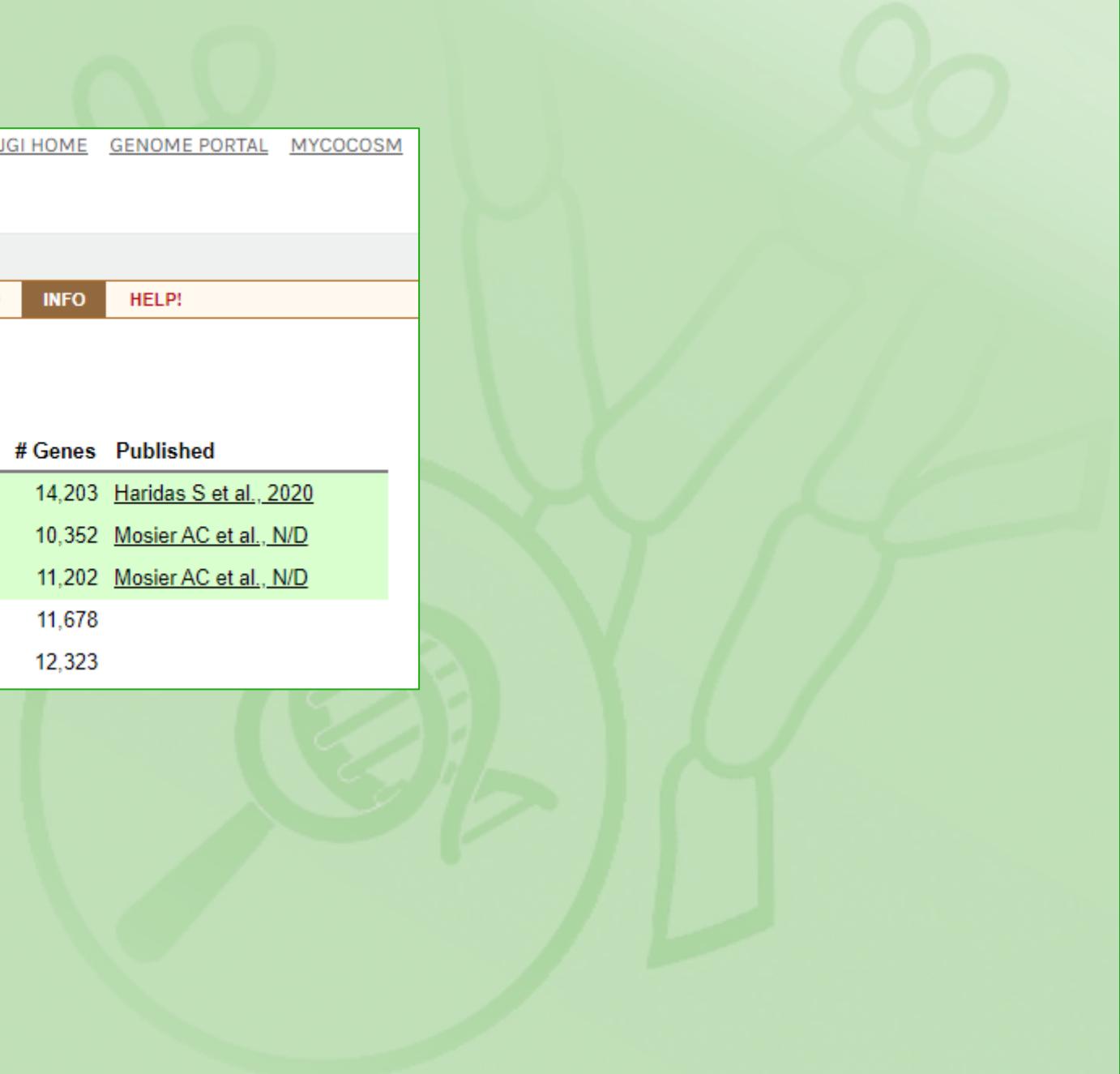
History ([Show revision history](#))

Comment

Global statistics

Total sequence length	29,759,385
Total ungapped length	29,757,884
Gaps between scaffolds	0
Number of scaffolds	17,844
Scaffold N50	3,015
Scaffold L50	3,002
Number of contigs	17,900
Contig N50	3,008
Contig L50	3,015
Total number of chromosomes and plasmids	0
Number of component sequences (WGS or clone)	17,844

Alguns exemplos



JGI MycoCosm THE FUNGAL GENOMICS RESOURCE

JGI HOME GENOME PORTAL MYCOCOSM

TREE SEARCH BLAST ANNOTATIONS ▾ MCL CLUSTERS DOWNLOAD INFO HELP!

Group Name: Dothideomycetes

##	Name	Assembly Length	# Genes	Published
1	Aaospshaeria arxii CBS 175.79 v1.0	38,901,049	14,203	Haridas S et al., 2020
2	Acidomyces richmondensis BFW	26,819,972	10,352	Mosier AC et al., N/D
3	Acidomyces richmondensis BFW	29,883,570	11,202	Mosier AC et al., N/D
4	Aliquandostipite khaoyaiensis CBS 118232 v1.0	119,520,317	11,678	
5	Alternaria alternata ATCC11680	33,569,022	12,323	

Alguns exemplos

JGI MycoCosm
THE FUNGAL GENOMICS RESOURCE

JGI HOME GENOME PORTAL MYCOCOSM

TREE SEARCH BLAST ANNOTATIONS ▾ MCL CLUSTERS DOWNLOAD INFO HELP!

Group Name: Dothideomycetes

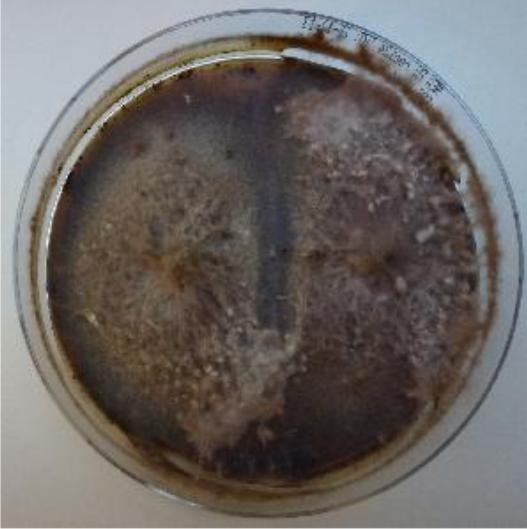
JGI MycoCosm
THE FUNGAL GENOMICS RESOURCE

JGI HOME GENOME PORTAL MYCOCOSM PHYCOCOSM LOGIN

We're soliciting feedback from JGI primary and data users on JGI Data Release and Utilization policies. Fill out our [Request for Information](#) by April 21.

Home • **Aaospshaeria arxii CBS 175.79 v1.0**

SEARCH BLAST BROWSE ANNOTATIONS ▾ MCL CLUSTERS SYNTENY DOWNLOAD INFO **HOME STATUS HELP!**



Aaospshaeria arxii (Aa) Aptroot (Decampiaceae) was originally isolated from corn (*Zea mays*) in Columbia and it has been found on various plants later. The ascocarps are scarce and have wide round ostioles filled with pseudoperiphyses. The outer wall is forming a *textura intricata* and the inner wall is composed of brown, thick-walled and compressed cells. The asci are numerous, bitunicate, and contain eight ascospores. The ascospores are brown, almost fusiform, with a finely pitted outer wall layer. The life style of *Aaospshaeria arxii* is unknown.

Genome Reference(s)

Please cite the following publication(s) if you use the data from this genome in your research:

Haridas S, Albert R, Binder M, Bloem J, LaButti K, Salamov A, Andreopoulos B, Baker SE, Barry K, Bills G, Bluhm BH, Cannon C, Castanera R, Culley DE, Daum C, Ezra D, González JB, Henrissat B, Kuo A, Liang C, Lipzen A, Lutzen F, Magnusson J, Mondo SJ, Nolan M, Ohm RA, Pangilinan J, Park HJ, Ramirez L, Alfaro M, Sun H, Tritt A, Yoshinaga Y, Zwiers LH, Turgeon BG, Goodwin SB, Spatafora JW, Crous PW, Grigoriev IV
[101 Dothideomycetes genomes: A test case for predicting lifestyles and emergence of pathogens](#).
Stud Mycol. 2020 Jun;96():141-153. doi: 10.1016/j.simyco.2020.01.003

Aaospshaeria arxii.
Image Credit: Manfred Binder

Alguns exemplos

JGI  **MycoCosm**
THE FUNGAL GENOMICS RESOURCE

[JGI HOME](#) [GENOME PORTAL](#) [MYCOCOSM](#)

JGI  **MycoCosm**
THE FUNGAL GENOMICS RESOURCE

[JGI HOME](#) [GENOME PORTAL](#) [MYCOCOSM](#) [PHYCOCOSM](#) [LOGIN](#)

We're soliciting feedback from JGI primary and data users on JGI Data Release and Utilization policies. Fill out our [Request for Information](#) by April 21.

[Home • *Colletotrichum nymphaeae* SA-01](#)

[SEARCH](#) [BLAST](#) [BROWSE](#) [ANNOTATIONS ▾](#) [MCL CLUSTERS](#) [SYNTENY](#) [DOWNLOAD](#) [INFO](#) [HOME](#) [HELP!](#)

The genome sequence and gene models of *Colletotrichum nymphaeae* SA-01 were provided by [Michael Thon](#) at the University of Salamanca, Spain. In order to allow comparative analyses with other fungal genomes sequenced by the Joint Genome Institute, a copy of this genome is incorporated into Mycocosm.

The genus *Colletotrichum* (phylum Ascomycota, subphylum Sordariomycetes, order Glomerellales) contains at least 150 species divided into ten major clades. One of the largest of these is the *Colletotrichum acutatum* species complex (CAsc), which includes fungal pathogens that infect a wide diversity of plants in natural and managed ecosystems. The species complex has a very wide host range, and strains have been associated with diseases of more than 90 genera of plants and at least three insect species. The complex taxonomy of CAsc reflects a complexity in evolutionary history, likely brought about by recent host jumps and/or changes in host range followed by adaptation. The CAsc also display great diversity of reproductive behaviors although most species seem to have lost their mating capability. Thus, the CAsc are excellent candidates for studying the process of speciation, host adaptation and the evolution of mating behavior.

Members of CAsc also show significant expansions in gene families associated carbohydrate metabolism, particularly in families of xyloglucanases and other plant cell wall degrading enzymes and possibly contain the largest diversity of carbohydrate active enzymes in the Ascomycetes. How this repertoire of enzymes has evolved, and why CAsc species maintain such diversity is unknown. Comparative analysis of these species will give us insight into the mode and tempo of gene duplications, selective pressures and other evolutionary processes that lead expansion of carbohydrate active enzymes.

Genome Reference(s)

Please cite the following publication(s) if you use the data from this genome in your research:

Baroncelli R, Amby DB, Zapparata A, Sarrocco S, Vannacci G, Le Floch G, Harrison RJ, Holub E, Sukno SA, Sreenivasaprasad S, Thon MR
[Gene family expansions and contractions are associated with host range in plant pathogens of the genus *Colletotrichum*.](#)
BMC Genomics. 2016 Aug 5;17():555. doi: 10.1186/s12864-016-2917-6

Alguns exemplos

May 12, 2020

Dataset Open Access

Assemblies and annotations from the paper "Genome compartmentalization predates species divergence in the plant pathogen genus *Zymoseptoria*"

Feurtey, Alice; Lorrain, Cécile; Croll, Daniel; Eschenbrenner, Christoph; Freitag, Michael; Habig, Michael; Haueisen, Janine; Möller, Mareike; Schotanus, Klaas; Stukenbrock, Eva

These files are the assemblies and annotations produced and analyzed in the revised version of the manuscript entitled "Genome compartmentalization predates species divergence in the plant pathogen genus *Zymoseptoria*".

Files (356.1 MB)		
Name	Size	
Annotations_2018_genomes_for_publication.tab	16.7 MB	Download
md5:04b083a07d68729d551d5555c8e7e095 ?		
Annotations_emapper_2018_genomes_for_publication.tab	39.1 MB	Download
md5:c19bd5609fded06ca11cb22f6672d27c ?		
Expression_data_Ztritici_inplanta.xlsx	3.9 MB	Download
md5:5db507efc693b329d23bec41b0e9c221 ?		

Alguns exemplos

May 12, 2020

Assemblies and annotations for "Genome compartmentalization predates species divergence in the plant pathogen *Zymoseptoria*"

Feurtey, Alice; Lorrain, Cécile; Croll, Daniel; Eschenbrenner, Christoph; Freitag, Möller, Mareike; Schotanus, Klaas; Stukenbrock, Eva

These files are the assemblies and annotations produced and analyzed in the "Genome compartmentalization predates species divergence in the plant pathogen *Zymoseptoria*".

Files (356.1 MB)

Name

Annotations_2018_genomes_for_publication.tab

md5:04b083a07d68729d551d5555c8e7e095 ?

Annotations_emapper_2018_genomes_for_publication.tab

md5:c19bd5609fded06ca11cb22f6672d27c ?

Expression_data_Ztritici_in planta.xlsx

md5:5db507efc693b329d23bec41b0e9c221 ?

Zpa63_softmasked_for_publication.fa	41.8 MB	 Download
md5:72c76aa814e2bf40d5e610aa9eaff398 ?		
Zt05_2019_for_publication.gff3	9.4 MB	 Download
md5:5651e9c675691a5b2bd5bfd8c36651bf ?		
Zt05_softmasked_for_publication.fa	41.7 MB	 Download
md5:32ab9ba9e846e3abf54dd5df1d9a8c34 ?		
Zt10_2019_for_publication.gff3	9.1 MB	 Download
md5:e4f98fb346afea5f18ff243fc6ff3db ?		
Zt10_softmasked_for_publication.fa	39.6 MB	 Download
md5:20001d4c65217071604010475-605-a1		

Availability of data and materials

The datasets generated and analyzed during the current study are available in the NCBI Short Read Archive (<https://www.ncbi.nlm.nih.gov/bioproject/?term=>) under the BioProject accession numbers: PRJNA638605, PRJNA639021 (Zpa63), PRJNA638553 (Zb87), PRJNA638515 (Zp13), PRJNA638382 (Za17), PRJNA414407 (Zt05 and Zt10). All the data supporting the findings of this study are openly available at <https://doi.org/10.5281/zenodo.3820378>. The assembled genomes can be found at <https://doi.org/10.5281/zenodo.3820378>. The gene annotations are deposited at <https://doi.org/10.5281/zenodo.3820378>. The functional annotation pipeline, additional scripts and command lines used to create the results presented in this manuscript can be found at https://gitlab.gwdg.de/alice.feurtey/genome_architecture_zymoseptoria. In planta

Alguns exemplos

August 18, 2020

Dataset Open Access

Mating-type locus rearrangements and shifts in thallism states in Citrus-associated *Phyllosticta* species

by Desirrê A. L. Petters-Vandresen

Files related to the manuscript published in *Fungal Genetics and Biology* (<https://doi.org/10.1016/j.fgb.2020.103444>), with preprint available in *bioRxiv* (<https://doi.org/10.1101/2020.04.14.2040725>).

Contains:

- 03 genome assemblies: *Phyllosticta capitalensis* LGMF01, *Phyllosticta citribraziliensis* LGMF08 and *Phyllosticta citricarpa* LGMF06;
- 14 protein-coding gene annotation based in 14 *Phyllosticta* genomes;
- Mating-type locus annotation from the 14 *Phyllosticta* strains in GBK format, and related FASTA sequences;

[Preview](#)



Alguns exemplos

August 18, 2020

Dataset Open Access

Mating-type locus rearrangements and shifts in thallism states in Citrus-associated *Phyllosticta* species

by Desirrê A. L. Petters-Vandresen

Files related to the manuscript published in *Fungal Genetics and Biology* with preprint available in bioRxiv (<https://doi.org/10.1101/2020.04.14.204101>)

Contains:

- 03 genome assemblies: *Phyllosticta capitalensis* LGMF01, *Phyllosticta* LGMF06;
- 14 protein-coding gene annotation based in 14 *Phyllosticta* genomes;
- Mating-type locus annotation from the 14 *Phyllosticta* strains in GBK fo

FastQC v 0.11.8 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) was employed to check the quality of reads after processing in Trimmomatic. *De novo* genome assemblies were generated with SPAdes v 3.13 (Bankevich et al., 2012), using default parameters. Contigs smaller than 500 bp were filtered and removed from the final assemblies, which were then evaluated with QUAST v 4.6.3 (Gurevich et al., 2013). Library and assembly statistics for the new assemblies are summarized in Table S1 and assemblies are available at Zenodo (<https://doi.org/10.5281/zenodo.3750350>).

with BLASTp. After confirmation, the gene annotations were further refined by comparison with the *MAT* loci from *Phyllosticta citricarpa* and some *Botryosphaeriaceae* species (Nagel et al., 2018). The sequences and annotations are deposited in Zenodo (<https://doi.org/10.5281/zenodo.3750350>). Functional domains in the annotated genes were determined using the Conserved Domain Search Service (CD-Search) from NCBI (Lu et al., 2020), using an e-value of 0.01 as threshold.

