

Organização de genomas microbianos

Profa. Dra. Chirlei Glienke

BIOGEMM-UFPR

GS Treinamentos e Consultoria

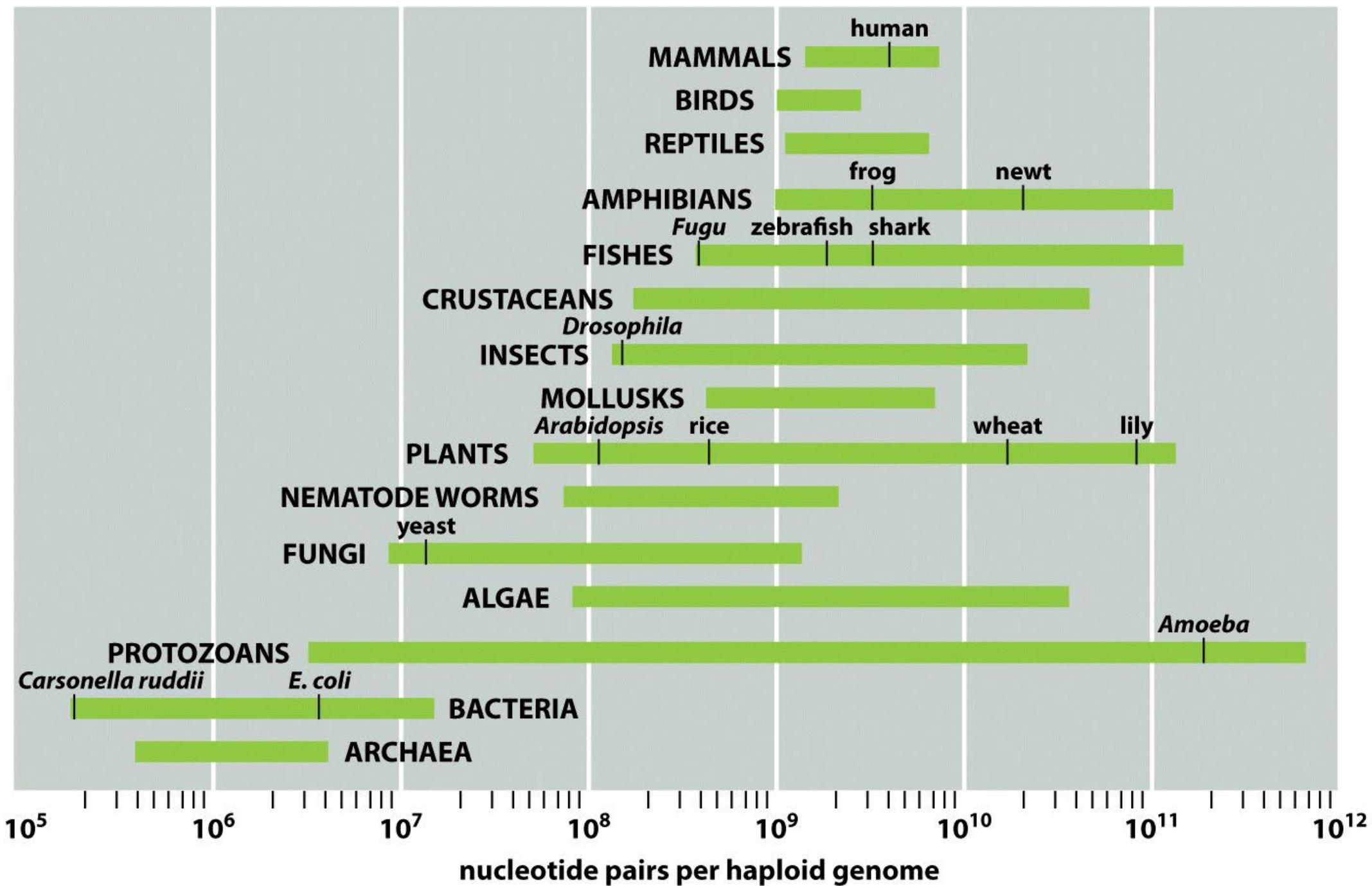
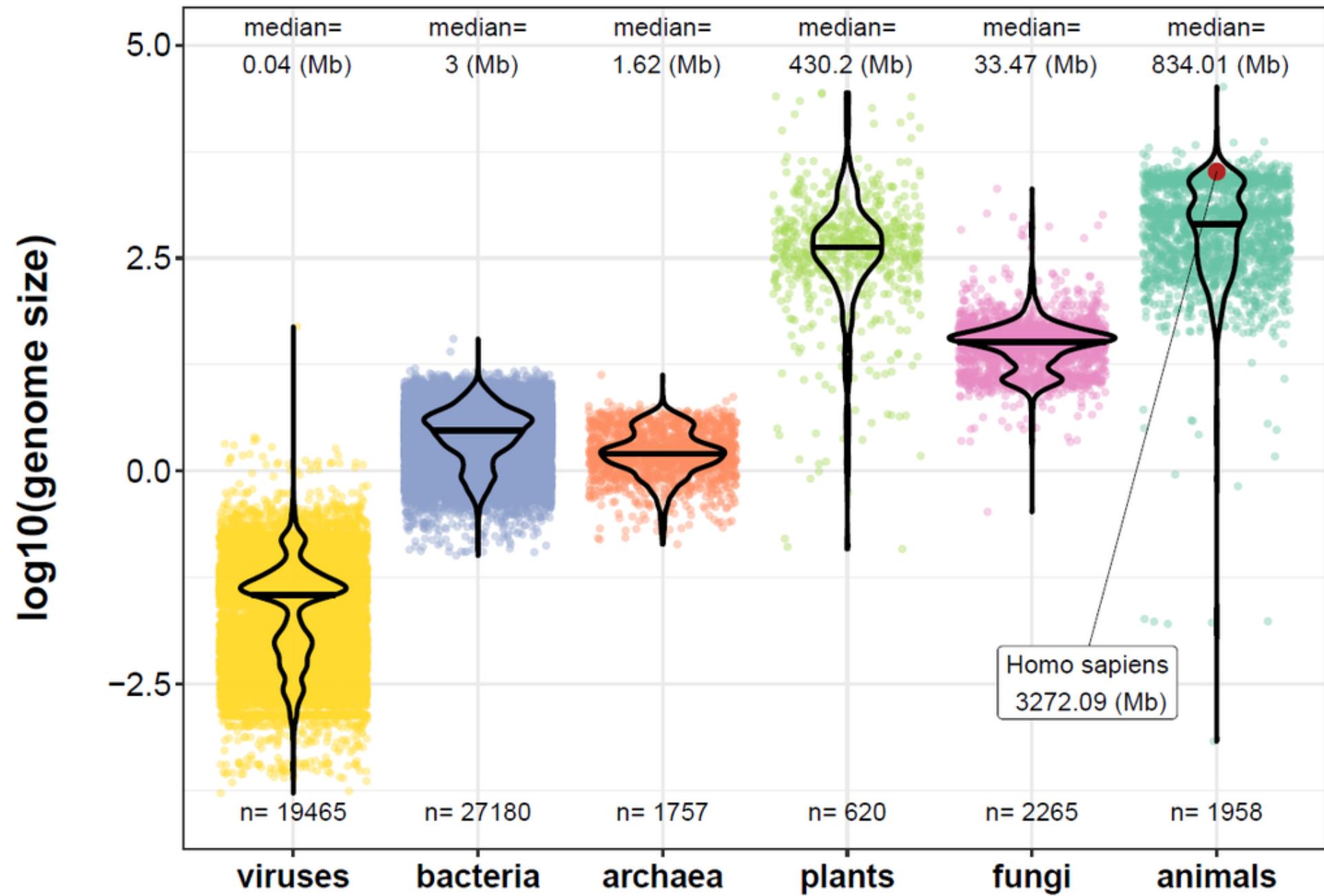


Figure 1-41 Essential Cell Biology 3/e (© Garland Science 2010)



Estrutura cromossômica e genoma de procariôntes

Estruturas cromossômicas em procariontes

- Procariontes são MONOPLÓIDES (tem apenas um conjunto de genes)
- A maioria dos procariontes estoca seu conjunto único de genes em um só cromossomo circular
- O DNA acumula-se no nucleóide

Cromossomo bacteriano

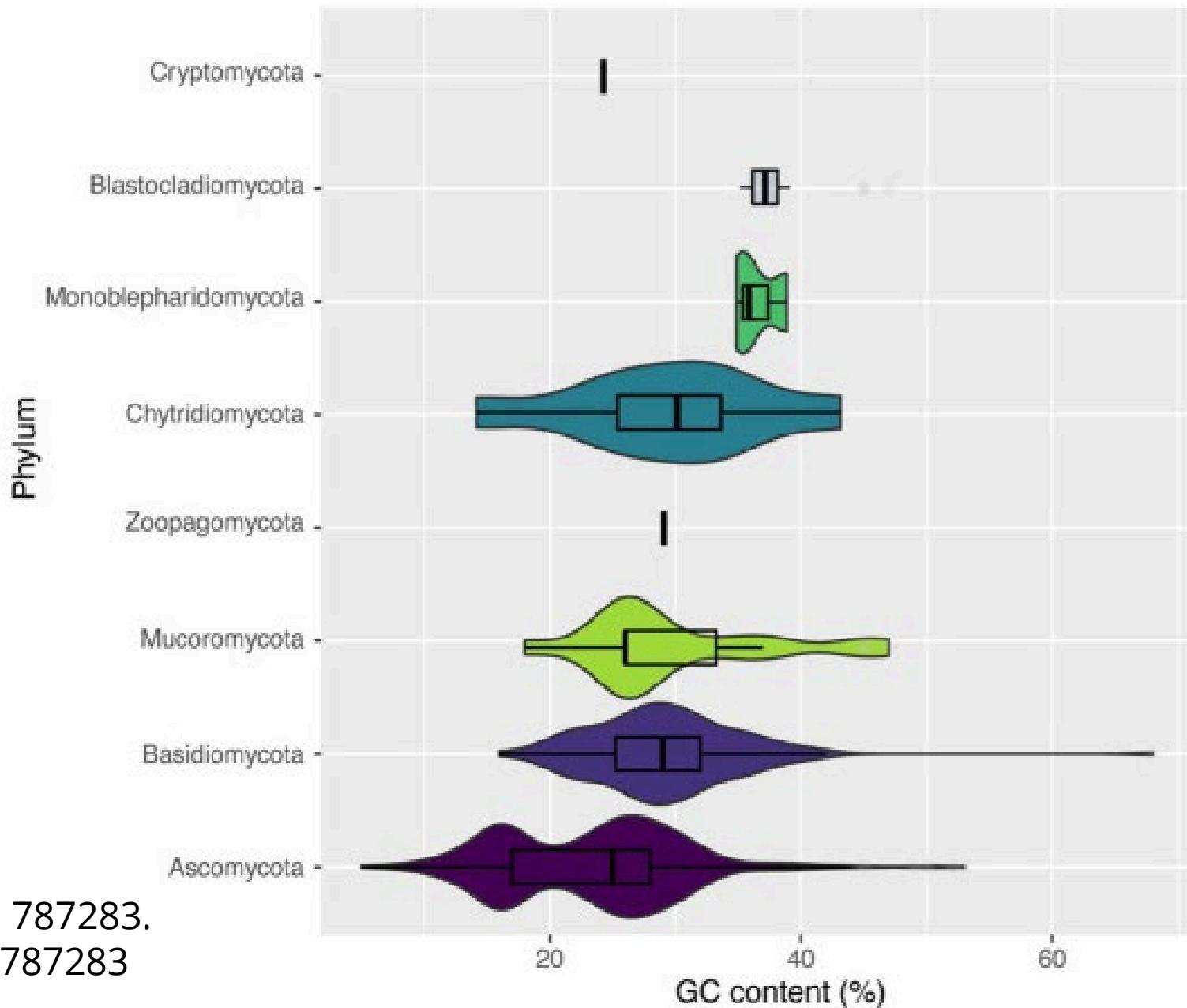
- Varia em tamanho: (~1.000 a 10.000 genes)
 - ~ 600 à 750 kb em *Mycoplasma* (482 CDS) - ~31.0 % de GC
 - ~ 5.000 kb em *Escherichia coli* (~3.000 genes) - 50.8 % de GC
 - ~ 8.000 – 9.000 kb em *Streptomyces* (6.994 genes) (actinobacteria)
 - ~ 72 % de GC

Comparativamente:

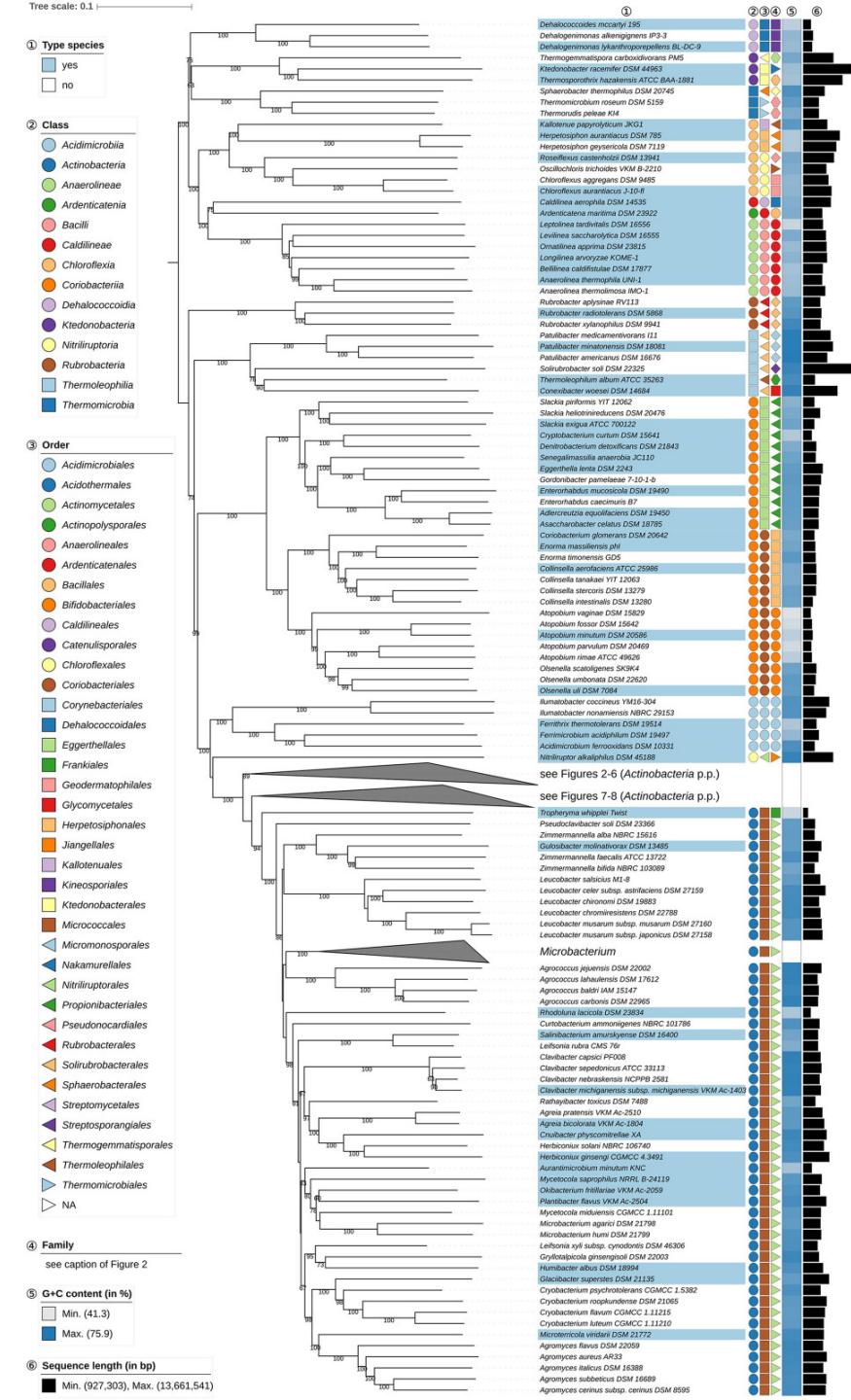
Saccharomyces cerevisiae – (12 Mb) 38.3 % de GC

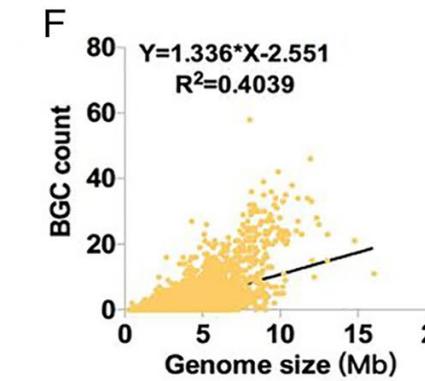
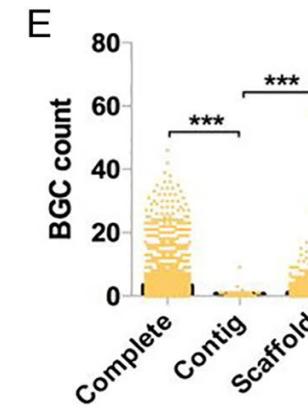
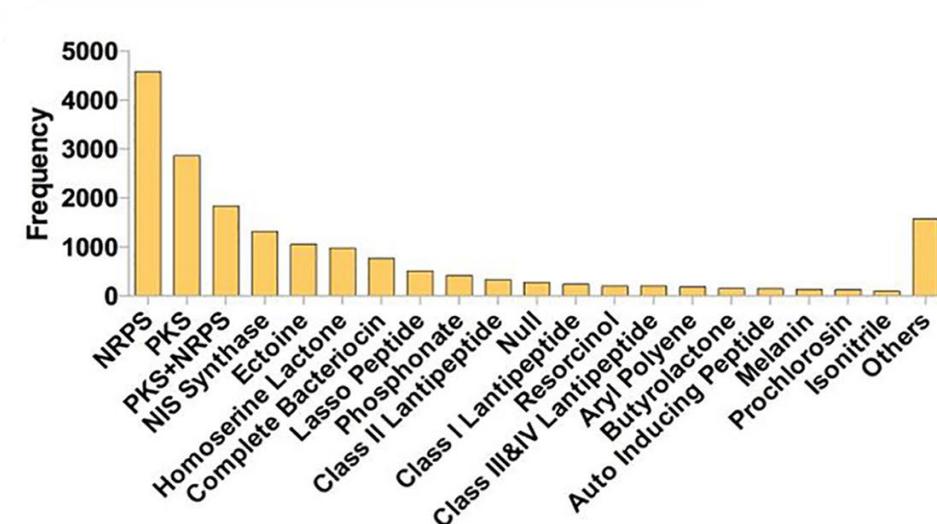
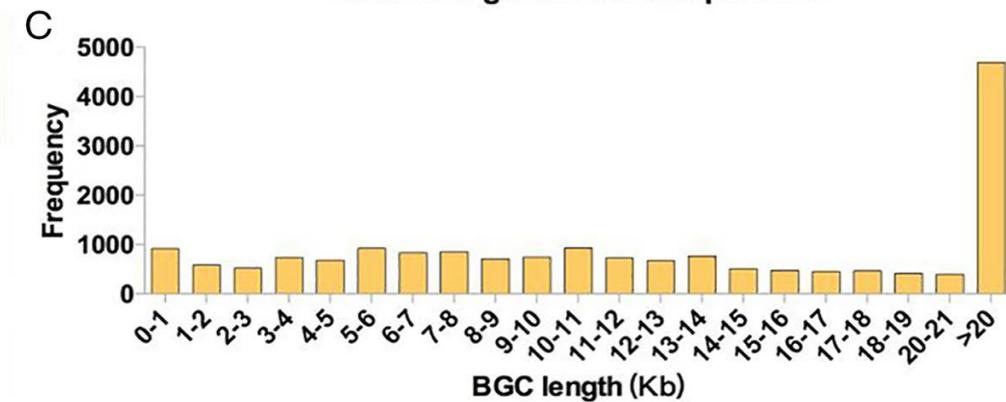
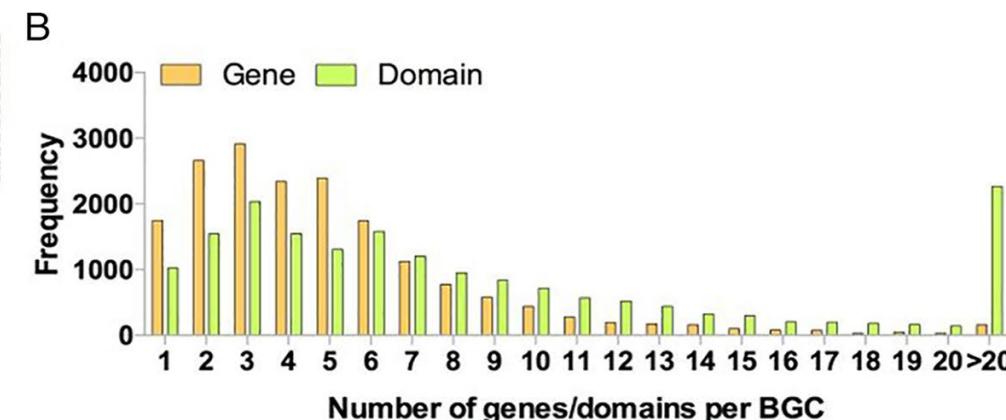
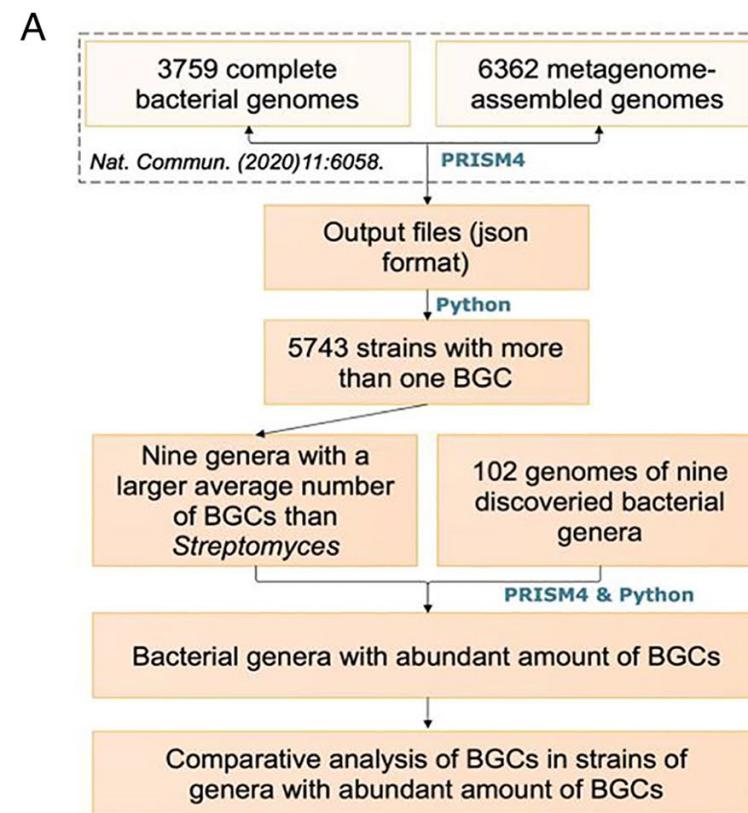
Fungos Filamentosos (30/50 - 180 Mb) – 47 – 55 % de GC

Genes mitocondriais de fungos – 31% a 35% de GC



Genome-Based Taxonomic Classification of the Phylum Actinobacteria.

