

Pangenomes

Profa. Dra. Chirlei Glienke

BIOGEMM-UFPR

GS Treinamentos e Consultoria

Problema/Observação 1:

Genomas de 6 linhagens de *Pyrenophora tritici-repentis*

	86-124	Biotrigo9-1	Ls13-192	90-2	M4*	DW5*
Race	2	2	4	4	1	5
Known NE	A	A	–	–	AC	B
			toxb	toxb x2		
Source	Canada	Brazil	USA	Canada	Australia	USA
GenBank accession	NRDI02	JAHCYZ00	JAHCSW00	JAAFOX00	NQIK02	MUXC02
Genes	14272	14450	12851	14224	15459	14276
Total protein (aa) length (Mb)	6.94	7.01	6.04	5.86	6.90	5.95
Predicted effectors‡	178 (1.2%)	169 (1.1%)	189 (1.4%)	380 (2.6%)	291 (1.8%)	314 (2.1%)

Problema/Observação 1:

Table 1 | Genome and lifestyle characteristics of different fungal plant pathogens

Species (isolate)	Isolate-specific host	Species-specific host	Lifestyle	Reproduction	Genome size (Mb)*	% Repetitive sequence*	Genome features linked to pathogenicity [‡]	Refs
<i>Blumeria graminis</i> f. sp. <i>tritici</i> (96224)	Wheat	Various	Biotroph	Sexual and asexual	180	90	Presence and/or absence of polymorphisms of candidate effectors	140, 141
<i>Blumeria graminis</i> f. sp. <i>hordei</i> (DH14)	Barley	Various	Biotroph	Sexual and asexual	120	64	Repeat-rich accessory regions that encode infection-specific transcribed genes	142
<i>Magnaporthe oryzae</i> (70–15)	Rice	Various crops and wild grasses	Hemi-biotroph	Sexual and asexual	41	10	Highly polymorphic effector candidates and translocations of effector genes	22, 143
<i>Ophiostoma novo-ulmi</i> (H327)	Elm	Elm	Necrotroph	Sexual and asexual	32	3.4	ND [#]	144
<i>Verticillium dahliae</i> (VdLs17)	Lettuce	Various	Necrotroph	Asexual	37	12	Enrichment of <i>in planta</i> -expressed effector candidates in LS [#] regions	16, 44
<i>Fusarium solani</i> / <i>Nectria haematococca</i> MPVI (77-13-4)	Pea	Various	Hemi-biotroph	Sexual and asexual	54	• <5 (core) • >10–25 (supernumerary)	LS chromosomes confer host specificity and virulence	46
<i>Fusarium graminearum</i> (PH-1)	Wheat	Wheat and barley	Hemi-biotroph	Sexual and asexual	36	<3	Enrichment of <i>in planta</i> -expressed and species-specific genes in regions of high SNP [#] density	145
<i>Fusarium oxysporum</i> f. sp. <i>lycopersici</i> (4287)	Tomato	Various	Hemi-biotroph	Asexual	60	28 (~74% of TE [#] located on LS [#] chromosomes)	LS chromosomes confer host specificity and virulence	43

LS, lineage-specific; ND, no data; TEs, transposable elements. *Genome size and repeat content refer to the respective reference isolate. Isolate-specific hosts refer to the host plant from which the reference isolate was collected. In some cases, other isolates of the same species infect other hosts. [‡]Genome characteristics have been inferred from comparative genomics analyses.

Problema/Observação 1:

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	<i>MAT1-1</i>	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	<i>MAT1-2</i>	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	<i>MAT1-1</i>	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	<i>MAT1-1</i>	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	<i>MAT1-2</i>	Scaffold 46	28,952,665	152	440,231	23	54.60%	1.30%	9108	96.1%	(Guarnaccia et al., 2019)
<i>P. citrichinaensis</i>	CPC 27913	<i>Citrus sinensis</i> (leaf litter)	Malta	Heterothallic	<i>MAT1-2</i>	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	<i>MAT1-2</i>	Contigs 1345, 2477, 3787	31,127,197	5748	21,637	411	53.05%	5.93%	9472	94.5%	(Wang et al., 2016)
	CGMCC 3.14348	N/A	China	Heterothallic	<i>MAT1-1</i>	Contigs 4070 and 4071	32,007,210	6716	11,522	811	52.57%	7.94%	9768	87.5%	GenBank Assembly: GCA_000382785.1
<i>P. citrichinaensis</i>	LGMF 06	<i>Citrus sinensis</i>	Brazil	Heterothallic	<i>MAT1-2</i>	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	<i>MAT1-2</i>	Scaffold 48	29,529,839	134	510,184	23	54.35%	2.45%	9083	96.2%	(Guarnaccia et al., 2019)

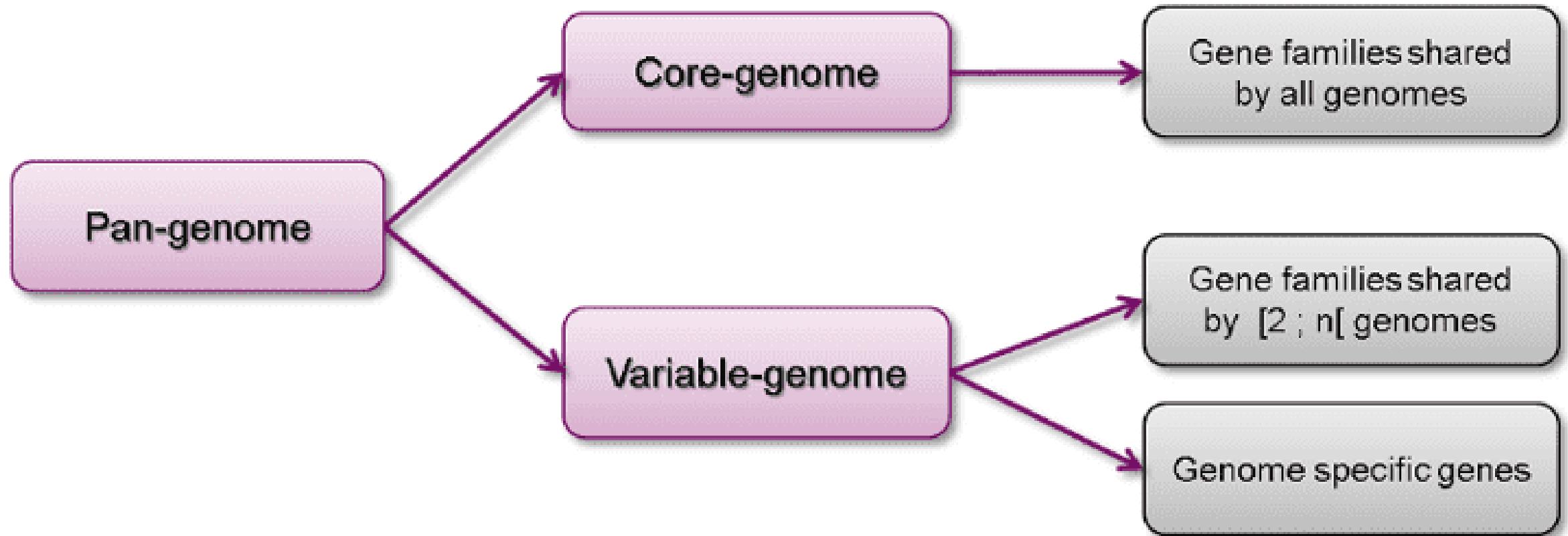
N/A: not available or not applicable to the present study.

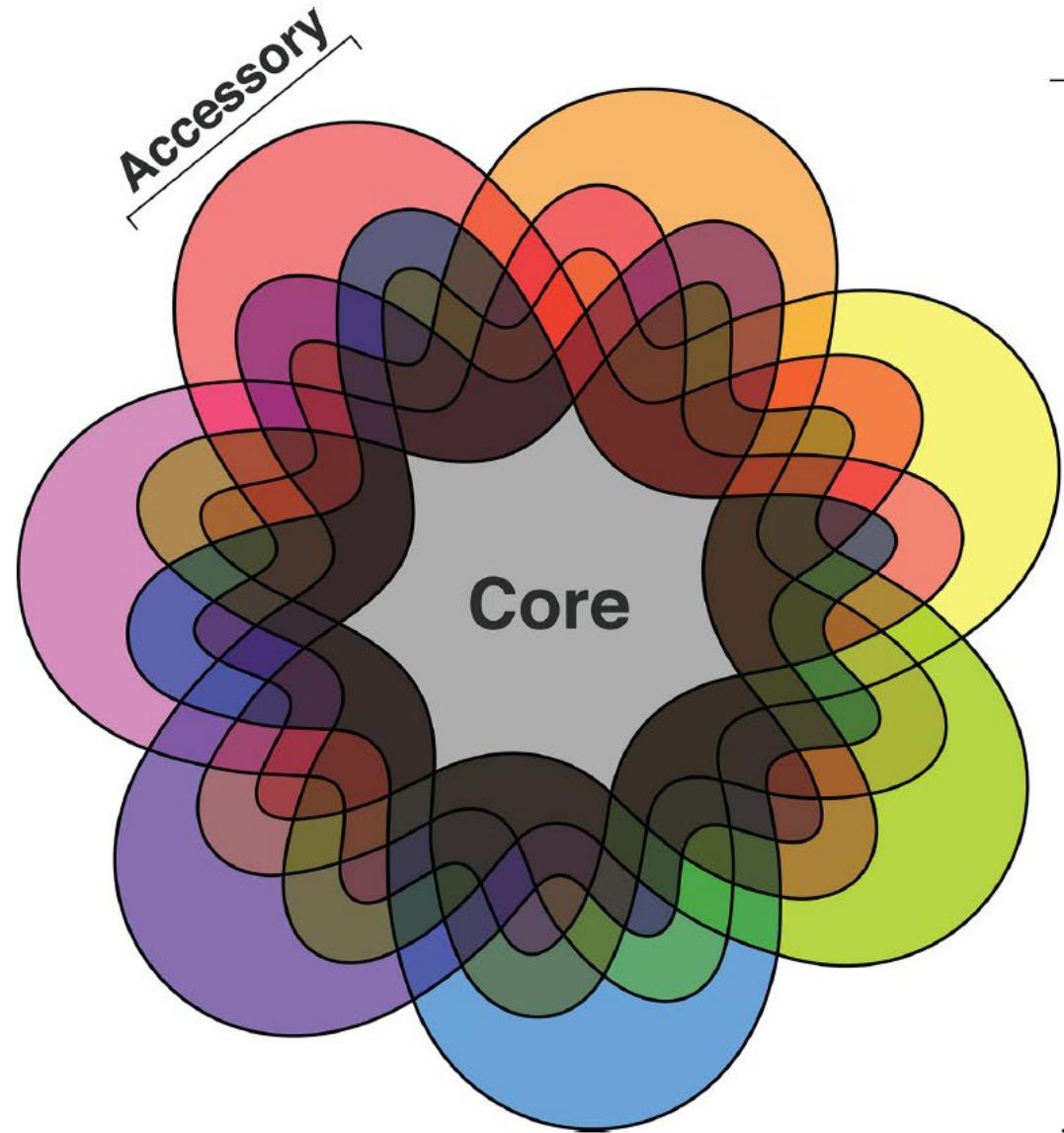
^a Type strains included in the analysis are indicated in **bold**. Culture collection abbreviations: CBS = Westerdijk Fungal Biodiversity Institute, Utrecht, Netherlands; CPC = Collection of Pedro W. Crous, held at the Westerdijk Fungal Biodiversity Institute; CGMCC = China General Microbiological Culture 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.