[Preview](#)

Código de programação

- Devem ser disponibilizados sempre que possível, devidamente comentados e de forma compreensível
- Garantem a reproduzibilidade do estudo e servem de base para vários estudos futuros



TEN YEARS REPRODUCIBILITY CHALLENGE

RESCIENCE SPECIAL ISSUE
FREE TO READ - FREE TO PUBLISH

Workshop
June 22, 2020
BORDEAUX



Would you dare to run the
code from your past self ?
(the one that does not answer mail)

SUBMISSION DEADLINE 01/04/2020

<http://rescience.github.io/ten-years>

In association with Inria, CNRS, Software Heritage, ReScience, Comité pour la Science Ouverte, URFIST Bordeaux & Mission de la pédagogie et du numérique pour l'enseignement supérieur.
Contact: nicolas.rougier@inria.fr

Alguns exemplos

PlantDr430 / CSU_scripts

Code Issues Pull requests Actions Projects Wiki Security Insights

master 1 branch 0 tags Go to file Add file Code

PlantDr430 Missed a line to comment out 8ed9e1d on 7 Jun 2020 88 commits

.gitignore Initial commit 2 years ago

Domain_enrichment.py Utilized two multi-correction tests 12 months ago

Fungal_recombination.py Missed a line to comment out 11 months ago

LICENSE Initial commit 2 years ago

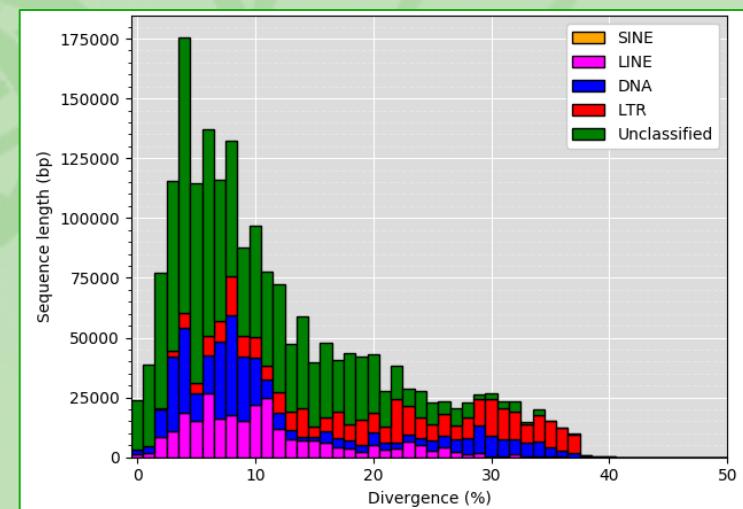
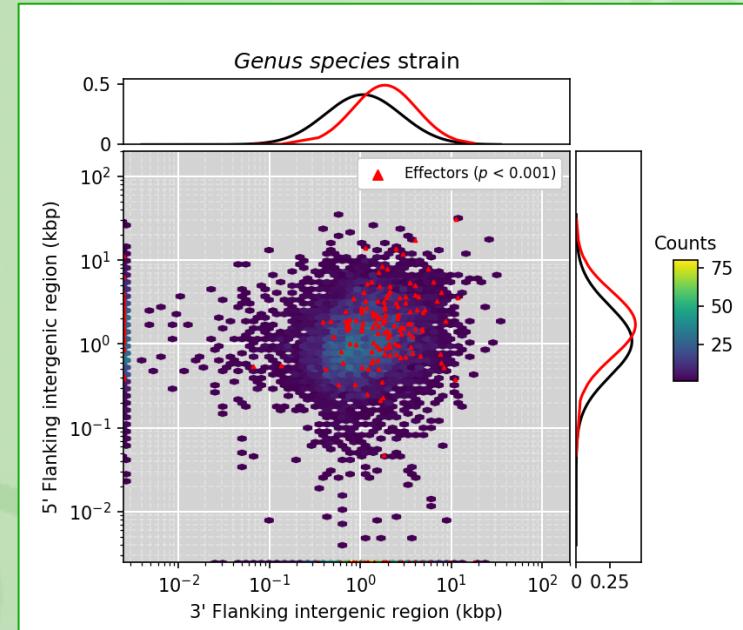
README.md Updated genome_speed_hexbin image example 2 years ago

RIP_blast_analysis.py Added gene counts function 15 months ago

Gene Density Compartmentalization

A custom script

(https://github.com/PlantDr430/CSU_scripts/blob/master/genome_speed_hexbins.py) was used to calculate local gene density measured as 5' and 3' flanking distances between neighboring genes (intergenic regions). To statistically



Alguns exemplos

G Genome_architecture_Zymoseptoria  Star 0

Project ID: 8197

26 Commits 1 Branch 0 Tags 1.9 MB Files 1.9 MB Storage

master genome_architecture_zymoseptoria History Find file Clone

 Second trial at clean up
alice.feurtey authored 8 months ago

 README  No license. All rights reserved

Name	Last commit
0_Assemblies	Second trial at clean up
1_Gene_prediction	Add schematic pipeline
2_Functional_predictions	Delete Parsing_functions_V1_0.py

<https://doi.org/10.5281/zenodo.3820378>. The functional annotation pipeline, additional scripts and command lines used to create the results presented in this manuscript can be found at https://gitlab.gwdg.de/alice.feurtey/genome_architecture_zymoseptoria. In planta



The figure is a treemap visualization showing the distribution of genes across various cell compartments for nine different samples. The compartments are color-coded according to a legend on the right. The samples are represented by columns at the bottom, labeled CB0940, Za17, Zb87, Zp13, Zpa63, Zt05, Zt09, and Zt10.

Cell_compartment	CB0940	Za17	Zb87	Zp13	Zpa63	Zt05	Zt09	Zt10
NA	2353	2207	2243	2210	2083	2443	2420	2357
Plastid	199	227	265	239	234	255	259	257
Peroxisome	479	353	348	389	332	383	352	377
Nucleus	2349	2274	2449	2399	2220	2624	2530	2450
Mitochondrion	891	582	652	622	605	675	752	669
Lysosome/Vacuole	139	114	100	127	94	109	120	114
Golgi_apparatus	50	39	45	45	44	42	49	39
Extracellular	1074	1328	876	1048	827	1143	941	1146
Endoplasmic_reticulum	570	516	523	515	475	521	500	515
Cytoplasm	3563	3219	3373	3435	3051	3543	3262	3425
Cell_membrane	768	604	606	632	563	648	654	642

Tabelas

- Armazemam grandes quantidades de informação com alto nível de detalhamento
- Úteis para leitores mais especializados e que precisem consultar detalhes específicos do estudo
- Auxiliam na seleção de genes candidatos para estudos funcionais
- Tabelas maiores geralmente incluídas como material suplementar

Alguns exemplos

Adaptado de:

PETTERS-VANDRESEN et al. 2020. *Fungal Genetics and Biology*. DOI: [10.1016/j.fgb.2020.103444](https://doi.org/10.1016/j.fgb.2020.103444)

Table 1
Genome and strain information of *Phyllosticta* spp. evaluated in this study.

Species	Strain ^a	Host	Origin	Thallism	Mating-type	Mating-locus location	Genome size (bp)	Contigs	N50	L50	GC content	Repeat content	Predicted genes	BUSCO ^b	Assembly
<i>P. capitalensis</i>	CBS 128856	<i>Stanhopea</i> sp.	Brazil	Homothallic	Both	Scaffold 10	32,461,131	14	2,860,346	5	54.58%	1.54%	9977	97.9%	(Guarnaccia et al., 2019)
	Gm33	<i>Citrus sinensis</i>	USA	Homothallic	Both	Contigs 513 and 1011	32,454,403	1341	51,729	184	54.56%	0.49%	10,183	94.4%	(Wang et al., 2016)
	LGMF 01	<i>Citrus latifolia</i> (leaf)	Brazil	Homothallic	Both	Contig 24	32,606,250	231	1,366,738	9	54.48%	0.15%	9953	98%	This study
<i>P. citriasihana</i>	CBS 120486	<i>Citrus maxima</i> (fruit)	Thailand	Heterothallic	MAT1-1	Scaffold 34	32,696,106	133	807,147	14	51.56%	8.32%	9282	98.1%	(Guarnaccia et al., 2019)
	CGMCC 3.14344	<i>Citrus</i> sp.	China	Heterothallic	MAT1-2	Unitig 22	34,225,214	92	968,885	10	51.42%	14.50%	9291	97.9%	(Wang et al., 2020)
<i>P. citribraziliensis</i>	CBS 100098	<i>Citrus</i> sp. (leaf)	Brazil	Heterothallic	MAT1-1	Scaffold 7	31,670,975	32	1,720,616	7	54.17%	6%	9574	98%	(Guarnaccia et al., 2019)
<i>P. citricarpa</i>	LGMF 08	<i>Citrus</i> sp. (leaf)	Brazil	Heterothallic	MAT1-1	Contig 61	31,002,620	563	263,424	40	54.34%	4.83%	9941	97.9%	This study
	CBS 127454	<i>Citrus limon</i>	Australia	Heterothallic	MAT1-2	Scaffold 46	28,952,665	152	440,231	23	54.60%	1.30%	9108	96.1%	(Guarnaccia et al., 2019)
	CPC 27913	<i>Citrus sinensis</i> (leaf litter)	Malta	Heterothallic	MAT1-2	Scaffold 20	32,267,666	82	900,945	13	52.56%	8.07%	9352	97.6%	(Guarnaccia et al., 2019)
	Gc 12	<i>Citrus sinensis</i>	USA	Heterothallic	MAT1-2	Contigs 1345, 2477, 3787	31,127,197	5748	21,637	411	53.05%	5.93%	9472	94.5%	(Wang et al., 2016)
<i>P. citrichinaensis</i>	CGMCC 3.14348	N/A	China	Heterothallic	MAT1-1	Contigs 4070 and 4071	32,007,210	6716	11,522	811	52.57%	7.94%	9768	87.5%	GenBank Assembly: GCA_000382785.1
	LGMF 06	<i>Citrus sinensis</i>	Brazil	Heterothallic	MAT1-2	Contig 51	32,021,677	354	421,474	28	52.68%	0.64%	11,217	97.8%	This study
	CBS 130529	<i>Citrus maxima</i> (leaf)	China	Homothallic	Both	Scaffolds 6 and 9	29,162,704	25	2,710,567	4	55.07%	2.99%	9131	97.5%	(Guarnaccia et al., 2019)
<i>P. paracitricarpa</i>	CBS 141357	<i>Citrus limon</i> (leaf litter)	Greece	Heterothallic	MAT1-2	Scaffold 48	29,529,839	134	510,184	23	54.35%	2.45%	9083	96.2%	(Guarnaccia et al., 2019)

N/A: not available or not applicable to the present study.

^a Type strains included in the analysis are indicated in bold. Culture collection abbreviations: CBS = Westerdijk Fungal Biodiversity Institute, Utrecht, Netherlands; CPC = Collection of Pedro W. Crous, held at the Westerdijk Fungal Biodiversity Institute; CGMCC = China General Microbiological Culture Collection Center, Chinese Academy of Sciences, Beijing, China; LGMF = Laboratório de Bioprospecção e Genética Molecular de Microrganismos, Universidade Federal do Paraná, Paraná, Brazil.

^b BUSCO completeness assessed with the Pezizomycotina dataset.

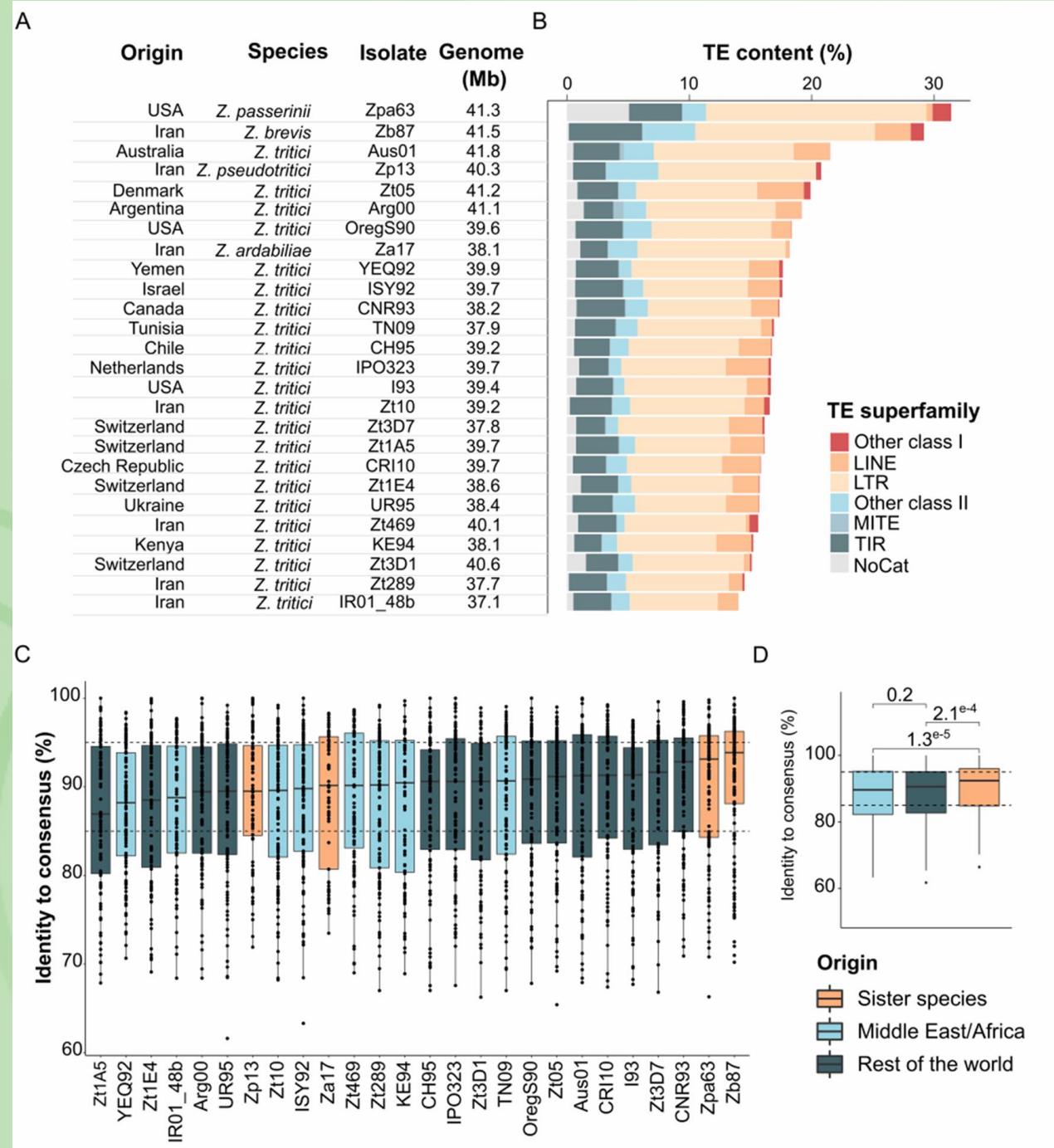
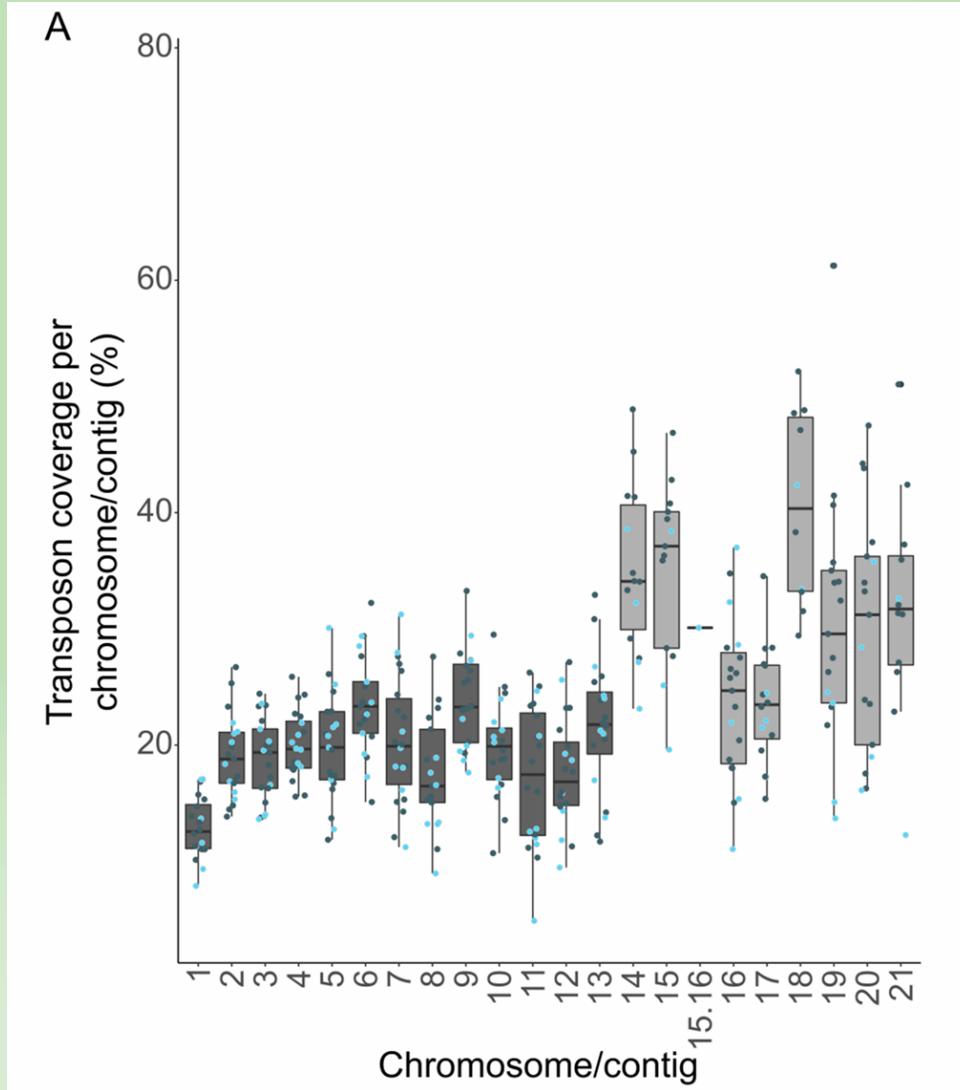
Alguns exemplos