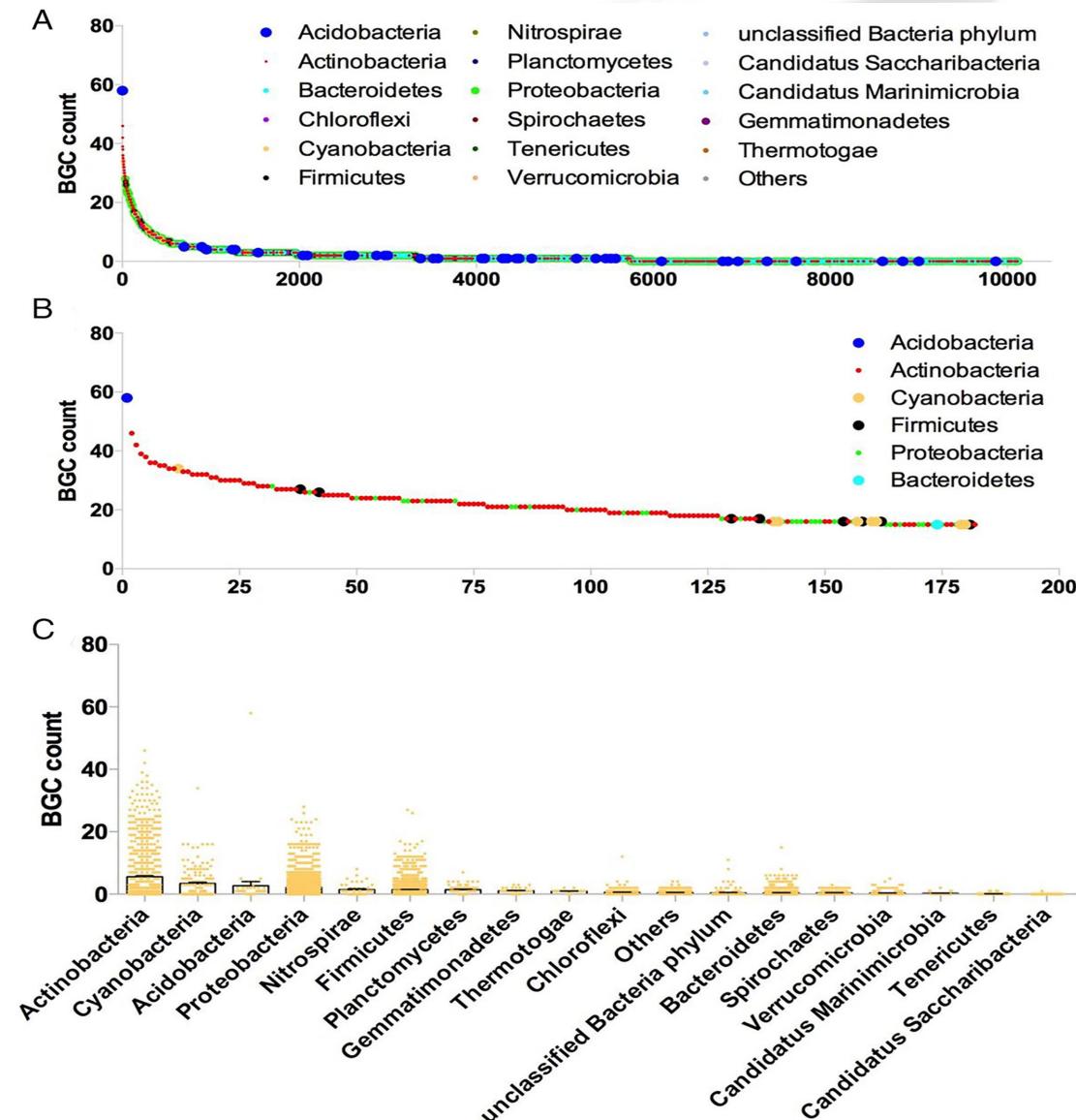




An atlas of bacterial secondary metabolite biosynthesis gene clusters

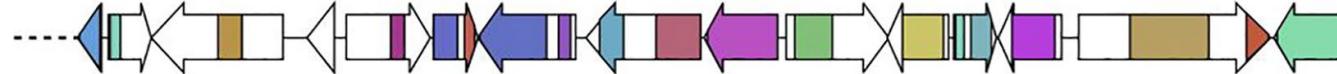


An atlas of bacterial secondary metabolite biosynthesis gene clusters

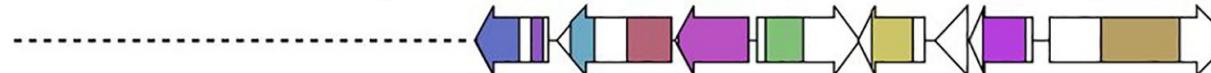




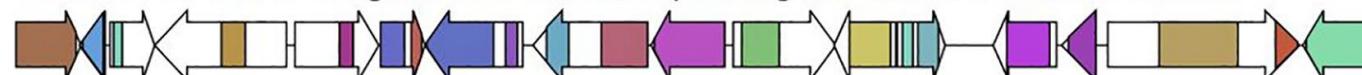
FWXV01000001.1.region005: *Kibdelosporangium aridum* strain DSM 43828



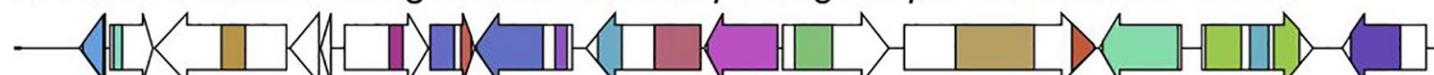
JNYM01001386.1.region001: *Kibdelosporangium aridum* subsp. *largum* strain NRRL B-24462



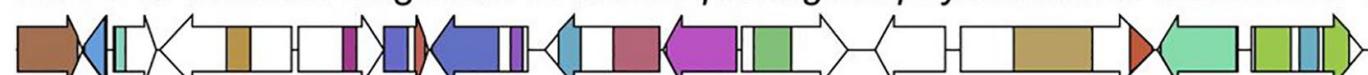
QHKI01000002.1.region001: *Kibdelosporangium aridum* strain A82846



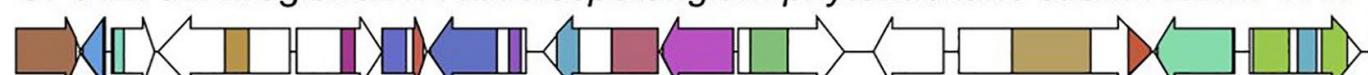
JAAATY010000015.region001: *Kibdelosporangium persicum* strain 4NS15



JADBEI010000002.region012: *Kibdelosporangium phytohabitans* strain DSM 104448



CP012752.1.region021: *Kibdelosporangium phytohabitans* strain KLBMP1111

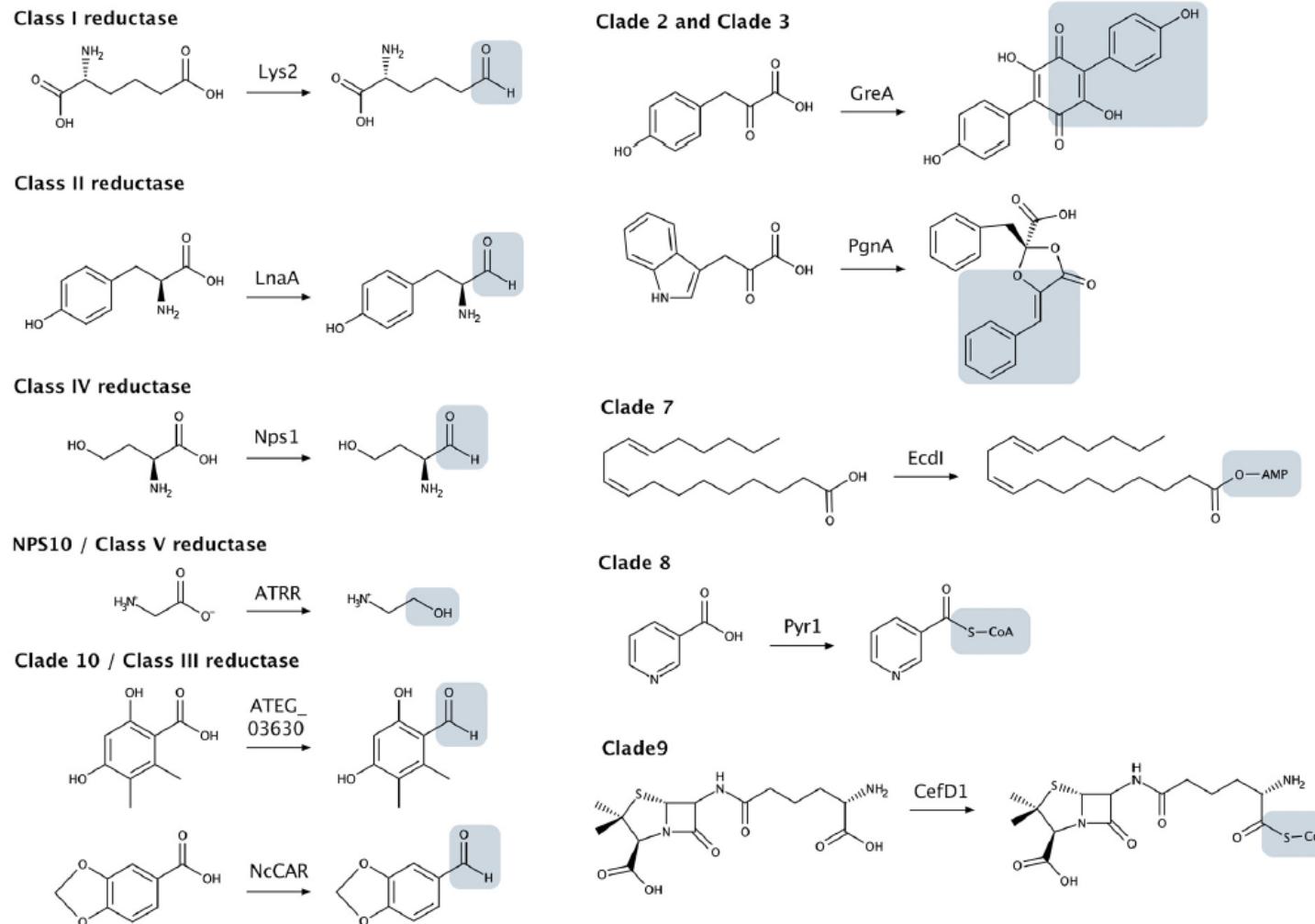


MTQO01000028.1.region001: *Actinosynnema* sp. ALI-1.44



CDME01000011.1.region007: *Kibdelosporangium* sp. MJ126-NF4

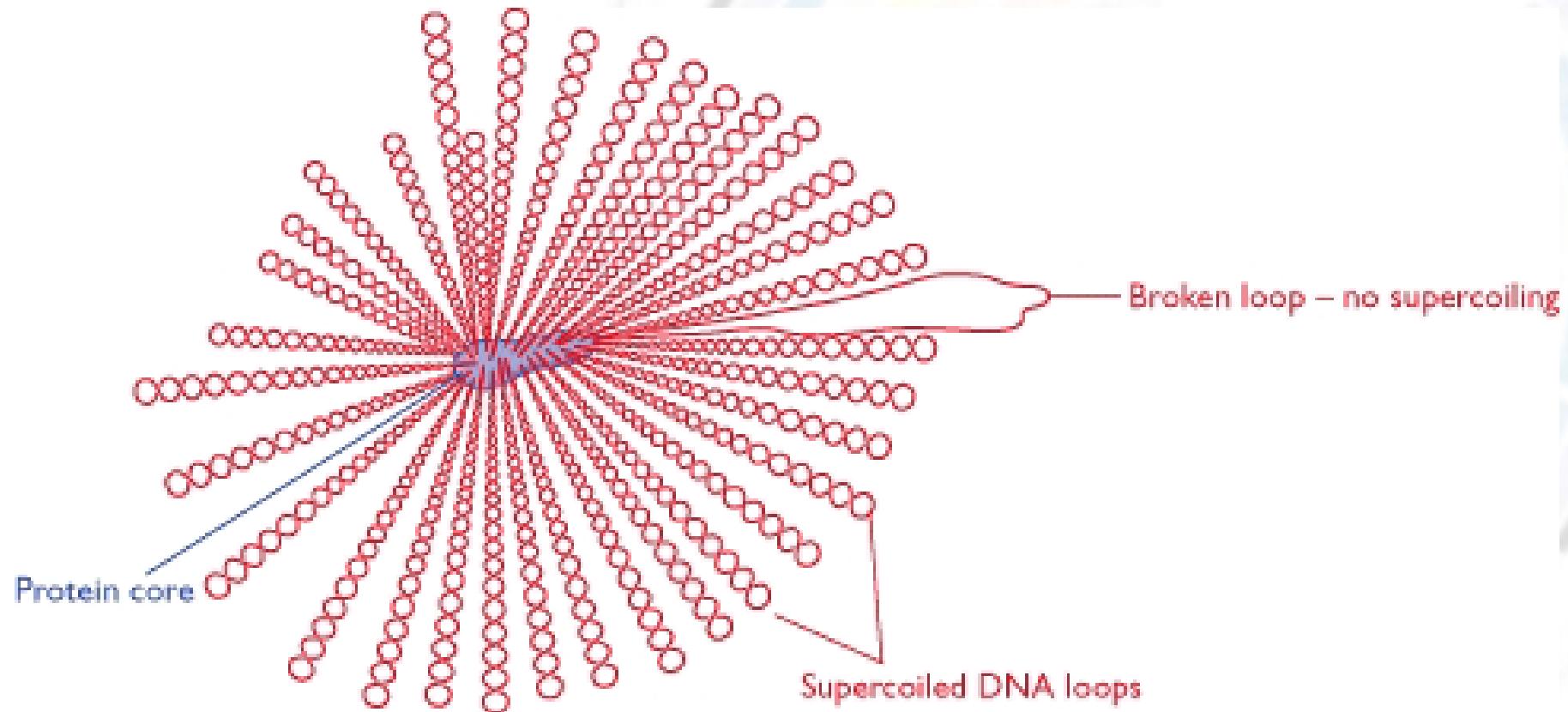




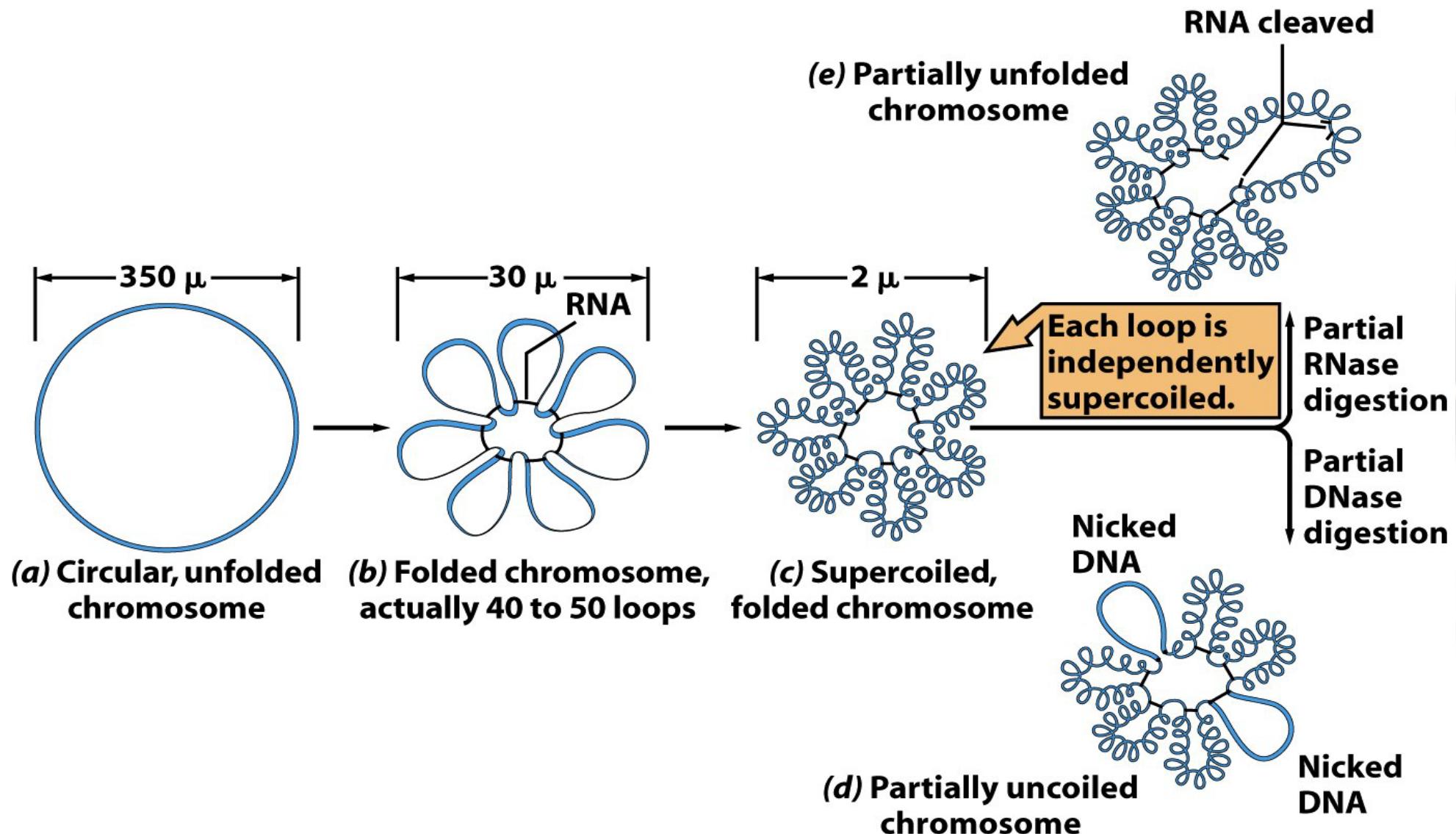
Genoma compacto

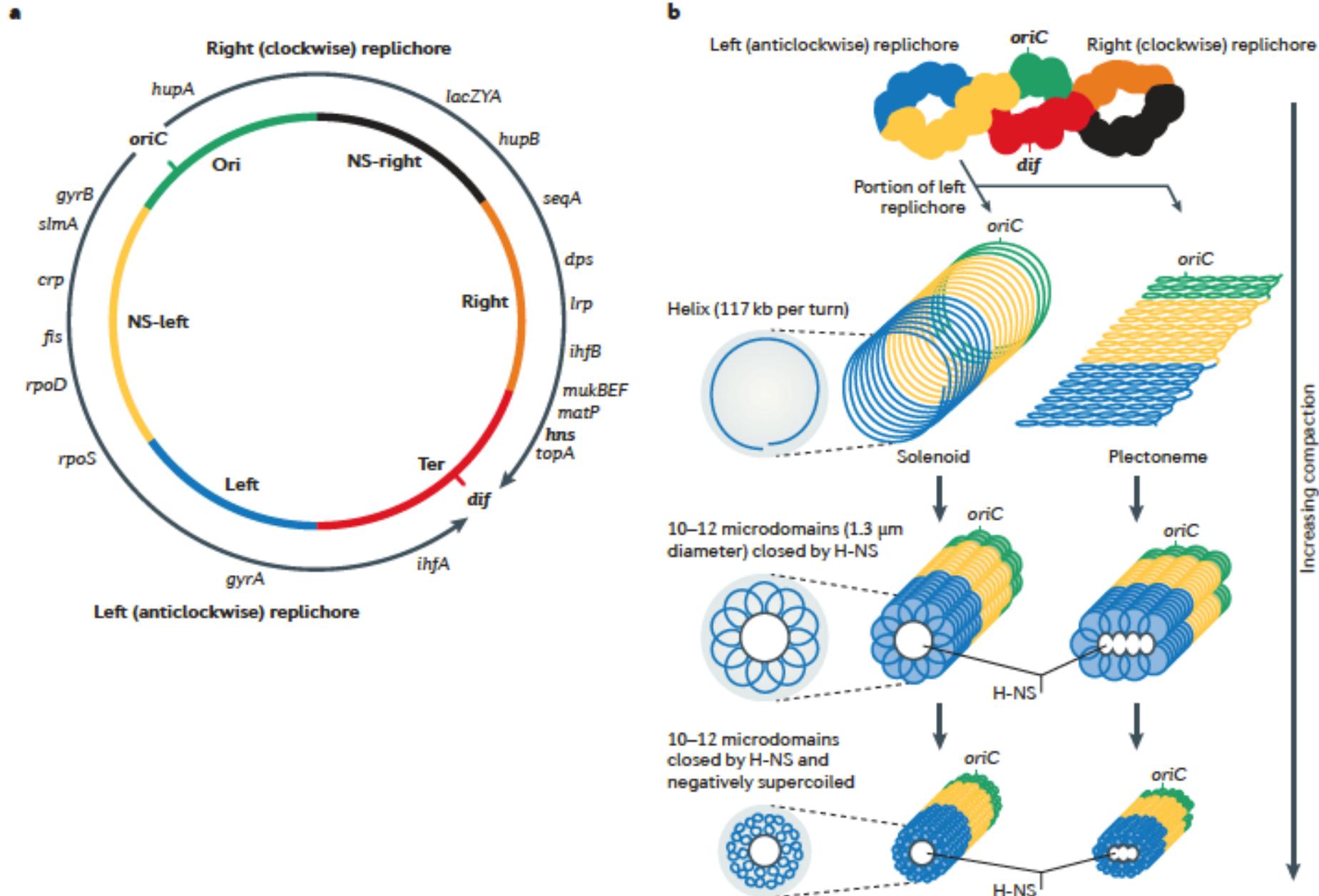
- Estado funcional
- 50 a 100 domínios ou alças (cada uma super-helicoidizada negativamente de modo independente)
- RNA (conectores) e proteínas são componentes do genoma compactado no nucleóide mas ~80% é DNA
- Genoma pode ser parcialmente relaxado por DNase e/ou RNase e/ou agentes que atuem nas proteínas
- As proteínas que atuam na condensação do DNA ainda não são totalmente conhecidas

MODELO DE ARRANJO DE DNA PROCARIÓTICO



Estrutura do estado funcional do cromossomo de *E. coli*





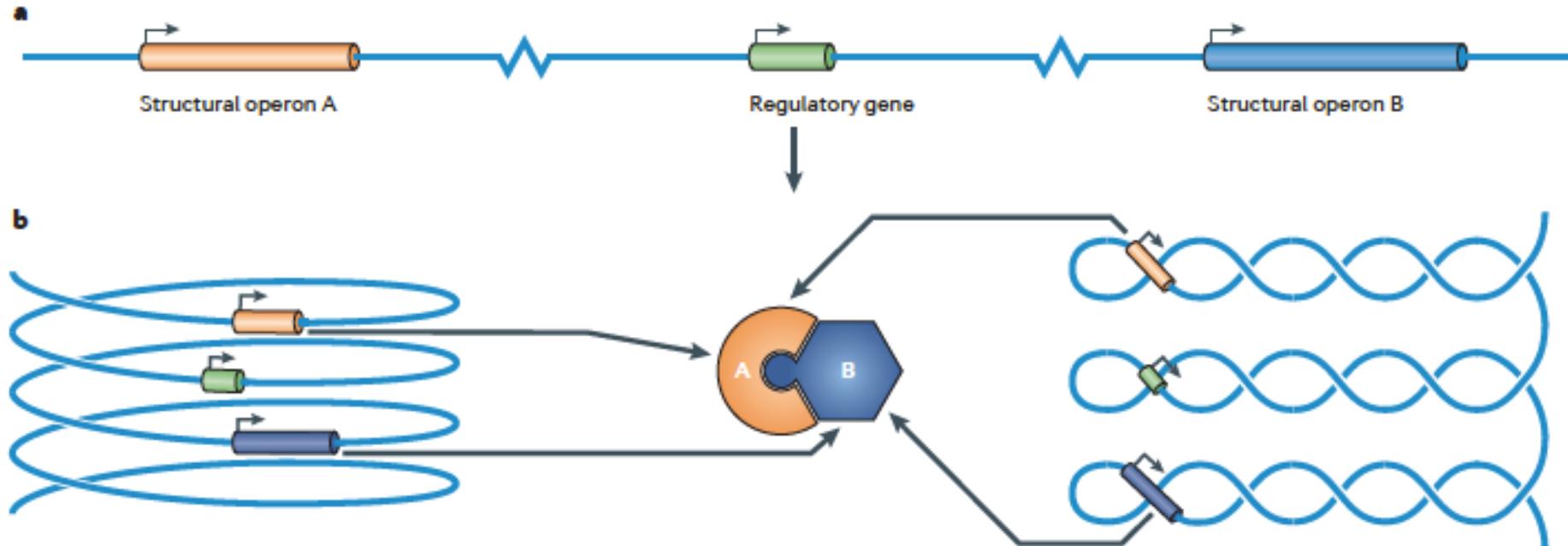


Figure 2 | Nucleoid folding and gene regulation. A simple regulon consisting of a regulatory gene and two structural operons, A and B, is illustrated in various conformations. **a** | When the chromosome is represented in a one-dimensional, linear form, the three genetic loci are separated by large distances in space. **b** | However, when the chromosome is reorganized

as a solenoid (left) or as a plectoneme (right), the periodicity of these structures brings the three genes close together, facilitating communication between the regulatory gene and its two target operons. Moreover, the products of the A and B operons are produced in close proximity, favouring their interaction.

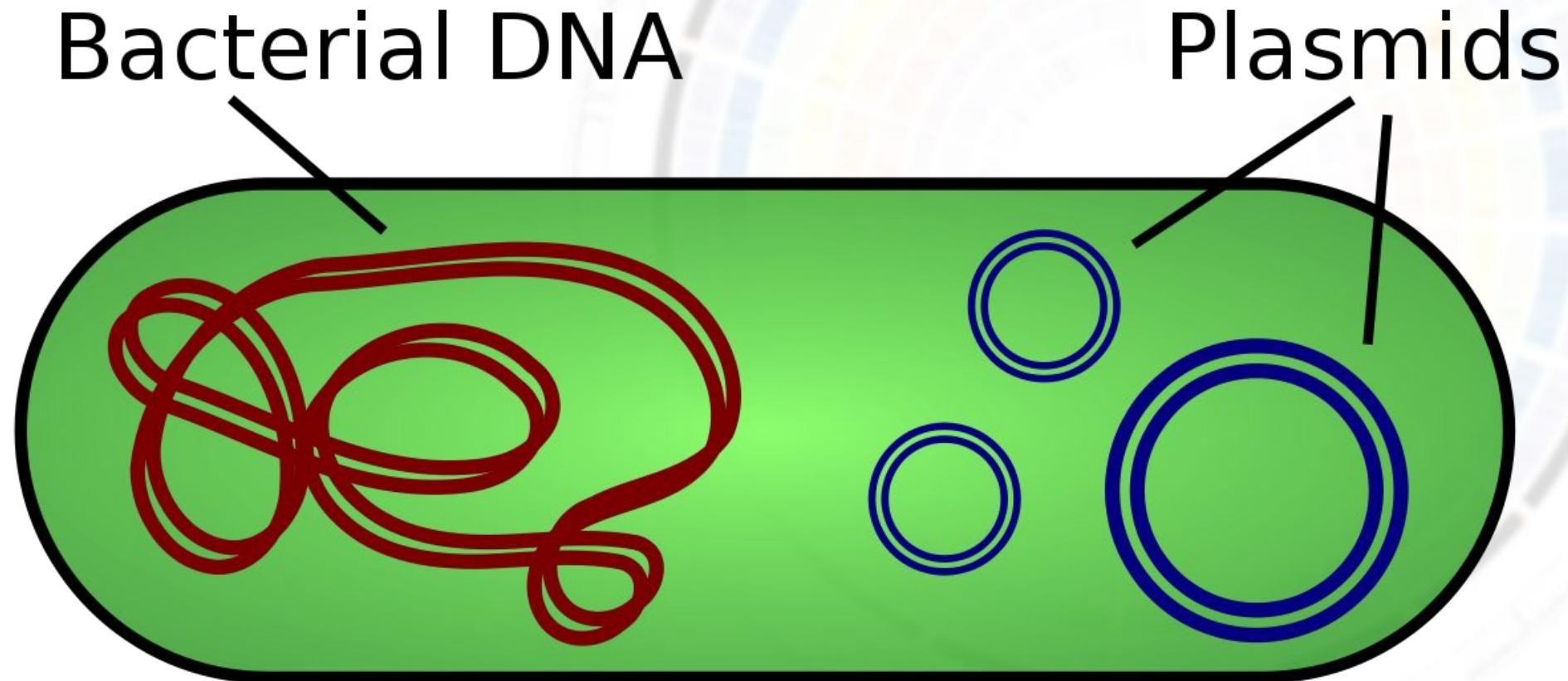
Cromossomo bacteriano

- Presença de proteínas associadas, semelhantes às histonas, envolvidas no empacotamento do cromossomo e regulação de transcrição

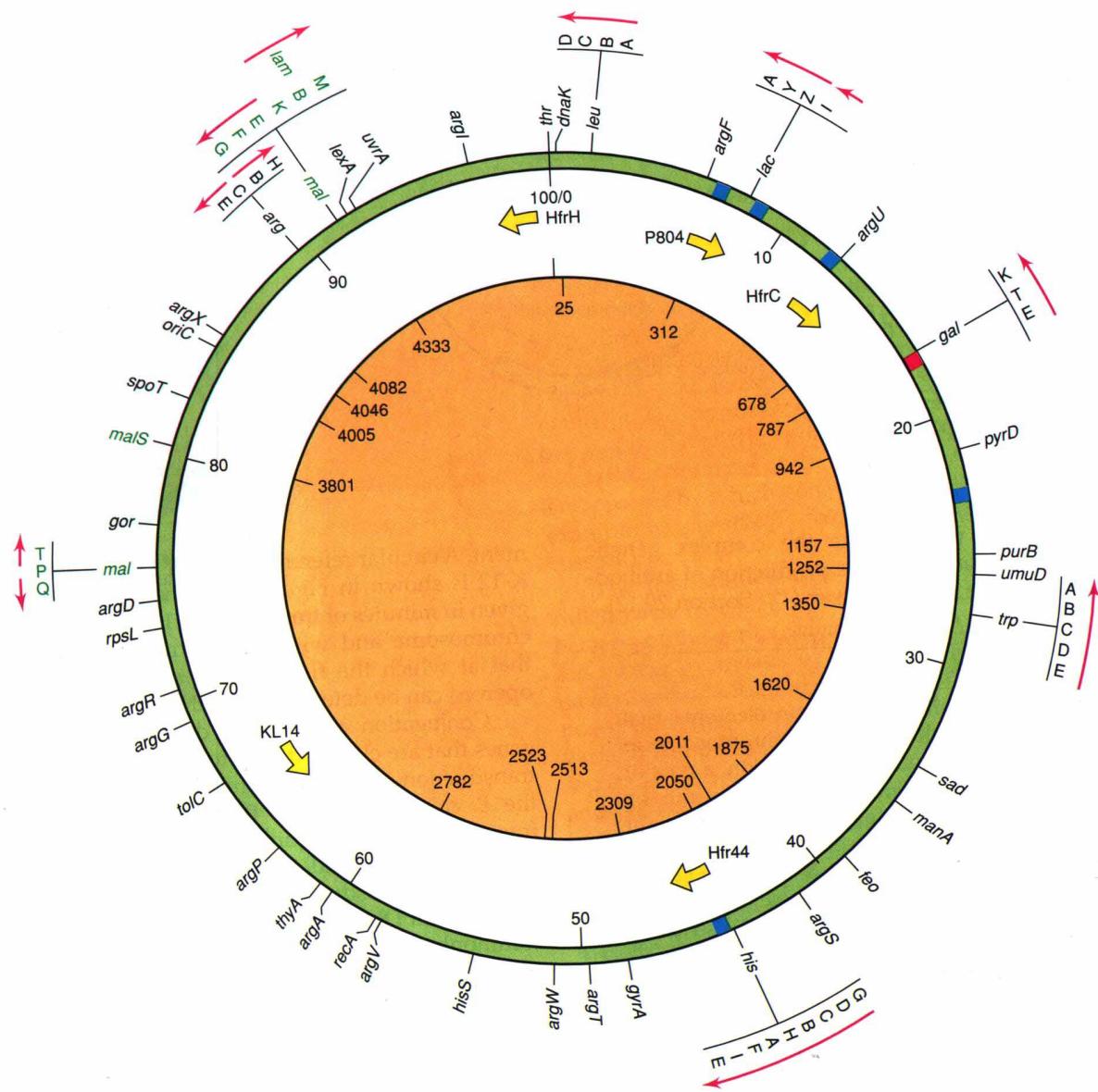
Proteínas que se ligam ao DNA (*E. coli*)