Variação intra-específica: pangenoma

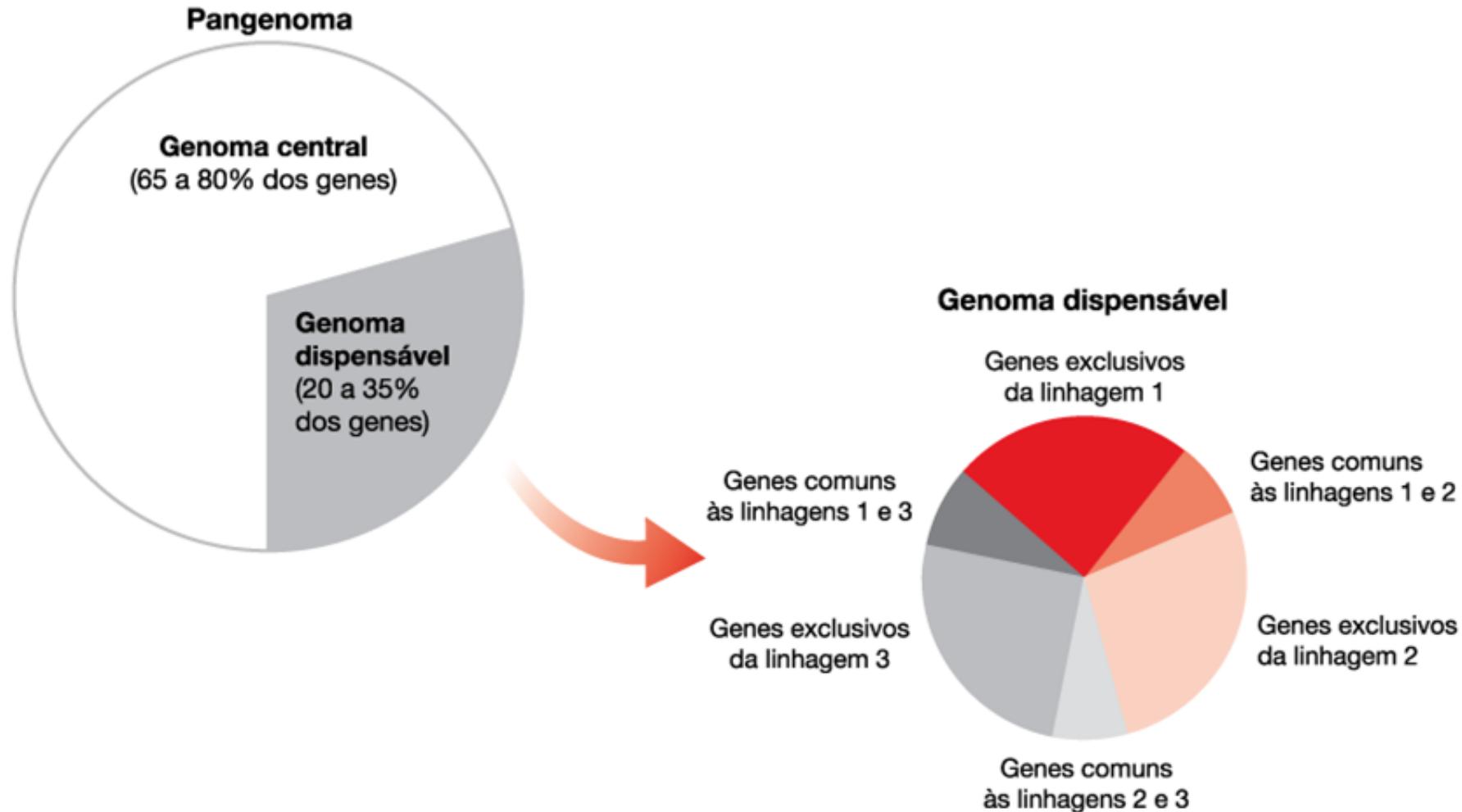
- Nem todos os genes presentes em um indivíduo estarão presentes em todos os indivíduos da mesma espécie
- Variações neste conteúdo podem estar associadas com aspectos biológicos (por exemplo: patogenicidade)
- Pangenoma: repertório gênico total de uma espécie
 - Genoma central (core): genes compartilhados por todos os indivíduos, isolados/linhagens ou populações da espécie
 - Genoma acessório ou dispensável (accessory): genes presentes em alguns indivíduos e ausentes em outros



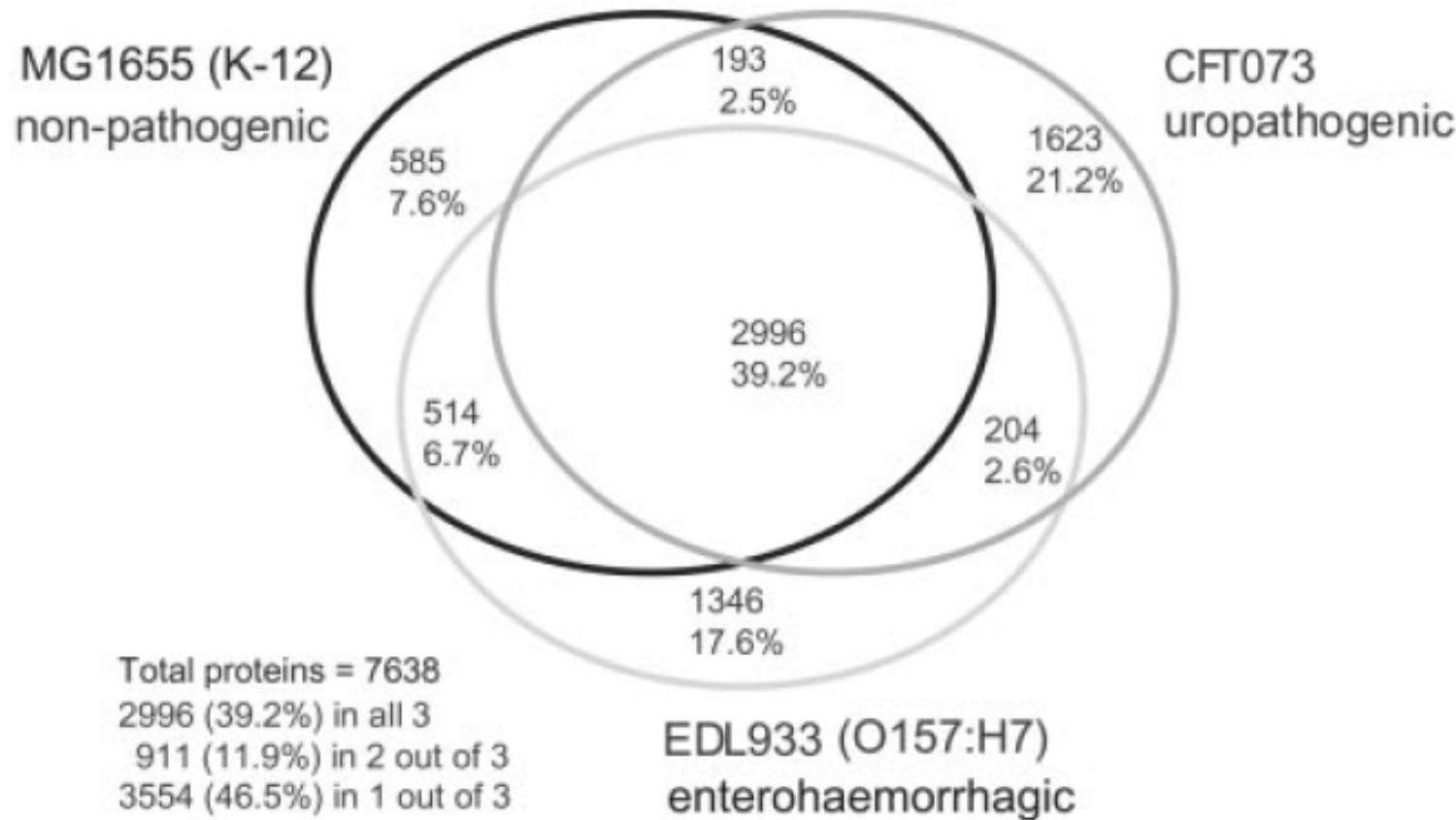


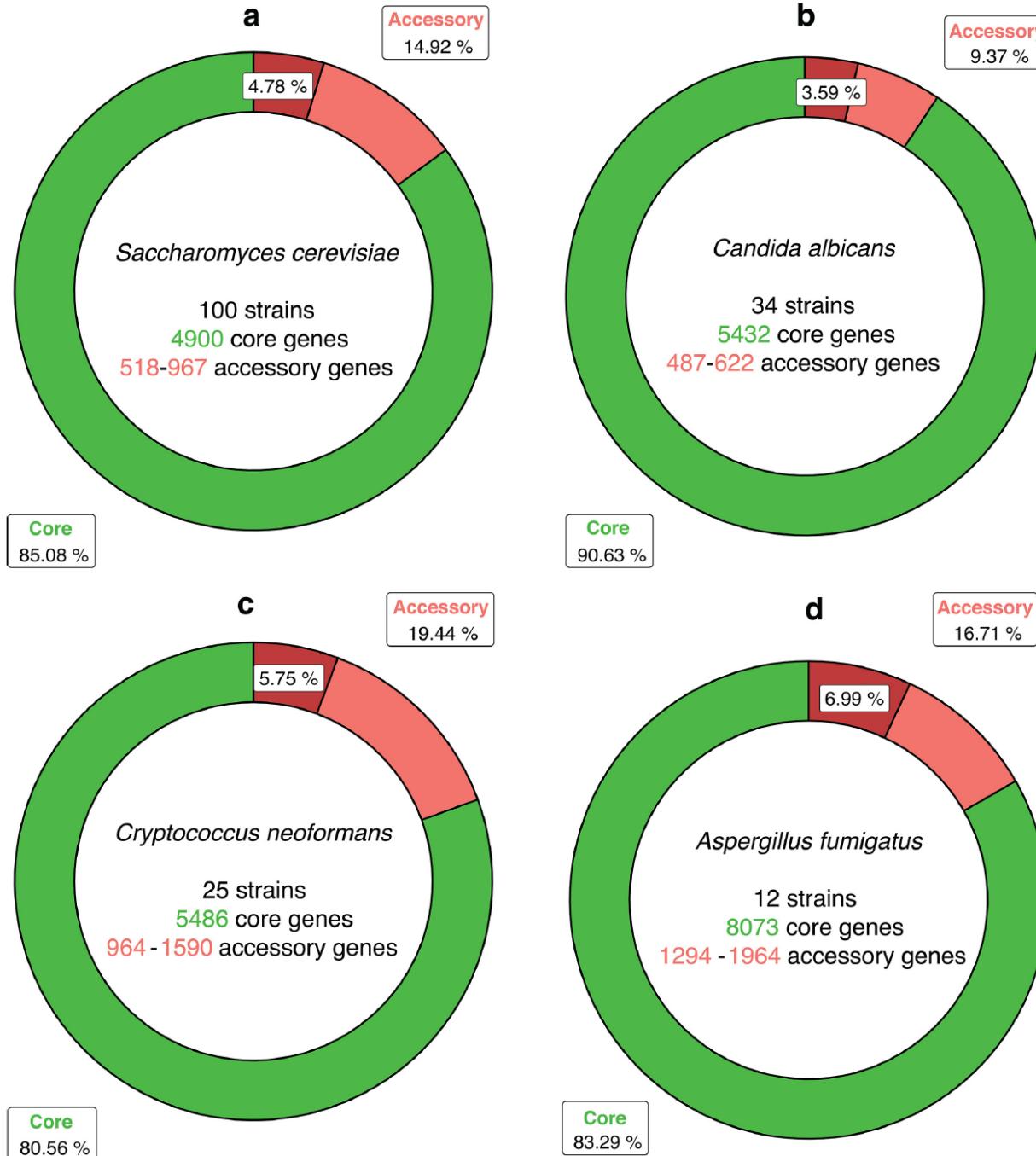
Pan-genome

Variação intra-específica: pangenoma



2002 - *Escherichia coli*: "amazingly"



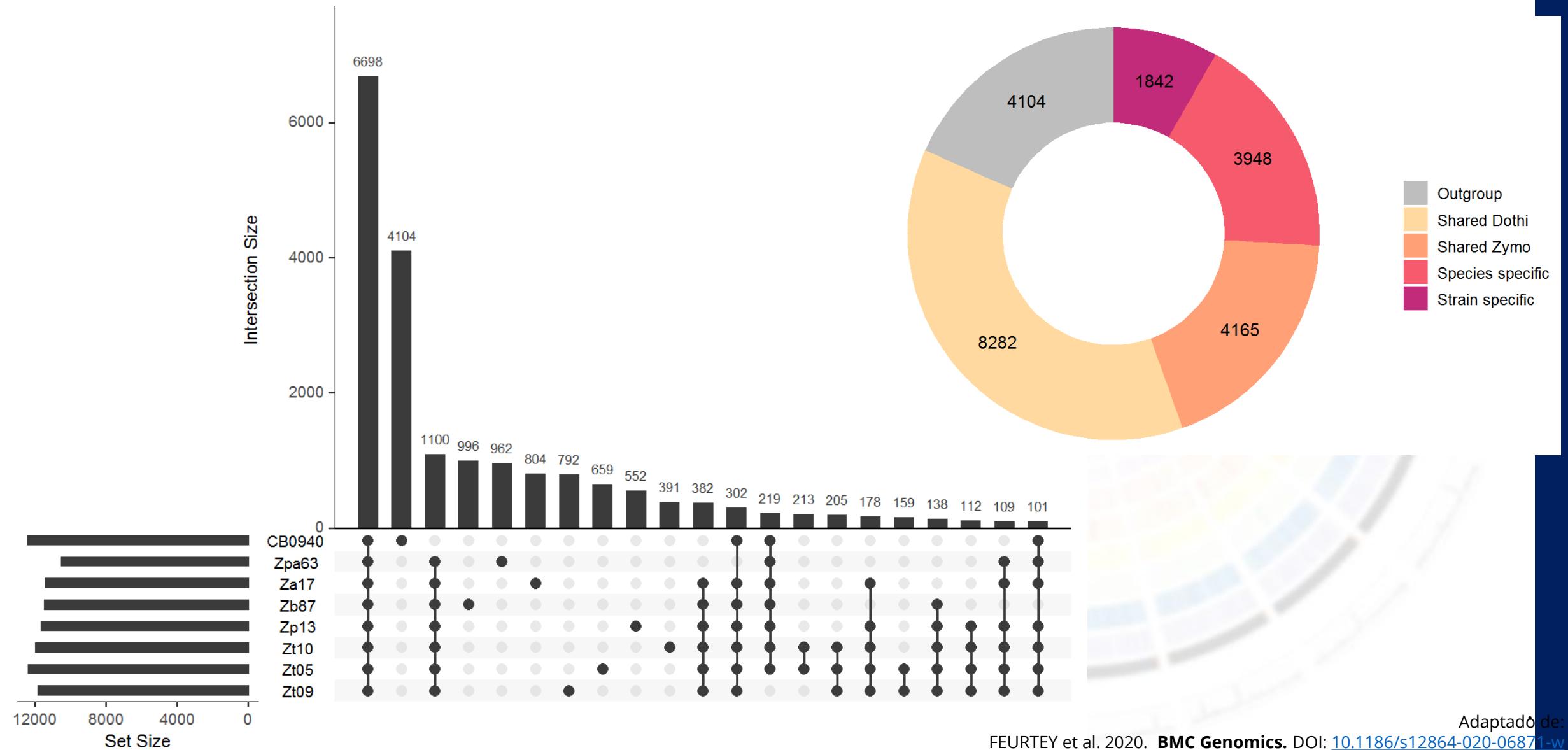


- Pangenomas de quatro espécies de fungos
 - Saccharomyces cerevisiae*
 - Candida albicans*
 - Cryptococcus neoformans* var. *grubii*
 - Aspergillus fumigatus*

Os gráficos de anéis representam o número total de modelos de genes em complementos de pan-genoma expressos como uma proporção do tamanho total do pan-genoma.