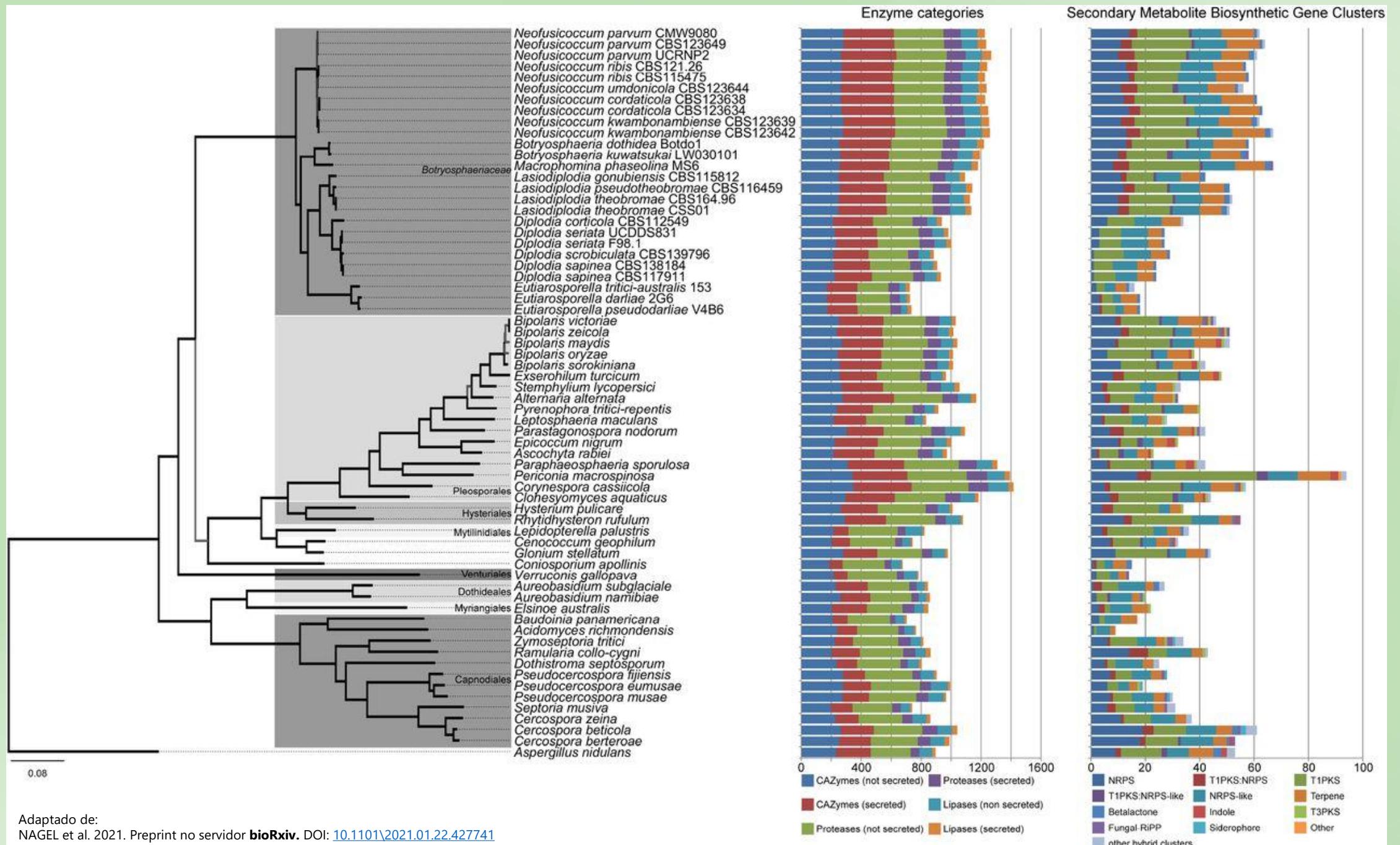
Species	<i>Zymoseptoria tritici</i>	<i>Zymoseptoria pseudotritici</i>	<i>Zymoseptoria brevis</i>	<i>Zymoseptoria ardabiliæ</i>	<i>Zymoseptoria passerinii</i>
Isolate	Zt05	Zt10	Zp13	Zb87	Za17
Origin	Denmark	Iran, Ilam province	Iran, Ardabil province	Iran	Iran, Ardabil province
Host	<i>Triticum aestivum</i>	<i>Triticum aestivum</i>	<i>Dactylis glomerata</i>	<i>Phalaris paradoxa</i>	<i>Lolium perenne</i>
Year (of isolation)	2004	2001	2004		2004
Contig number	30	19	42	29	50
Total length (bp)	41,240,984	39,248,105	40,312,446	41,586,671	38,100,668
Mean contig size (bp)	1,374,699	2,065,690	59,820	1,434,023	762,013
N50	2,454,671	2,925,395	2,115,121	2,744,794	1,156,695
L50	6	5	7	7	11
Contigs with telomeric repeats on both ends	12	19	5	9	0
Complete BUSCO genes (%)	98.4	98.7	98.5	97.0	98.2
Number of genes	12,386	11,991	11,661	11,480	11,463
Repeat content (%)	19.9	16.5	20.8	29.2	18.2
Percentage of non-core sequences compared to the IPO323 reference	12.49	8.69	4.27	4.6	3.45
NCBI Biosample	SAMN04494882	SAMN02981321	SAMN02981322	SAMN03294124	SAMN02981326
					SAMN02981330

Figuras

- Úteis para todos os tipos de leitores do trabalho: entendimento de tendências gerais e dos principais resultados de um estudo
- Presentes no texto principal e material suplementar
- **Cenário ideal:** figuras que reúnam e apresentem os diversos resultados e aspectos importantes de um estudo

Gráfico de barras e box-plot





Adaptado de:
NAGEL et al. 2021. Preprint no servidor **bioRxiv**. DOI: [10.1101/2021.01.22.427741](https://doi.org/10.1101/2021.01.22.427741)

Diagrama de Venn

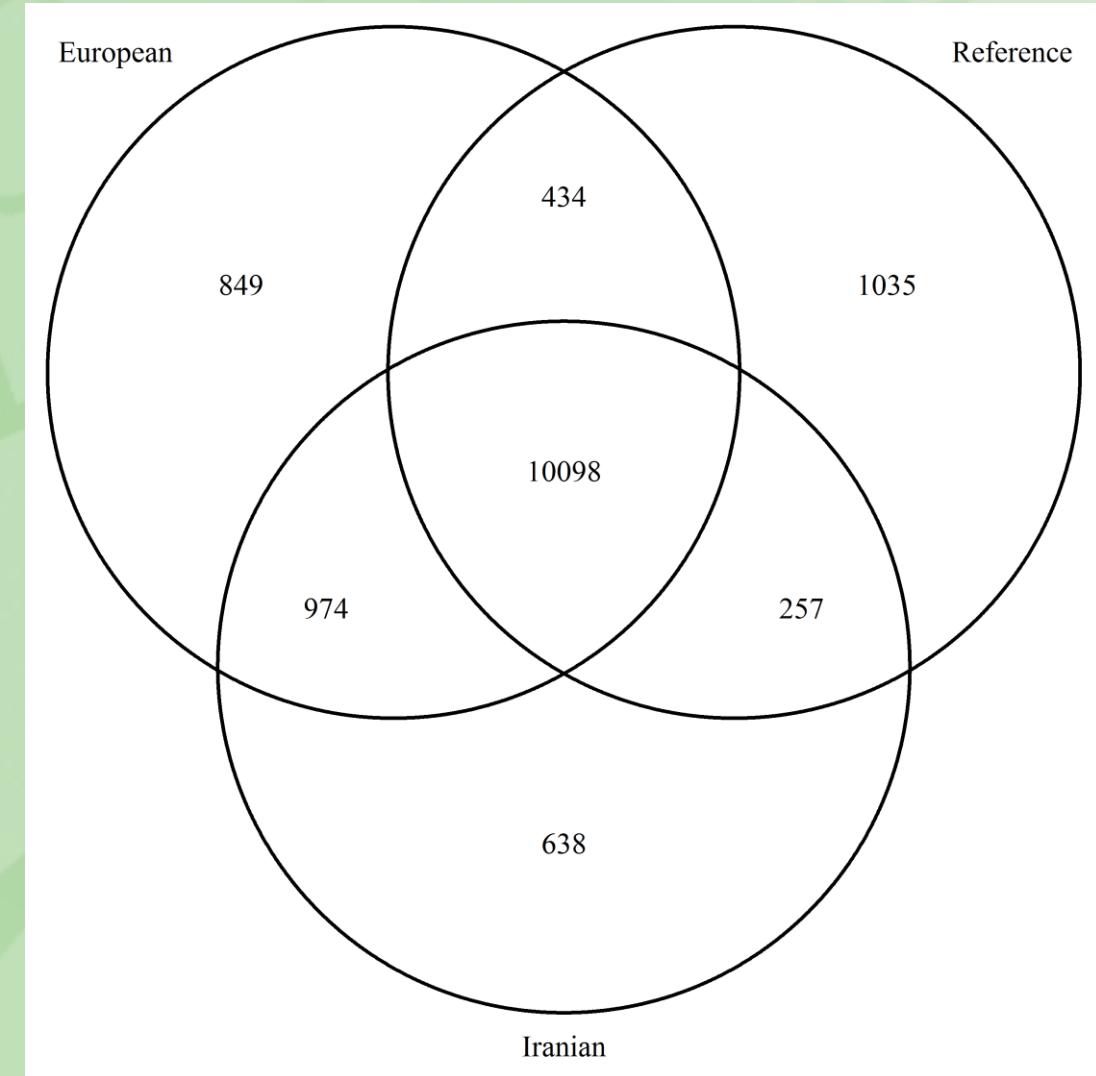
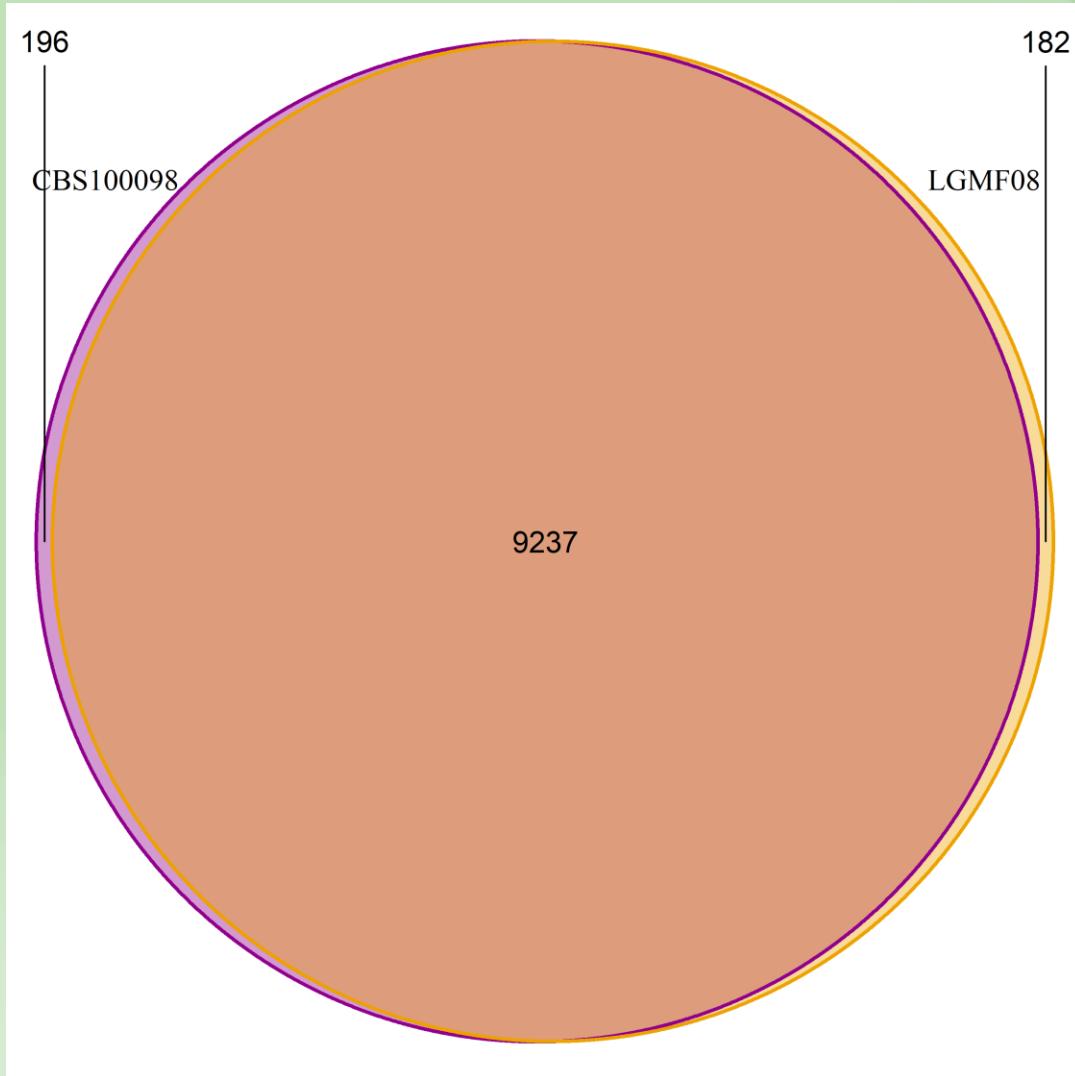