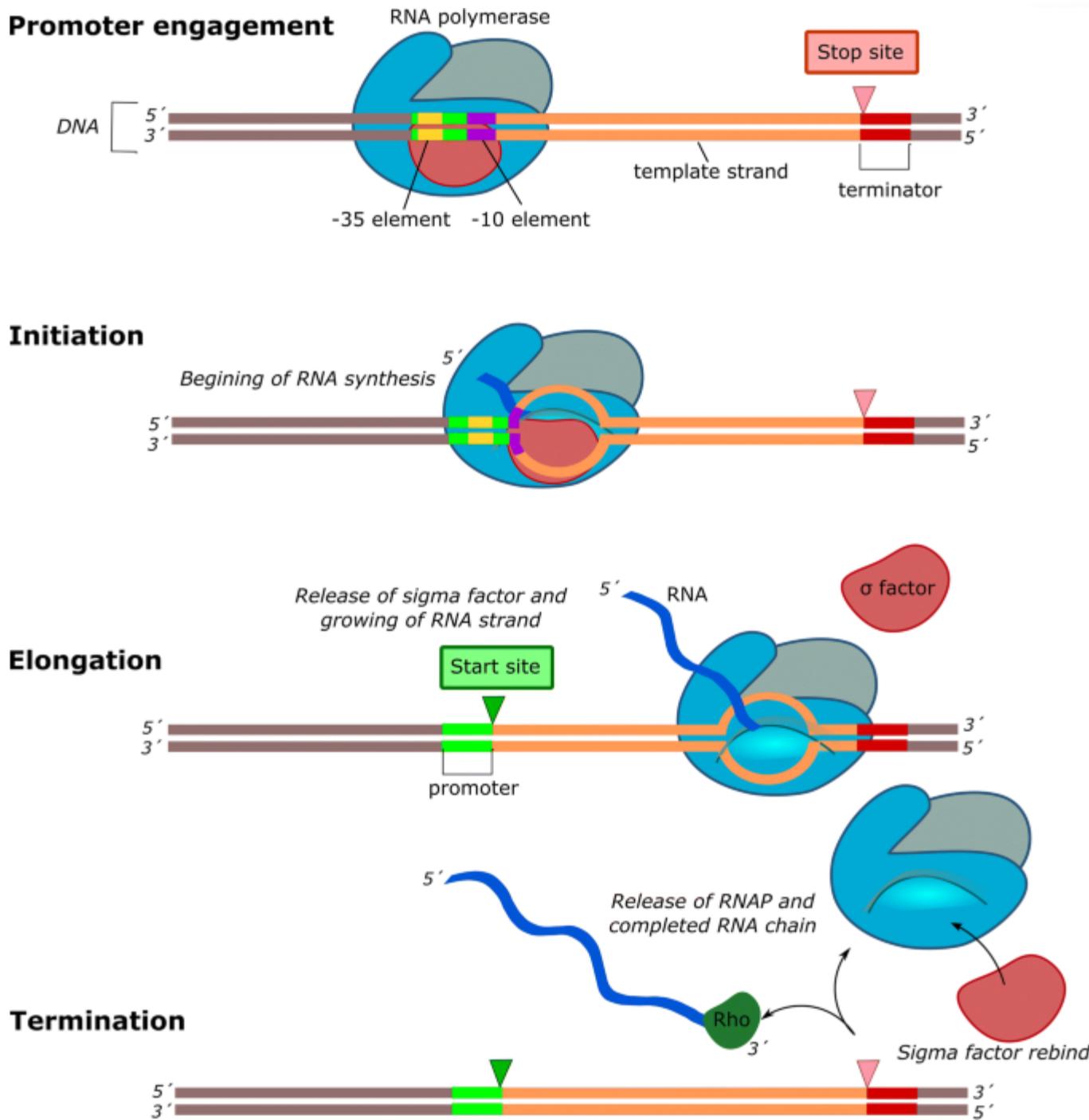
Proteína	Composição	No./ célula	Semelhança
HU	Dímero: α e β (9 kDa)	40 000	Histona H2B
H	Dímero (28 kDa)	30 000	Histona H2A
IHF	Dímero: α e β (10,5 e 9,5 kDa)	?	HU
H1/H-NS	Monómero (15 kDa)	10 000	?
HLP1	Monómero (17 kDa)	20 000	?
P	Monómero (3 kDa)	?	Protaminas

E. coli



Cromossomo de *E. coli*





- 10 Schaller-Pribnow box (purple rectangle) TATAAT
- 35 element (yellow box) TTGACA
- sigma factor (reddish oval)
- promoter (green rectangle)

Transcrição

- Decidir: o que deve ser transcrito?
- RNA polimerase + Fator Sigma
 - Duas principais famílias:
 - σ^{54} (RpON): genes associados ao metabolismo do nitrogênio
 - σ^{70} (RpOD or σ^A): demais genes

reconhecimento da região promotora

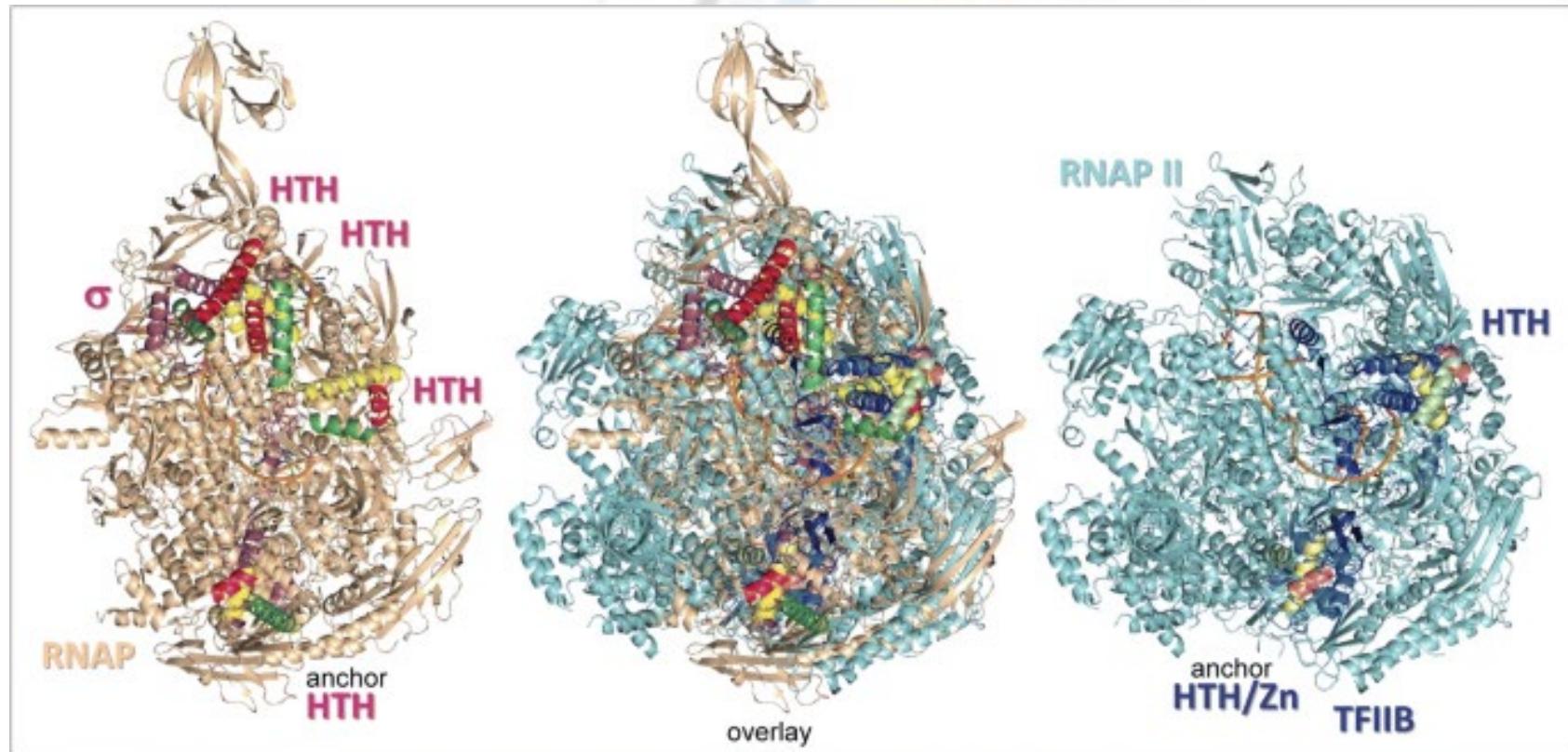
- -10 Schaller-Pribnow box TATAAT
- -35 element (yellow box) TTGACA

abertura da fita de DNA

- Pela região -10 TATAAT

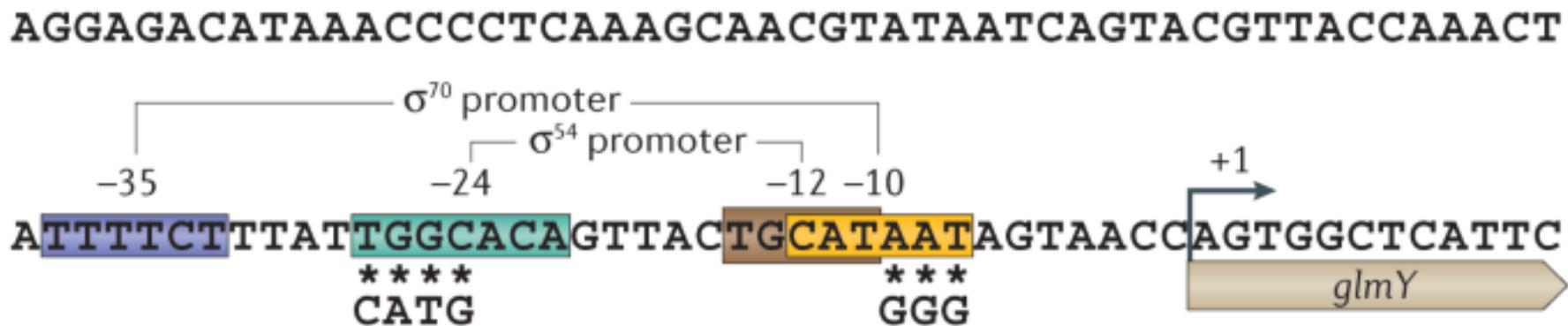
alongamento: 5'- 3' (fita sense)

Comparação entre RNAP (procariótica) e RNAPII (levedura) e a sobreposição

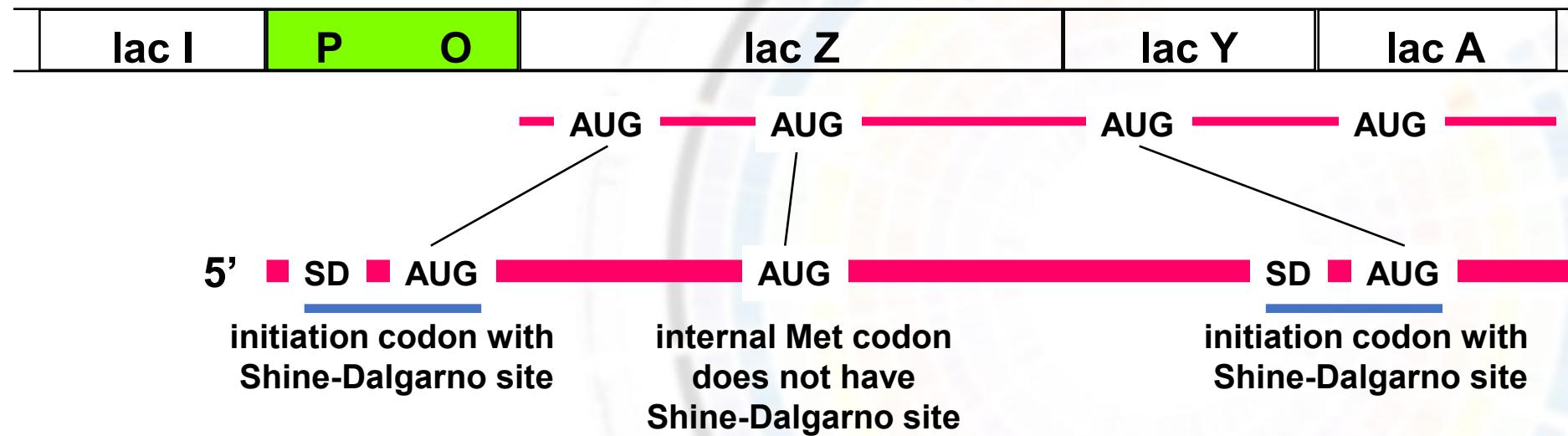


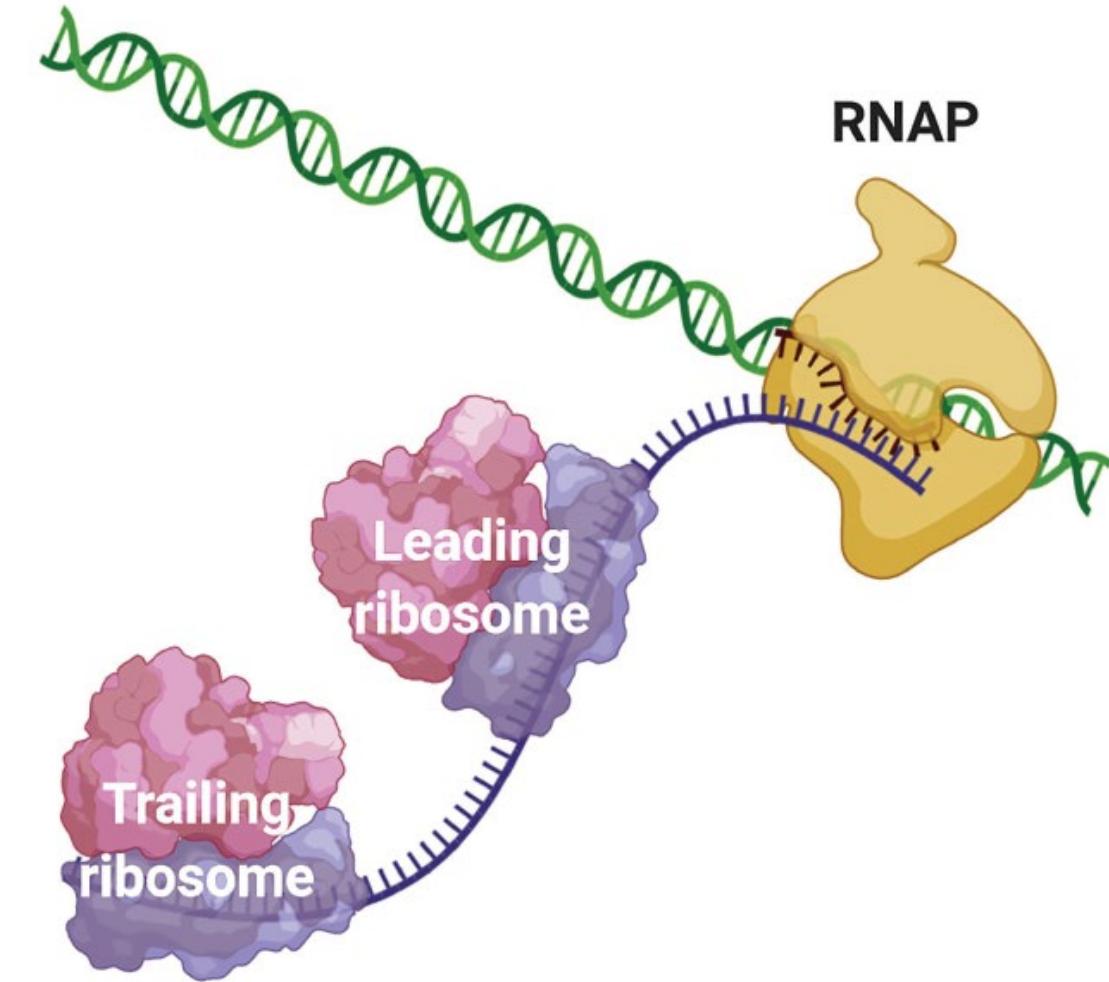
Promotores reconhecidos pelas duas famílias de fator Sigma

a

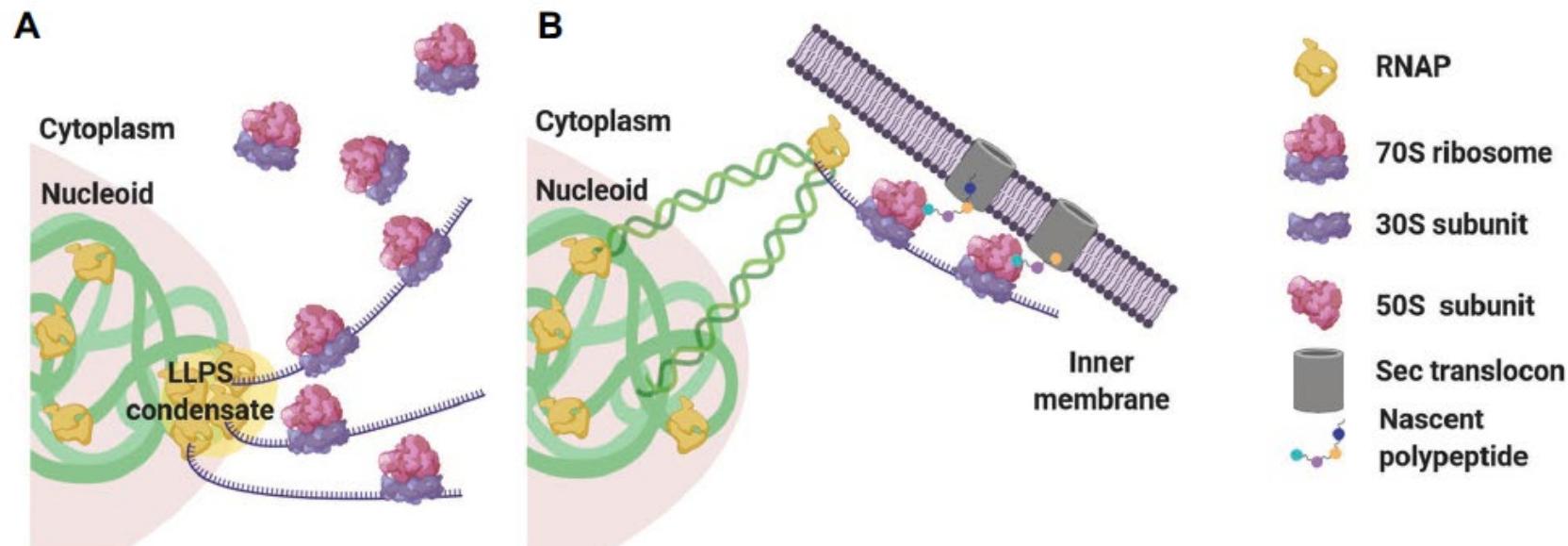


Operon Lac de E. coli: transcrição mRNA multigênico



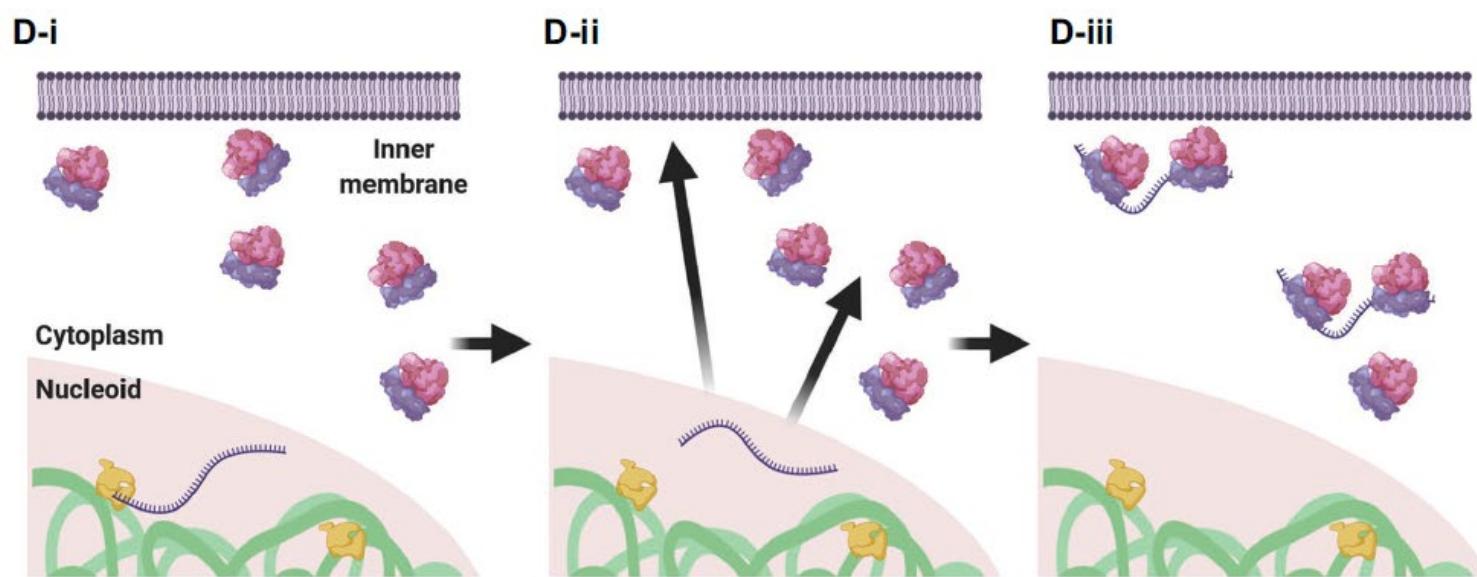
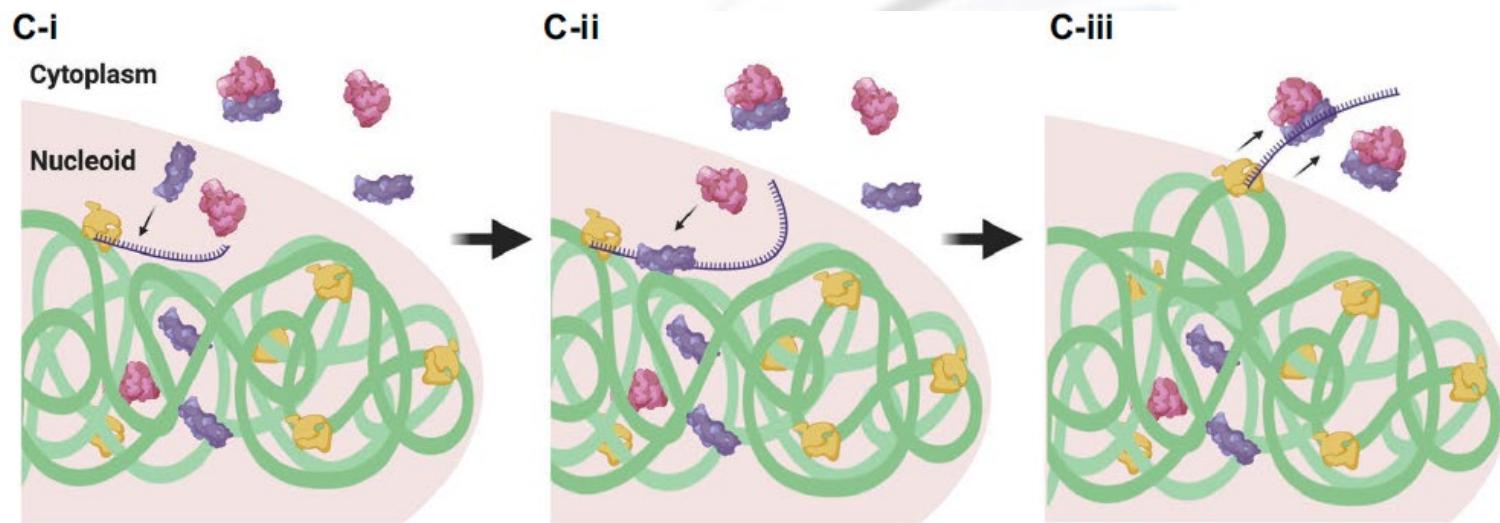


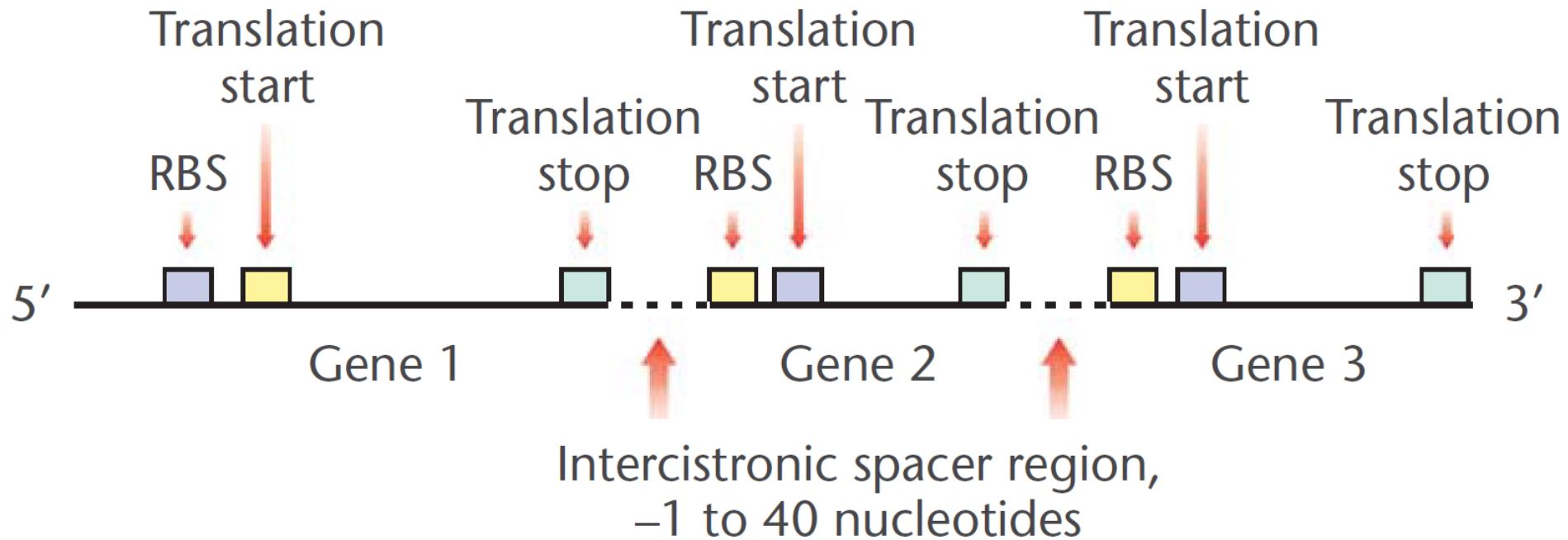
Alternativas para co-transcrição e tradução em espécies como *E. coli*, *Bacillus*, *Bdellovibrio* e *Pseudomonas putida*

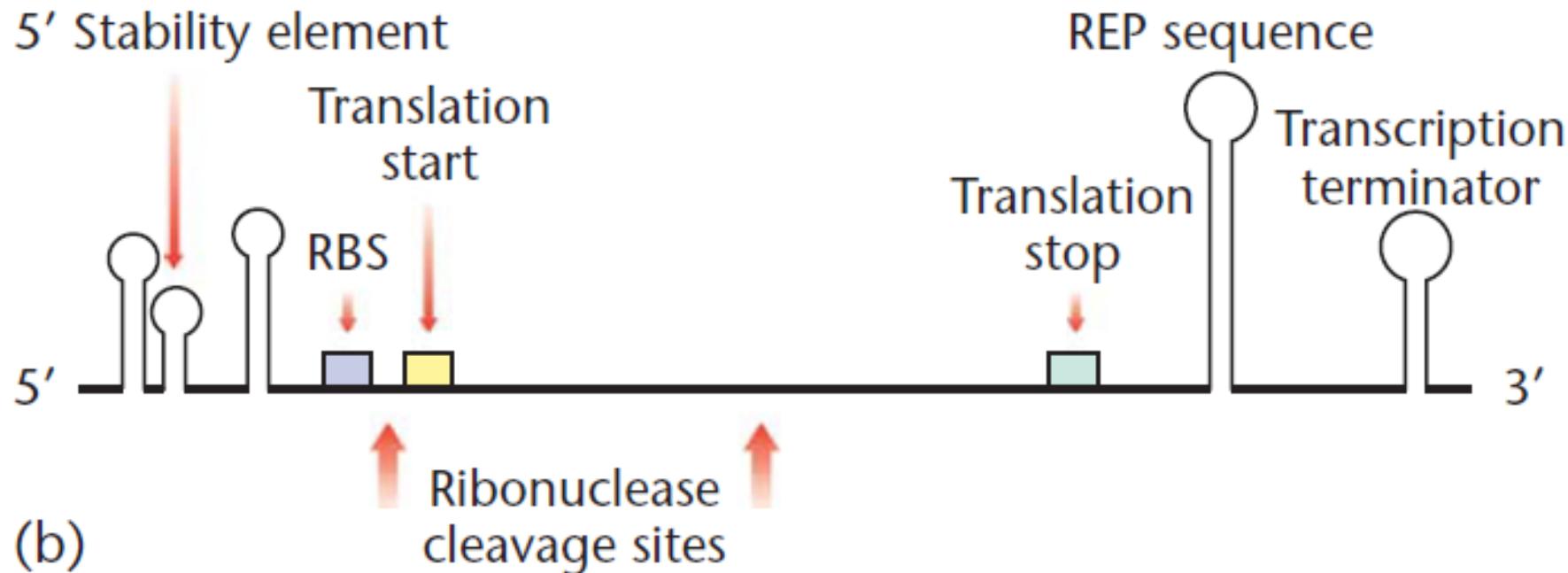


Ribossomos são segregados do nucleoide

Irazorreta-Olaziregi M and
Amster-Choder O (2021) Coupled
Transcription-Translation in
Prokaryotes: An Old Couple With
New Surprises.
Front. Microbiol. 11:624830.
doi: 10.3389/fmicb.2020.624830







REP elements (repetitive extragenic palindromes): Estabilidade do mRNA

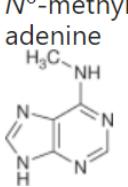
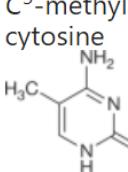
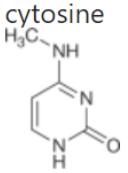
REP: cerca de 40 pares de bases.