As seções em vermelho escuro representam modelos de genes “core”(centrais) duplicados nos genomas acessórios

Variação intra-específica: pangenoma



Pangenomas fúngicos

- Genes centrais (core):
 - enriquecidos em funções de manutenção
 - agrupados em áreas protegidas de cromossomos
 - conservados por maior eficiência de seleção purificadora
- Genes acessórios:
 - mais propensos a codificar proteínas específicas de linhagem que facilitam a adaptação ambiental como genes envolvidos no metabolismo secundário e especificidade de nicho
 - tendem a se concentrar em regiões subteloméricas altamente variáveis e em rápida evolução e adjacentes a elementos transponíveis

Pangenomas fúngicos

- Genes acessórios:
 - variação é essencial para facilitar a seleção diversificada e a adaptação às pressões ambientais específicas da população
 - pode ser particularmente importante para a adaptação do patógeno fúngico

Ilhas genômicas

- Segmentos que diferem do padrão geral do genoma
 - Conteúdo GC
 - Padrões de sintenia
 - Orientação gênica
- Geralmente associados a elementos transponíveis e eventos de transferência horizontal de genes
- Contém genes não essenciais e em geral conferem vantagens adaptativas ao organismo
 - Ilhas de patogenicidade: capacidade de infectar e causar doença em uma espécie hospedeira

Importância da pangenômica

- A pangenômica de patógenos em rápida evolução pode ser usada para identificar conjuntos de genes potencialmente específicos, que podem ser explorados no futuro
- Potencial para aplicações médicas e biotecnológicas, incluindo o desenvolvimento de novas estratégias terapêuticas e ferramentas de manipulação de genes para o controle de doenças

Importância da pangenômica

- Análise de pangenoma é uma das abordagens eficazes para entender melhor um clado de bactérias patogênicas porque ajuda a elaborar tratamentos terapêuticos que são eficazes para todo o clado e eficazes apenas para uma espécie ou cepa específica dentro um clado, com base nas semelhanças e diferenças biológicas no clado

Genes acessórios

- Podem influenciar diferenças fenotípicas entre isolados
- Exemplos
 - em bactérias, isolados resistentes e sensíveis a antibióticos da mesma espécie podem ter genomas acessórios diferentes

Evolução de Pangenomas

- O pangenoma de uma espécie pode evoluir como consequência do modo de vida:
 - espécies simpátricas podem ter grandes pangenomas (e, portanto, um grande grau de variação intraespecífica), enquanto espécies ambientalmente isoladas ou altamente especializadas têm pangenomas menores
- Em procariotos
 - transferência horizontal de genes
- Em eucariotos
 - Variam entre as espécies e pode incluir eventos de rearranjo do genoma ou processos de evolução adaptativa mais discretos

Evolução de Pangenomas

- *Zymoseptoria tritici*
 - Altos níveis de conteúdo de genoma acessório funcionalmente redundante (40%)
 - Mecanismos de defesa do genoma da própria espécie induzindo polimorfismos em oposição a eventos de duplicação de genes
 - A maioria dos ganhos genéticos é resultado de eventos de duplicação, seguidos de diversificação e neofuncionalização
 - Os ganhos genéticos foram atribuídos a eventos de transferência horizontal de genes ou à emergência de genes de novo a partir de DNA não codificante
 - As perdas genéticas podem ocorrer através de recombinação não homóloga
 - Ganhos e perdas genéticas são críticos para a rápida adaptação de patógenos de plantas a diferentes hospedeiros

Pangenoma de 5 linhagens de *Zymoseptoria tritici*

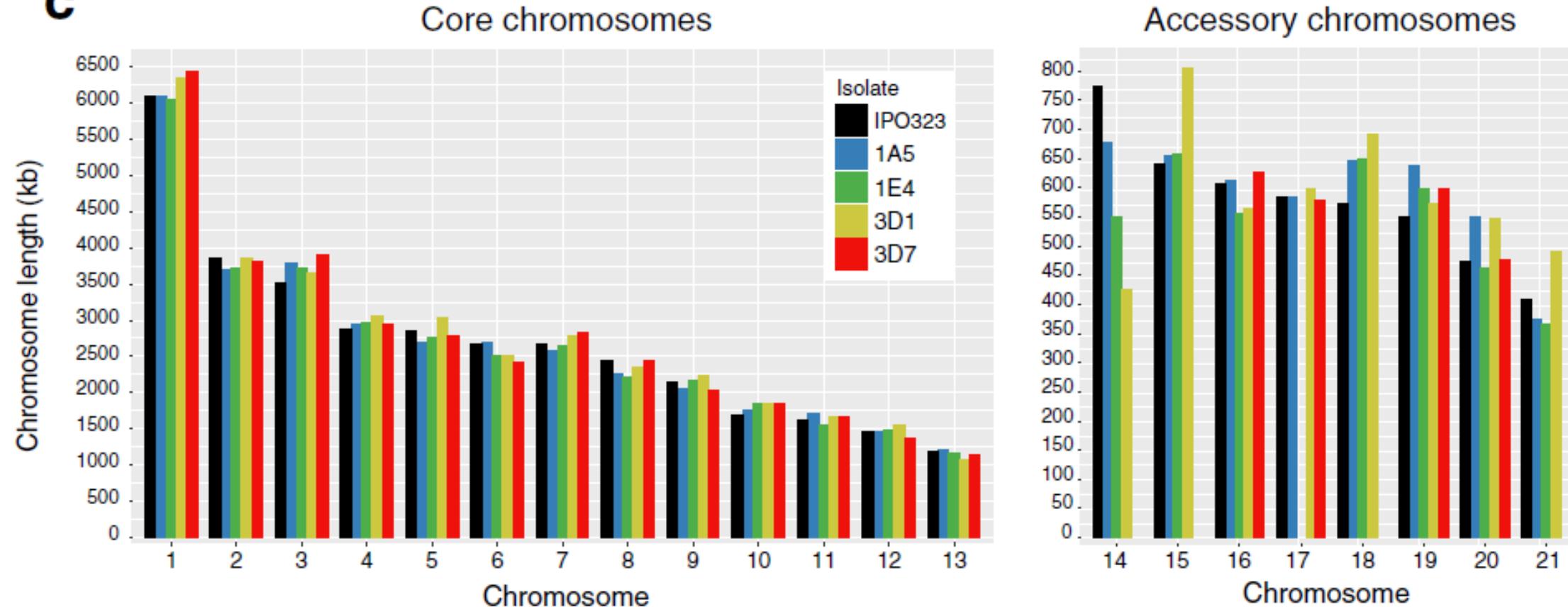
Table 1 Genome assembly and annotation statistics of the five complete *Zymoseptoria tritici* genomes

Genomes assembly	1A5	1E4	3D1	3D7	IPO323
Genome size, Mb	39.7	38.6	40.7	37.9	39.7
Chromosomes, n	21	20	21	17	21
Gene annotation					
Predicted genes	12,092	12,033	12,006	11,737	11,839
Average gene length, bp	1520.2	1524.2	1514.1	1502.6	1620.9
Average protein length, aa	467.5	468.3	466.4	457.2	487.8
Number of exons	29,716	30,015	29,033	30,399	30,068
Average exons per gene	2.46	2.49	2.42	2.59	2.54
Average exon length, bp	570.7	563.3	579.8	530.7	575.2
Number of introns	17,628	17,984	17,027	18,663	18,226
Average introns per gene	2.18	2.21	2.15	2.29	2.54
Average intron length, bp	80.7	79.7	79	80.6	91.6
Genes with introns	8099	8129	7925	8142	8044
Gene density, Mb ⁻¹	304.6	311.7	294.9	309.7	298.3
Protein functions					
Proteins with Pfam domain	7403	7439	7431	7353	7375
Secreted protein	1025	1059	1014	1034	932
Predicted effectors ^a	309	286	296	330	221

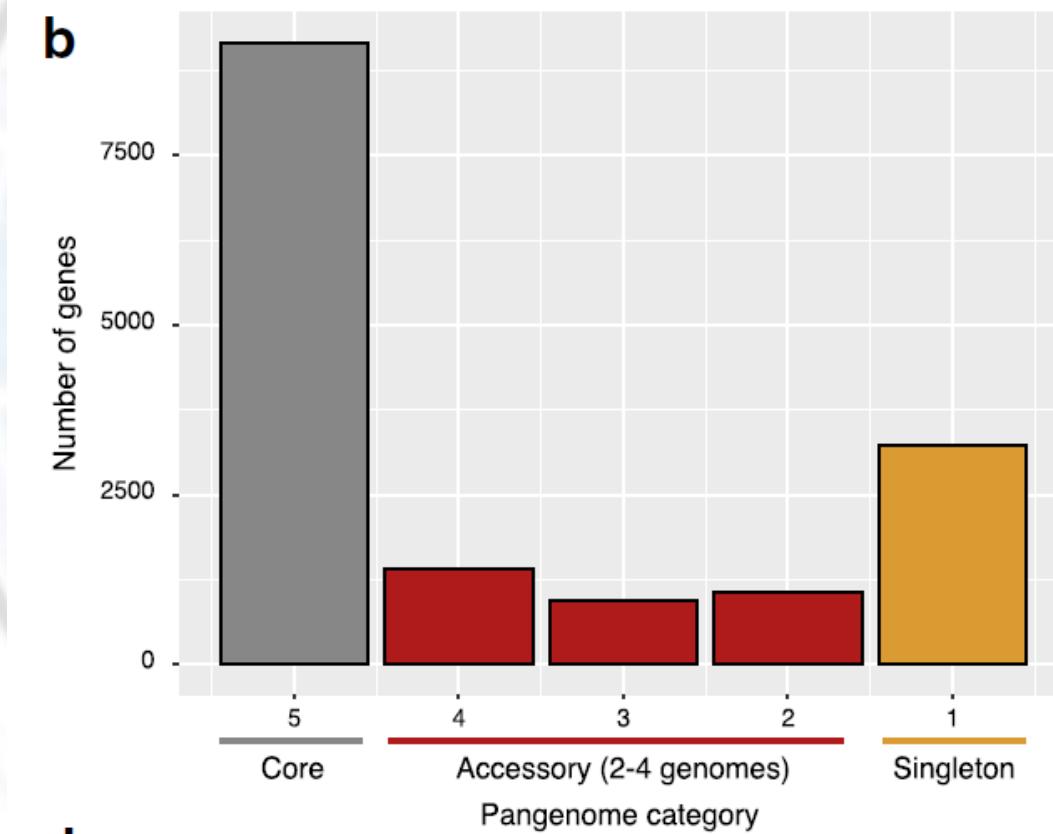
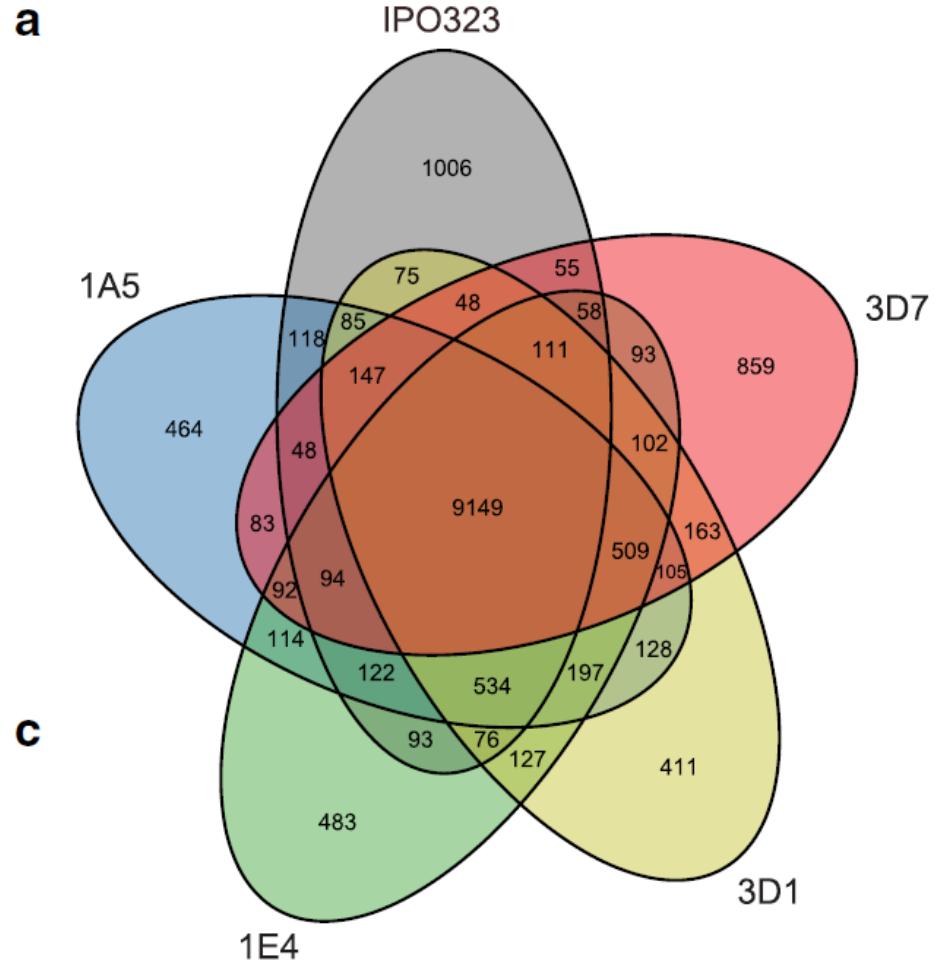
^aEffectors were predicted using the software EffectorP [50]