Diagrama de Venn

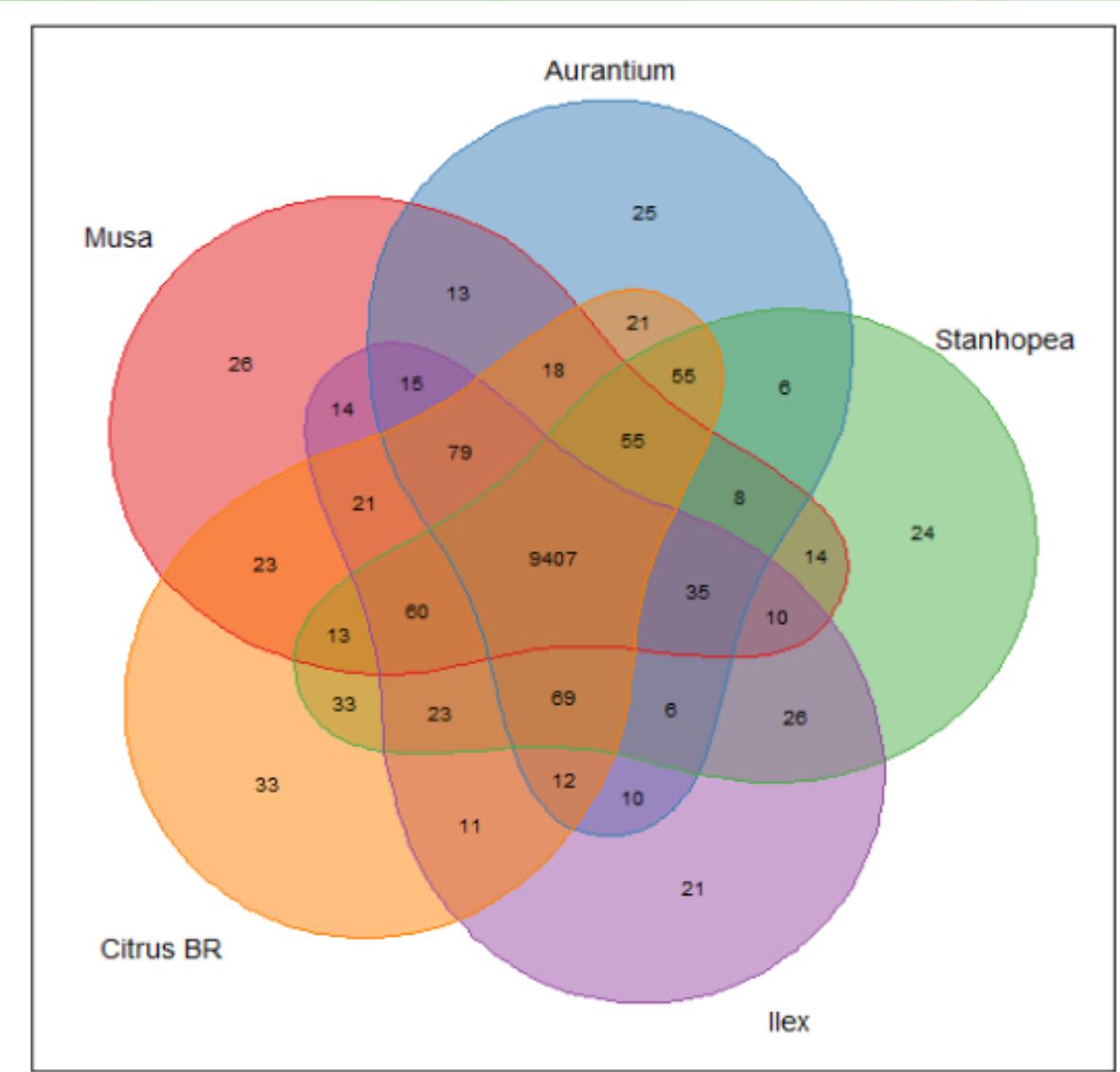
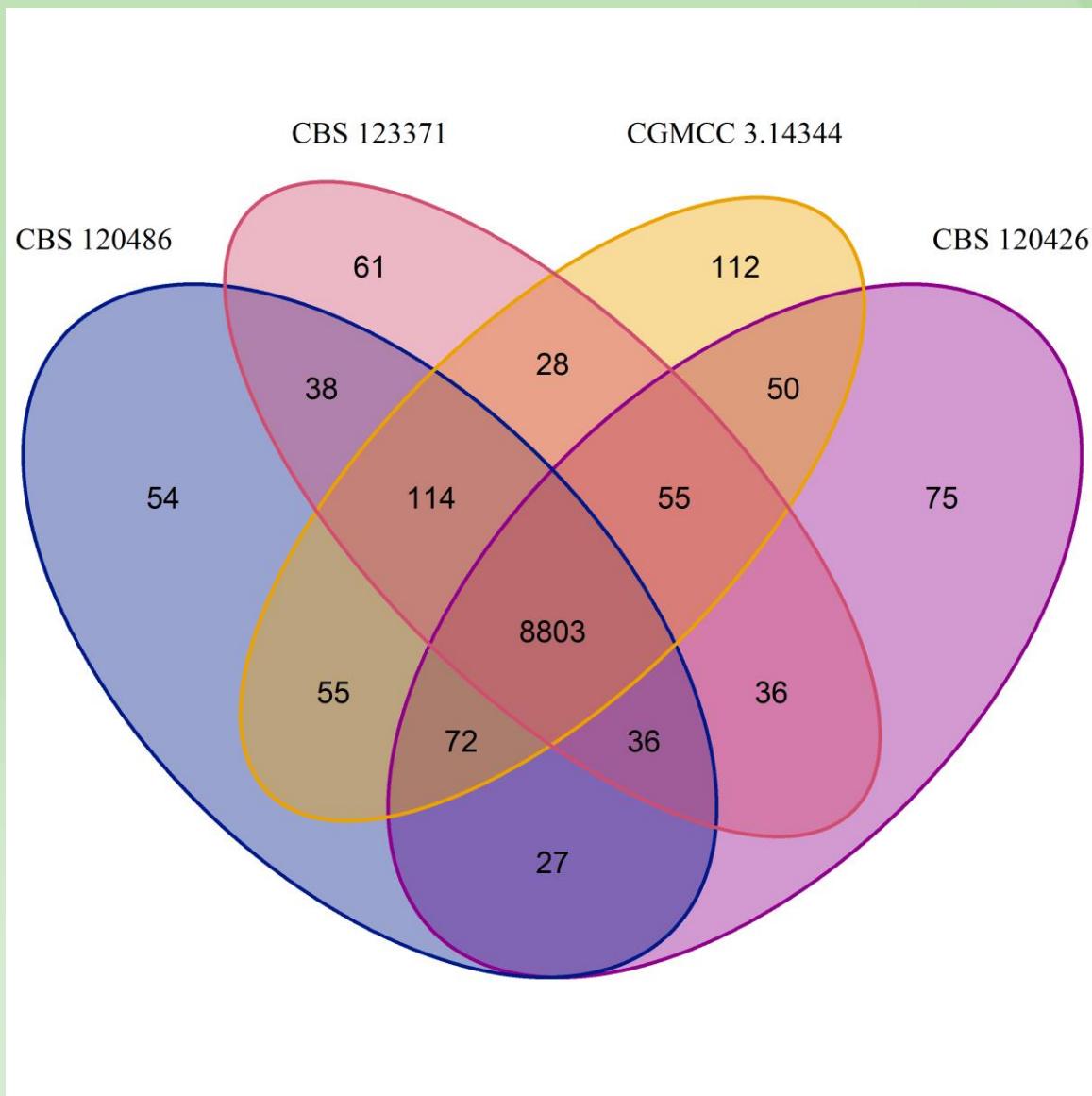
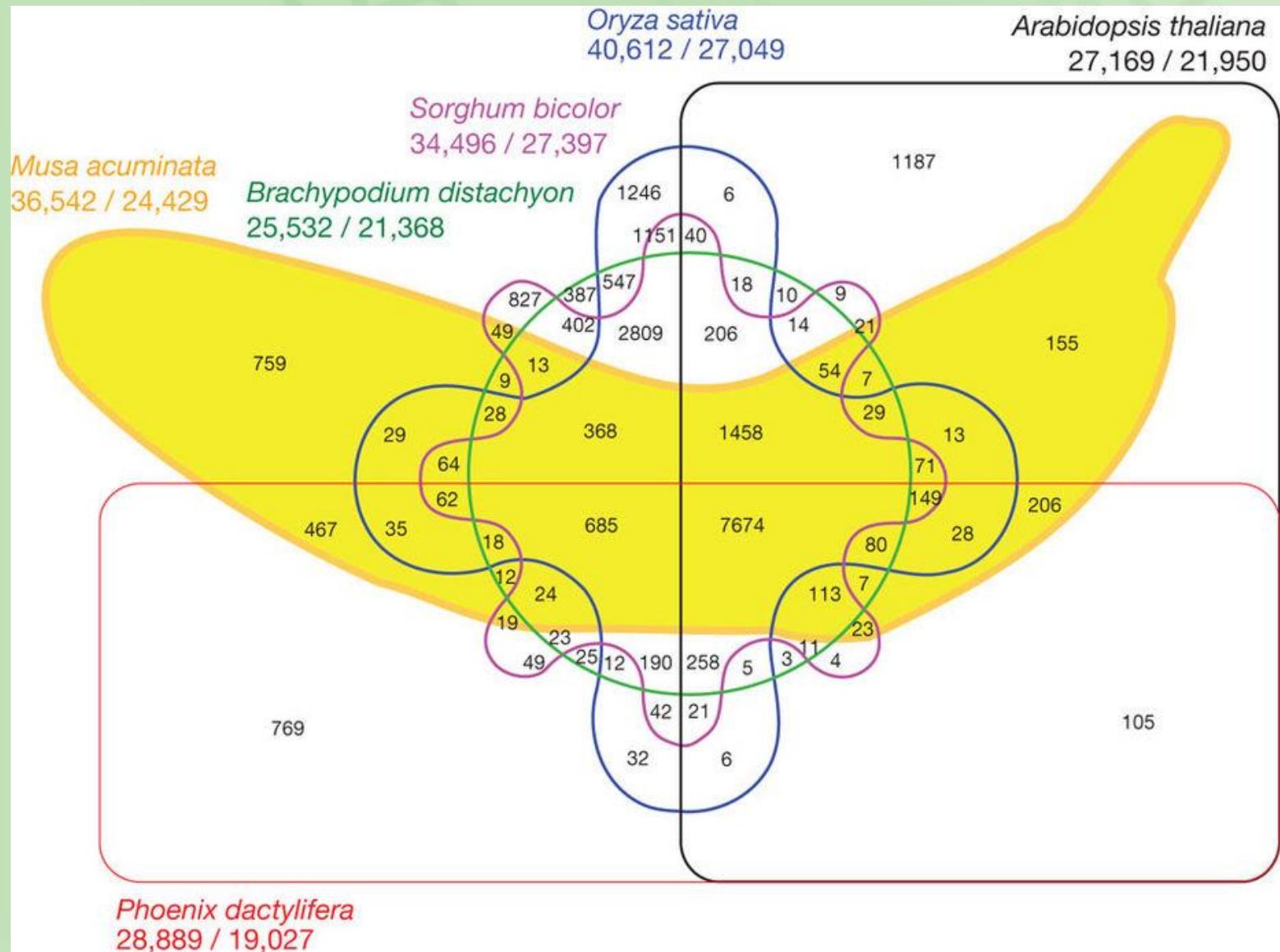
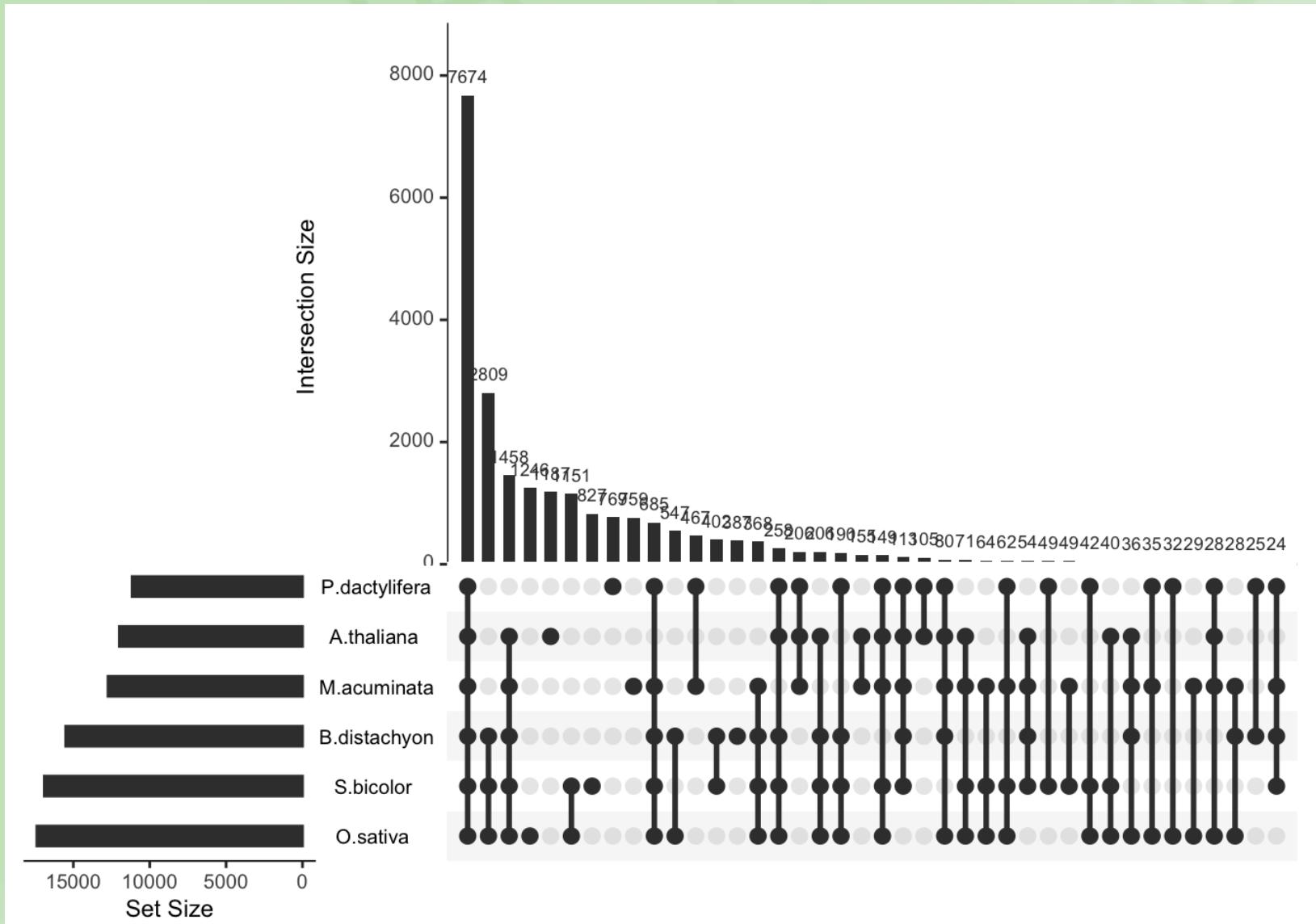