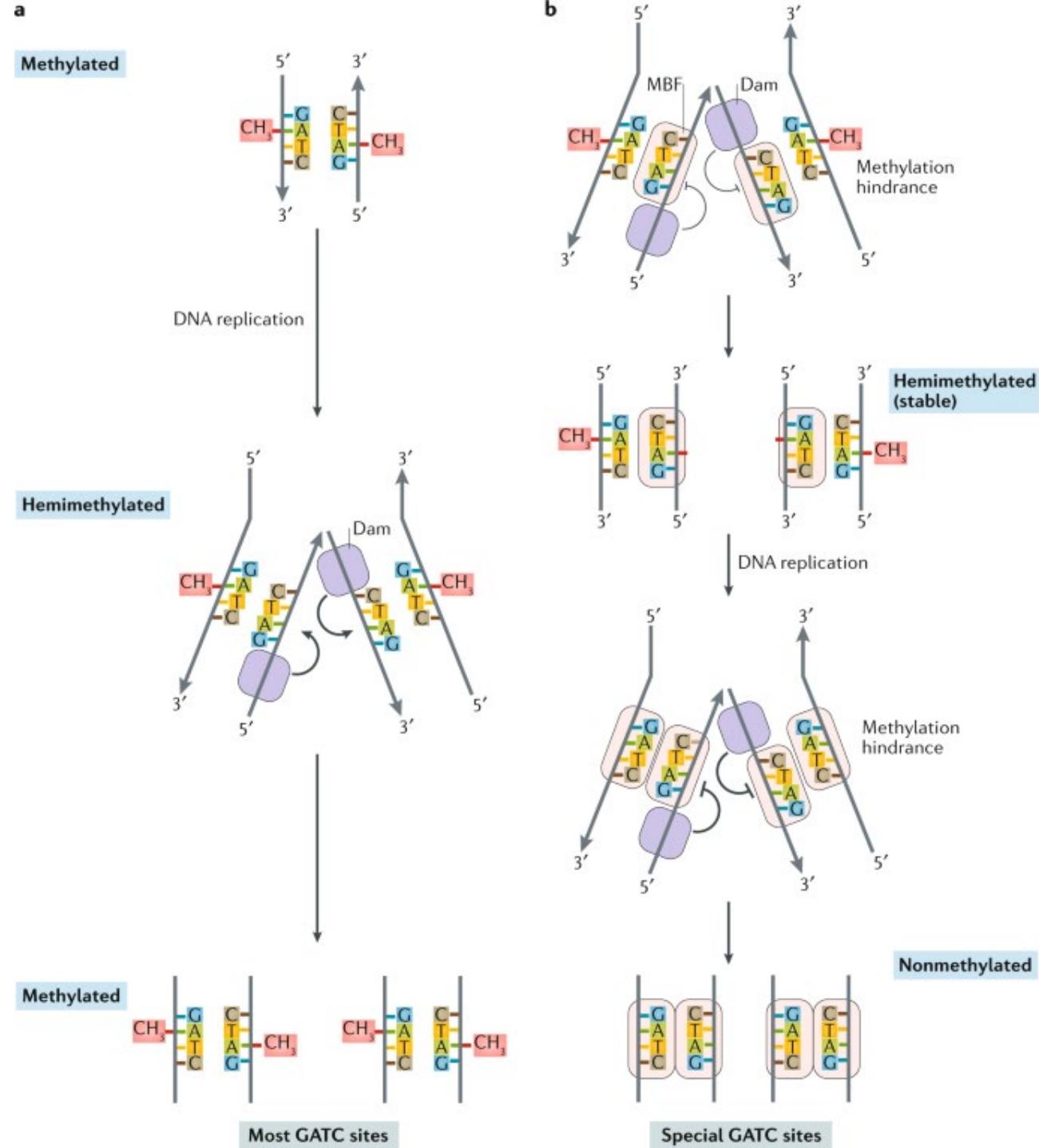
Aproximadamente 500-1000 cópias de sequências REP intercaladas no DNA cromossômico de *E. coli*, representando cerca de 1% do genoma

Regulação em bactérias

- Metilação do DNA – atua em:
 - defesa do genoma
 - replicação e segregação cromossômica
 - organização de nucleóides
 - controle do ciclo celular
 - reparo do DNA
 - regulação da transcrição
 - controle da mudança reversível (variação de fase) da expressão gênica
- Permite a adaptação de populações bacterianas a ambientes hostis ou em mudança
- Modula a interação de patógenos com seus hospedeiros eucarióticos.

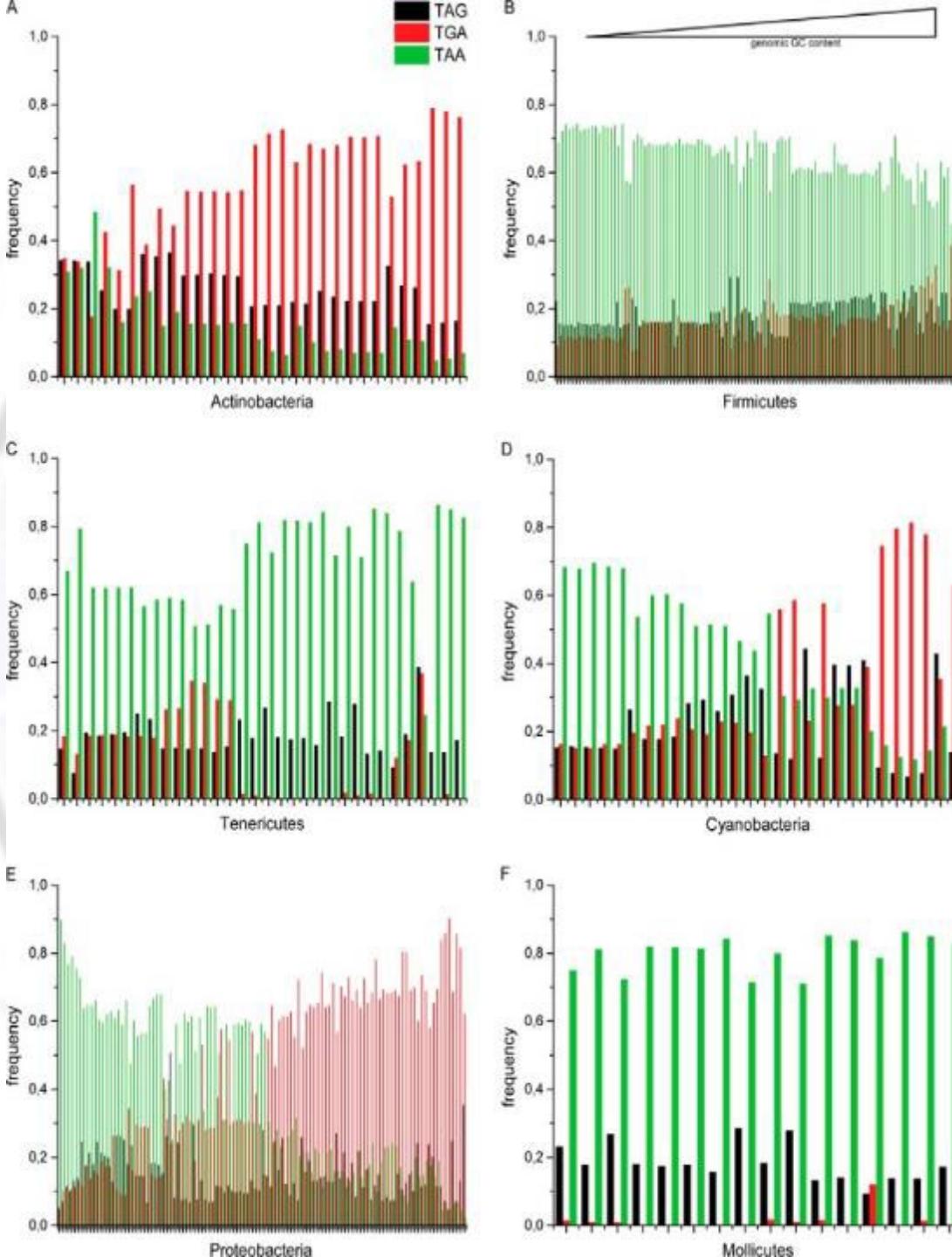
Exemplos de escritores e leitores do epigenoma bacteriano

Methylated base	Writer	Target sequence (5'-3')	Readers	Physiological process or phenotype under control
 <i>N</i> ⁶ -methyladenine	Dam	GATC	DnaA ¹⁰² , SeqA ⁸⁶ , MutHLS ¹⁰⁵ , RNA polymerase ⁹¹ , transposase ^{91,134} , Lrp ^{69,88,89,163} , OxyR ^{70,71,72,94} , Fur ⁷³ and HdfR ^{74,75}	Chromosome replication and segregation ^{86,102,104} , nucleoid organization ^{87,164} , mismatch repair ¹⁰⁵ , transposition ^{91,134} , conjugal transfer ⁸⁸ , motility ⁵⁶ , synthesis and secretion of virulence determinants ^{11,12,56,62,165} , envelope structure ¹⁶⁶ , bacteriophage resistance ¹⁰⁷ and antibiotic resistance ¹⁵⁵
	CcrM	GANTC	GcrA ⁹⁰ , MucR ^{76,167} and RNA polymerase ^{53,57}	Cell cycle control ^{32,34}
	ModA13	AGAAA	Unknown	Antibiotic resistance, epithelial cell invasion and biofilm formation ¹¹³
	SpnD39III	Allele-specific (six variants)	Unknown	Opacity (capsule synthesis) and virulence ¹²³
 <i>C</i> ⁵ -methylcytosine	Dcm	CCAGG and CCTGG	Unknown	General stress response ¹²⁸ and drug transport ¹²⁹
	JHP1050	GCGC	Unknown	Cell morphology, competence, outer membrane composition and copper resistance ¹²⁶
 <i>N</i> ⁴ -methylcytosine	M2.Hpy.All	TCTTC	Unknown	Virulence (adhesion to host cells) ¹³⁰

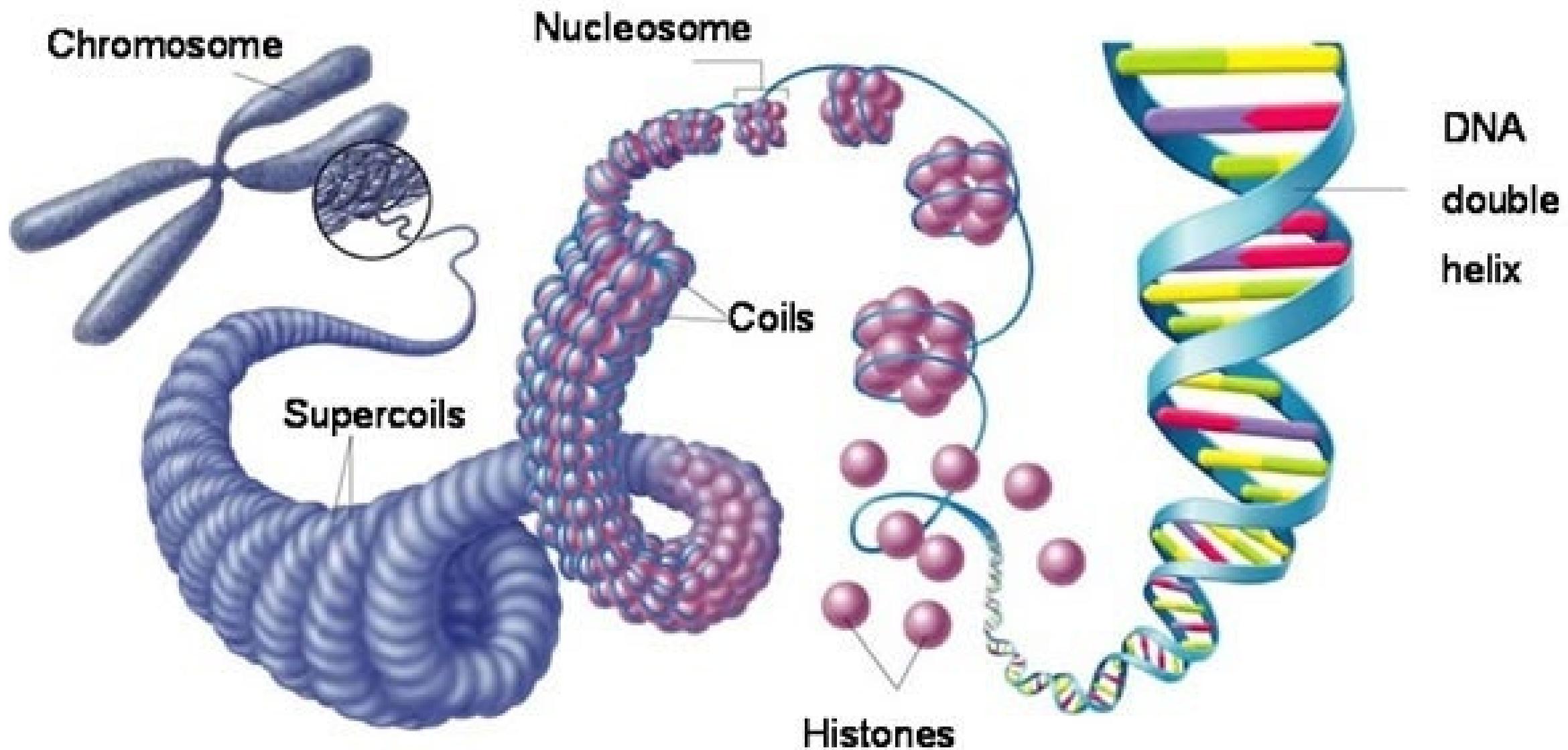


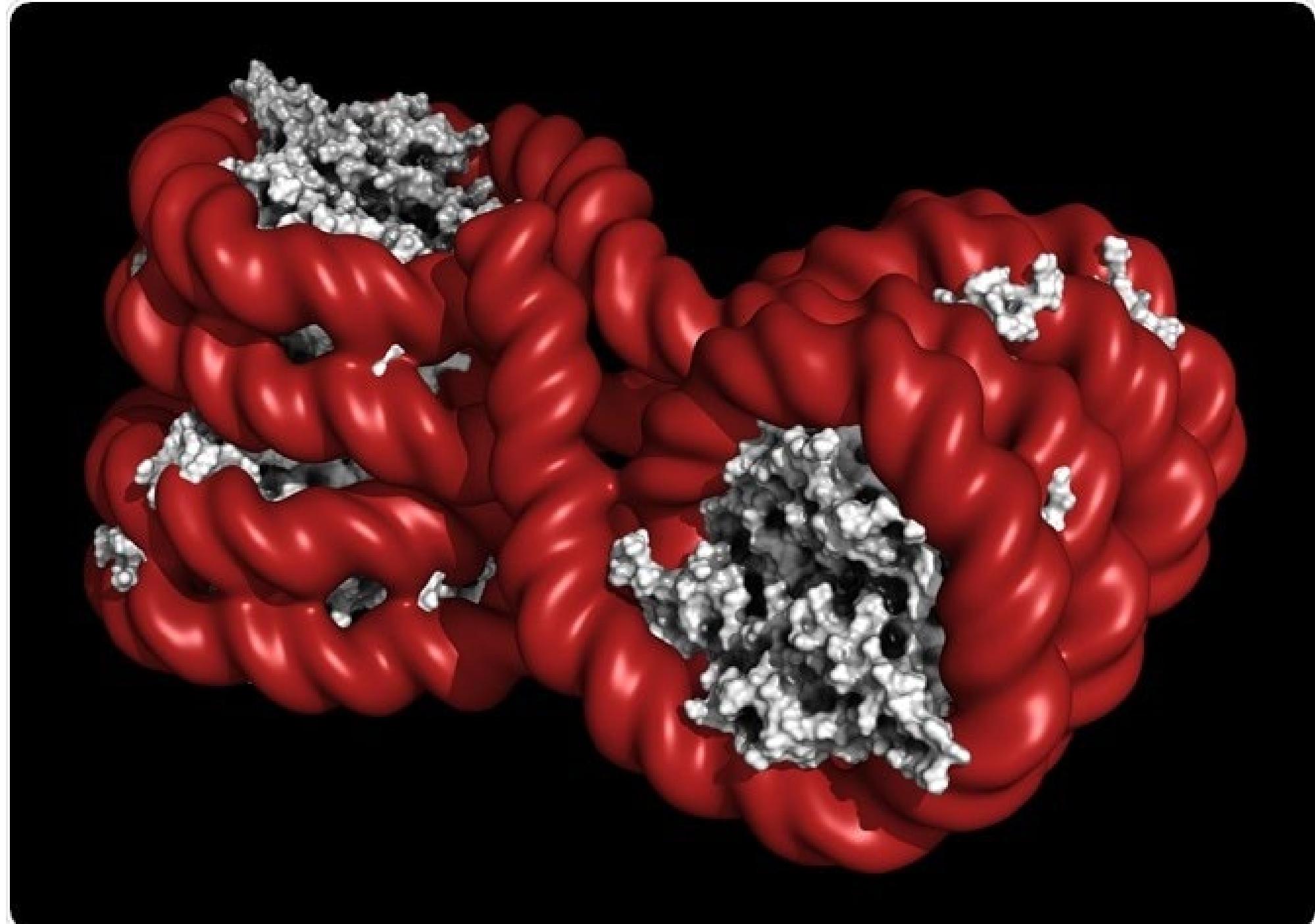
Taxas de diferentes stop codons em bactérias

TAG (UAG)
TGA (UGA)
TAA(UAA)

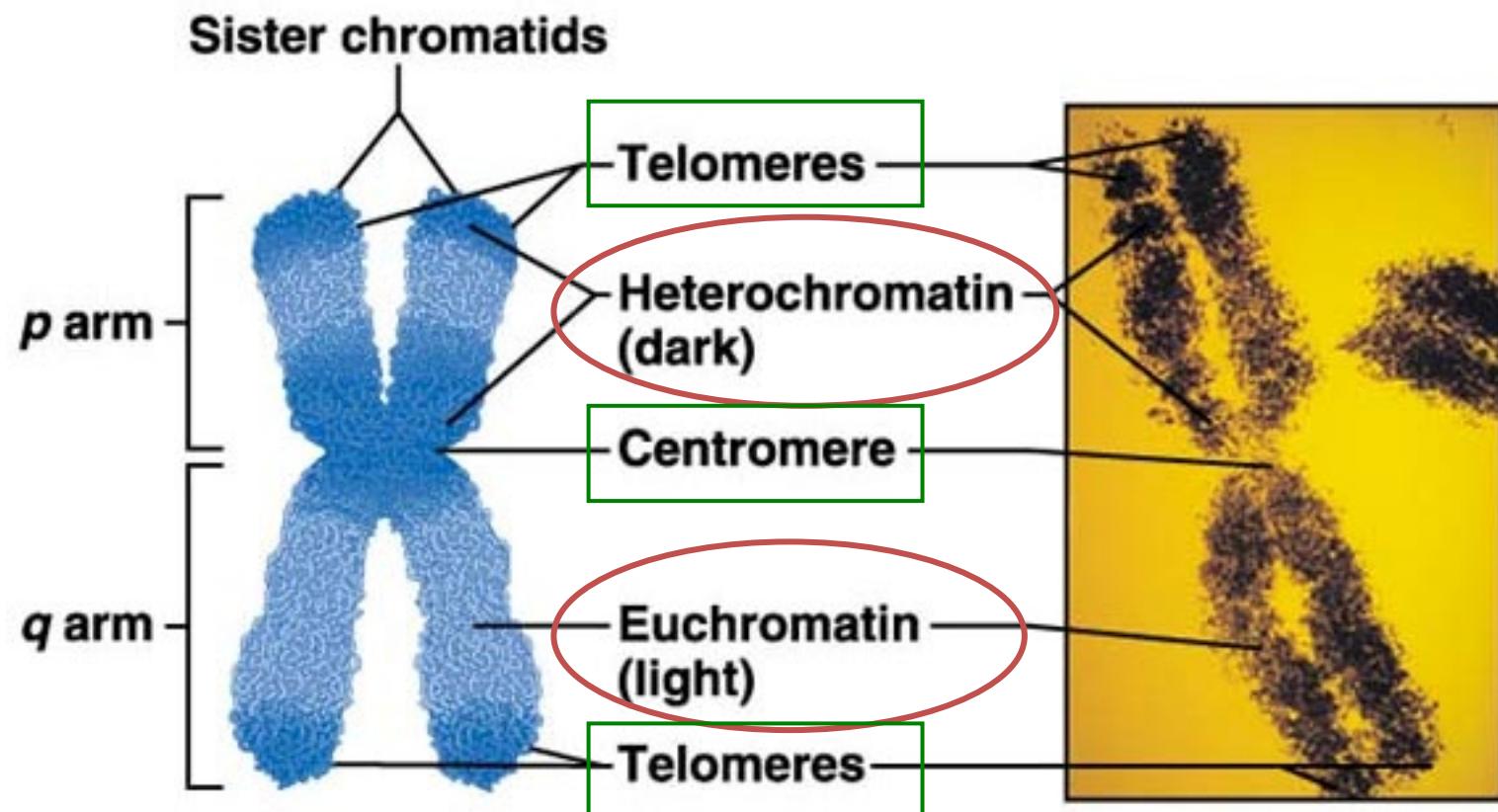


Genomas eucarióticos



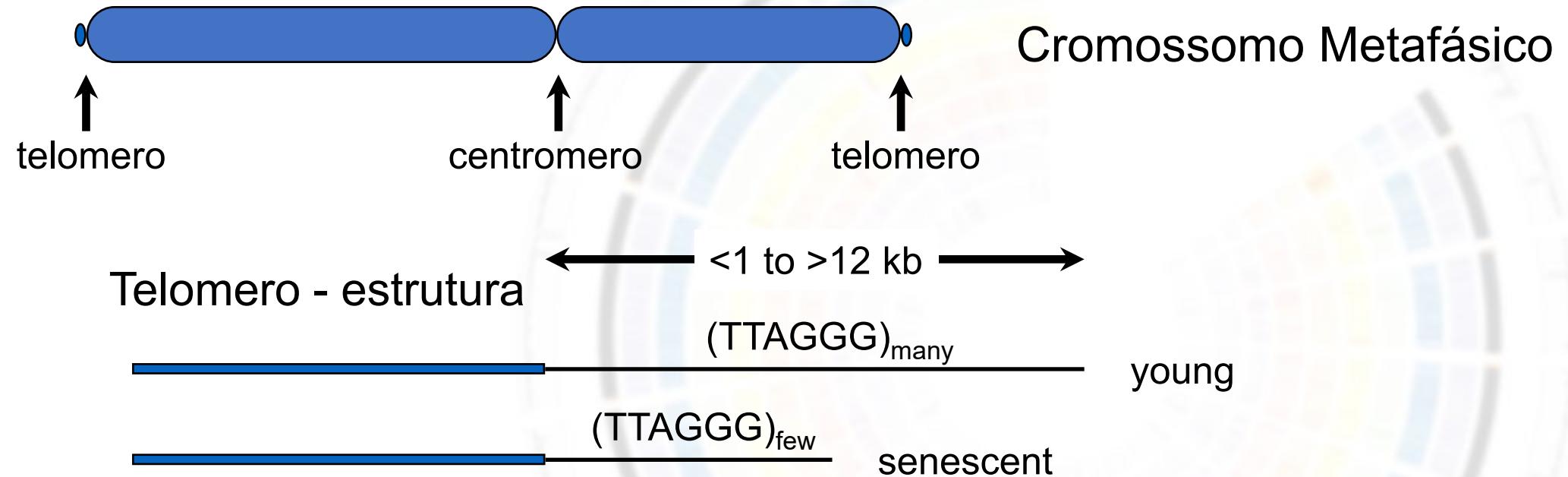


Anatomia de um cromossomo eucariótico



Heterocromatina: cromatina densa / transcrecionalmente inativa
Eucromatina: cromatina menos densa / transcrecionalmente ativa

Telômeros e centrômeros



Sequencia telomérica de muitos eucariotos incluindo muitos fungos: TTAGGG
Sequencia telomérica de *Sacharomyces cerevisiae*: TG1-3/C1-3A)

Species	Sequence	Reference
<i>Neurospora crassa</i>	TTAGGG	Schechtman 1990
<i>Fusarium oxysporum</i>	TTAGGG	Powell and Kistler 1990
<i>Histoplasma capsulatum</i>	TTAGGG	Woods and Goldman 1992
<i>Cladosporium fulvum</i>	TTAGGG	Coleman et al. 1993
<i>Podospora anserina</i>	TTAGGG	Javerzat et al. 1993
<i>Ustilago maydis</i>	TTAGGG	Guzman and Sanchez 1994
<i>Magnaporthe grisea</i>	TTAGGG	Farman and Leong 1995
<i>Pneumocystis carinii</i>	TTAGGG	Underwood et al. 1996
<i>Aspergillus nidulans</i>	TTAGGG	Bhattacharyya and Blackbum 1997
<i>Aspergillus oryzae</i>	TTAGGGTCAACA	Kusumoto et al. 2003
<i>Cryptococcus neoformans</i>	TTAG ₃₋₅	Edman 1992
<i>Schizosaccharomyces pombe</i>	T ₁₋₂ ACA 0-1C 0-1G 1-6	Matsumoto et al. 1987
<i>Saccharomyces cerevisiae</i>	TG ₂₋₃ (TG) ₁₋₆	Shampay et al. 1984
<i>Saccharomyces exiguus</i>	TG ₂₋₃ (TG) ₁₋₆	Cohn et al. 1998
<i>Saccharomyces castellii</i>	TCTGGGTG	Cohn et al. 1998
<i>Saccharomyces dairensensis</i>	TCTGGGTG	Cohn et al. 1998
<i>Saccharomyces kluuyveri</i>	GACATGCGTACTGTGAGGTCT GGGTG	Cohn et al. 1998
<i>Candida albicans</i>	TCTAACTCTTGGTGTACGGA TG	McEachern and Hicks 1993
<i>Candida guilliermondii</i>	TACTGGTG	McEachern and Blackbum 1994
<i>Candida glabrata</i>	CTGTGGGTCTGGTG	McEachern and Blackbum 1994
<i>Candida maltosa</i>	CAGACTCGCTTGGTGTACGGA TG	McEachern and Blackbum 1994
<i>Candida tropicalis</i>	TCACGATCATTGGTGTA(A/C)G GATG	McEachern and Blackbum 1994
<i>Candida pseudotropicalis</i>	TGATTAGTTATGTGGTGTACG GATT	McEachern and Blackbum 1994
<i>Kluyveromyces lactis</i>	TGATTAGGTATGTGGTGTACG GATT	McEachern and Blackbum 1994

The sequence corresponding to the template region of the *S. cerevisiae* telomerase RNA, ACCAACACCCACACA, is also found in the *Saccharomyces* species *S. paradoxus*, *S. cariocanus*, *S. mikatae*, *S. kudriavzevii*, *S. bayanus*, and *S. pastorianus* (Dandjinou et al. 2004).