Pangenoma de 5 linhagens de *Zymoseptoria tritici*

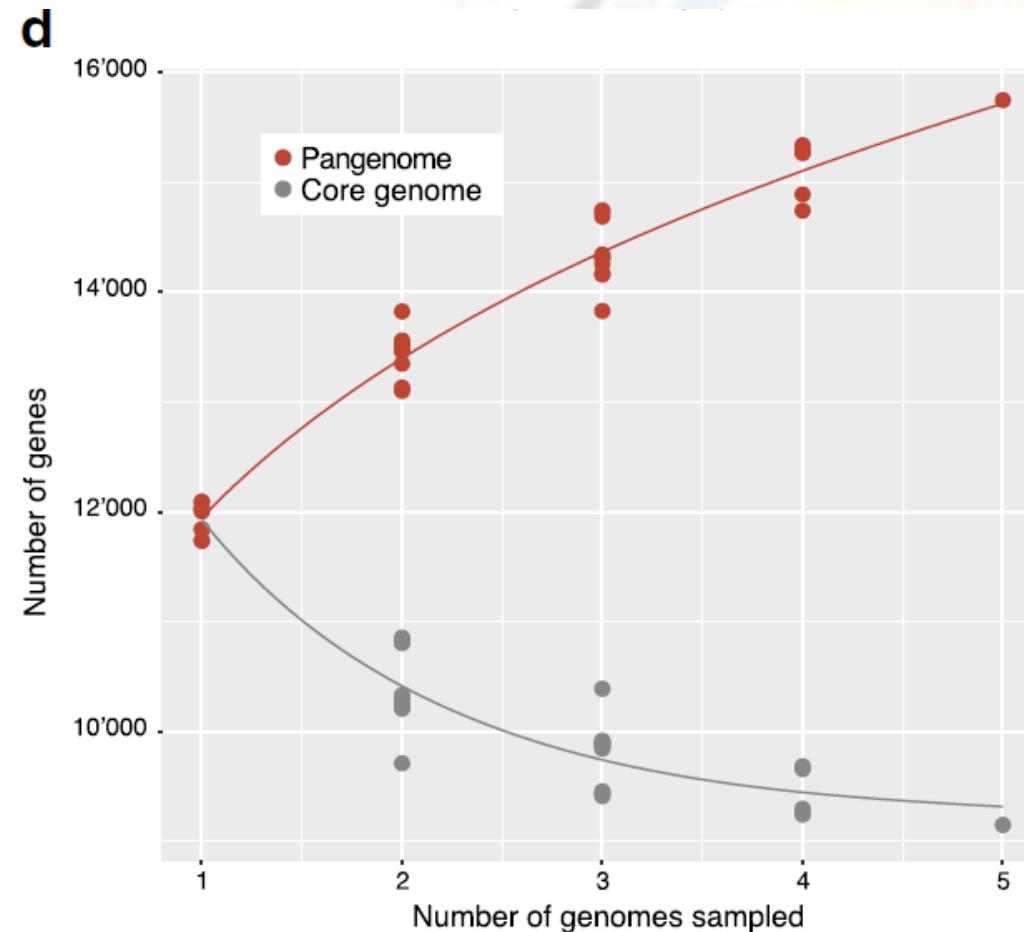
C



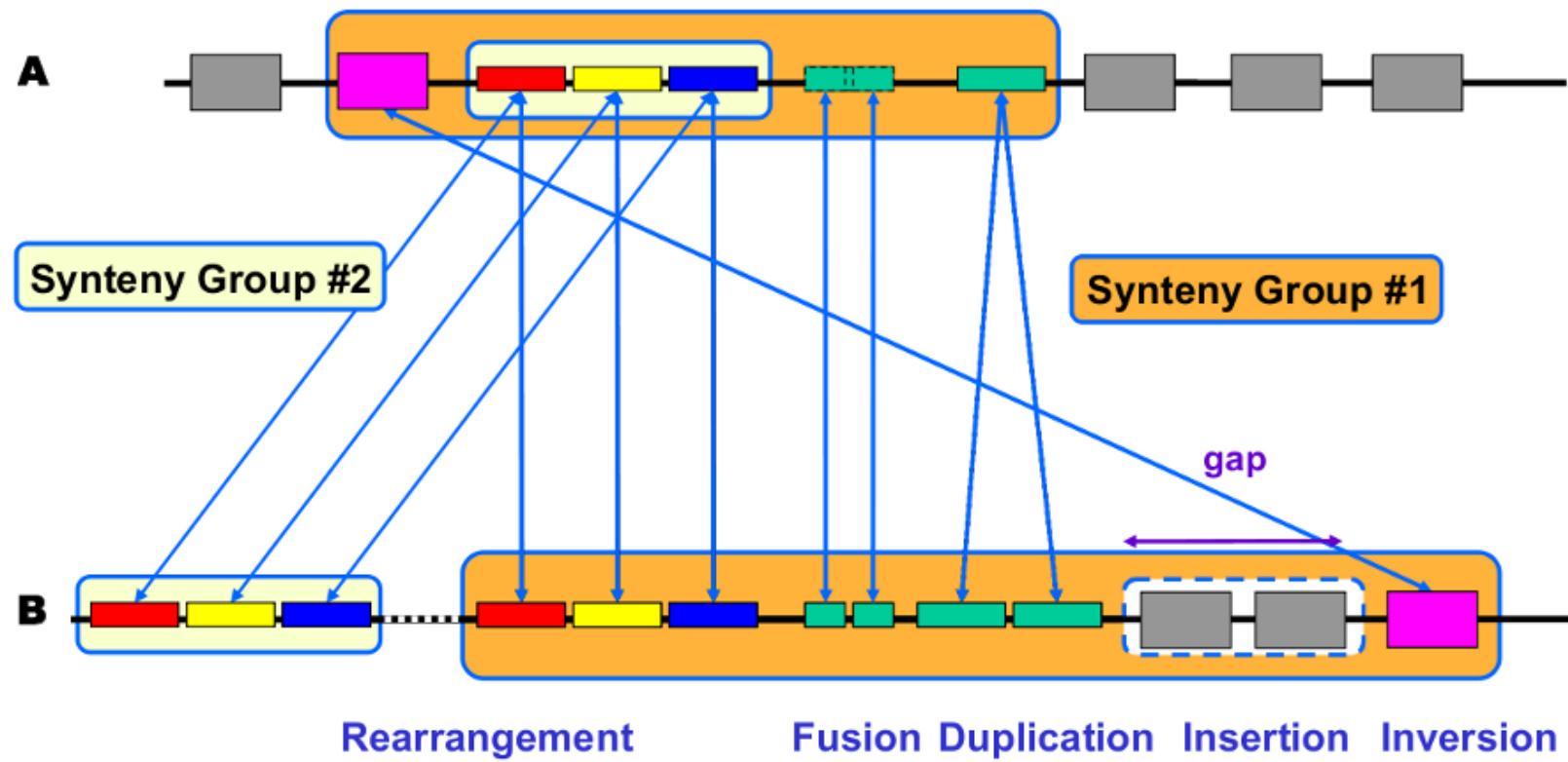
Pangenoma de 5 linhagens de *Zymoseptoria tritici*



Problema 2: Pangenoma de 5 linhagens de *Zymoseptoria tritici*

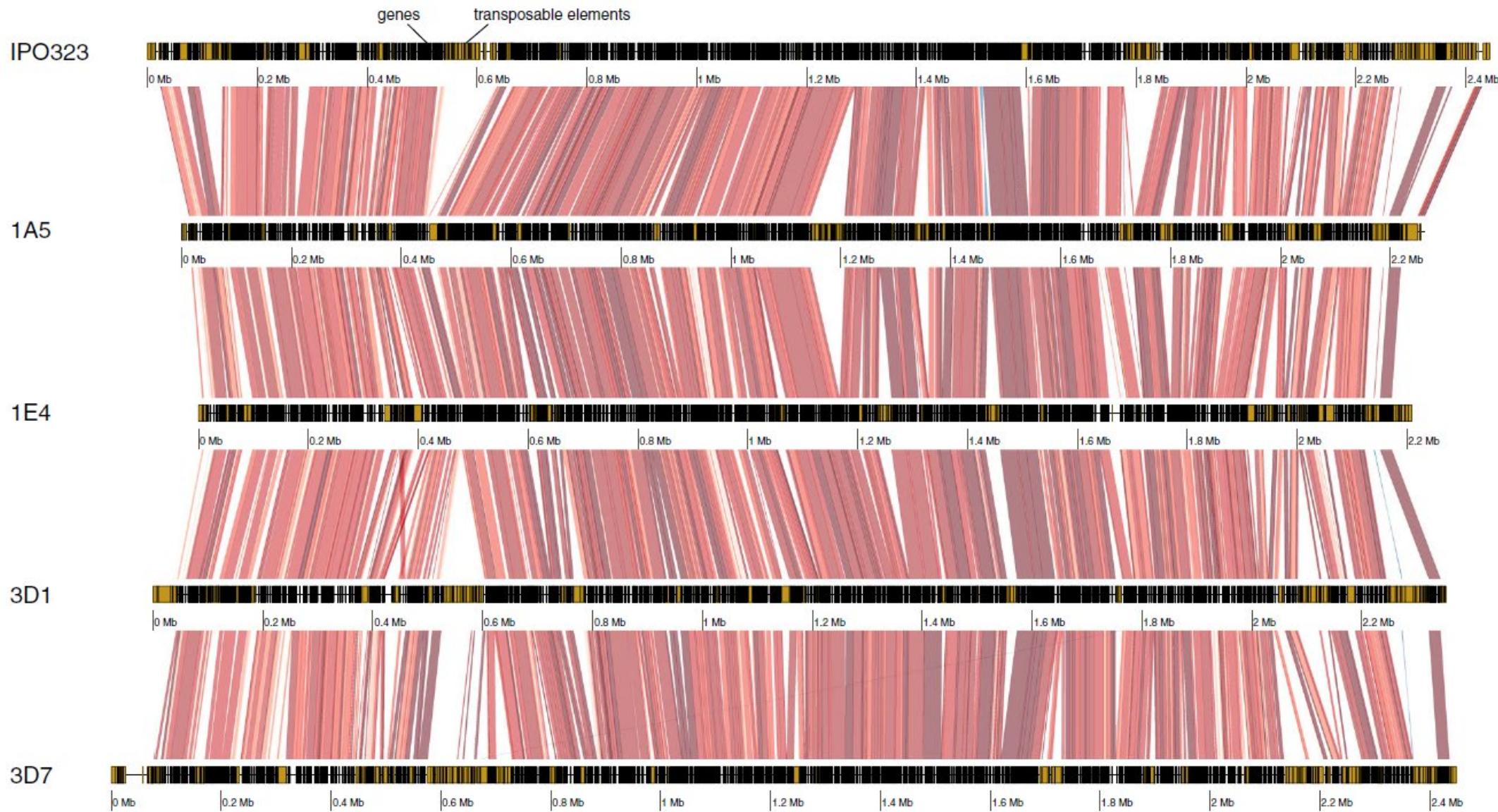


Sintenia



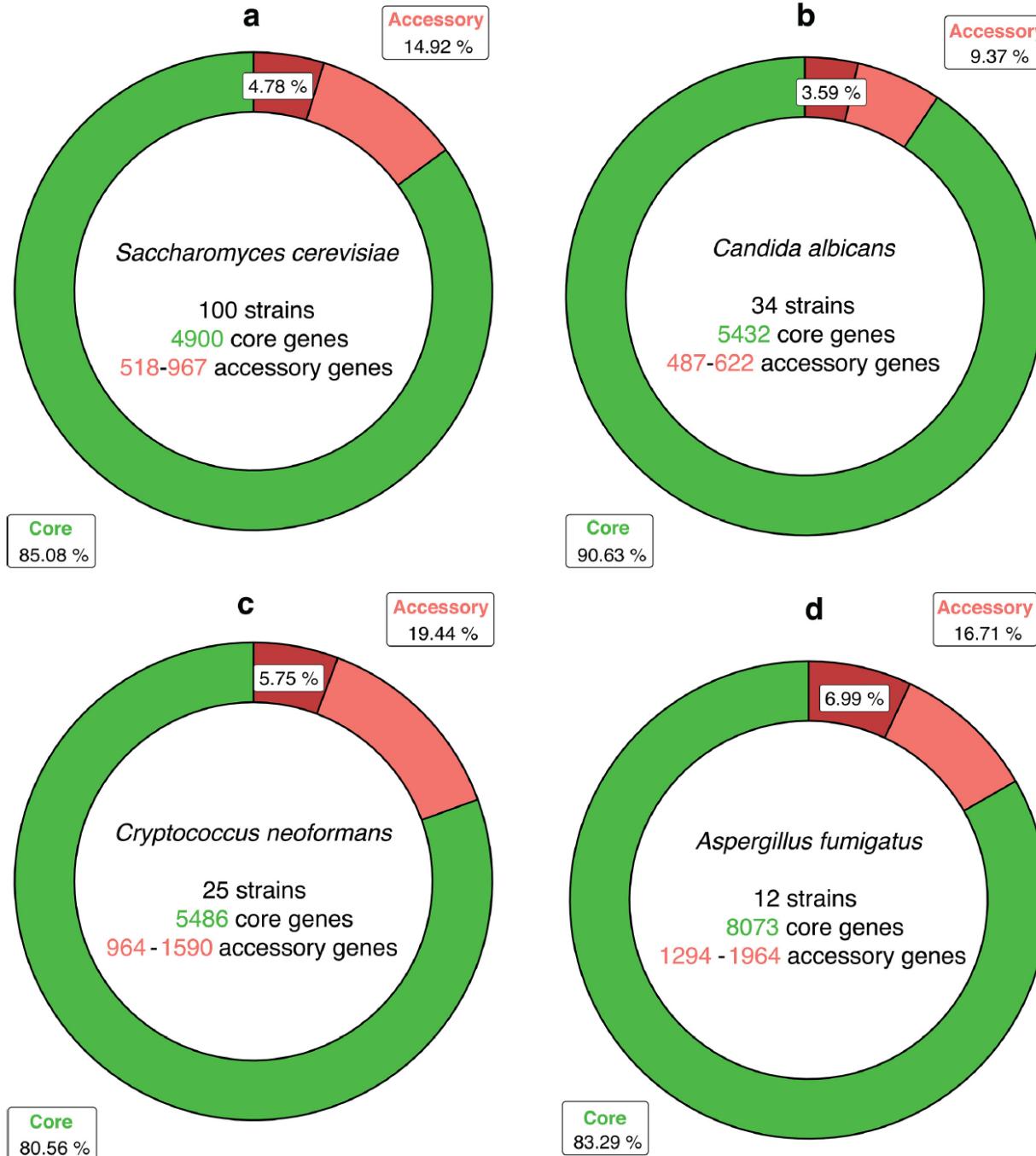
<https://www.genoscope.cns.fr/agc/website/spip.php?article588>