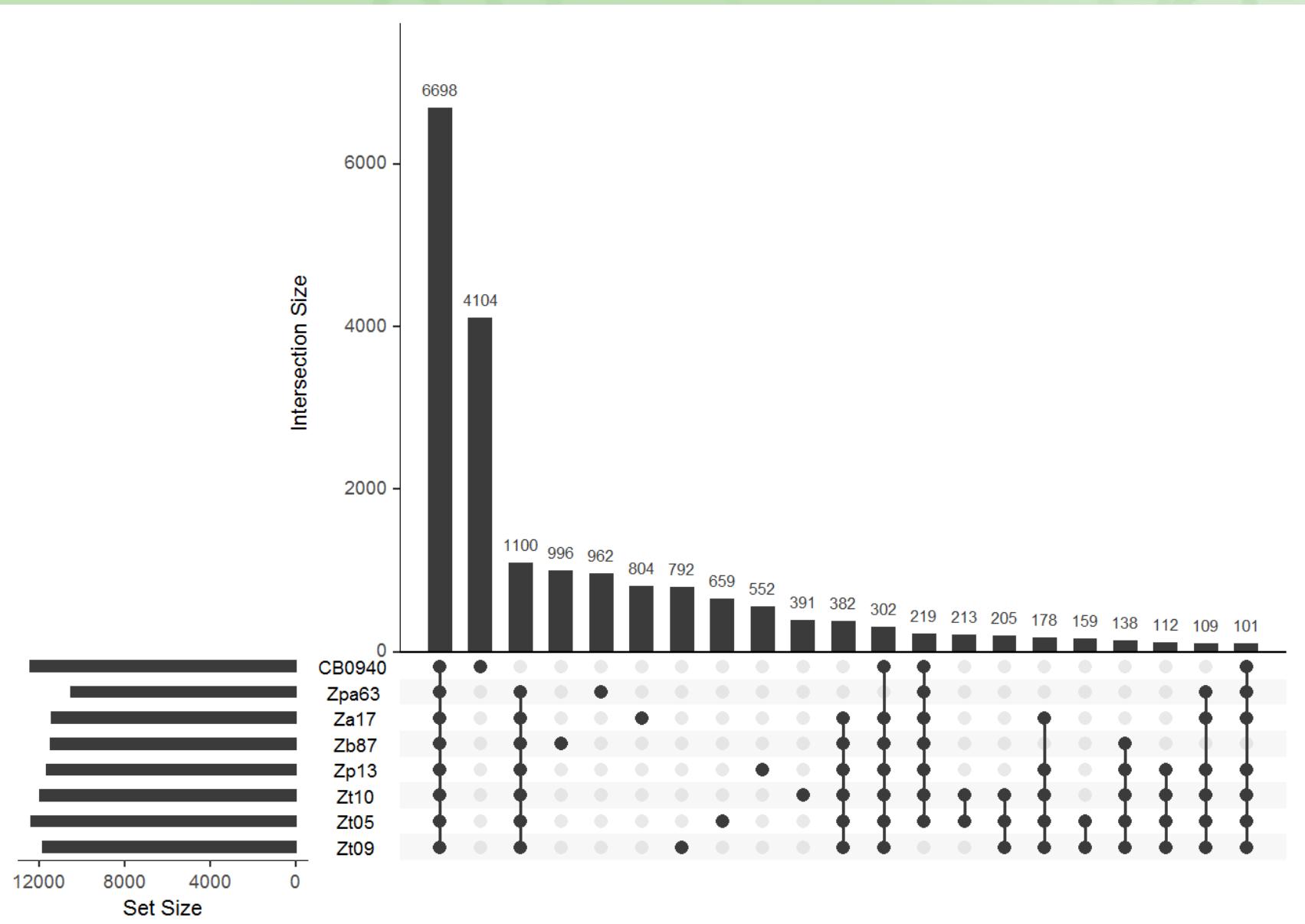
Diagrama de Venn



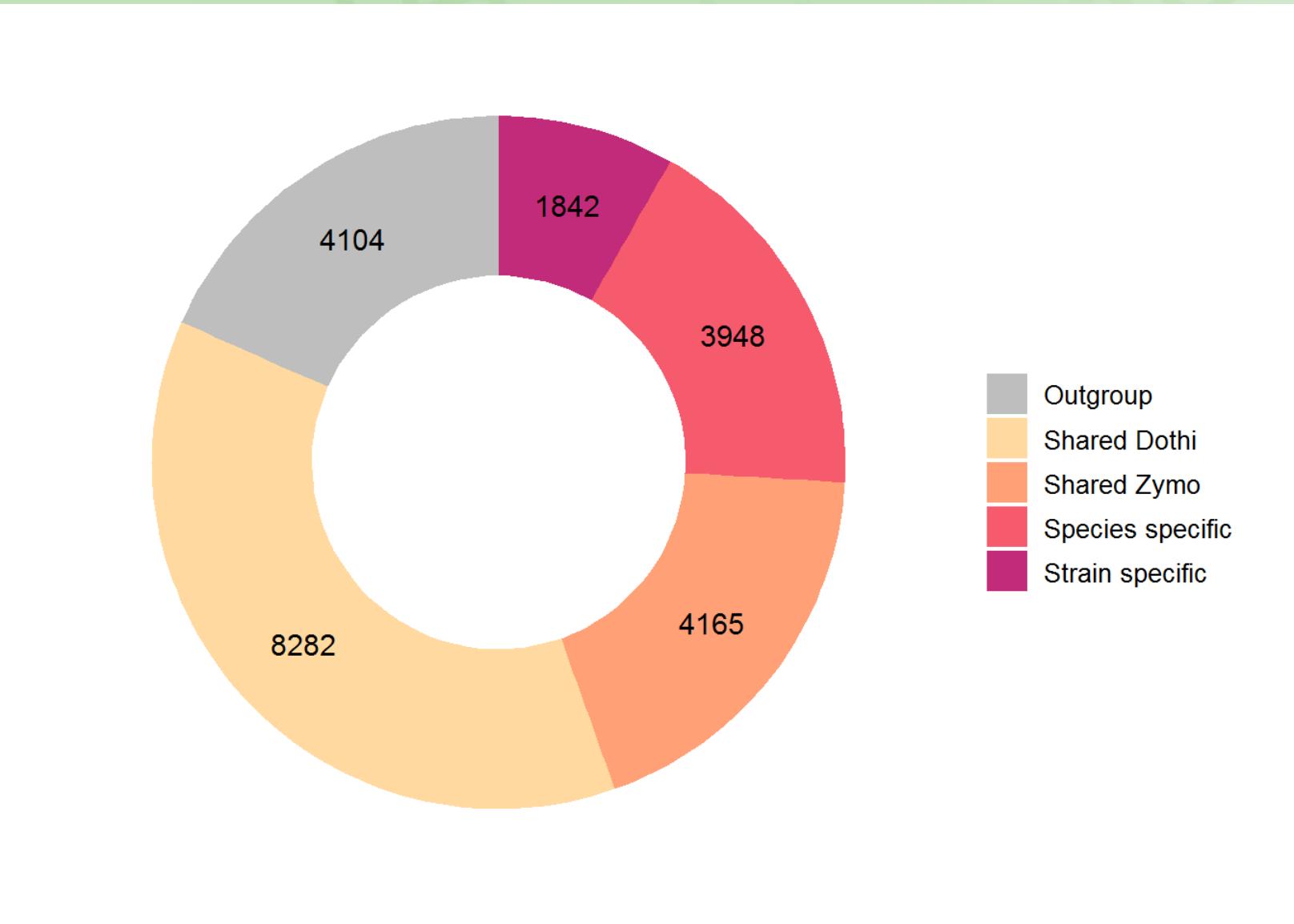
Upset plot



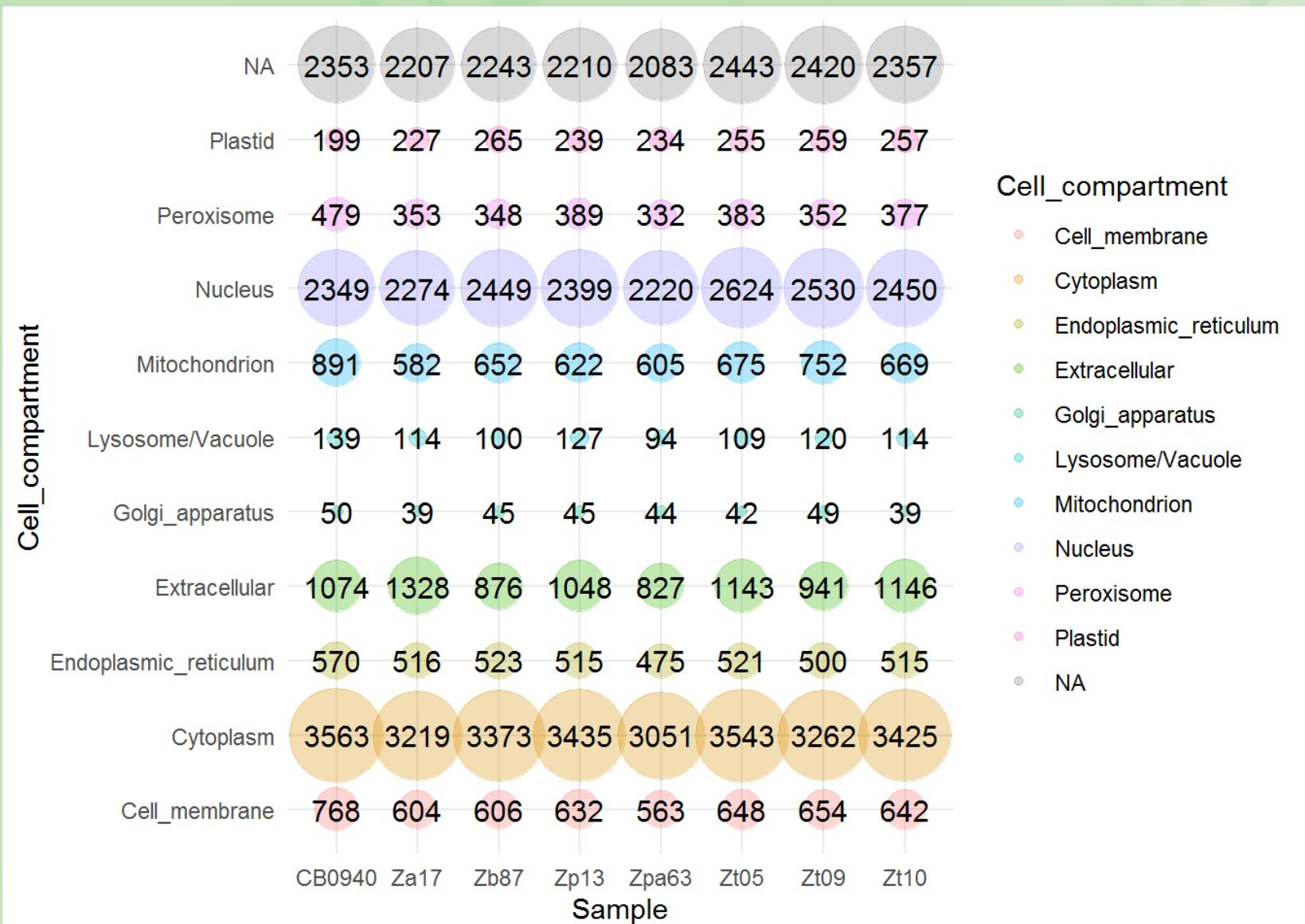
UpSet plot



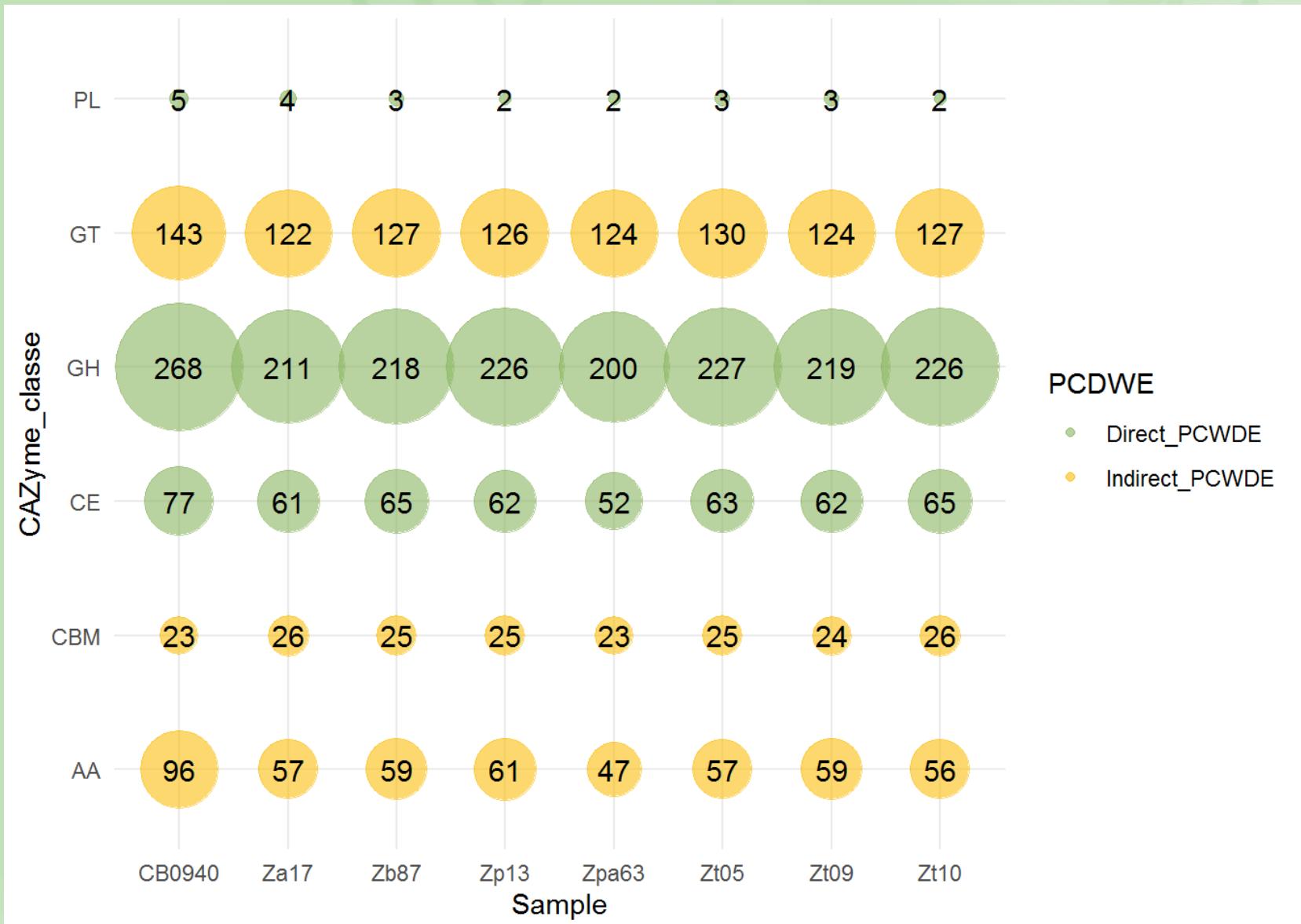
Donut plot



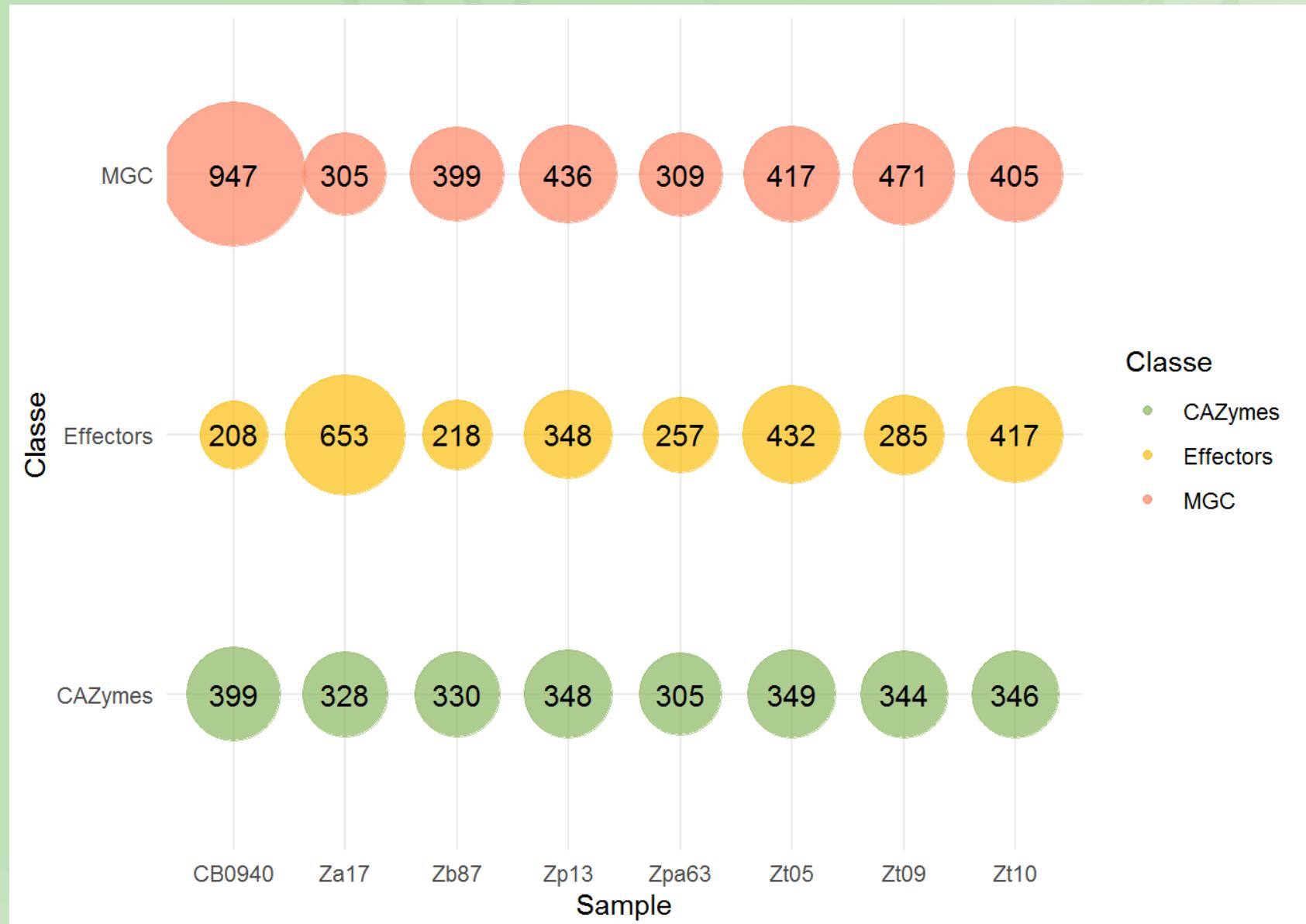
Bubble plot



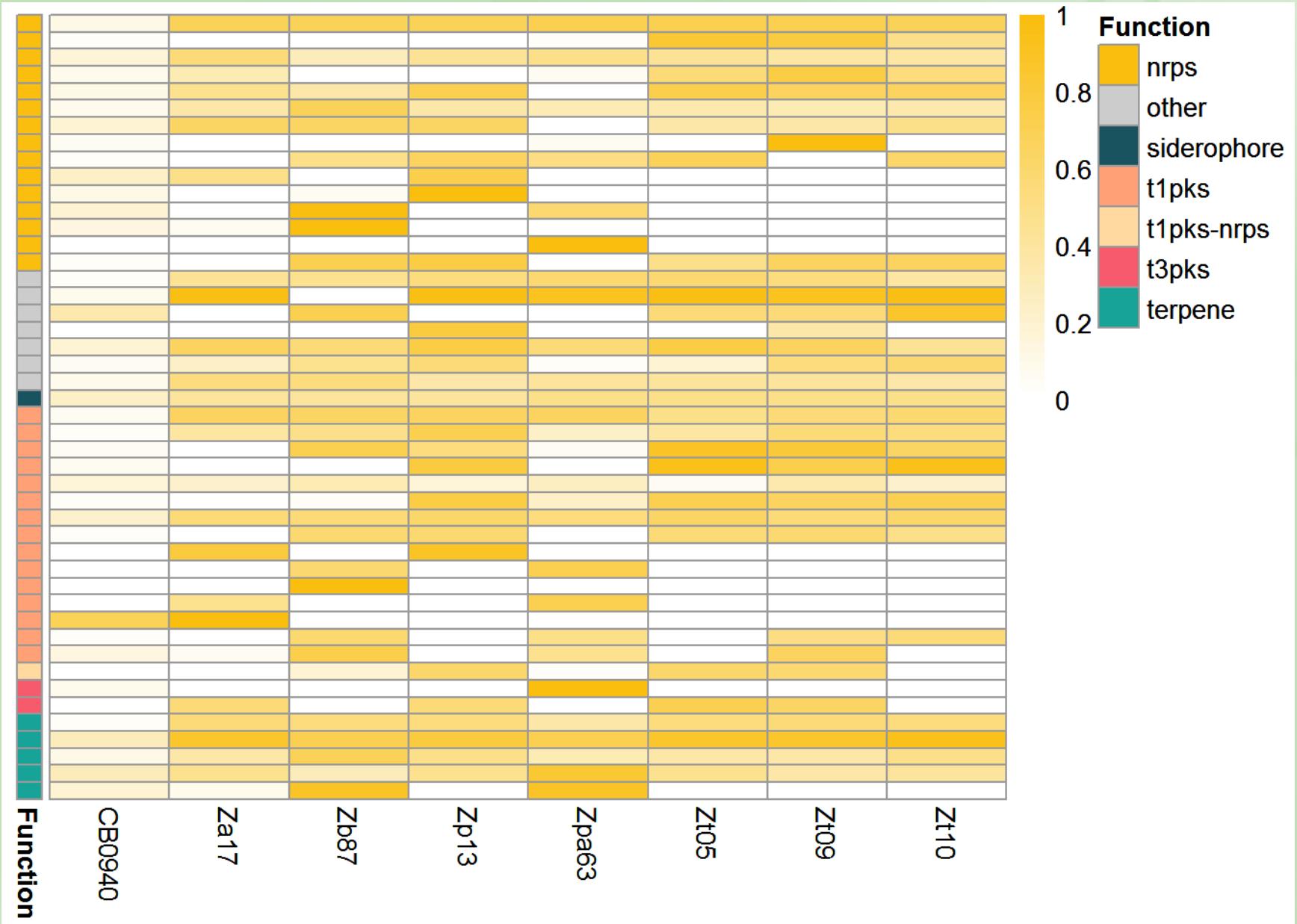
Bubble plot