Sequências teloméricas em fungos

Topics in Current Genetics, Vol. 15

P. Sunnerhagen, J. Piškur (Eds.): Comparative Genomics

DOI 10.1007/4735_108 / Published online: 11 October 2005

© Springer-Verlag Berlin Heidelberg 2005

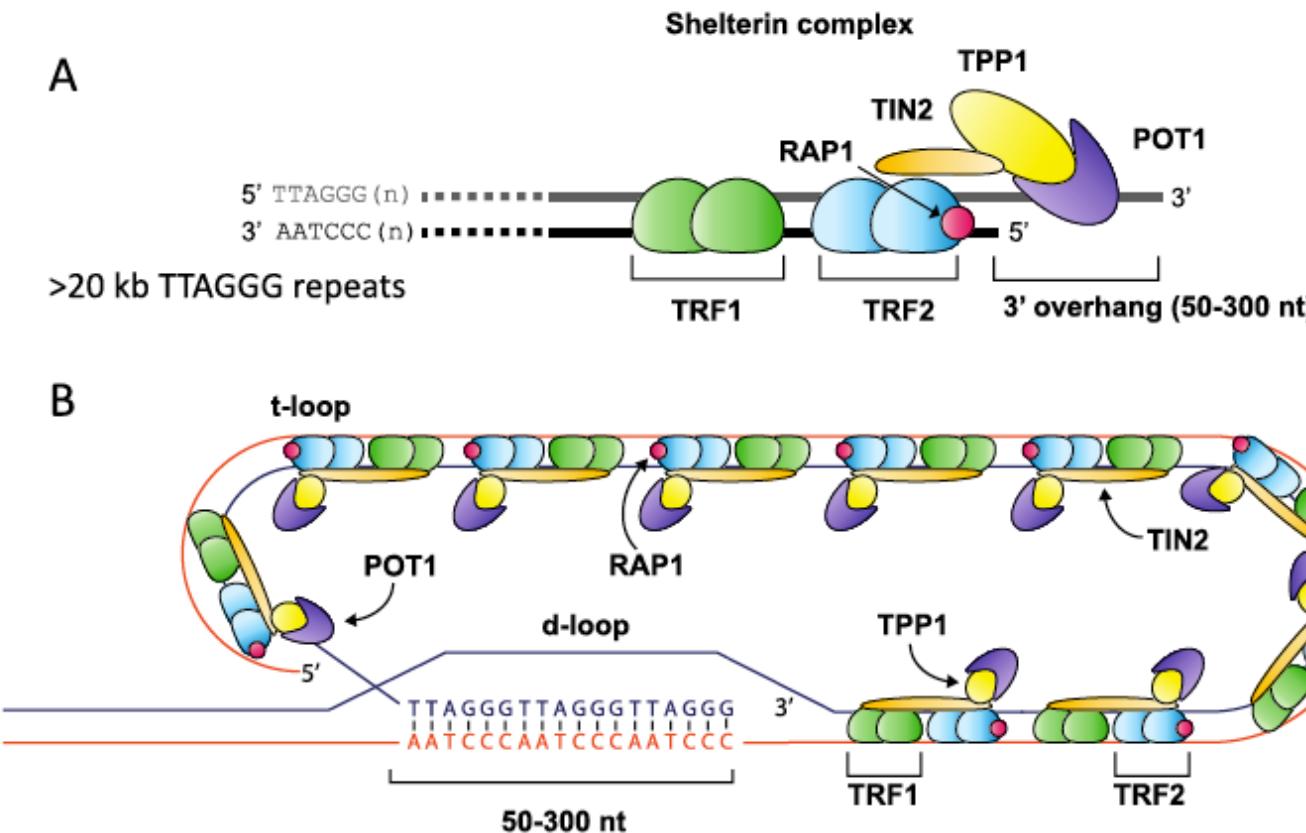
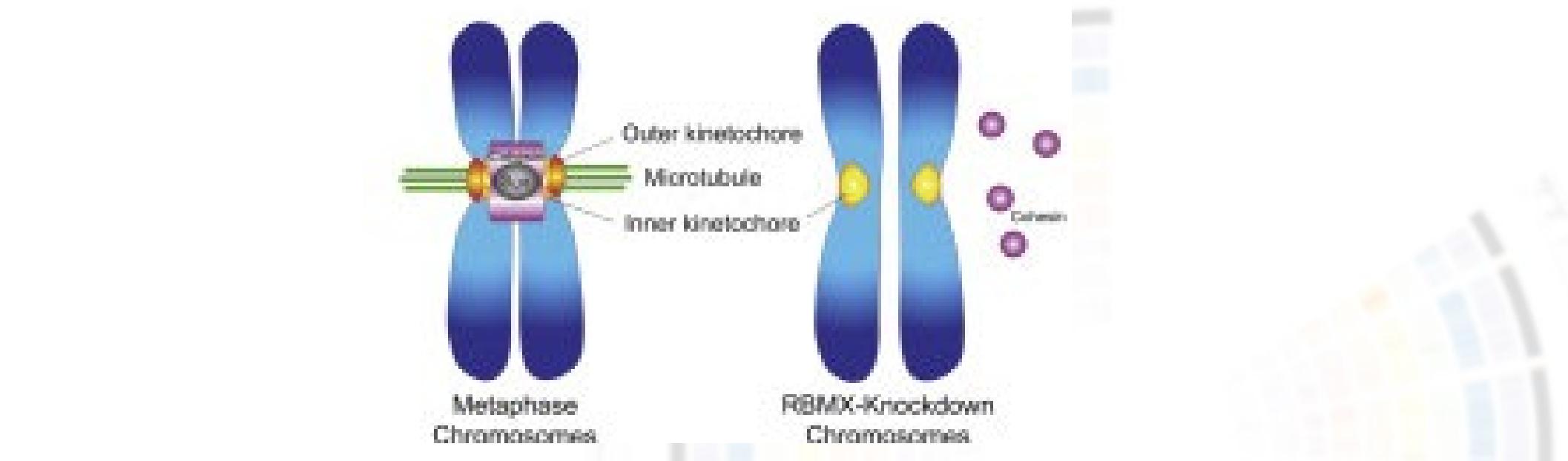
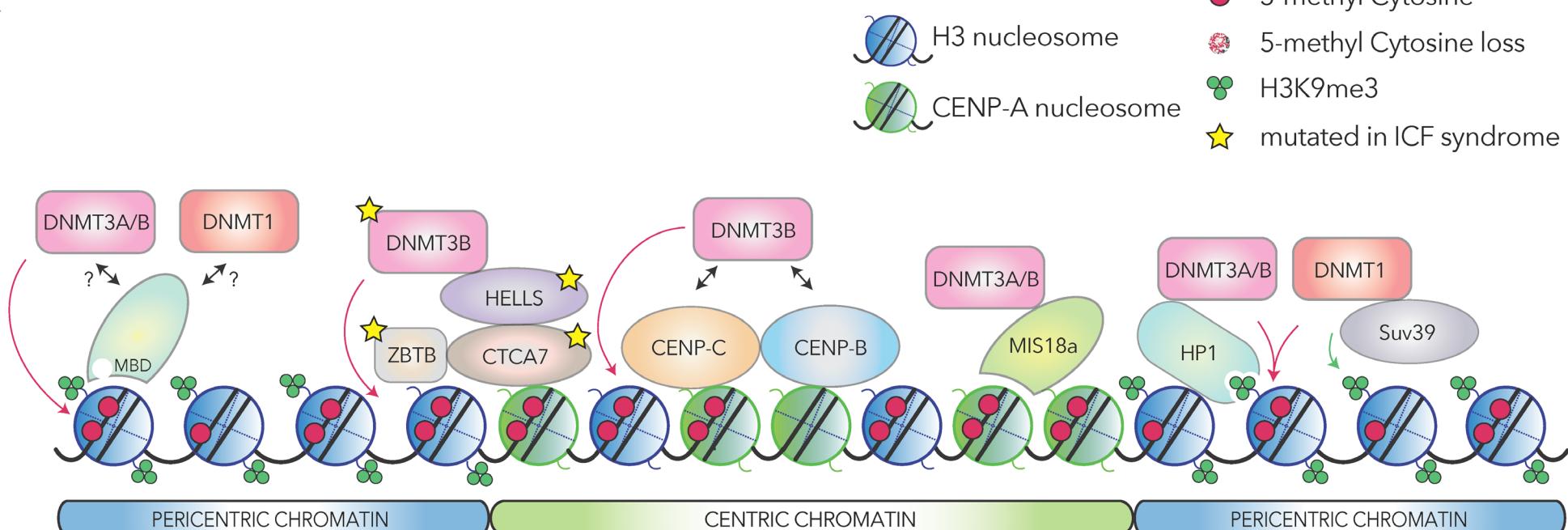


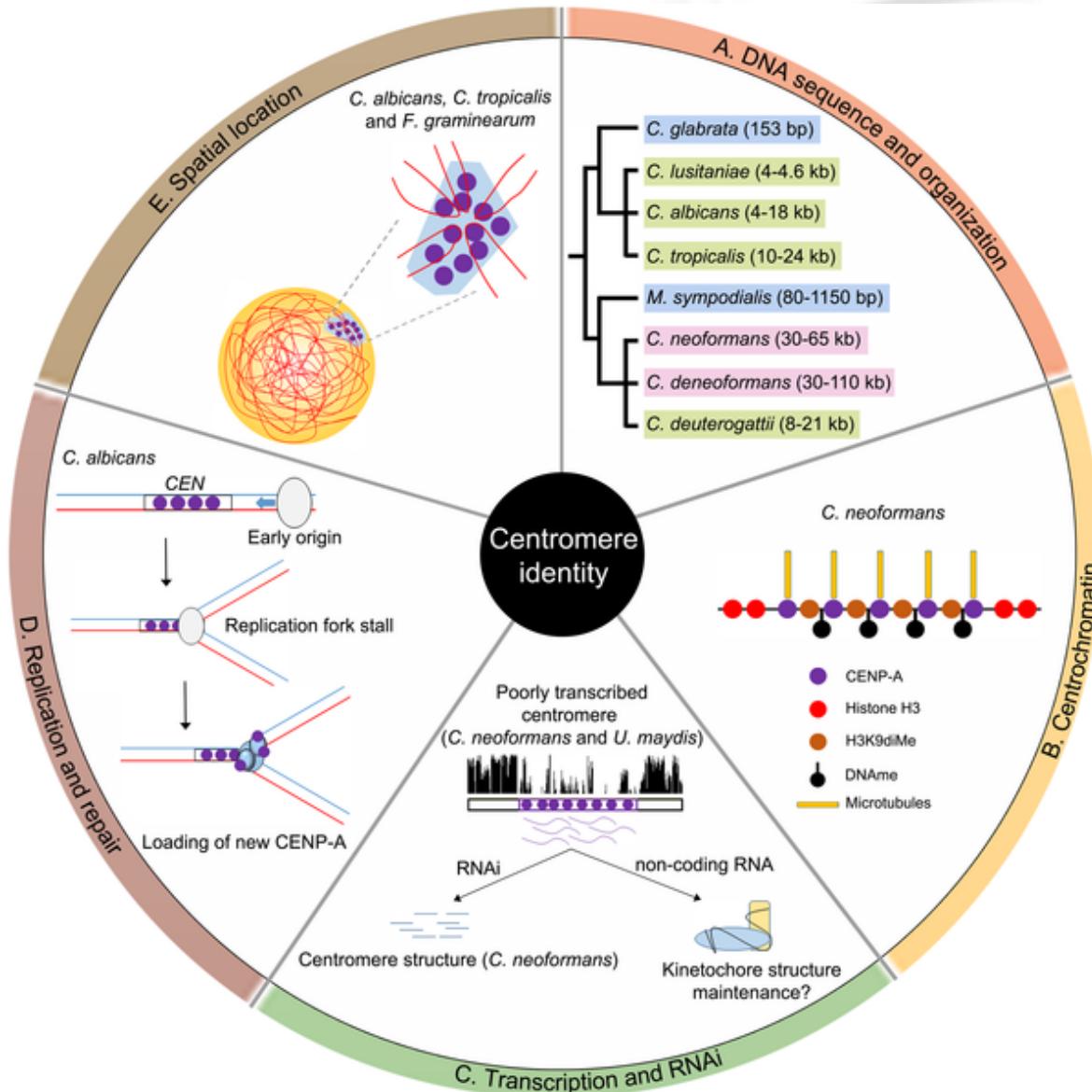
Figure 1. Overview of telomere structure. The terminal ends of mammalian chromosomes consist of an array of $(TTAGGG)_n$ repeats ending with a 3' overhang of between 50 and 300 nt in length (A). This array is bound by many protein components including members of the shelterin complex, which anchors to the repeat array through Telomere Repeat binding Factors 1 & 2 (TRF1 and TRF2), binding repeats as a homodimer, and forming a complex with TIN2 (TRF1-interacting factor), RAP1 (Repressor Activator Protein 1), TPP1 and POT1 (Protection of Telomere 1) (A). The repeat array folds into a higher-order t-loop structure where the 3' overhang displaces a portion of the forward strand to create a d-loop, thereby sequestering the free chromosome end from the DNA repair machinery (B).



A



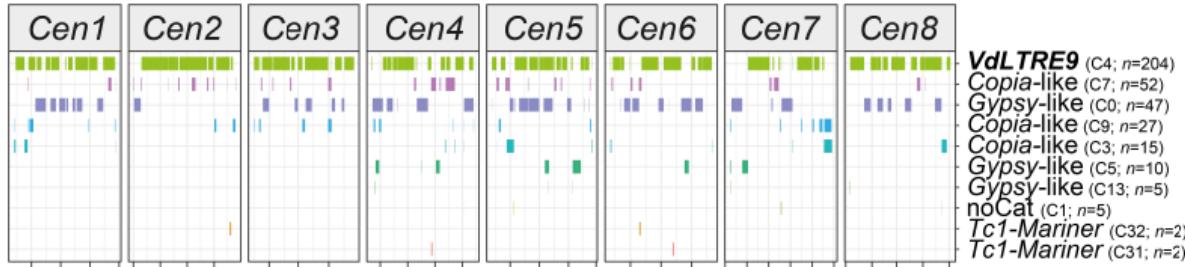
Cinco principais determinantes da identidade do centrômero em fungos patogênicos



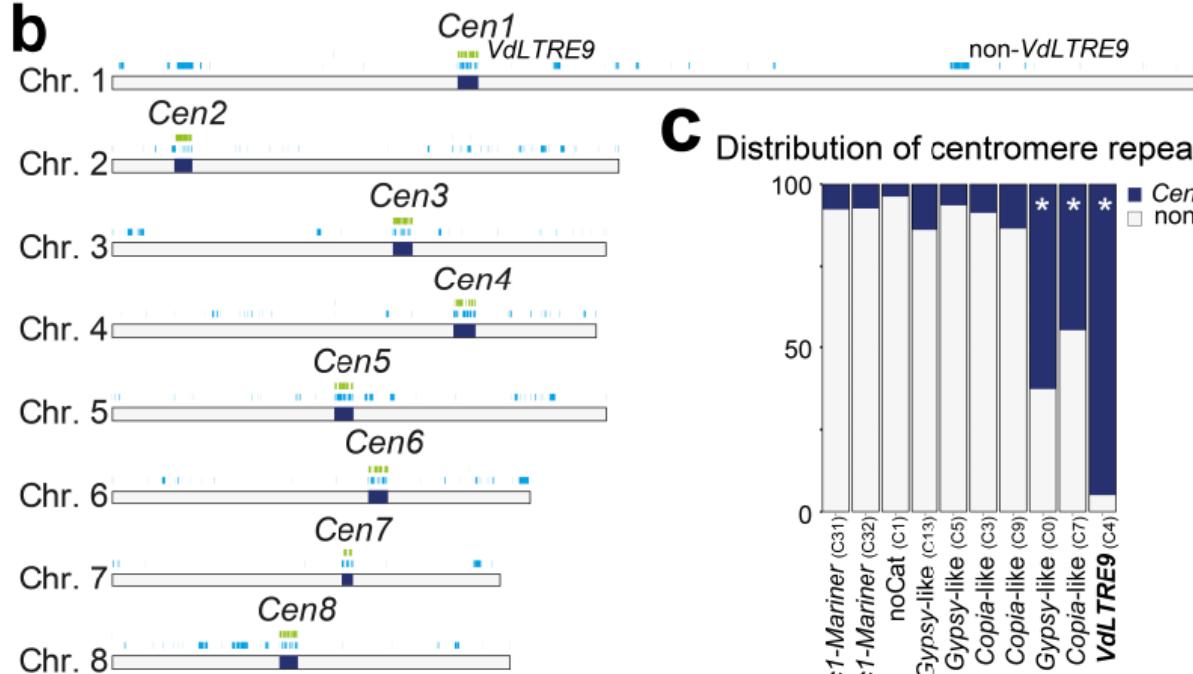
Yadav V, Sreekumar L, Guin K, Sanyal K (2018) Five pillars of centromeric chromatin in fungal pathogens. PLOS Pathogens 14(8): e1007150. <https://doi.org/10.1371/journal.ppat.1007150>
<https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1007150>

Sequencias repetitivas em centrômeros de *Verticillium dahliae*

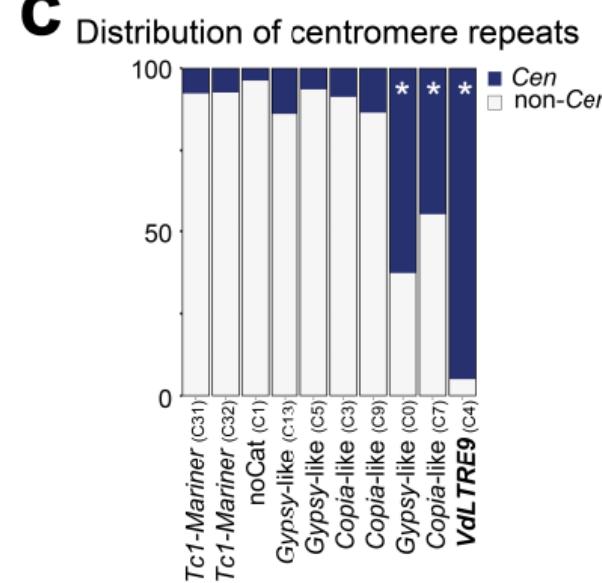
a



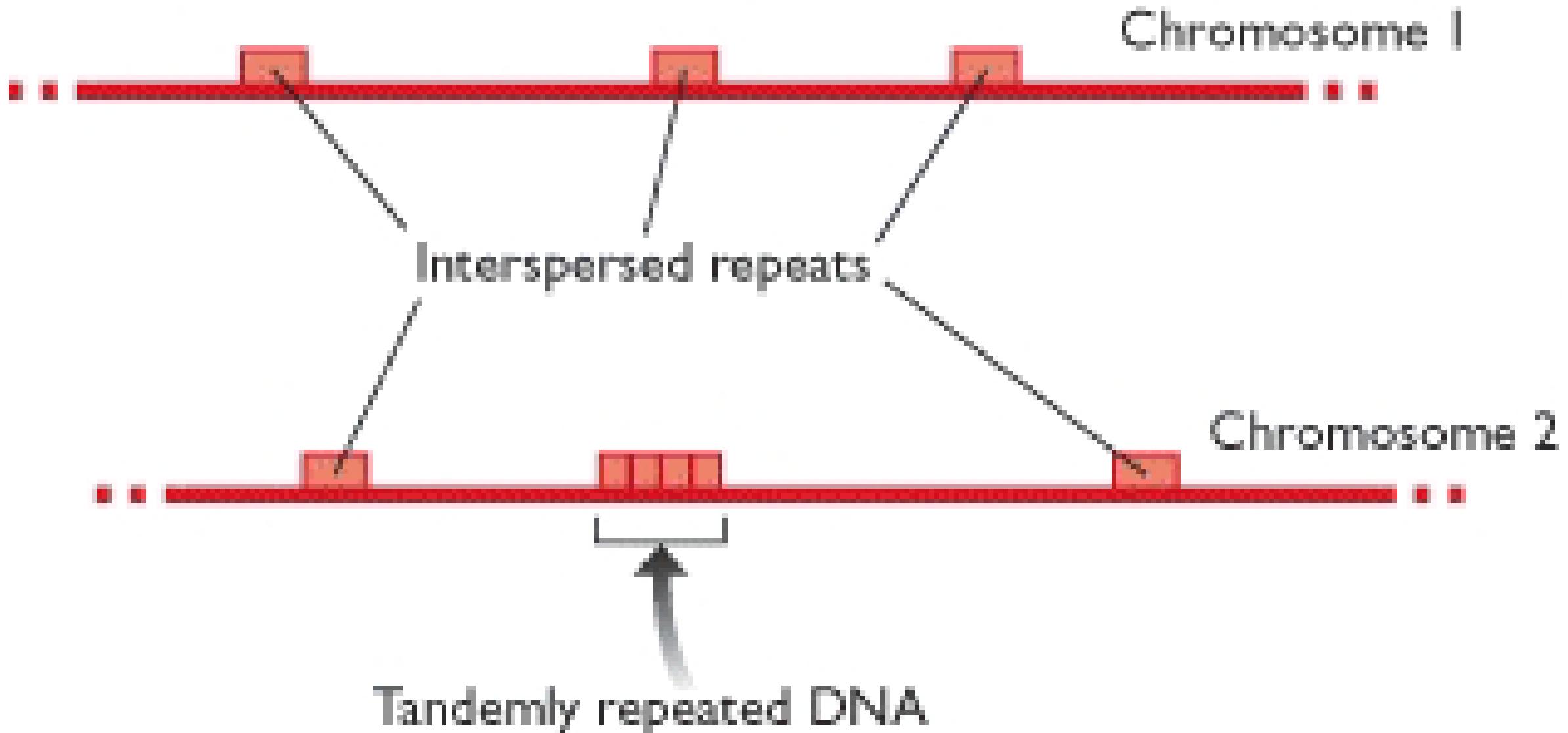
b

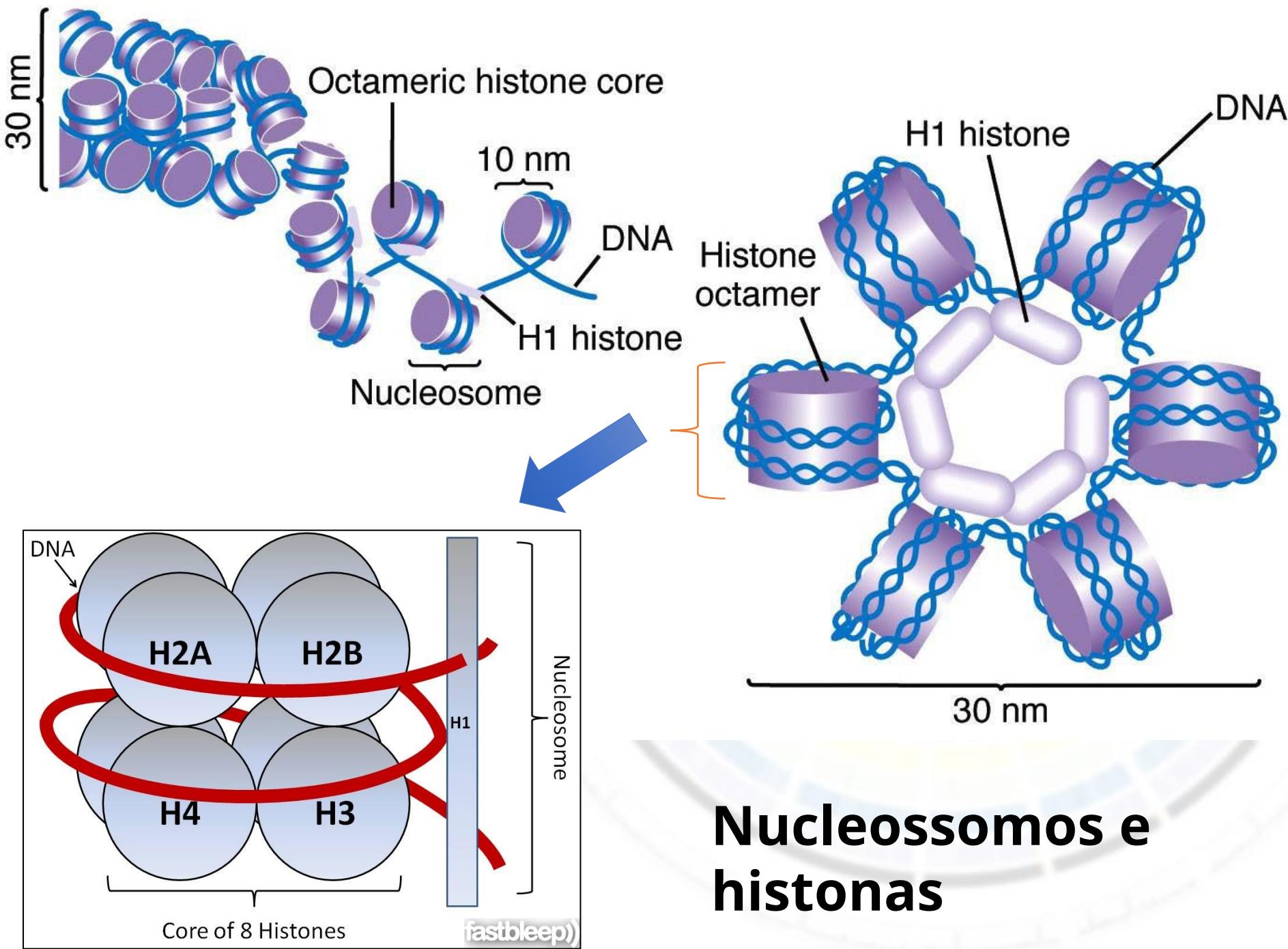


c



Sequencias repetitivas: dispersas e *in tandem*

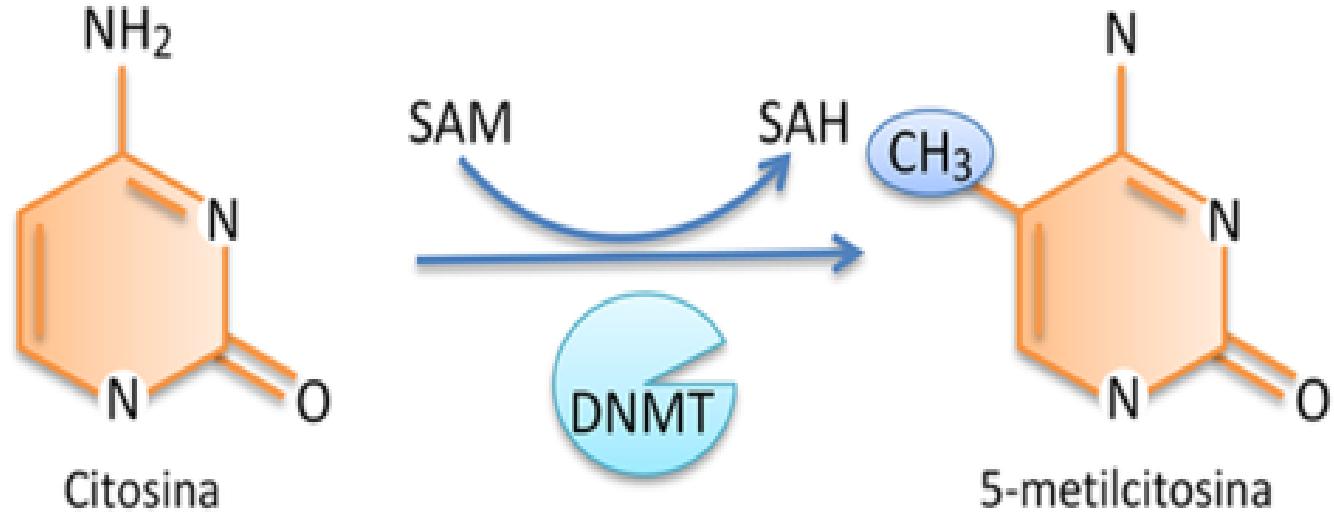




Mecanismos epigenéticos

1. Metilação do DNA
2. Modificação das histonas
3. Modificação por RNAi

Metilação de DNA



SAM: S-adenosil-metionina (doador do grupo metil)