Pangenoma de 5 linhagens de *Zymoseptoria tritici* Sintenia cromossômica pareada entre homólogos do cromossomo 8



Evolução de Pangenomas

- *Saccharomyces cerevisiae*
 - parece ter surgido via introgressão de espécies de *Saccharomyces* intimamente relacionadas, com um número menor originado de eventos HGT com outras leveduras



- Pangenomas de quatro espécies de fungos
 - Saccharomyces cerevisiae*
 - Candida albicans*
 - Cryptococcus neoformans* var. *grubii*
 - Aspergillus fumigatus*

Os gráficos de anéis representam o número total de modelos de genes em complementos de pan-genoma expressos como uma proporção do tamanho total do pan-genoma.

As seções em vermelho escuro representam modelos de genes centrais duplicados nos genomas acessórios

Pangenomas fúngicos – genomas acessórios

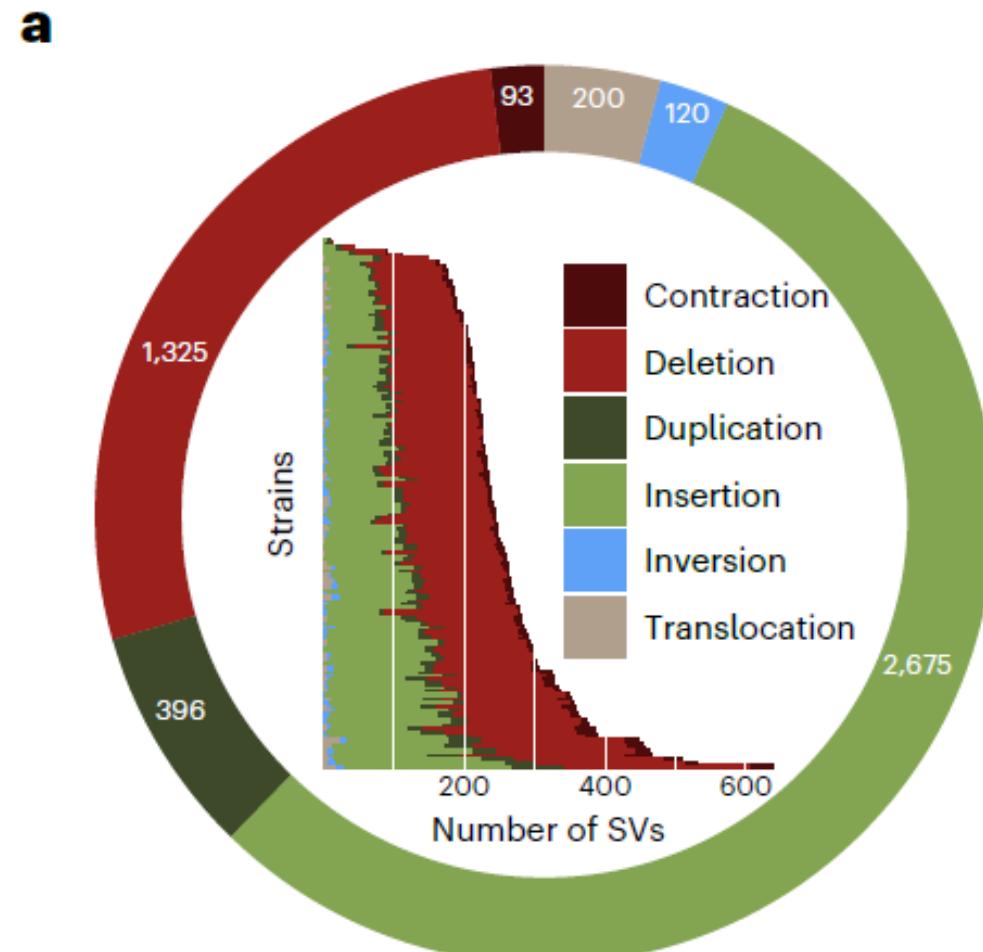
- Enriquecidos por genes de origem eucariótica e surgem através de inovações eucarióticas, como duplicações gênicas
- Sem evidências de HGT em larga escala
- Encontrados em grupos nas regiões terminais e subterminais dos genomas
- Enriquecidos para processos que podem estar implicados na patogenicidade fúngica ou resistência antimicrobiana

Conclusões

Species	Strains	Core genome		Accessory genome		Pan-genome	
		GMs	Clusters	GMs	Clusters	GMs	Clusters
<i>Saccharomyces cerevisiae</i>	100	490 000	4900	85 940 (27 511)	2850 (776)	575 940	7750
<i>Candida albicans</i>	34	184 688	5432	19 098 (7312)	1893 (1013)	203 786	7325
<i>Cryptococcus neoformans</i>	25	137 150	5486	33 091 (9974)	2698 (776)	170 241	8193
<i>Aspergillus fumigatus</i>	12	96 876	8073	19 435 (8127)	3002 (1170)	116 311	11 075

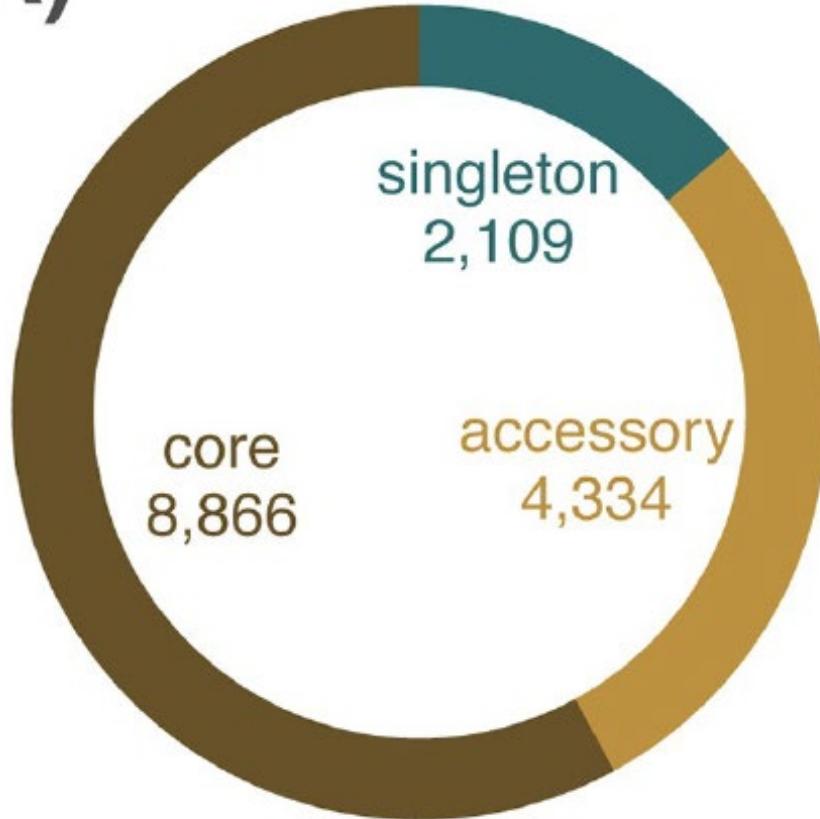
- Os genomas centrais dos fungos são enriquecidos com genes de origem antiga e facilitam muitos processos metabólicos, reguladores e de sobrevivência essenciais em espécies comensais e patogênicas
- Genomas acessórios de fungos são enriquecidos para genes de origem mais recente, parecem evoluir e variam em tamanho por processos como duplicação de genes e eventos de ganho/perda dentro de linhagens, e são enriquecidos para genes envolvidos no transporte de moléculas e metabolismo de carboidratos, entre outras funções
- HGT pode desempenhar apenas um papel limitado na evolução do pangenoma fúngico
- Processos eucarióticos, como a duplicação de genes, podem influenciar a evolução do pangenoma fúngico
- Regiões subterminais de genomas de fungos podem abrigar o conteúdo de genoma acessório
- Os genomas centrais e acessórios fúngicos abrangem várias vias biológicas e fenótipos

Conjuntos genômico de telômero a telômero de 142 linhagens de *Saccharomyces cerevisiae*



Aspergillus fumigatus

A)

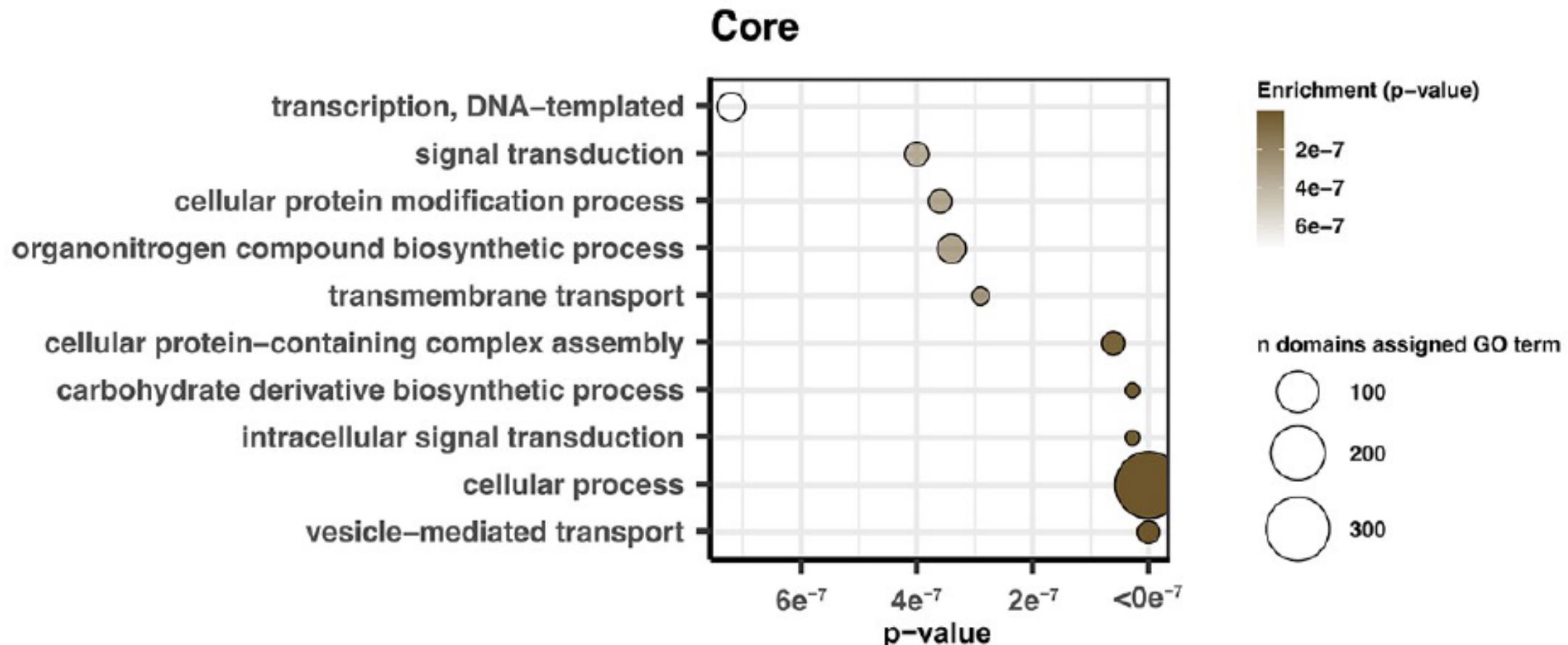


Pangenoma de 260 cepas de *A. fumigatus*:

- Total: 15.508 famílias de genes
- 8.595 (55,42%) genes centrais (presentes em >95% das cepas)
- 3.660 (23,60%) genes acessórios (presentes em >2 e <248 cepas)
- 3.253 (20,98%) singletons (presentes em apenas 1 isolado)