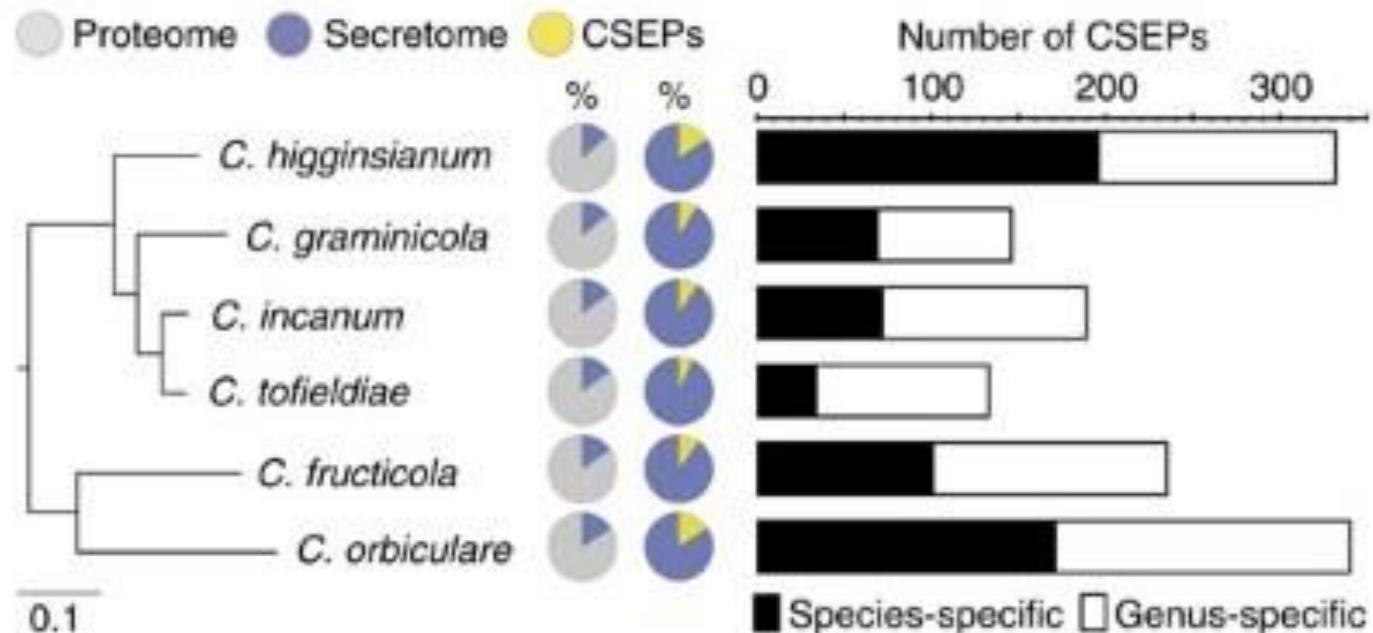
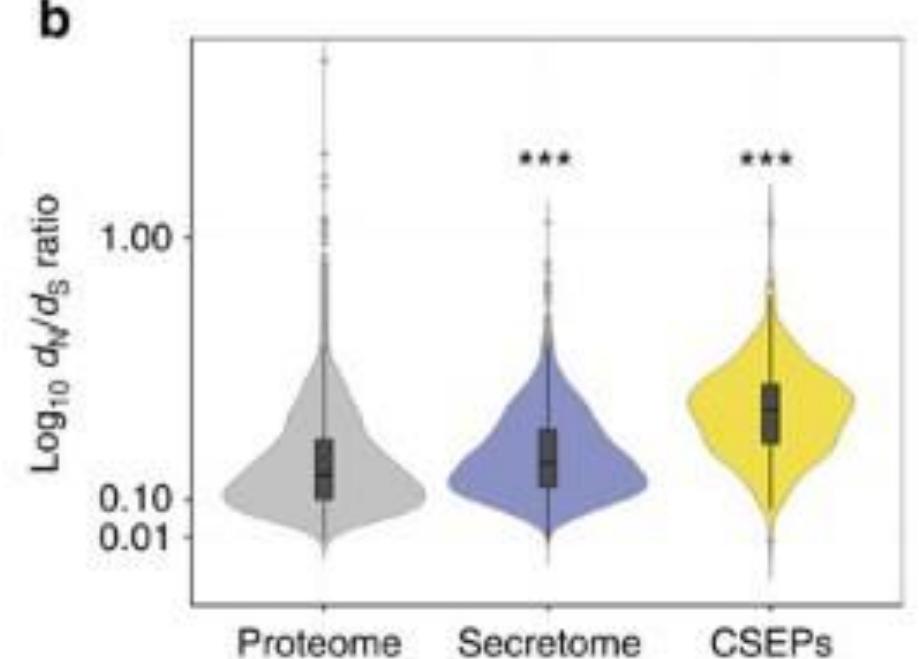


Bubble plot



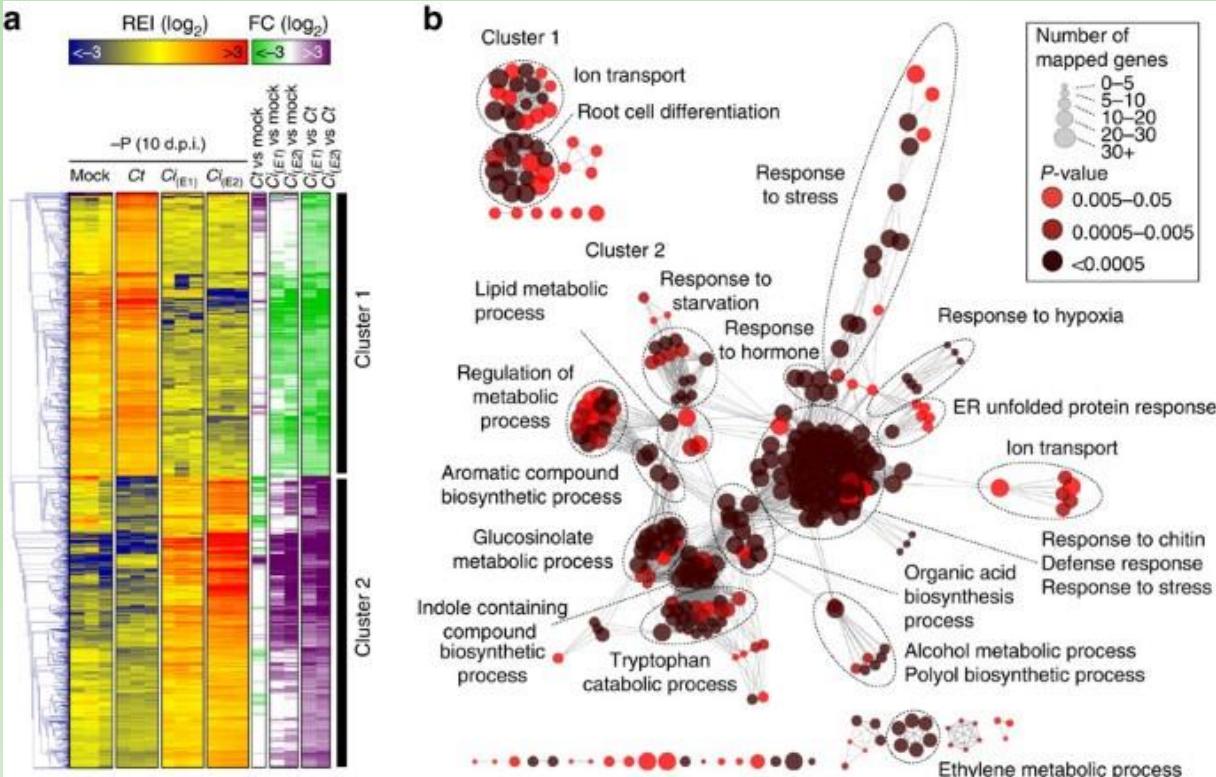
Heatmap



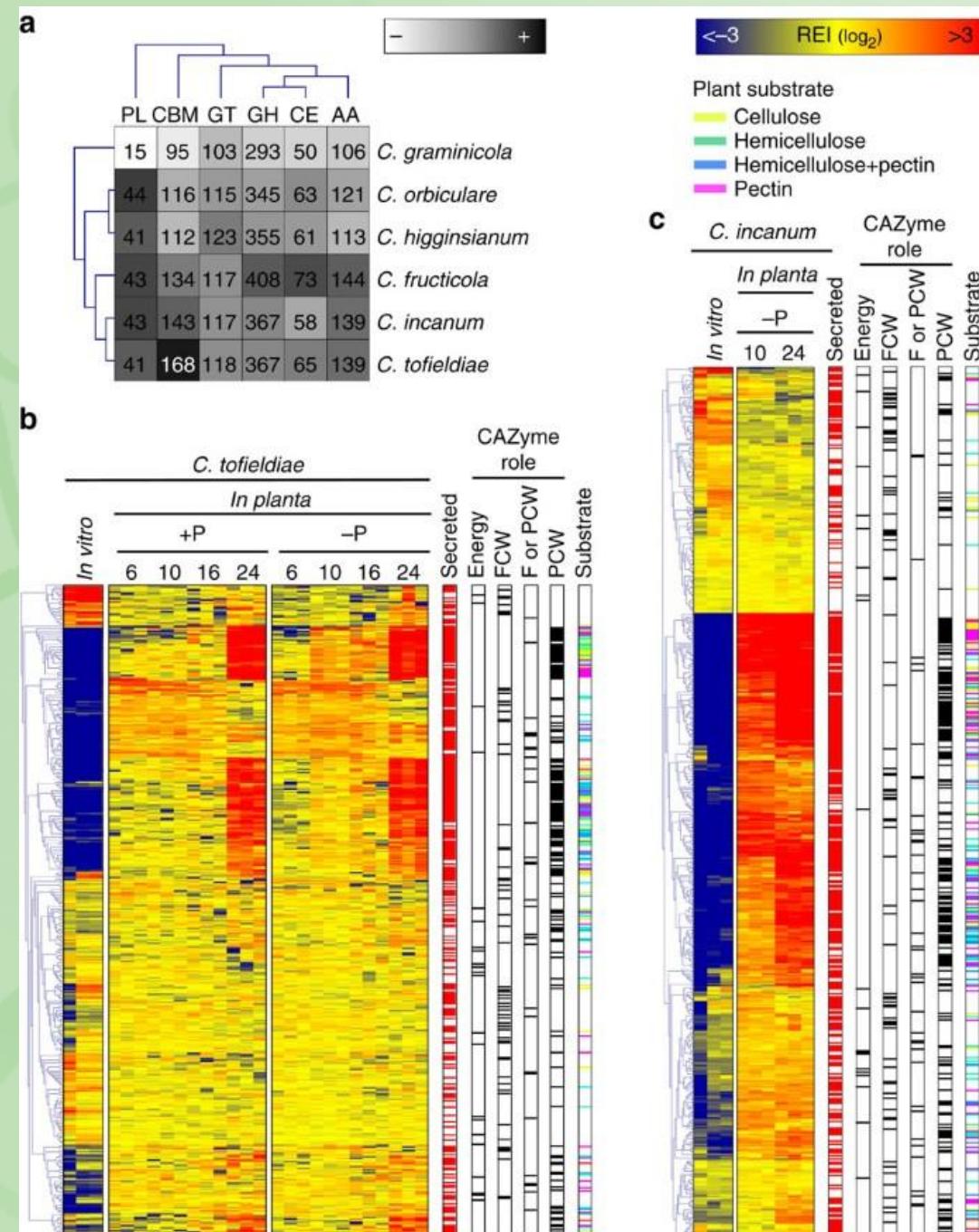
a**b**

Adaptado de:
HACQUARD et al. 2016. **Nature Communications**. DOI: [10.1038/ncomms11362](https://doi.org/10.1038/ncomms11362)