SAH: S-adenosil-homocisteína

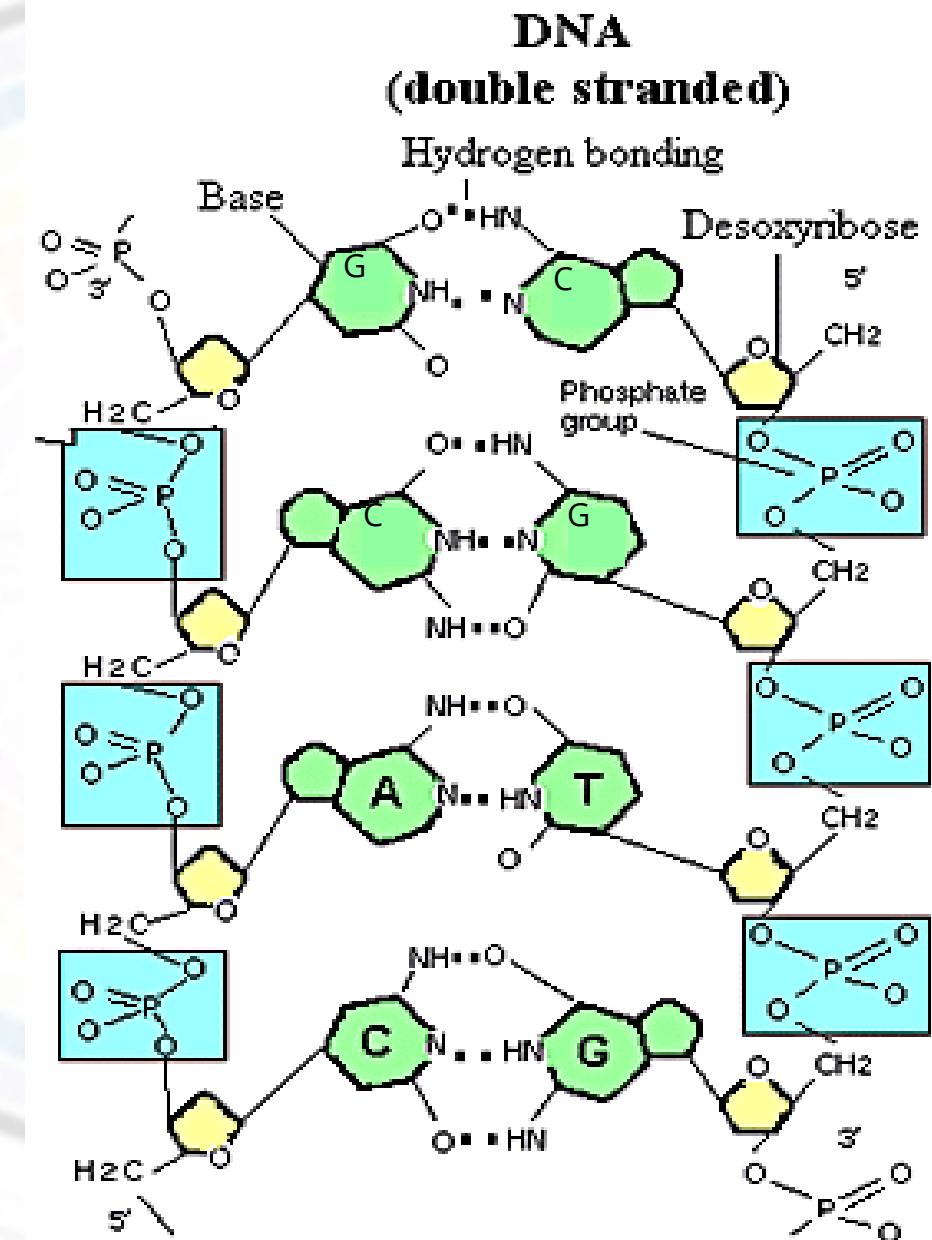
DNMT: DNA metiltransferase

Metilação do DNA

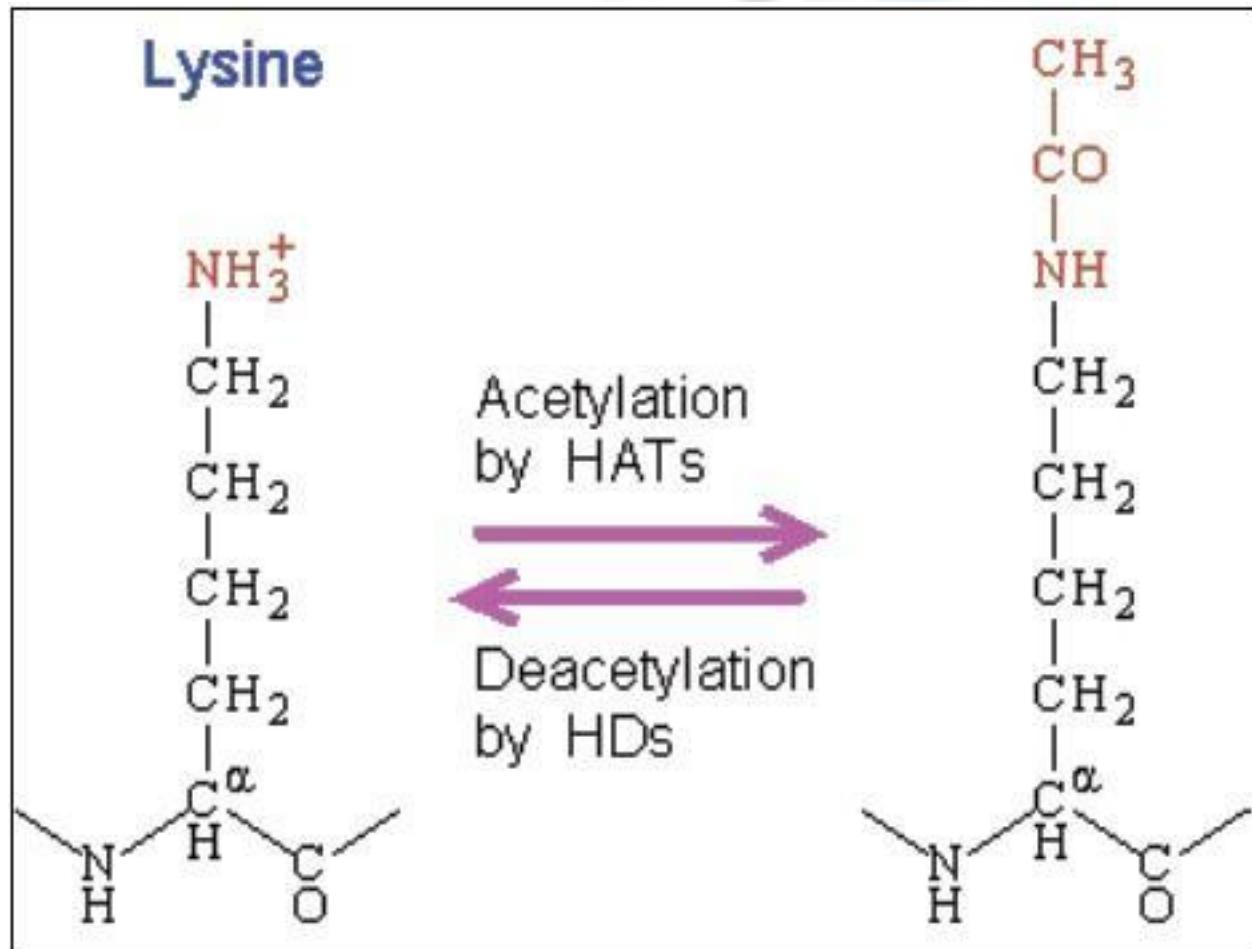
- Ocorre no carbono 5' de uma citosina que geralmente precede uma guanina (CpG)

5' - CpG - 3'

ligação
fosfodiéster

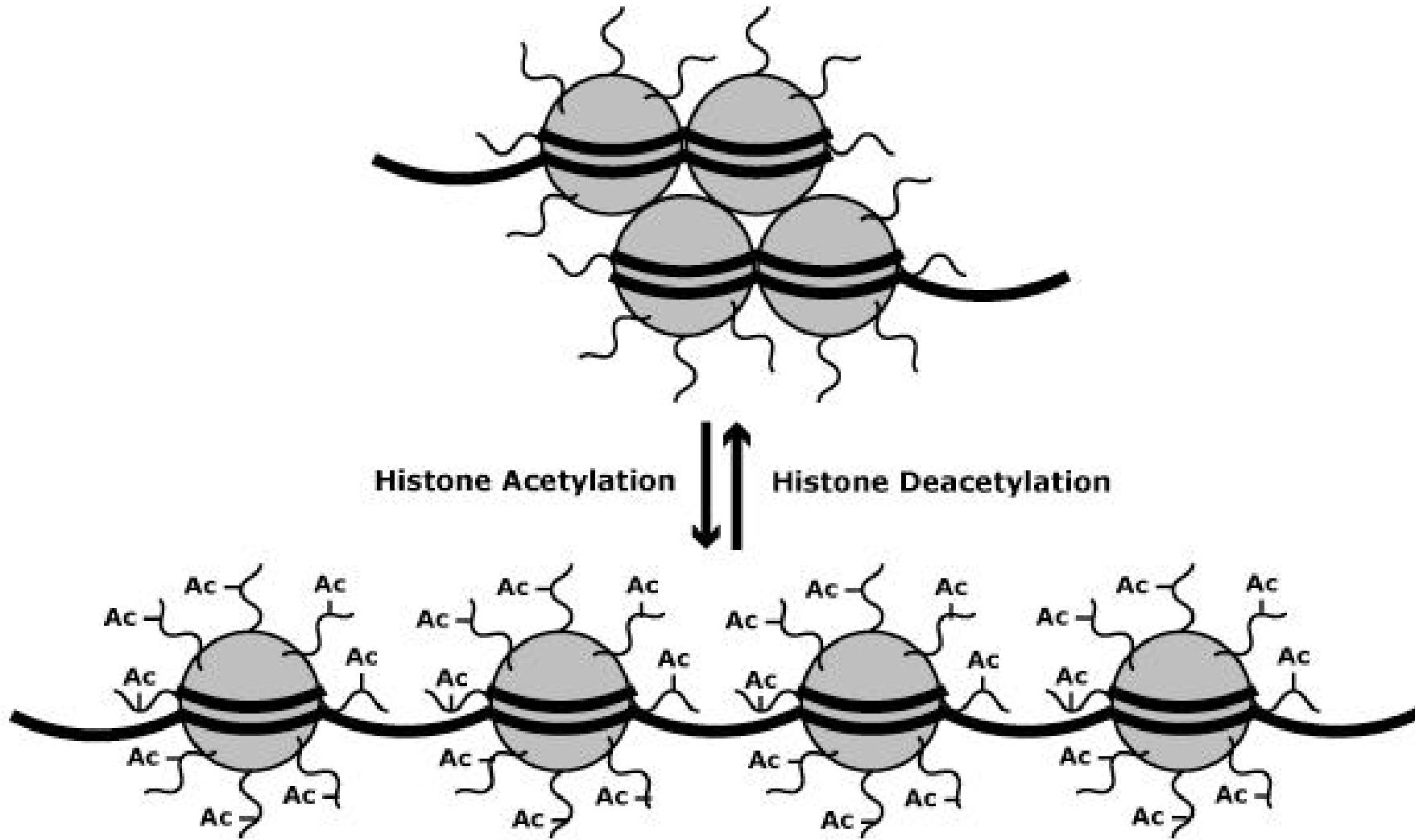


Modificação de histonas

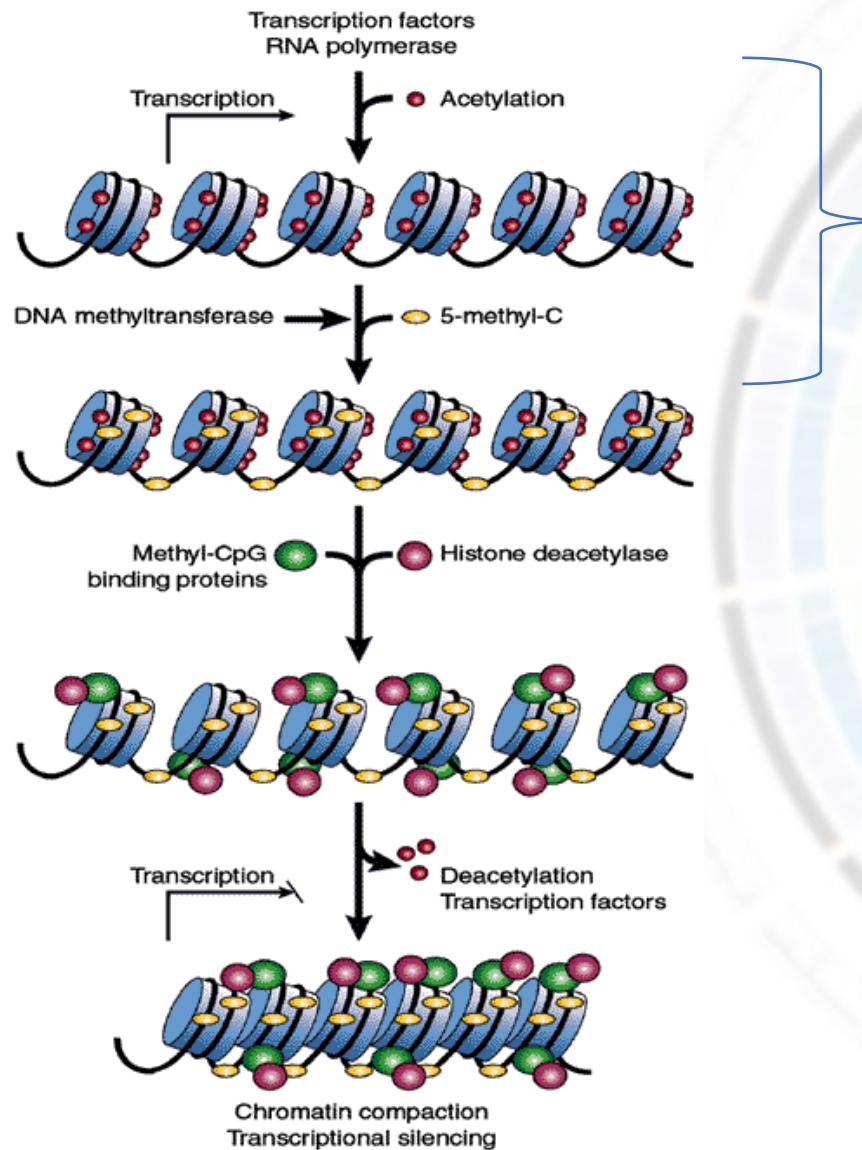


Enzimas: histonas acetilases (HATs)
histonas desacetilases (HDs ou HDACs)

Modificação de histonas

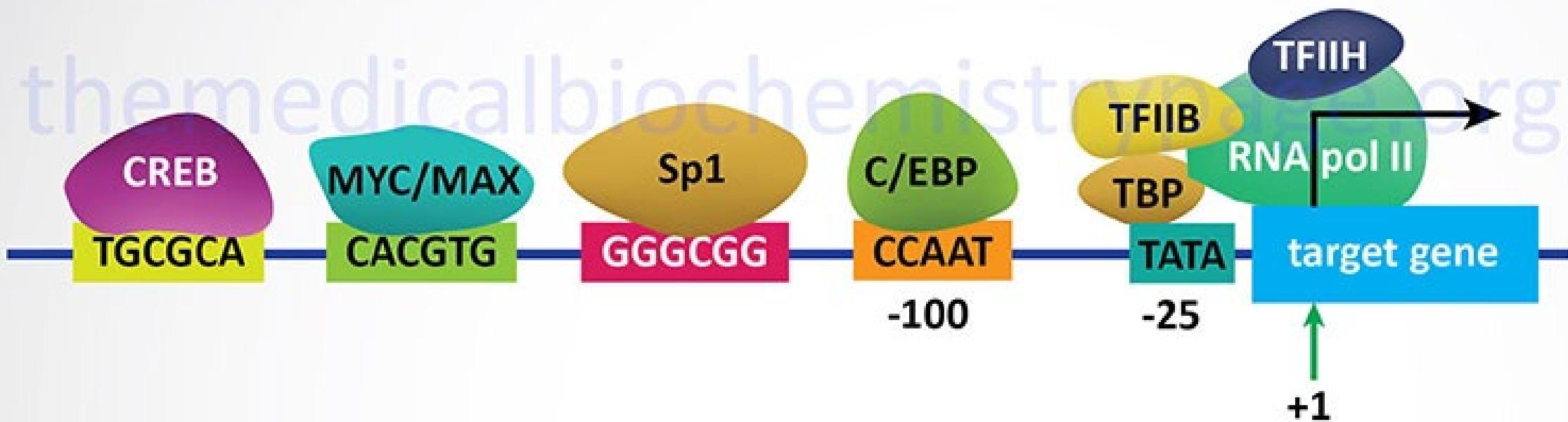


Modificação de histonas

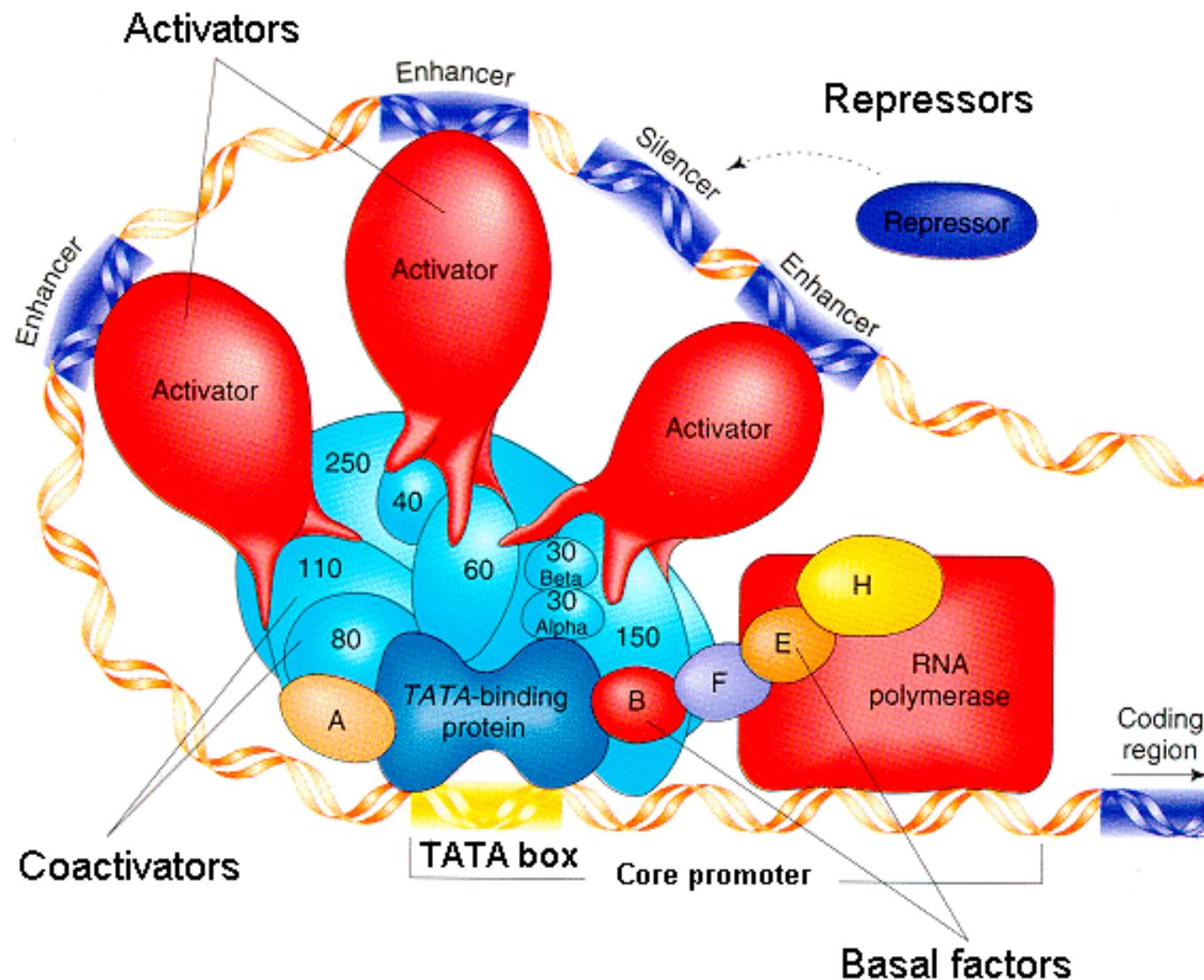


Regiões da cromatina transcrevionalmente ativas tendem a estar hiperacetiladas e hipometiladas.