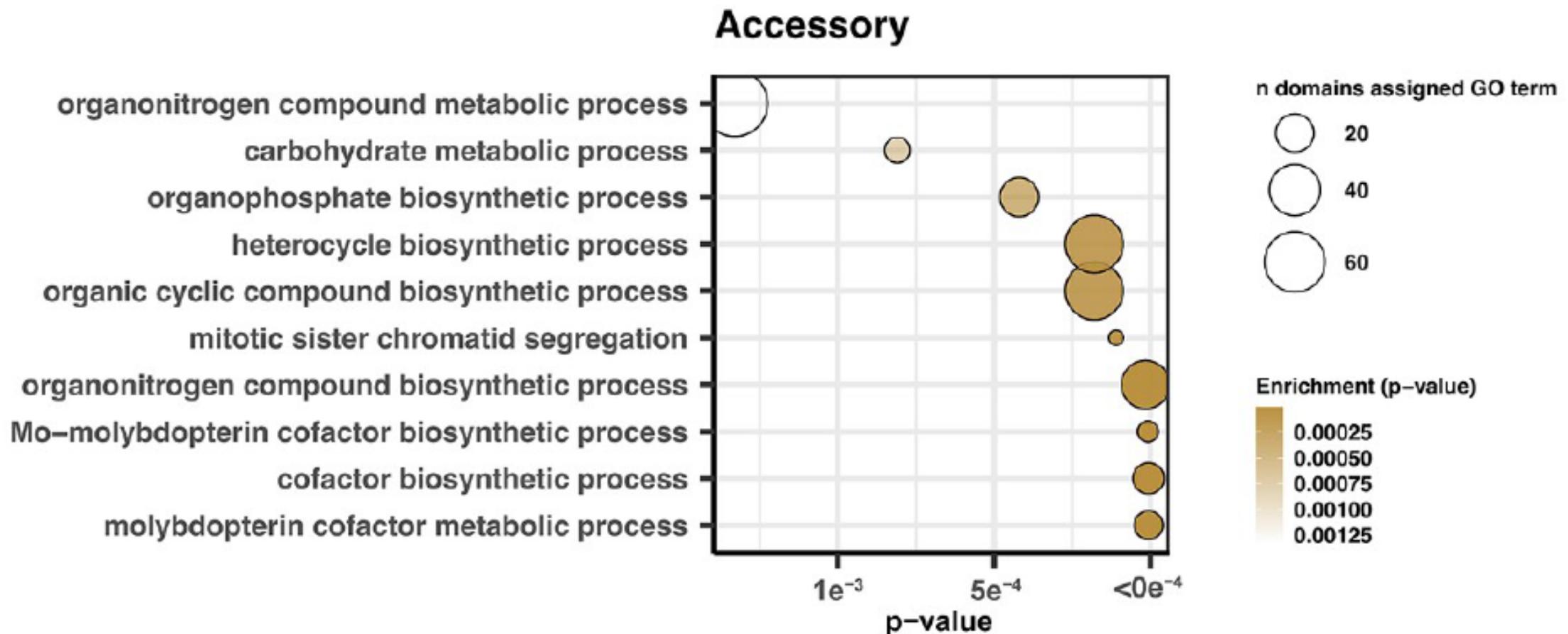
Enriquecimento GO (Gene Orthology)

A



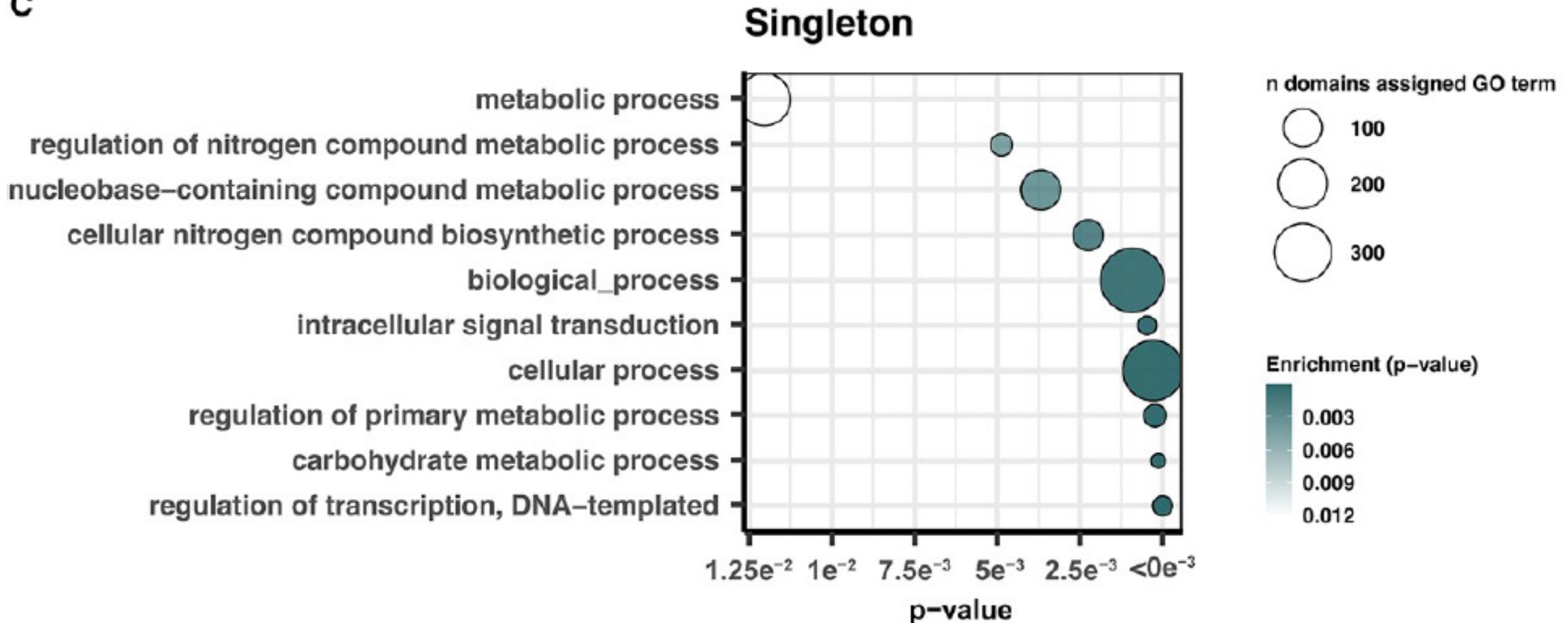
Enriquecimento GO (Gene Orthology)

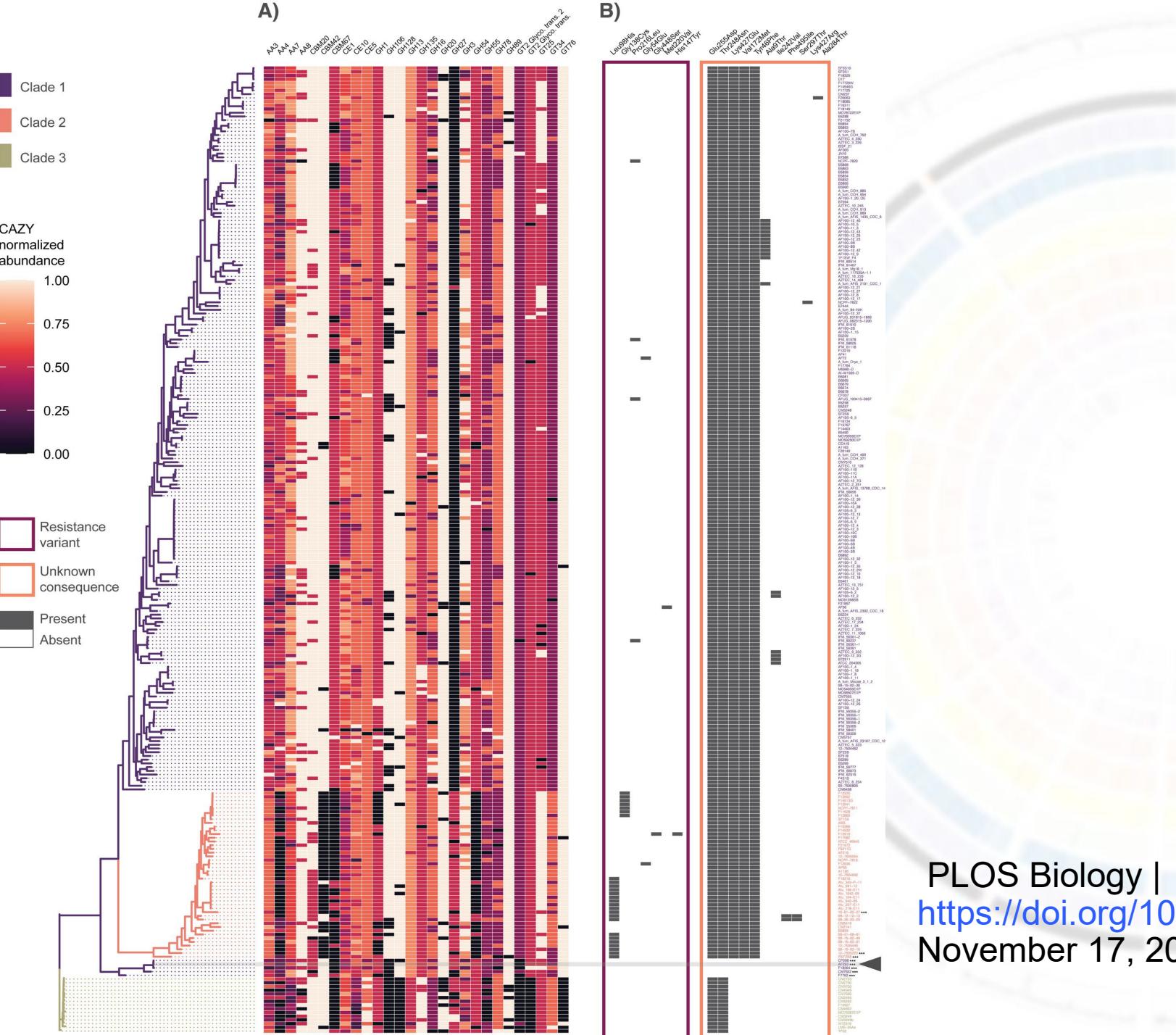
B



Enriquecimento GO (Gene Orthology)

C



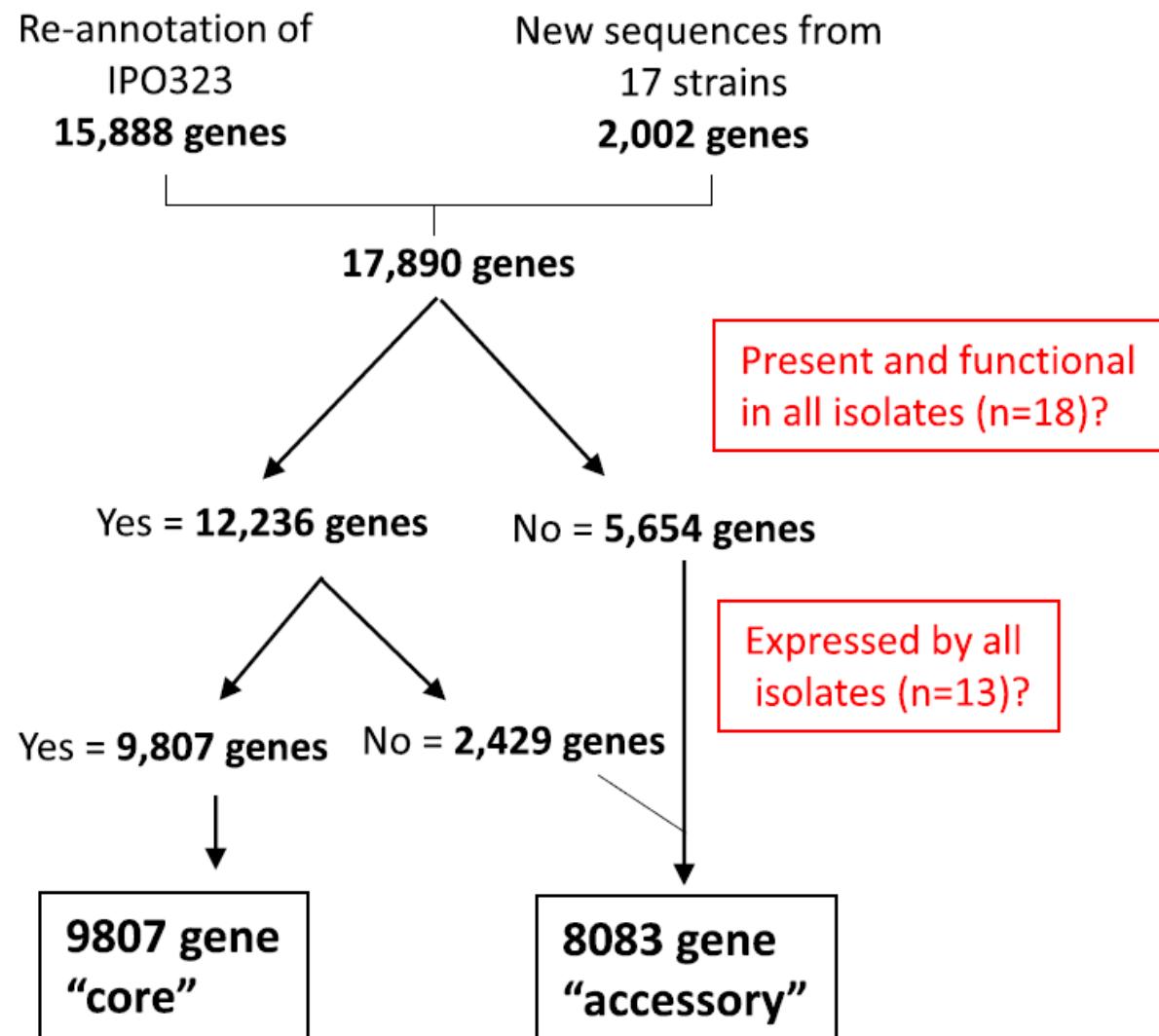


Problema 3

Observação: análise de expressão é importante para espécies patogênicas:

- se um gene central, previsto por meio da análise da sequência de DNA genômico, não for expresso por uma cepa/isolado patogênico bem-sucedido durante a infecção, ele pode ser considerado acessório de forma mais apropriada