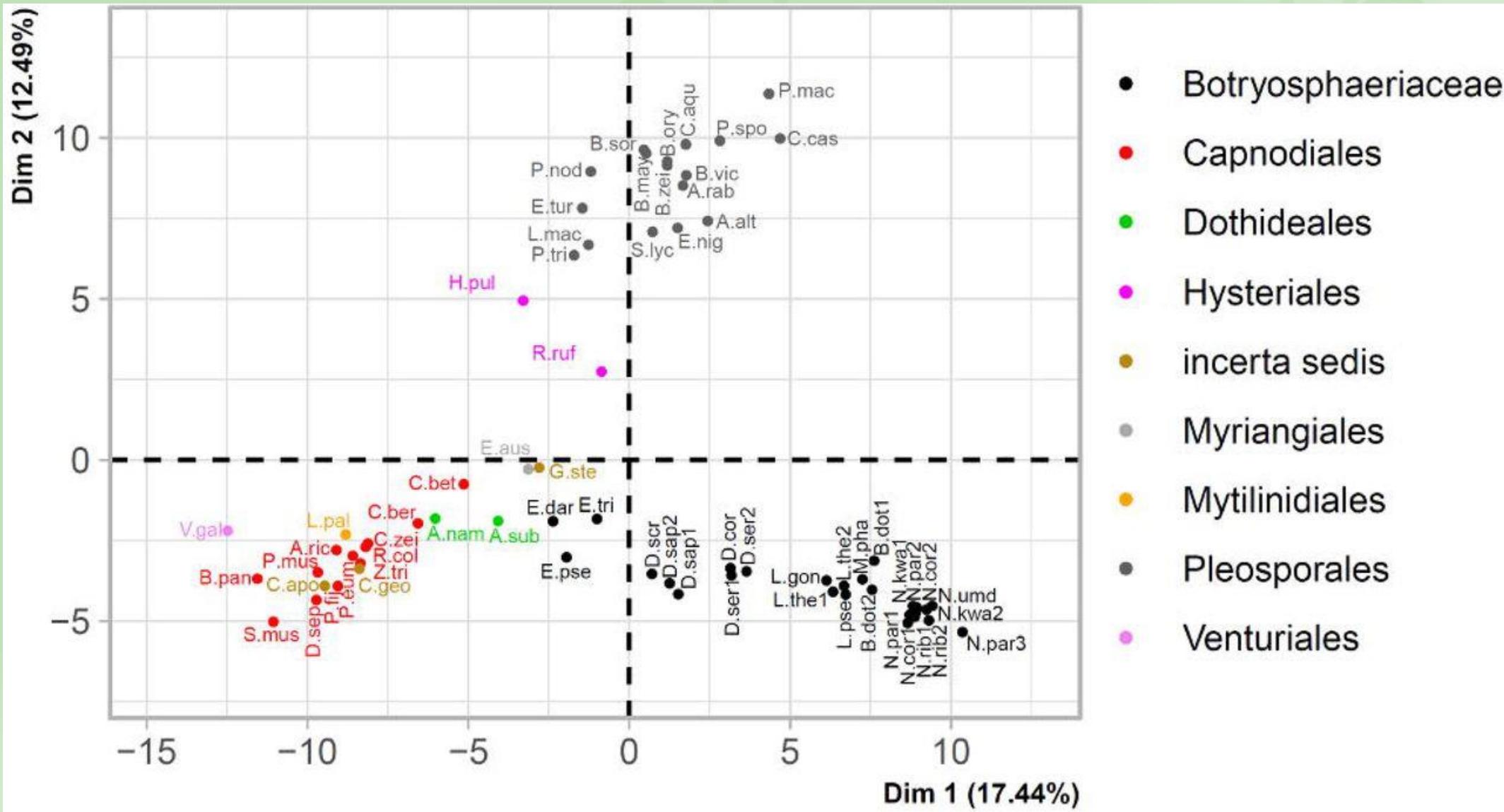
Combinando dados de expressão gênica



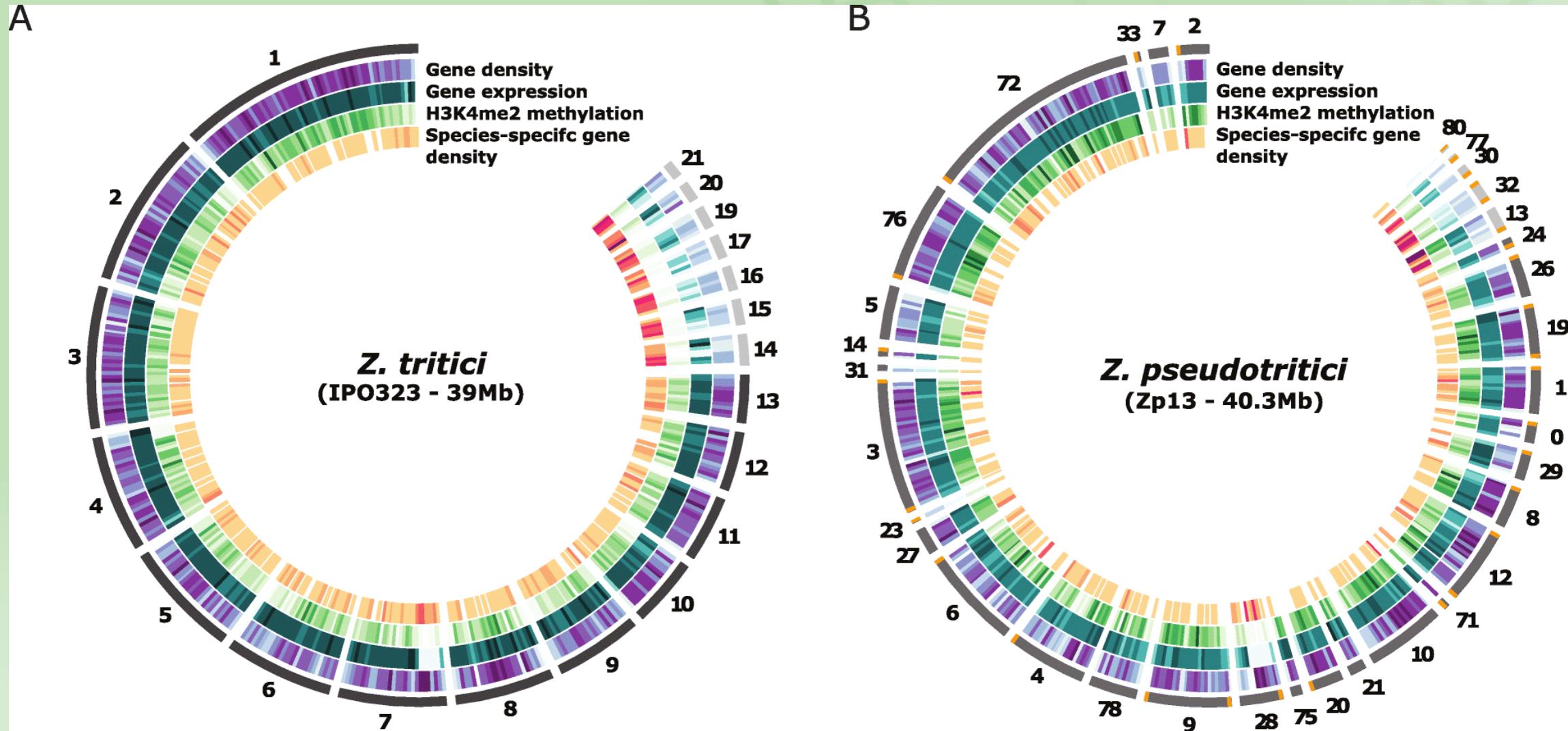
Adaptado de:
HACQUARD et al. 2016. *Nature Communications*. DOI: [10.1038/ncomms11362](https://doi.org/10.1038/ncomms11362)



PCA

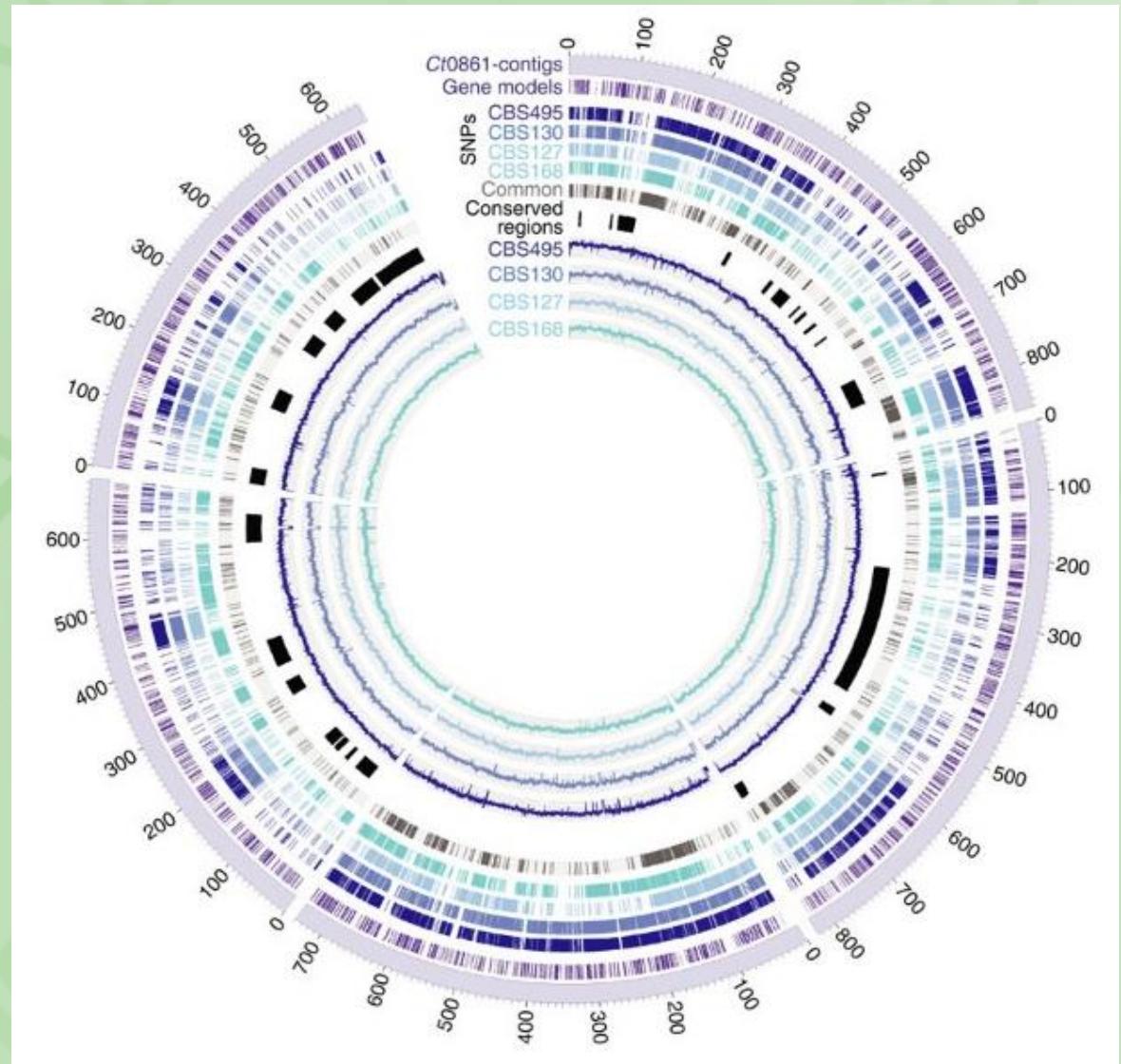
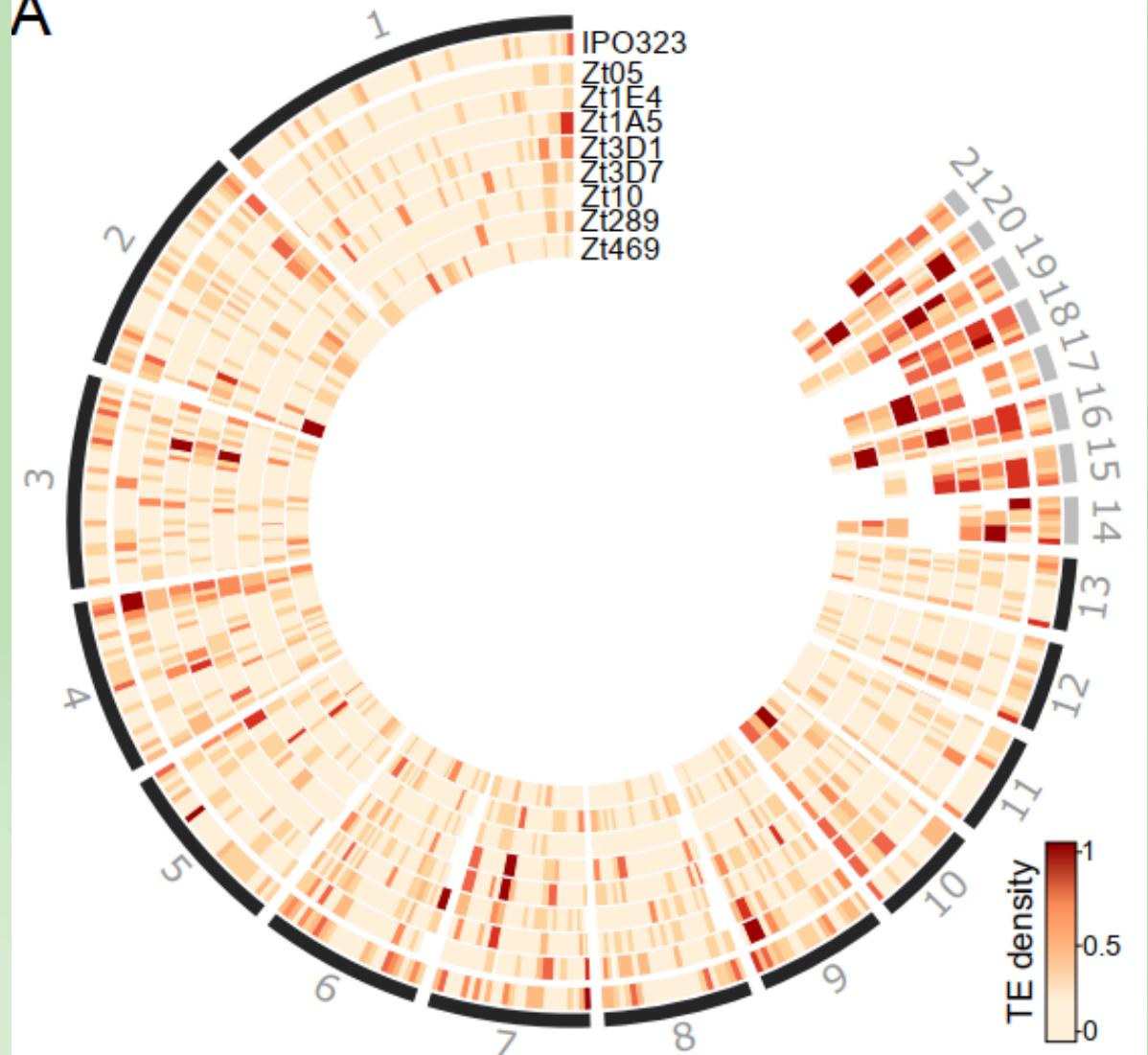


Circos plot (heatmap, diferentes características)

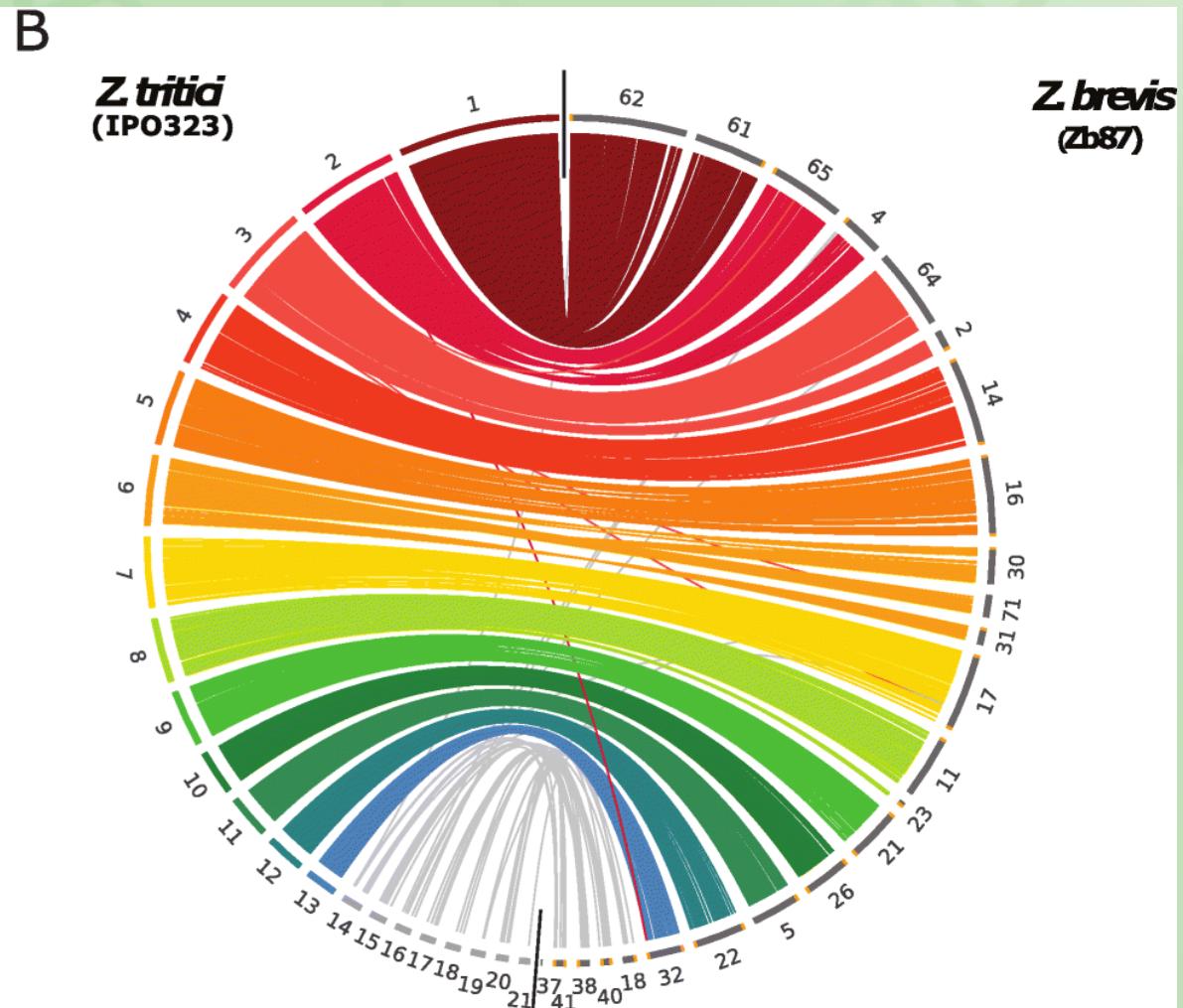
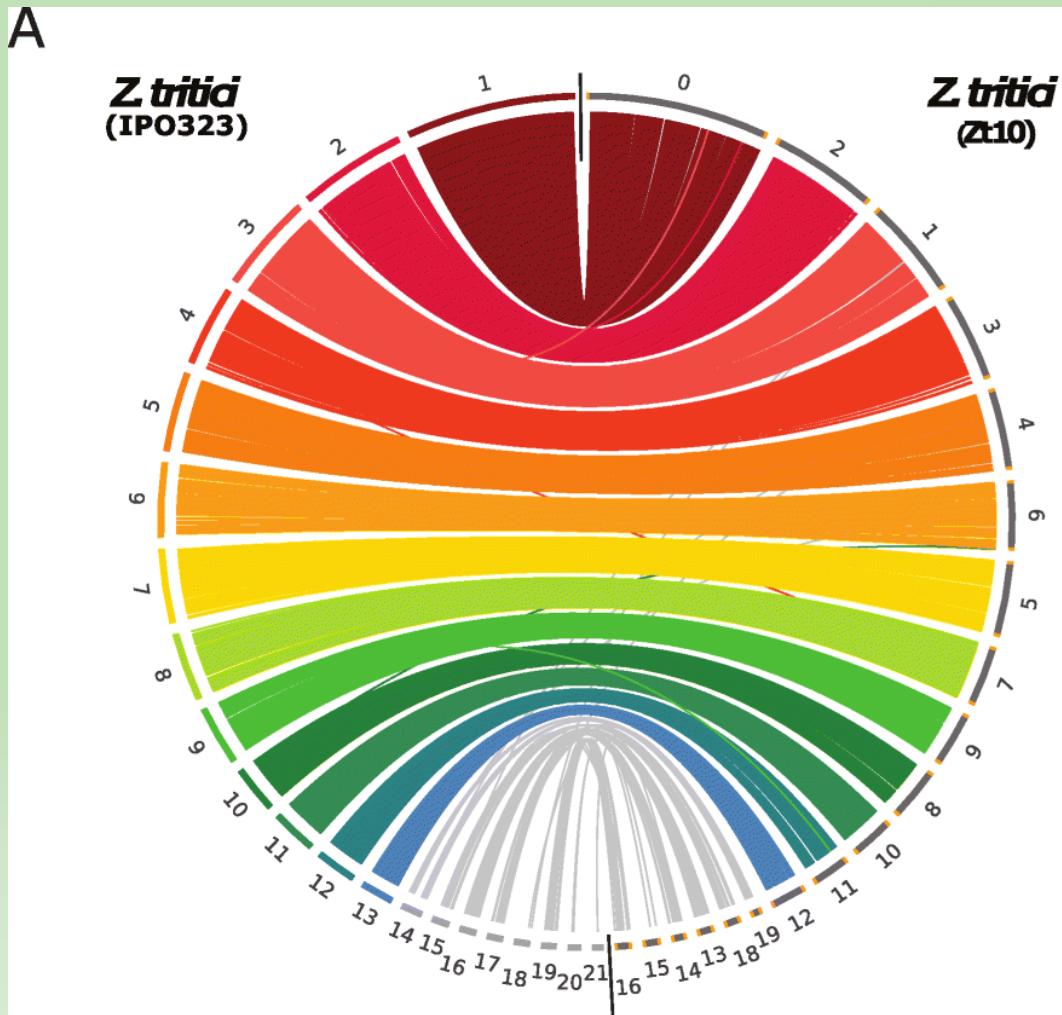