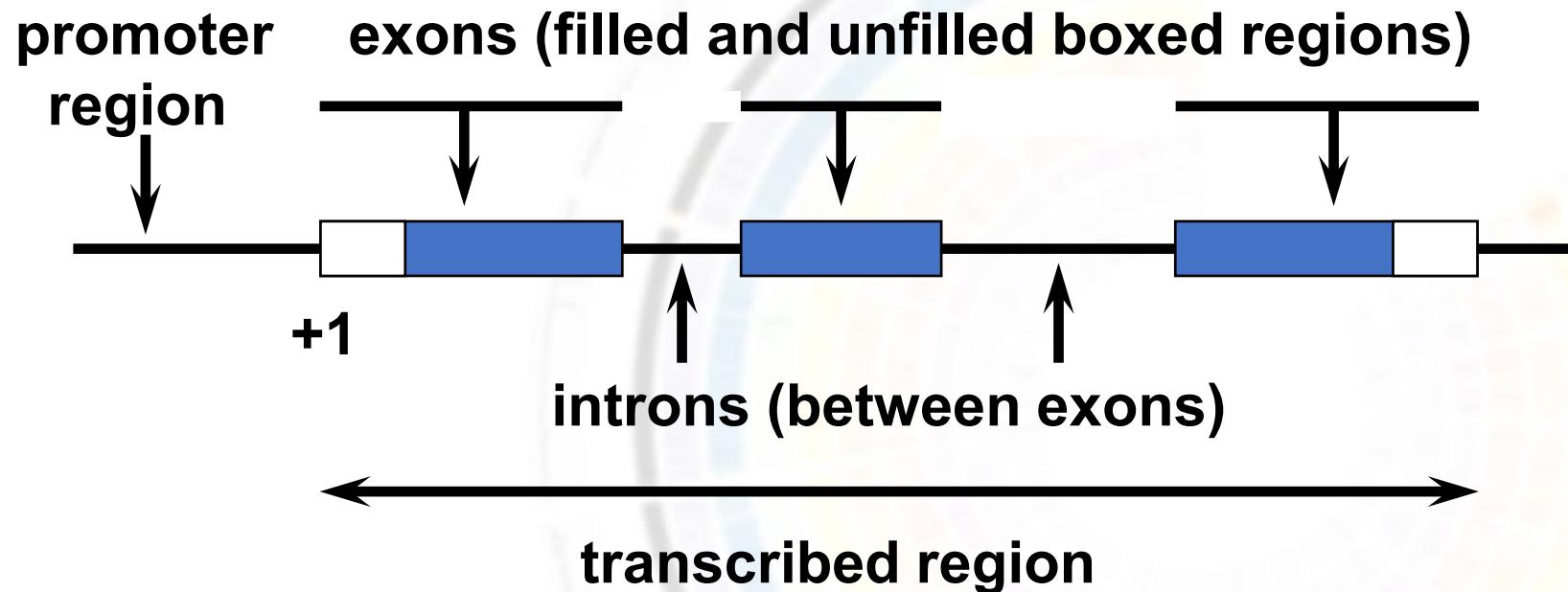
Organização de genes eucariotos



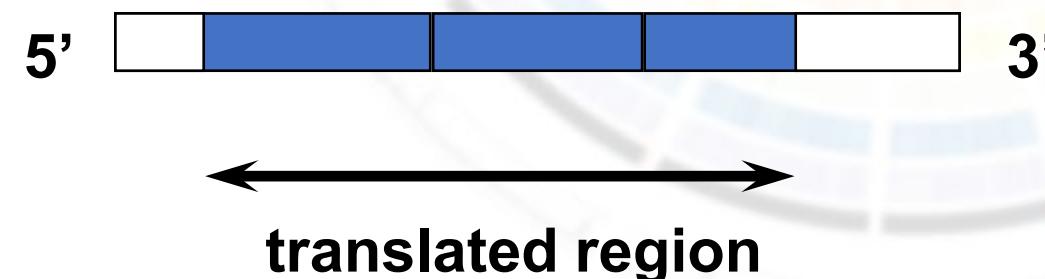
Organização de genes eucariotos



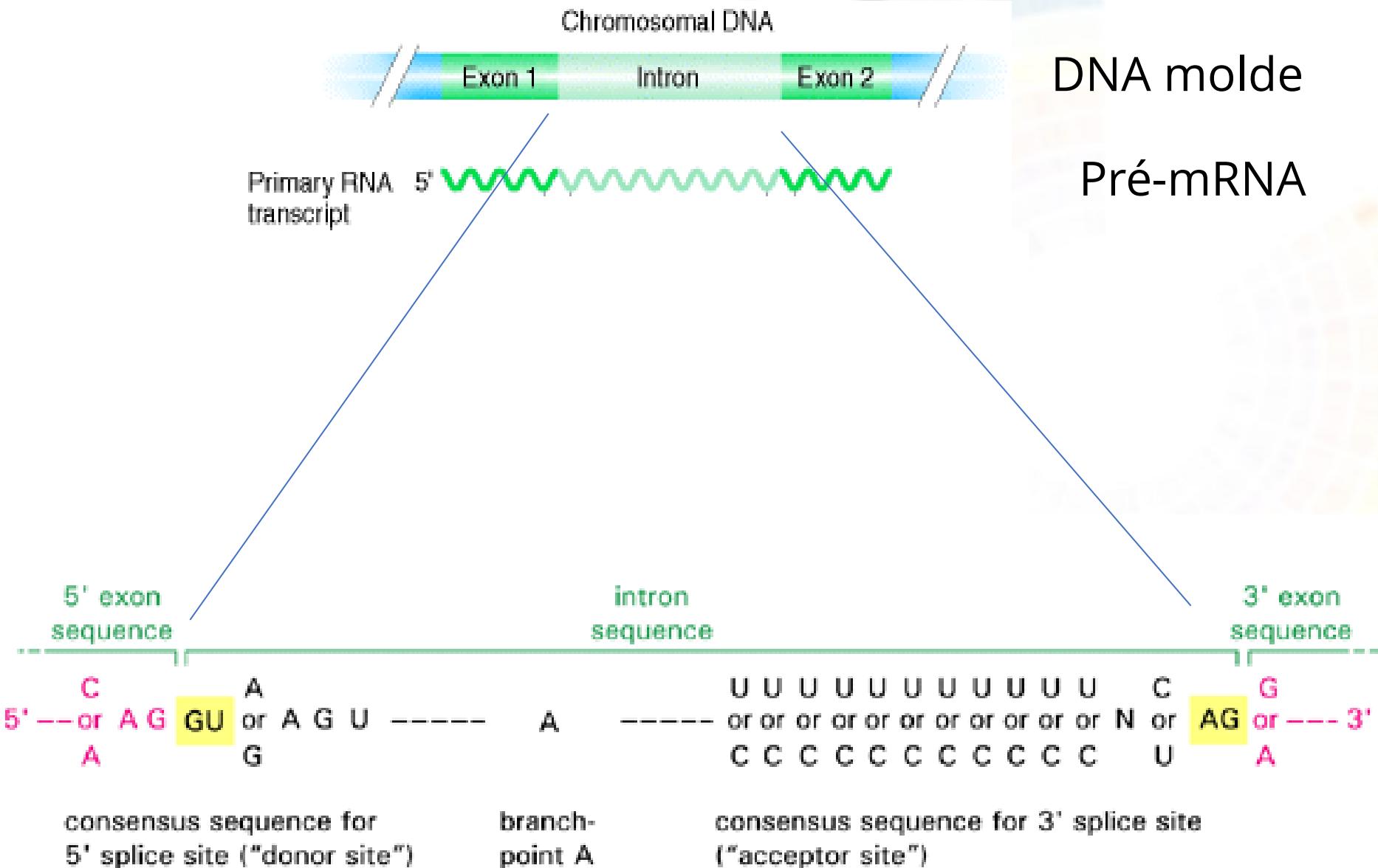
Estrutura de genes eucarióticos



mRNA structure

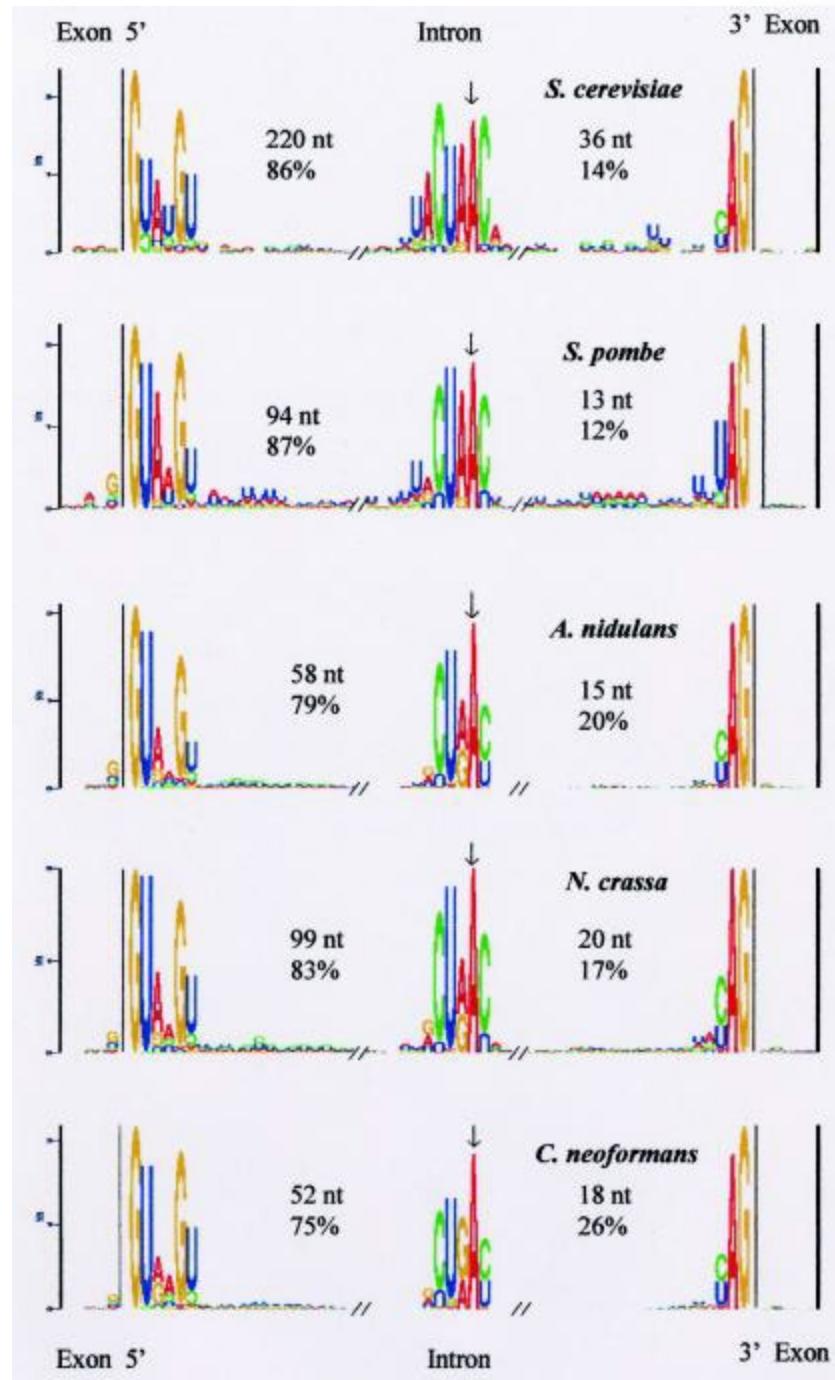


Recomposição de mRNA – remoção de introns



Sítios canônicos e não canônicos de splicing em animais, fungos e plantas

	GT-AG	GC-AG	AT-AC	Others
Animals	98.334%	0.983%	0.106%	0.577%
Fungi	98.715%	1.009%	0.019%	0.257%
Plants	97.886%	1.488%	0.092%	0.534%



Eukaryot Cell. 2004 Oct; 3(5):
1088-1100.doi:
10.1128/EC.3.5.1088-
1100.2004

Sítios adjacentes ao sítio de splicing GA-AG

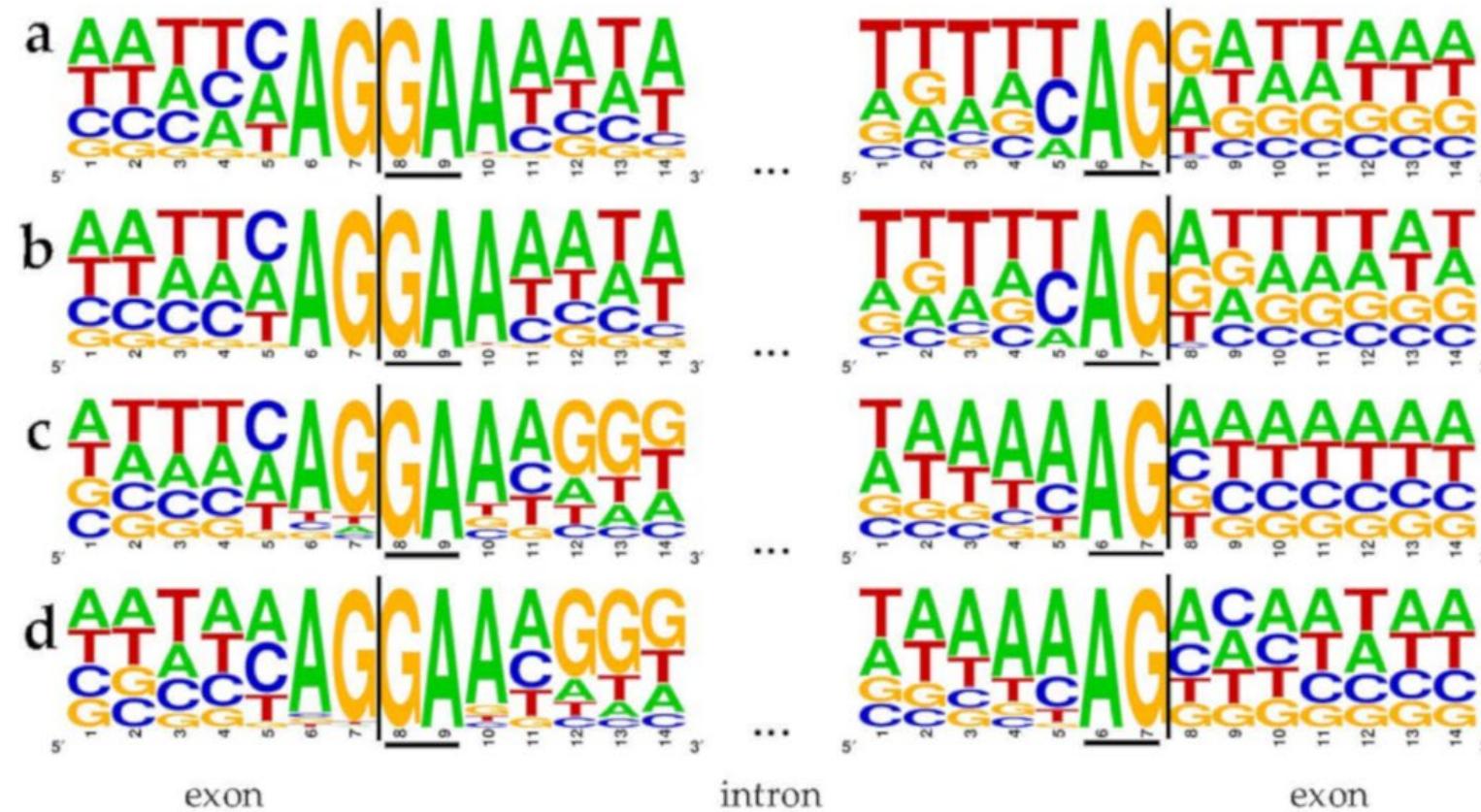
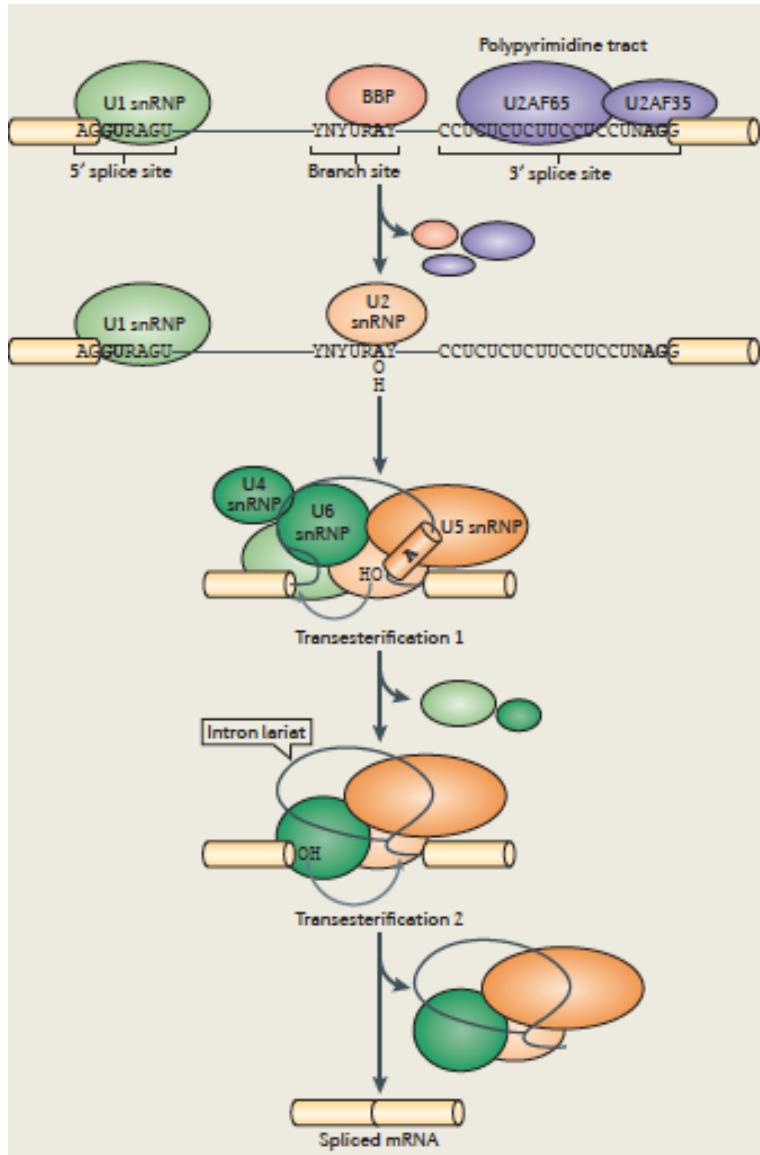


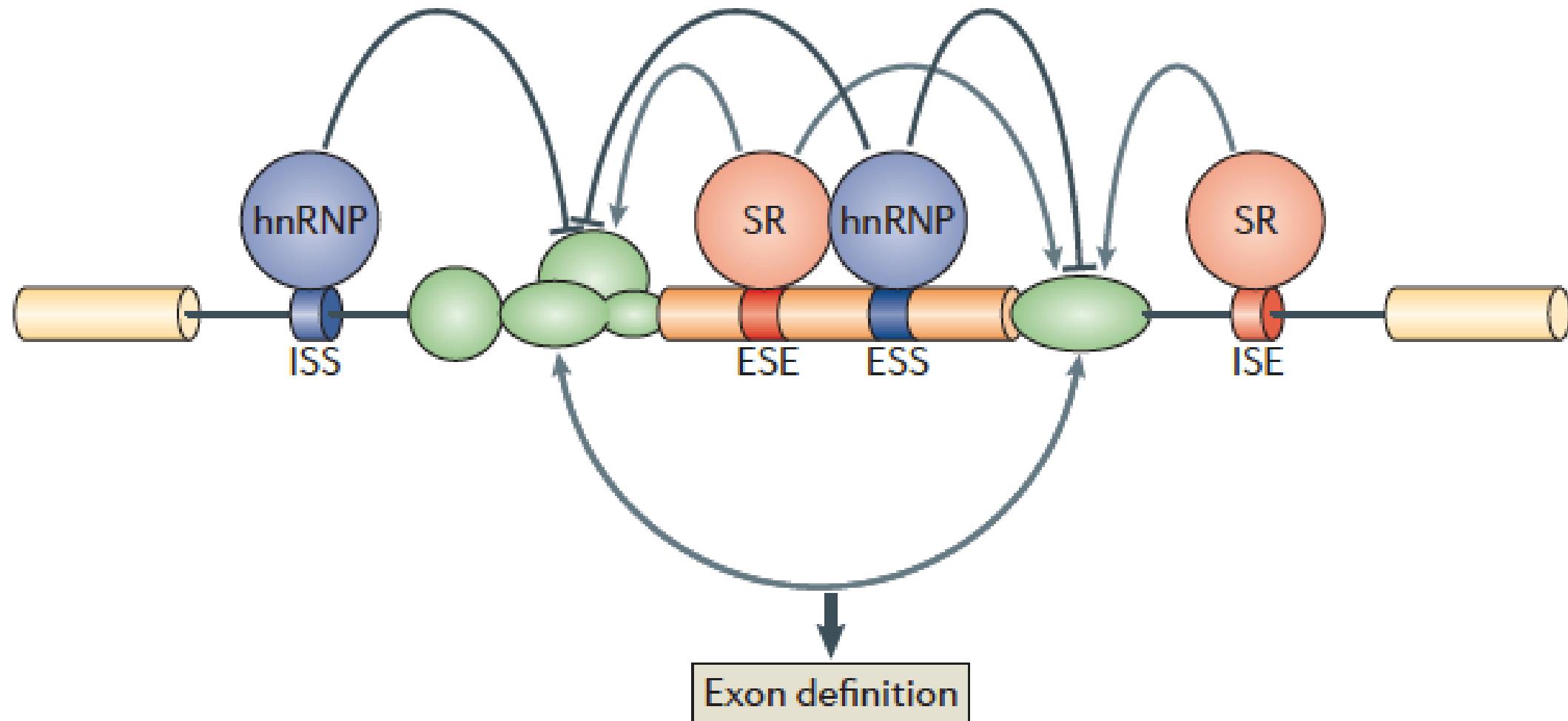
Figure 2. Flanking positions of GA-AG splice site combinations in *Eurytemora affinis* (**a,b**) and *Oikopleura dioica* (**c,d**). All splice site combinations (**a,c**) as well as all 5795 with RNA-Seq data supported splice site combinations (**b,d**) of these two species were investigated. Seven exonic and seven intronic positions are displayed at the 5' and 3' splice sites. Underlined bases represent the terminal dinucleotides of the intron, i.e., the 5' and 3' splice site.

Recomposição de mRNA – remoção de introns

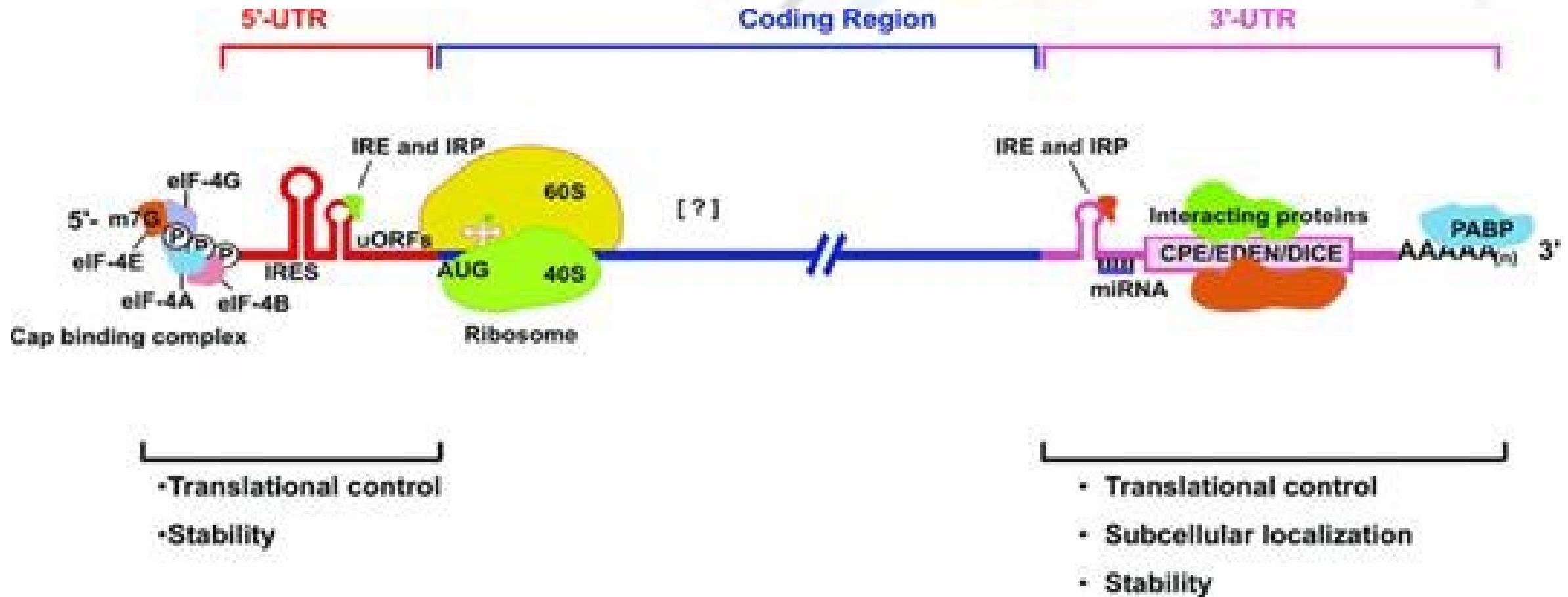


- Ocorre no núcleo
- Mediado pelo spliceosoma
- Unidades do spliceosoma:
 - Ribonucleoproteínas
 - Proteínas e RNAs

Recomposição de mRNA – remoção de introns



Estrutura e organização de mRNA



Introns em fungos

Evolução dos íntrons ao longo da maioria das linhagens fúngicas:

- perda frequente de íntrons
- relativamente poucos casos de ganho de íntrons
- retenção de íntrons ancestrais

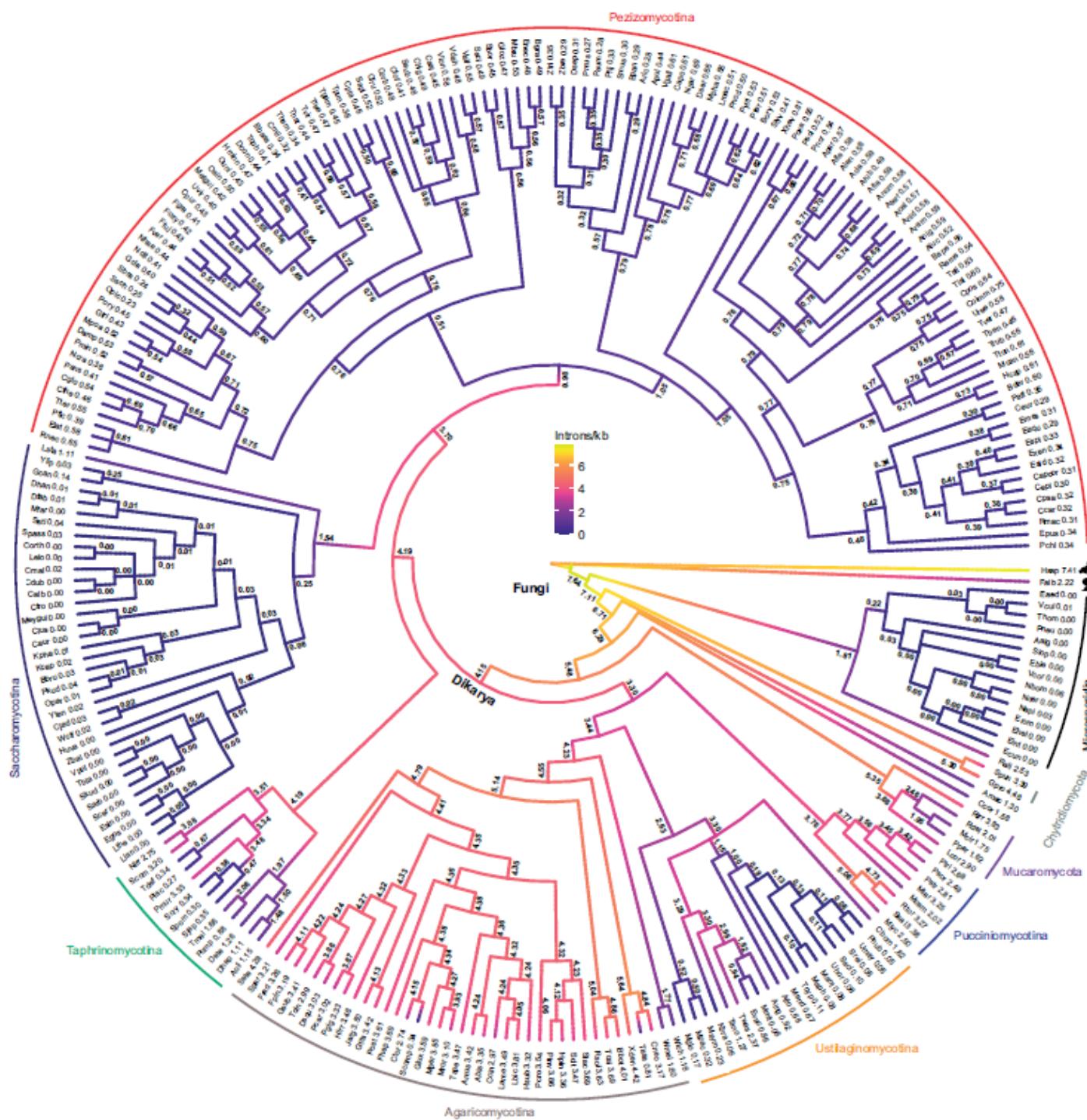
Exemplo:

Apenas 4% dos genes de *S. cerevisiae* possuem íntrons.