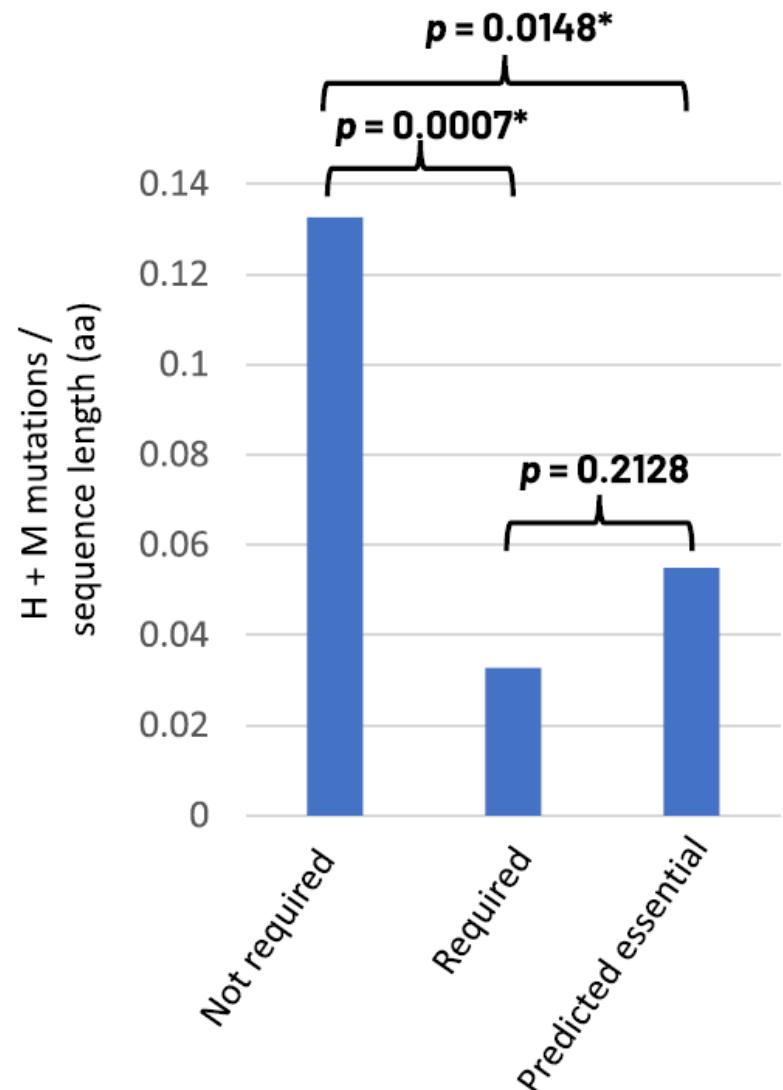
Pangenoma de linhagens europeias de *Zymoseptoria tritici*



Resumo das etapas usadas para gerar o número de genes “centrais” e de genes “acessórios”

Os números representam o número de genes identificados em cada estágio do pipeline

Pangenoma de linhagens europeias de *Zymoseptoria tritici*

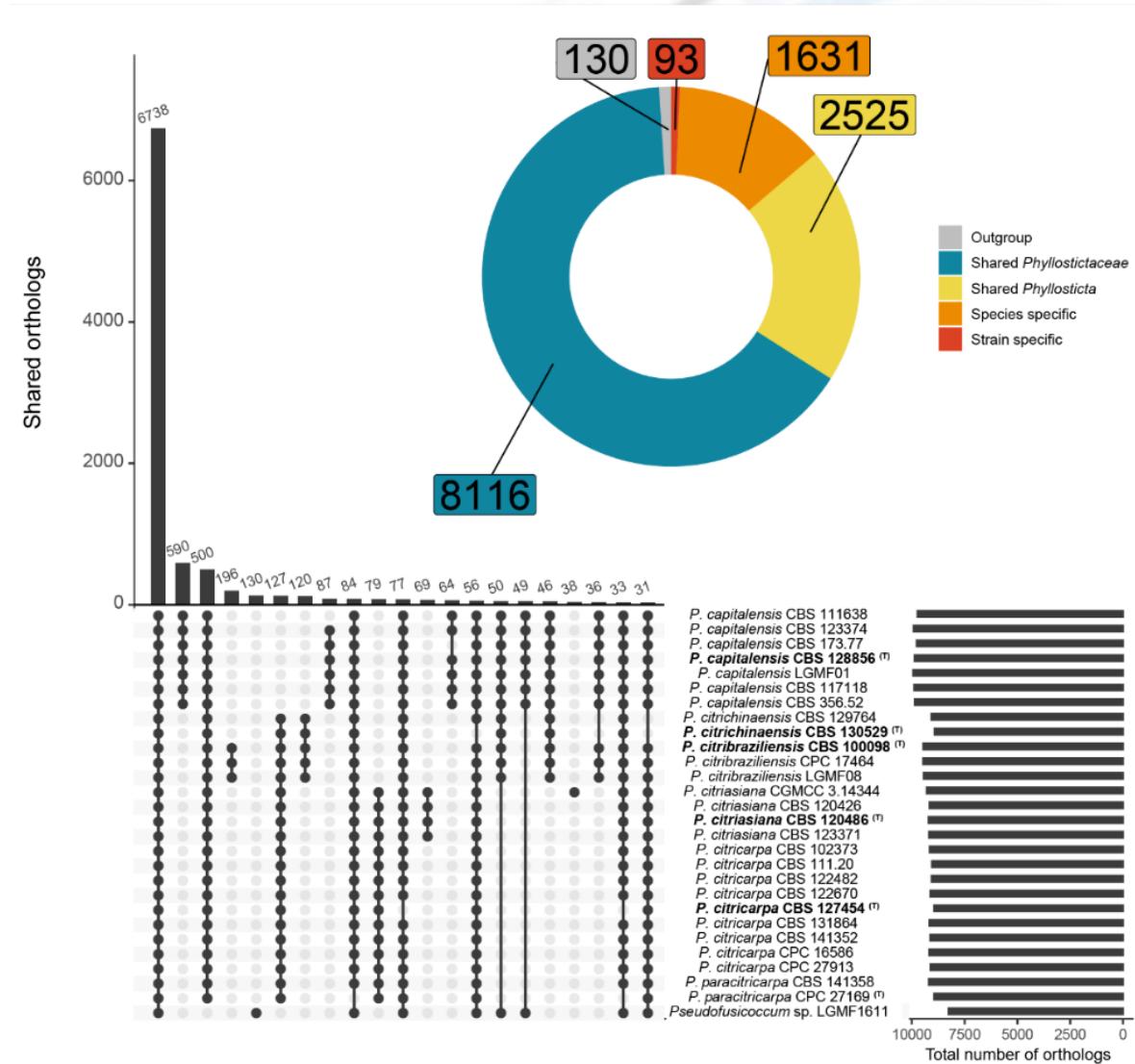


- Análise de conjuntos de genes centrais de patogenicidade validada experimentalmente e/ou genes essenciais:
 - taxas de mutação mais baixas do que as observadas em genes não essenciais

Principais observações

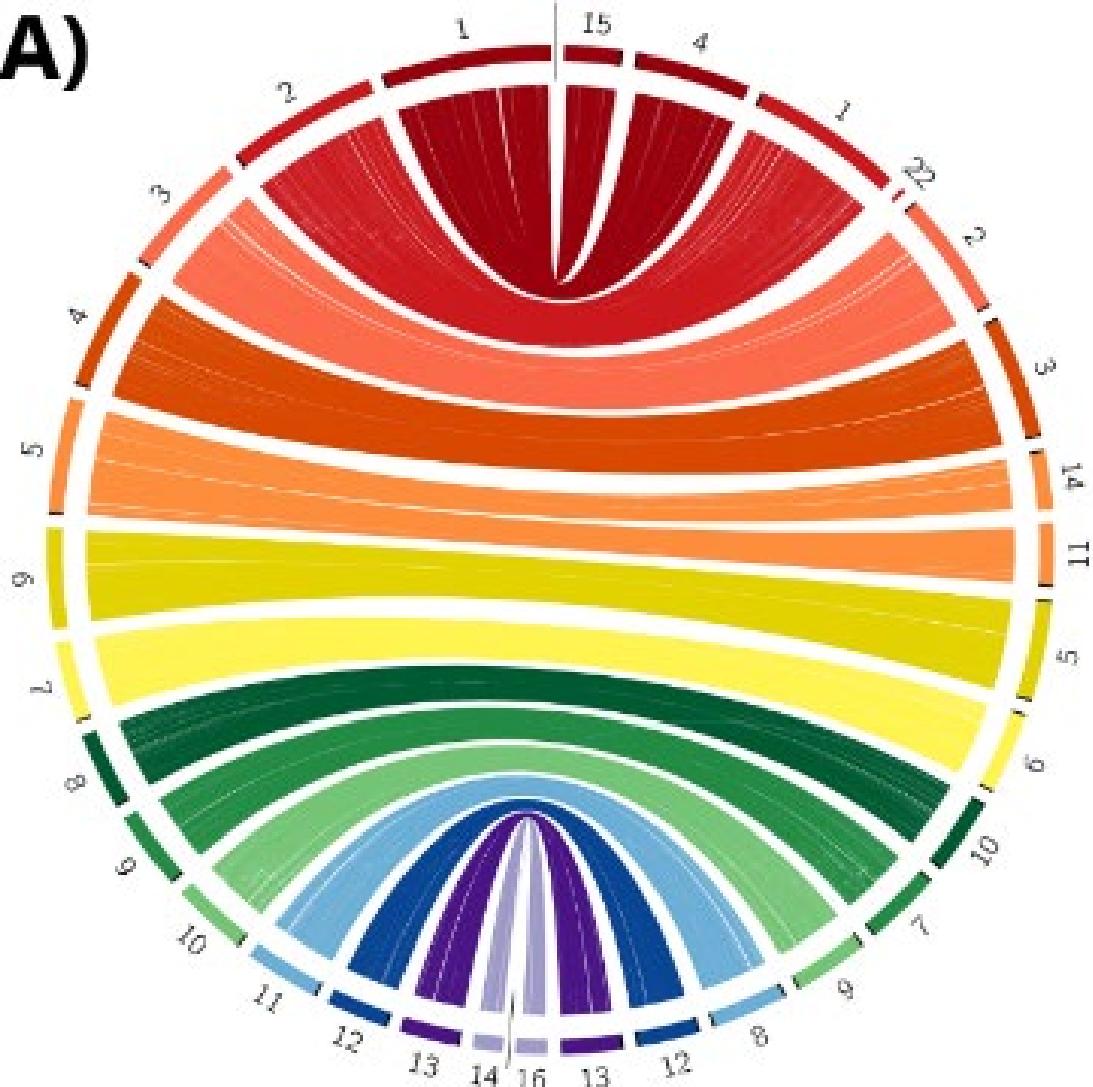
- Predições de localização de proteínas revelam genes no pangenoma acessório que podem funcionar na adaptação a ambientes em mudança
- A expressão de proteínas efetoras no genoma acessório é mais variável entre as linhagens do que aquelas no genoma central
- Os genes do genoma central que codificam proteínas com funções de virulência essenciais ou essenciais têm níveis gerais mais baixos de polimorfismo de aminoácidos
- Pangenômica e transcriptômica combinadas podem ser usadas como armas contra patógenos de rápida evolução