Circos plot (heatmap, diferentes individuos)

A



Circos plot (sintenia)

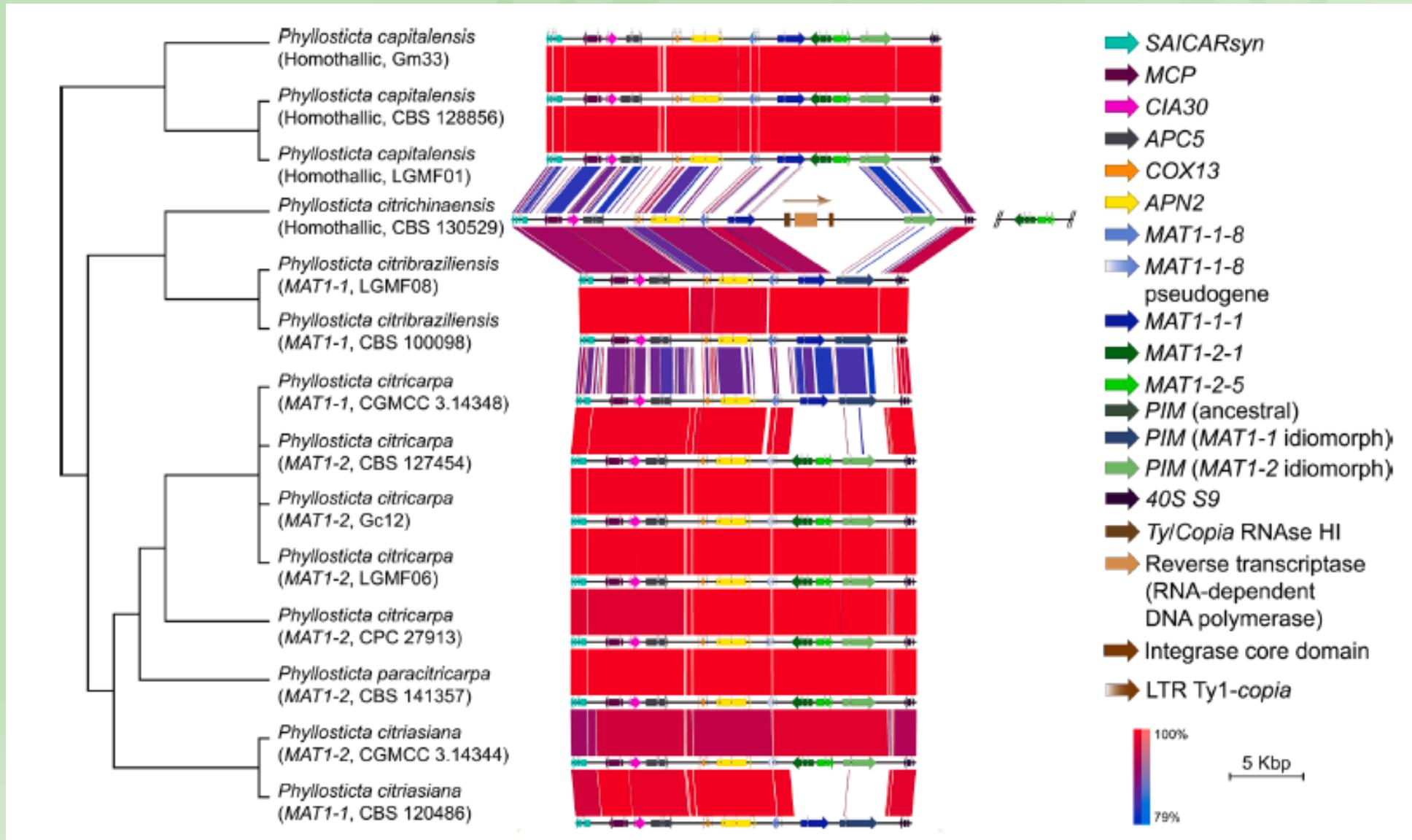


Adaptado de:
FEURTEY et al. 2020. BMC Genomics. DOI: [10.1186/s12864-020-06871-w](https://doi.org/10.1186/s12864-020-06871-w)

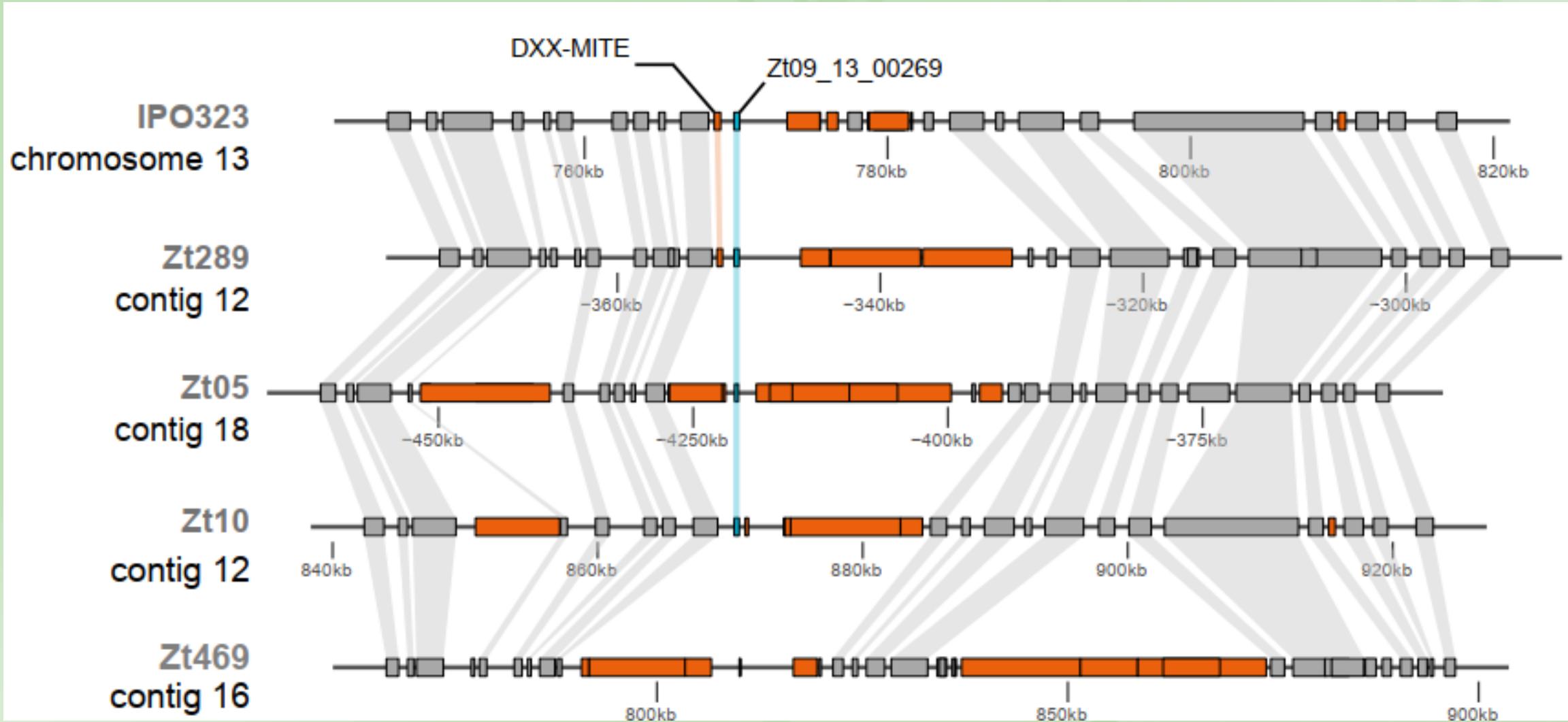
Sintenia (regiões específicas)

Adaptado de:

PETTERS-VANDRESEN et al. 2020. *Fungal Genetics and Biology*. DOI: [10.1016/j.fgb.2020.103444](https://doi.org/10.1016/j.fgb.2020.103444)

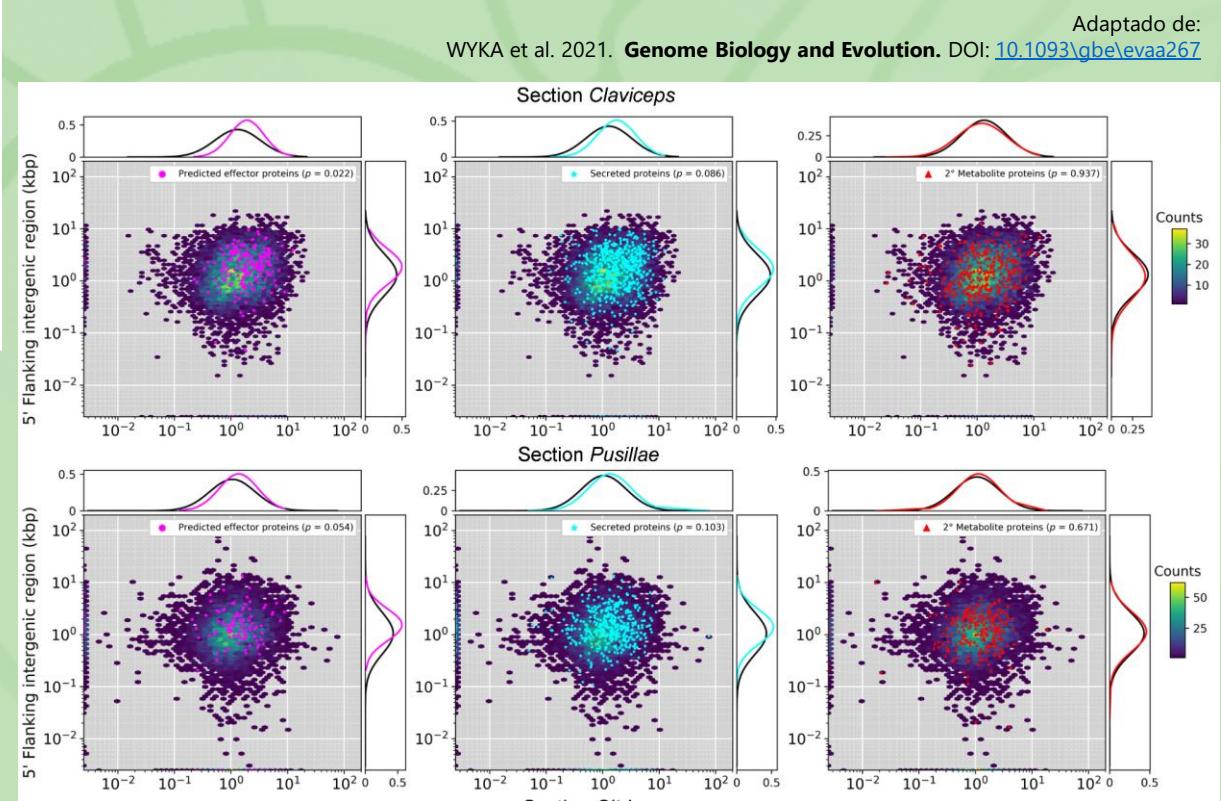
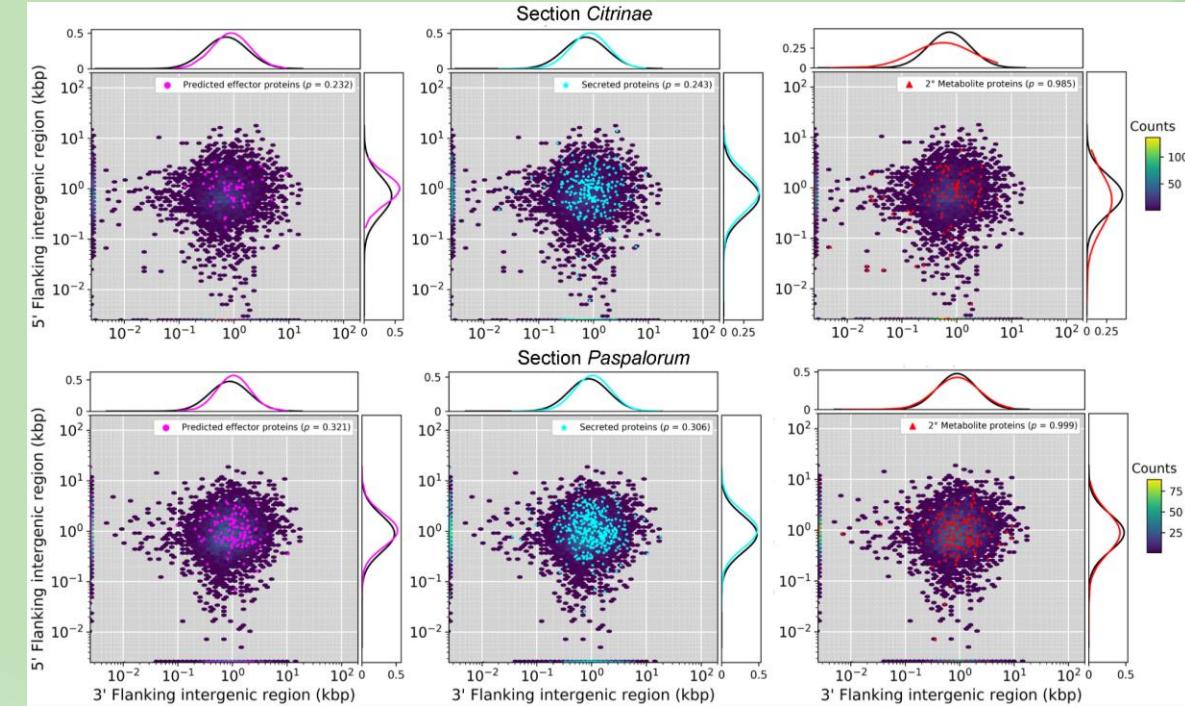


Sintenia (regiões específicas)



Adaptado de:
LORRAIN et al. 2020. Preprint no servidor **bioRxiv**. DOI: [10.1101/2020.05.13.092635](https://doi.org/10.1101/2020.05.13.092635)

Hexbin plot



Como escolher o formato (ou combinação ideal)?

- Tentativa e erro
- Inspirações em artigos similares
- Consultando material especializado para compreender aplicações, vantagens e desvantagens



The R Graph Gallery



storytelling data

Checklist de submissão de um artigo

- Os dados brutos de sequenciamento estão disponíveis em uma base de dados pública?
- As montagens de genoma/transcriptoma e anotações estão disponíveis em uma base de dados pública?
- Arquivos de código utilizados para processamento dos dados ou produção de tabelas/figuras estão disponíveis em um repositório público?
- As informações de acesso a todos esses dados estão claras e disponíveis no artigo?