Exemplo: Espécies de *Phyllosticta*

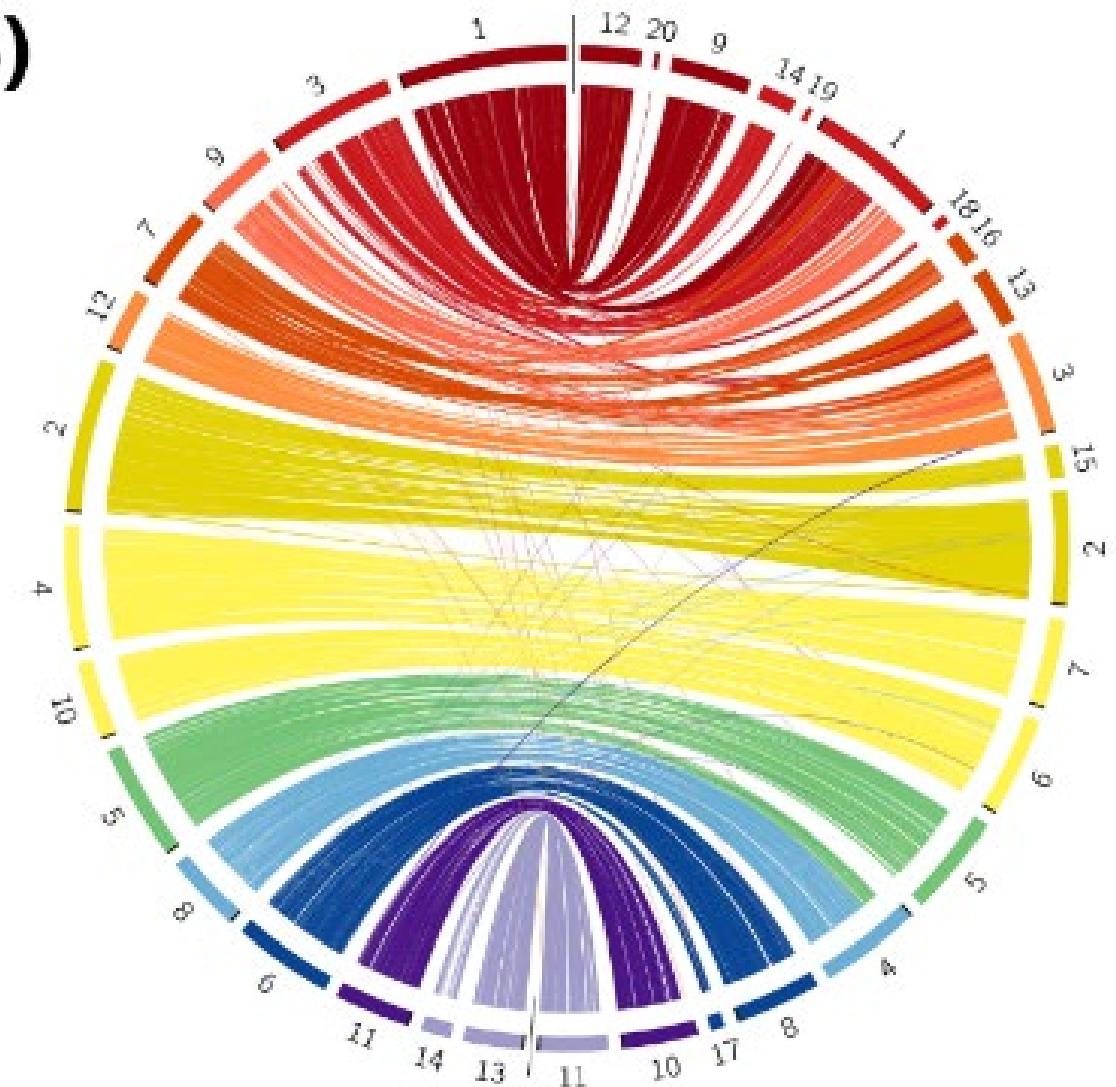


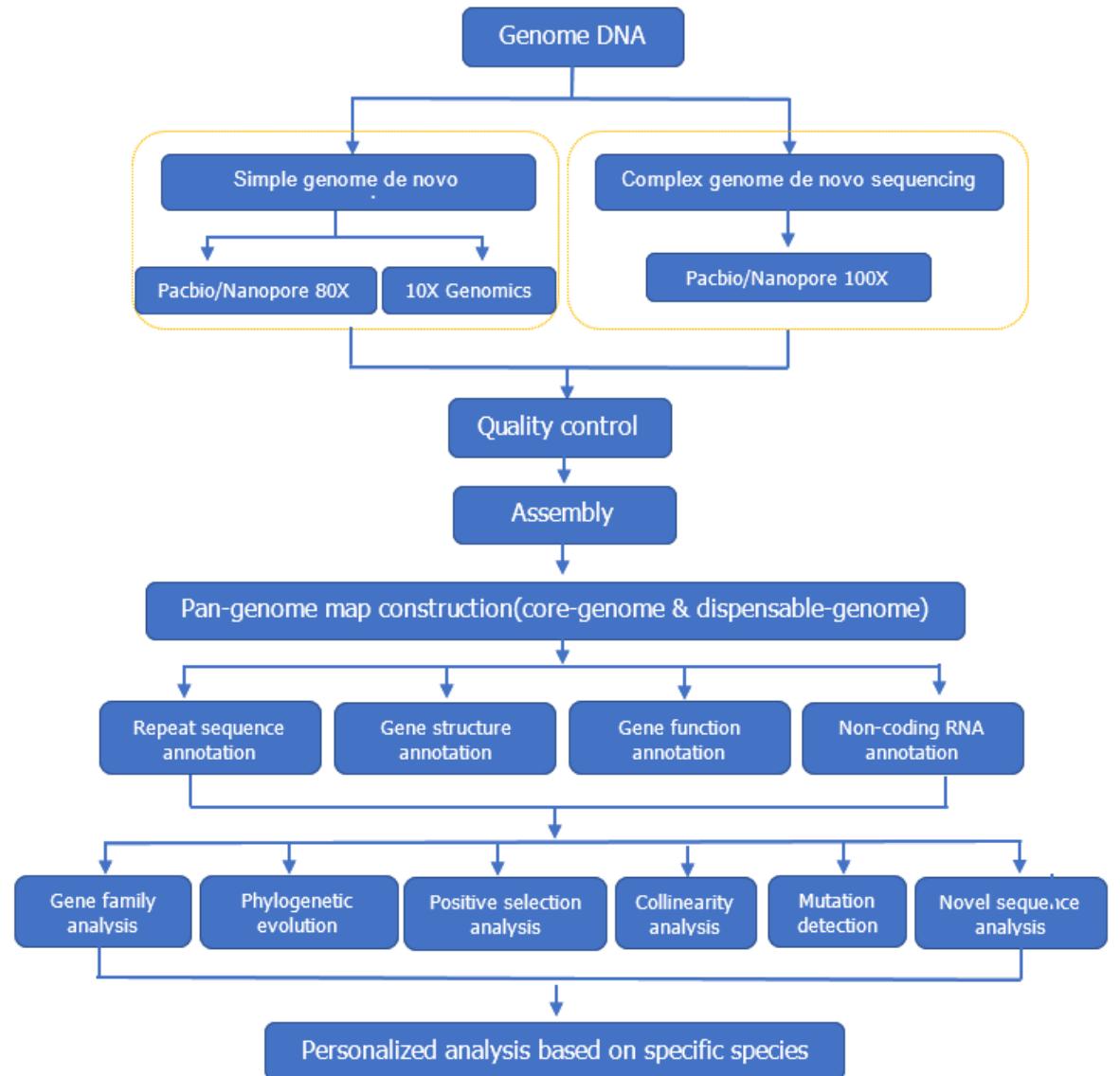
Sintenia - Alto grau de conservação e sintenia em comparações intra- e interespecíficas em espécies de *Phyllosticta* associadas a citros

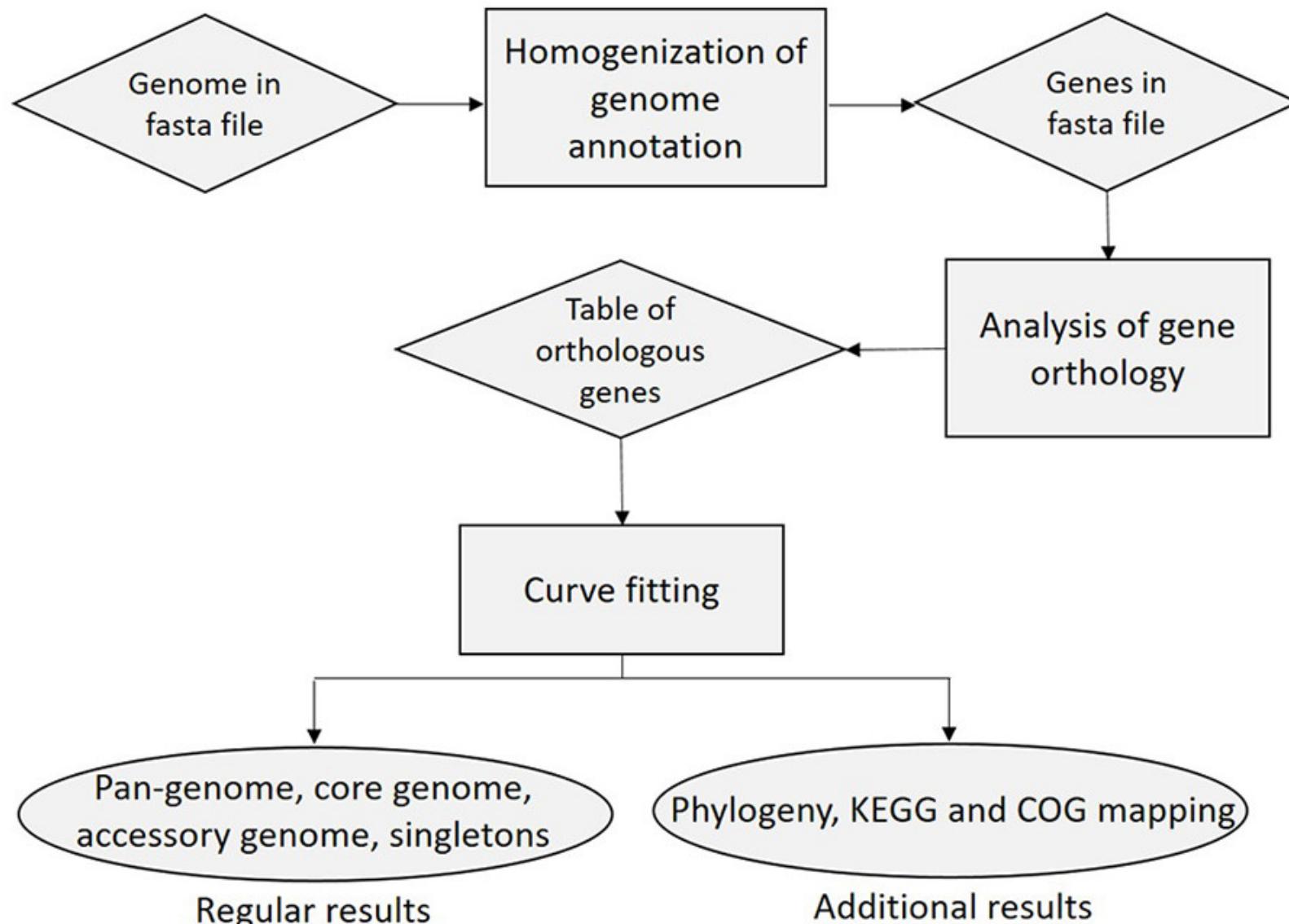
A)



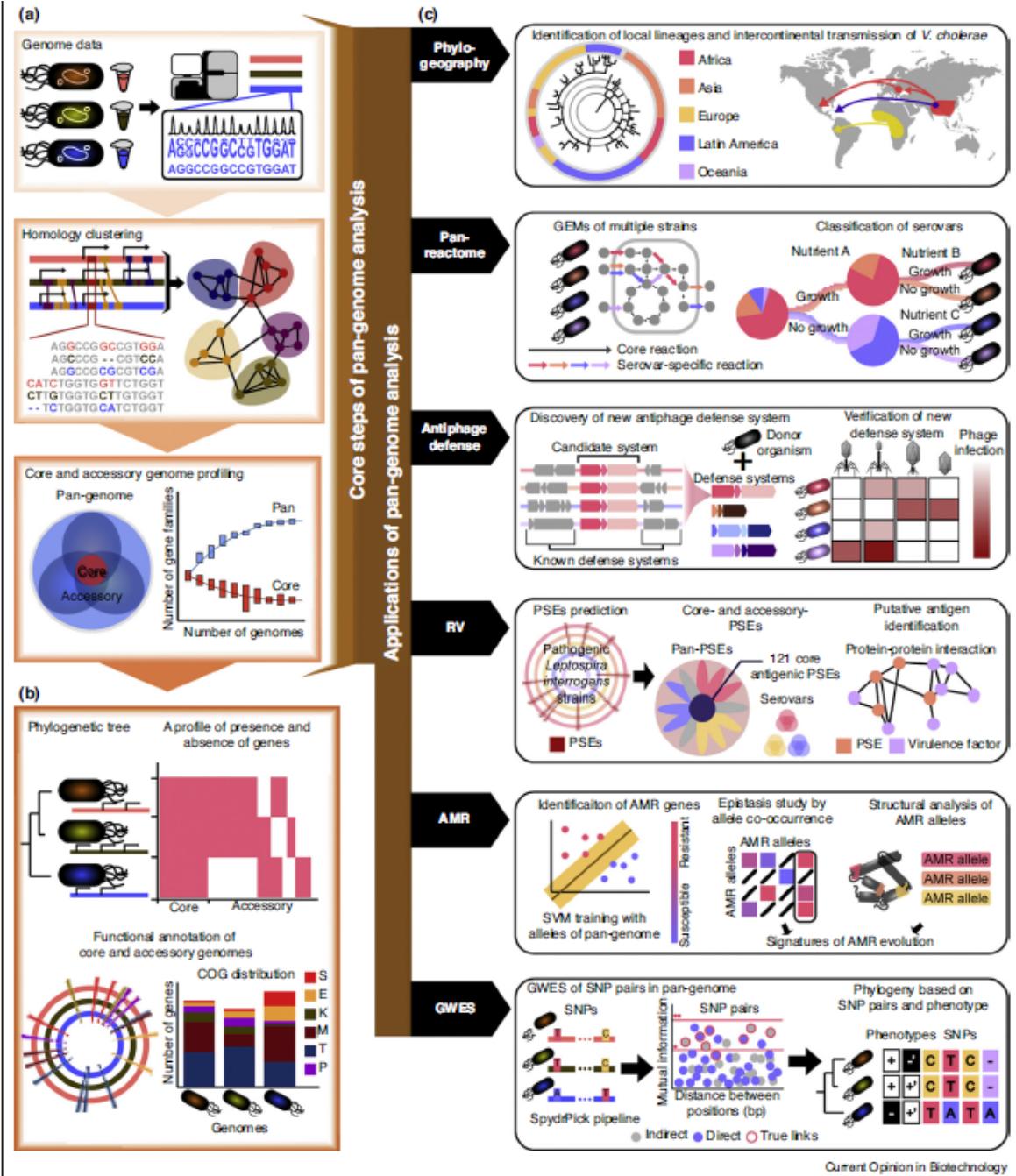
B)







Esquema geral e aplicações da análise pan-genômica



(a) Fluxo de trabalho esquemático de análise pan-genômica, que compreende a preparação de dados do genoma, agrupamento de homologia e perfil de genomas centrais e acessórios

(b) Informação biológica antecipada obtida do pángeno incluindo uma árvore filogenética, presença ou ausência de genes em um cláudo microbiano e anotação funcional de genes centrais acessórios

(c) Aplicações de análise pan-genômica para melhor compreensão e controle de bactérias patogênicas