Cryptococcus neoformans: densidade de íntrons relativamente alta de 4 íntrons/kb

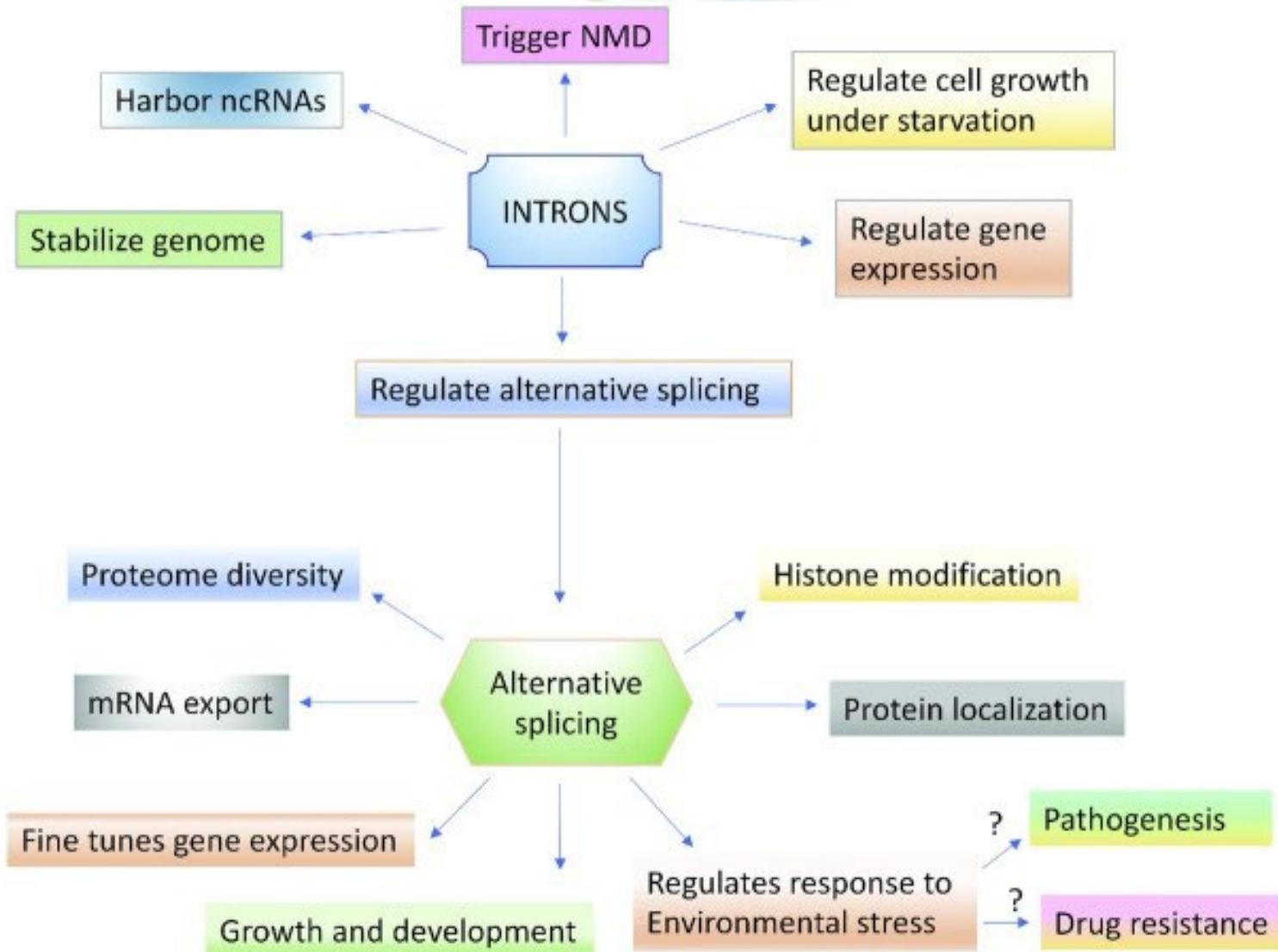
Número total de íntrons juntamente com as porcentagens correspondentes de genes submetidos a AS (ou expressando mais de uma isoforma) em diferentes genomas fúngicos

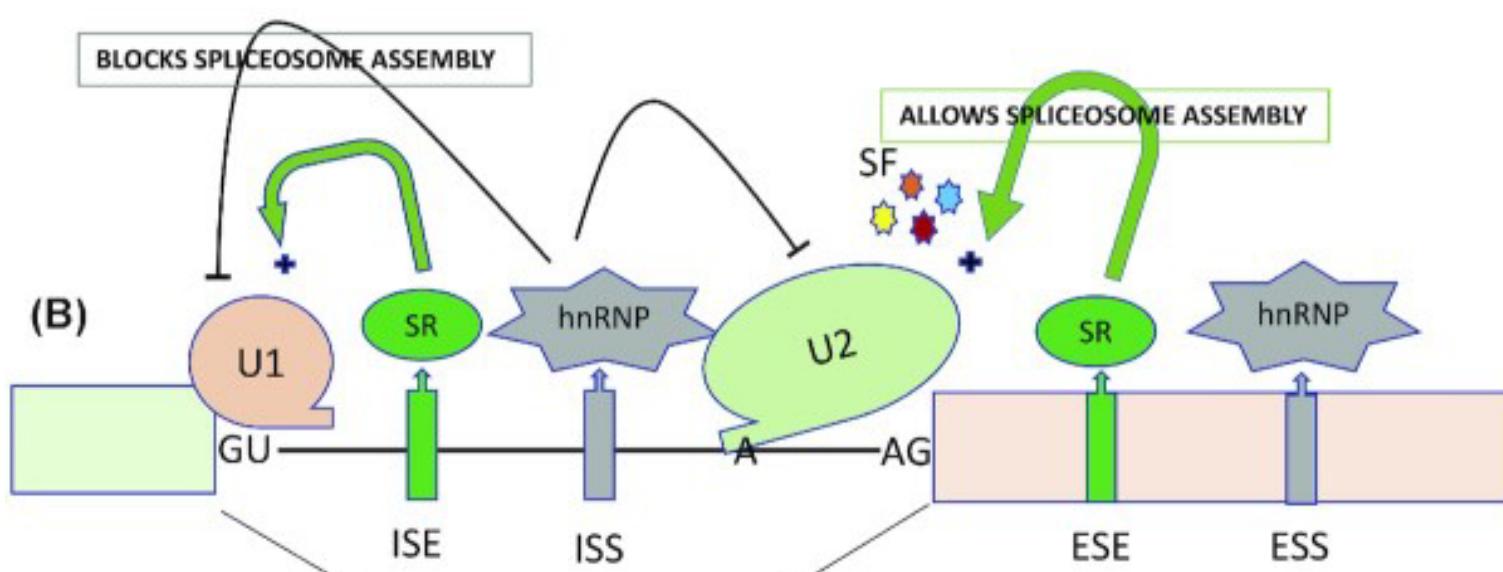
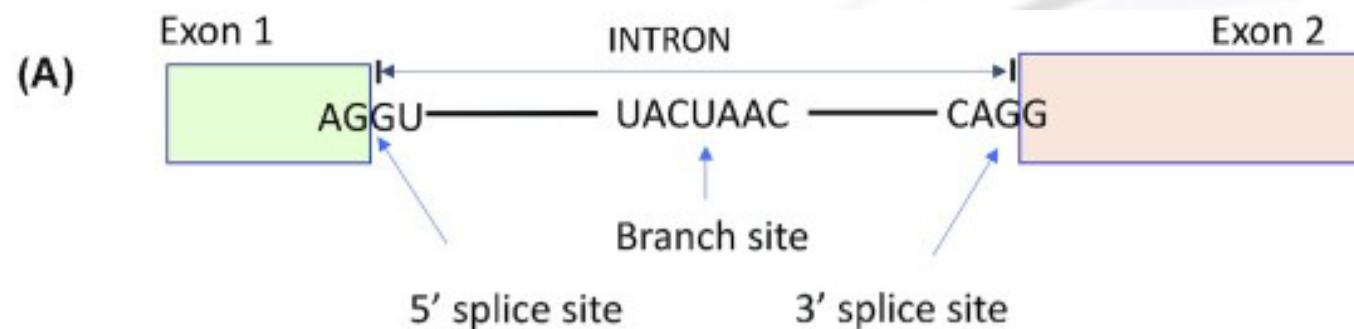
Fungi	Number of genes	Number of introns	% of genes undergoing AS
<i>Ascomycota</i>			
<i>Arthroderma benhamiae</i>	7984	10 332	8.2
<i>Histoplasma capsulatum</i>	11 216	42 485	38.82
<i>Paracoccidioides brasiliensis</i>	9132	28 179	15.4
<i>Coccidioides immitis</i>	10 440	17 815	13.4
<i>Aspergillus nidulans</i>	9541	16 797	7.3
<i>Aspergillus niger</i>	10 597	17 668	9.5
<i>Aspergillus fumigatus</i>	10 144	30 312	30.21
<i>Neurospora crassa</i>	9841	14 323	8.8
<i>Podospora anserina</i>	10 257	11 261	4.8
<i>Saccharomyces cerevisiae</i>	5781	358	0.18
<i>Candida albicans</i>	6620	867	4.08
<i>Candida glabrata</i>	5632	292	1.95
<i>Schizosaccharomyces pombe</i>	5073	3878	0.6
<i>Trichoderma reesei</i>	9143	18 802	2.5
<i>Trichoderma longibrachiatum</i>	10 792	16 465	48.9
<i>Botryotinia fuckeliana</i>	16 389	22 334	2.7
<i>Gibberella zeae</i>	23 218	38 261	5.9
<i>Magnaporthe grisea</i>	14 010	18 795	7.9
<i>Mycosphaerella graminicola</i>	10 952	17 661	6.1
<i>Phaeosphaeria nodorum</i>	15 983	21 371	2.4
<i>Pichia stipitis</i>	5807	2580	0
<i>Basidiomycota</i>			
<i>Cryptococcus neoformans</i>	6604	40 000	59
<i>Coprinopsis cinerea</i>	13 544	30 180	8.6
<i>Laccaria bicolor</i>	18 216	36 757	5.9
<i>Phanerochaete chrysosporium</i>	10 048	48 688	7.7
<i>Ustilago maydis</i>	6522	4279	2.3
<i>Mucoromycotina</i>			
<i>Rhizopus oryzae</i>	17 459	40 515	2.3
<i>Lichtheimia corymbifera</i>	11 350	54 131	18.98

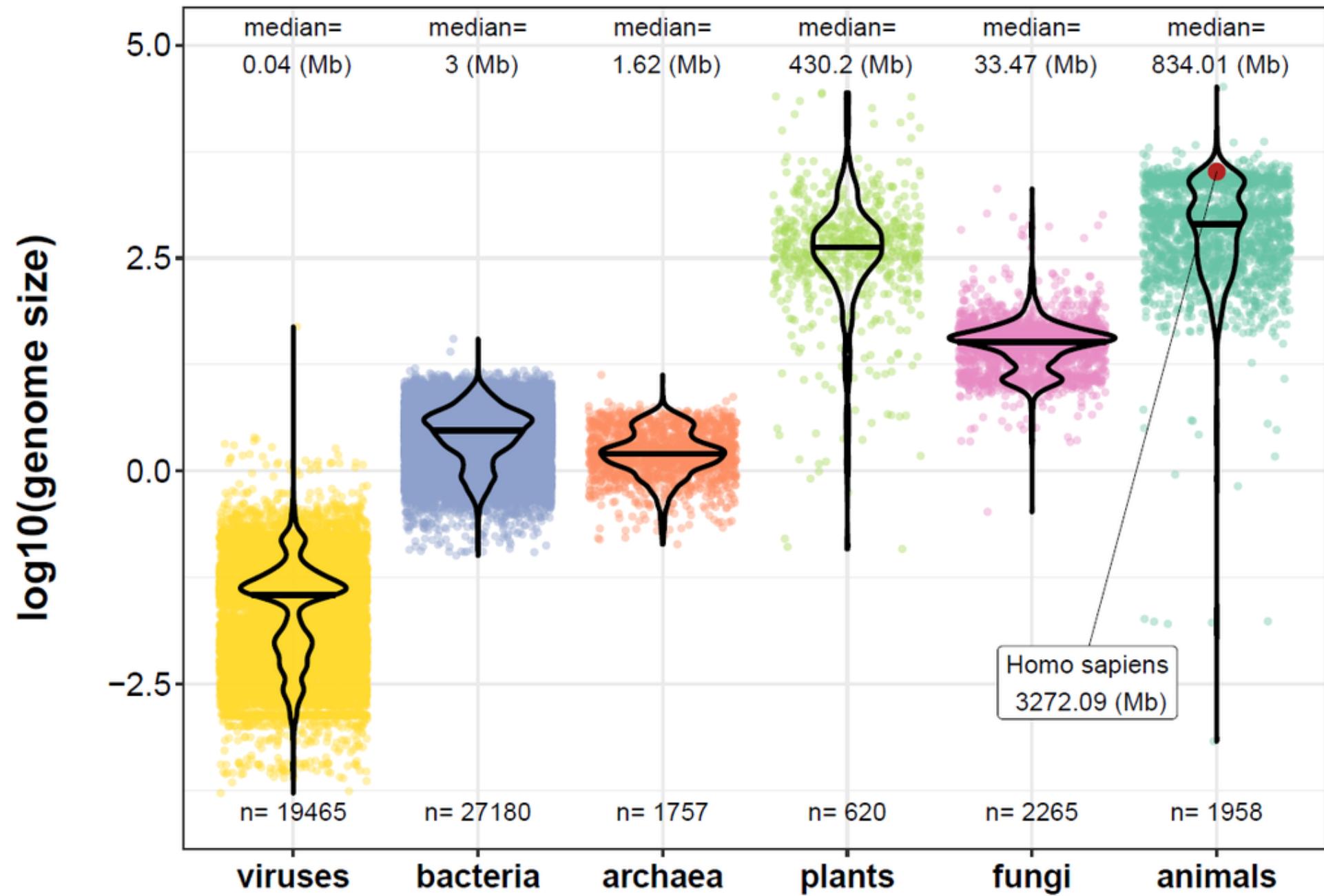


Mol. Biol. Evol.
38(10):4166–4186
doi:10.1093/molbev/msab094 Advance Access
publication March 27, 2021

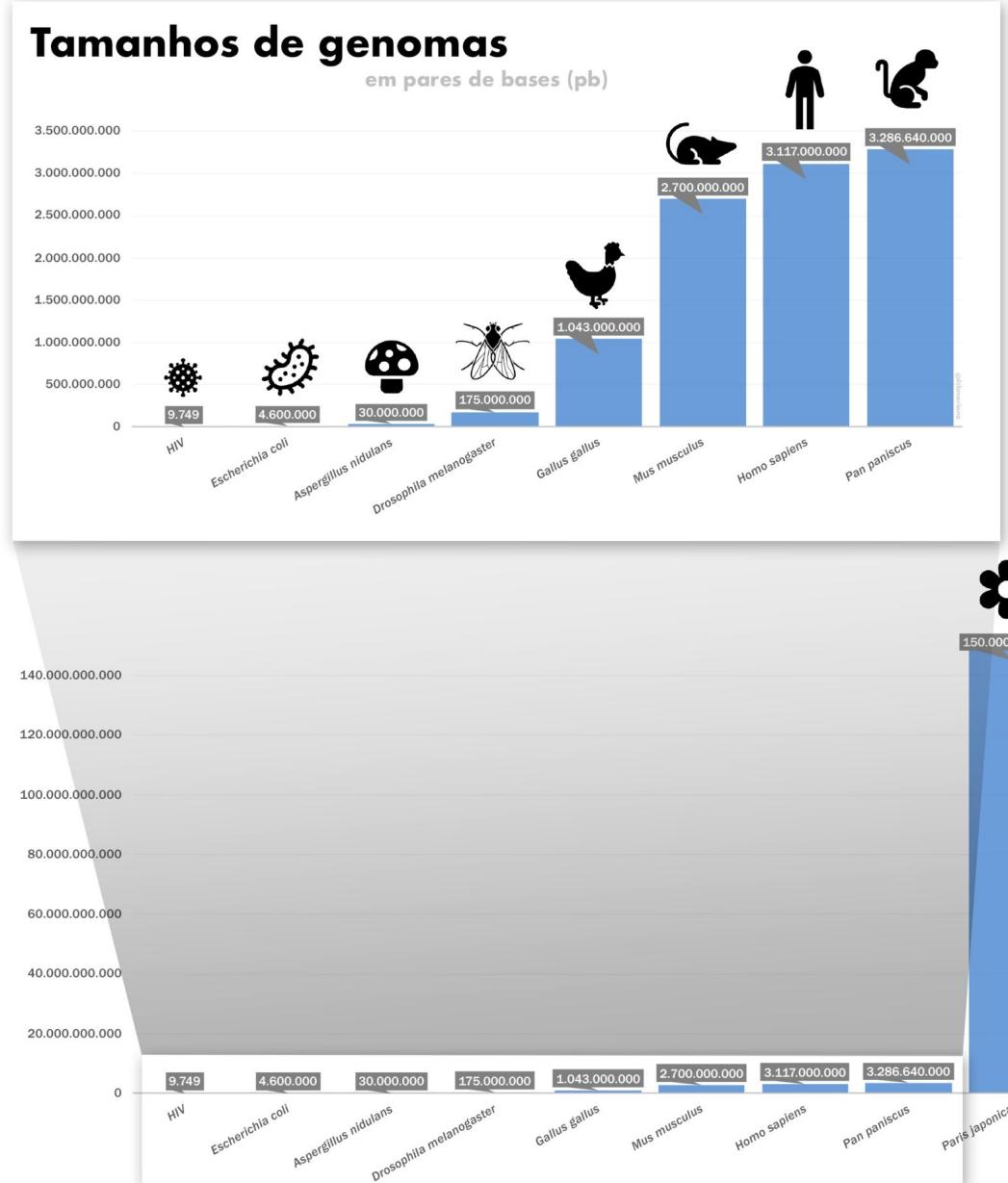
Representação esquemática da regulação intrônica de vários processos, incluindo AS em várias espécies de fungos







Variação no tamanho dos genomas



Genome size

(A species comparison)

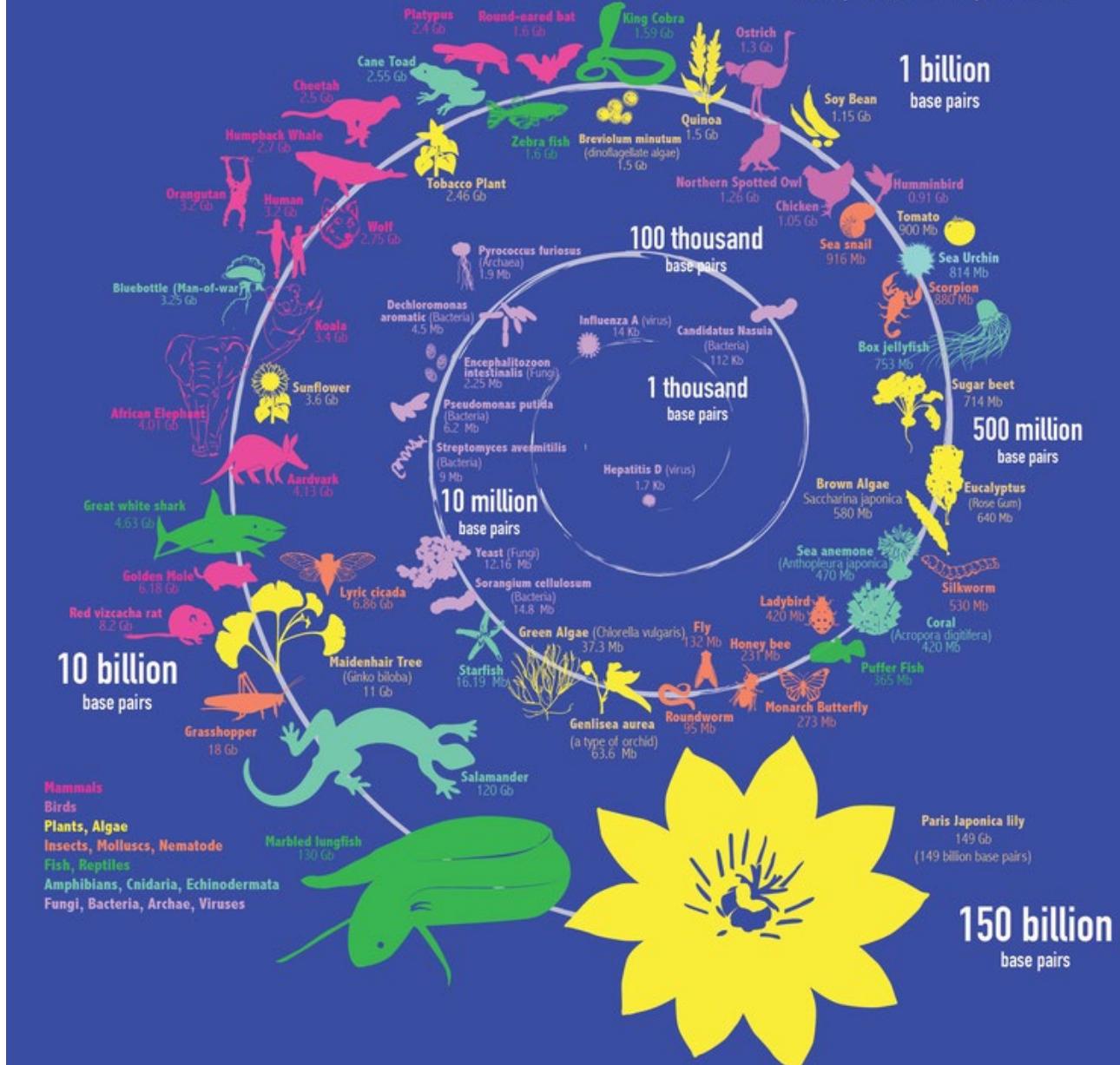


Table 1 | Genome and lifestyle characteristics of different fungal plant pathogens

Species (isolate)	Isolate-specific host	Species-specific host	Lifestyle	Reproduction	Genome size (Mb)*	% Repetitive sequence*	Genome features linked to pathogenicity‡	Refs
<i>Sporisorium scitamineum</i> (Sscl8)	Sugar cane	Sugar cane	Biotroph	Sexual	20	6.7	Repeat-rich gene clusters that encode effector candidates	48
<i>Ustilago maydis</i> (521)	Maize	Maize	Biotroph	Sexual	20	6.7	Repeat-rich gene clusters that encode effector candidates	48, 66
<i>Microbotryum lychnidis-dioicae</i> (p1A1 Lamole)	Red campion	Red campion	Biotroph	Sexual	33	14	Repeat-rich gene clusters that encode effector candidates	103, 104
<i>Melampsora larici-populina</i> (98AG31)	Poplar and larch	Poplar and larch	Biotroph	Sexual and asexual	101	45	ND#	73
<i>Puccinia graminis</i> f. sp. <i>tritici</i> (CDL75-36-700-3, race SCCL)	Wheat	Wheat and barley	Biotroph	Sexual and asexual	89	45	Highly polymorphic effector candidates	73
<i>Zymoseptoria tritici</i> (IPO323)	Wheat	Wheat	Hemi-biotroph	Sexual and asexual	40	<ul style="list-style-type: none"> • 18.6 (genome mean) • 16.6 (core) • 33.6 (accessory) 	<ul style="list-style-type: none"> • Orphan regions are enriched in <i>in planta</i>-expressed genes • Possible function of accessory chromosomes in virulence 	45, 57, 58
<i>Leptosphaeria biglobosa</i> 'canadensis' (J154)	Mustard	Crucifers	Necrotroph	Sexual and asexual	32	3.9	ND#	71
<i>Leptosphaeria maculans</i> 'brassicae' (v23.1.3)	Oilseed rape	Crucifers	Hemi-biotroph	Sexual and asexual	45	35.5 (99.8% of all repeats located in AT-isochores)	<ul style="list-style-type: none"> • Enrichment of effector candidates and chromatin-mediated effector candidate regulation in AT isochores • Conditionally dispensable chromosome contains avirulence-encoding gene 	29, 60

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

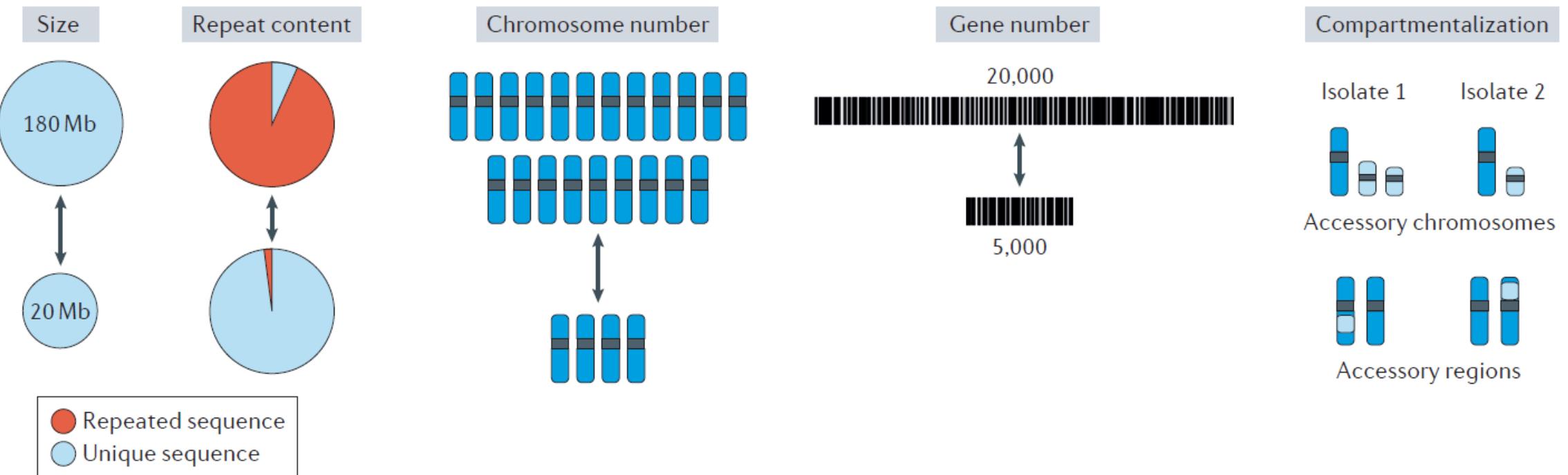
Species (isolate)	Isolate-specific host	Species-specific host	Lifestyle	Reproduction	Genome size (Mb)*	% Repetitive sequence*	Genome features linked to pathogenicity [‡]	Refs
<i>Blumeria graminis</i> f. sp. <i>tritici</i> (96224)	Wheat	Various	Biotroph	Sexual and asexual	180	90	Presence and/or absence 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> /Nectria haematococca 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.

Genomas de fungos fitopatogênicos

Variação em tamanho de genoma, sequências repetitivas, número de cromossomos, número de genes e compatimentalização

a Fungal plant pathogen genomes are highly diverse



3D Genome of *Penicillium oxalicum*

TABLE 1 Chromosome-level genome features of *P. oxalicum* strain HP7-1^a

Genome features	Value
Chromosome features	
No. of scaffolds	8
No. of contigs	13
Size of scaffolds (Mb)	30.64
GC content of scaffolds (%)	50.69
N_{50} scaffold length (bp)	4,068,681
N_{90} scaffold length (bp)	2,593,408
Maximum scaffold length (bp)	5,870,233
Minimum scaffold length (bp)	1,795,220
Avg no. of contigs per scaffold	1.63
Protein-coding genes of scaffolds	9,728
Avg gene length (bp)	1,615.07
Avg CDS length (bp)	1,412.63
GC content of CDS sequences (%)	54.44
Genome features	
Avg no. of exons per gene	2.94
Avg no. of introns per gene	1.94
Complete BUSCOs of genome (%)	98.6
rRNAs	45
tRNAs	201

^aBUSCO, Benchmarking Universal Single-Copy Orthologs; CDS, coding DNA sequence.

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. citriasiiana</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.

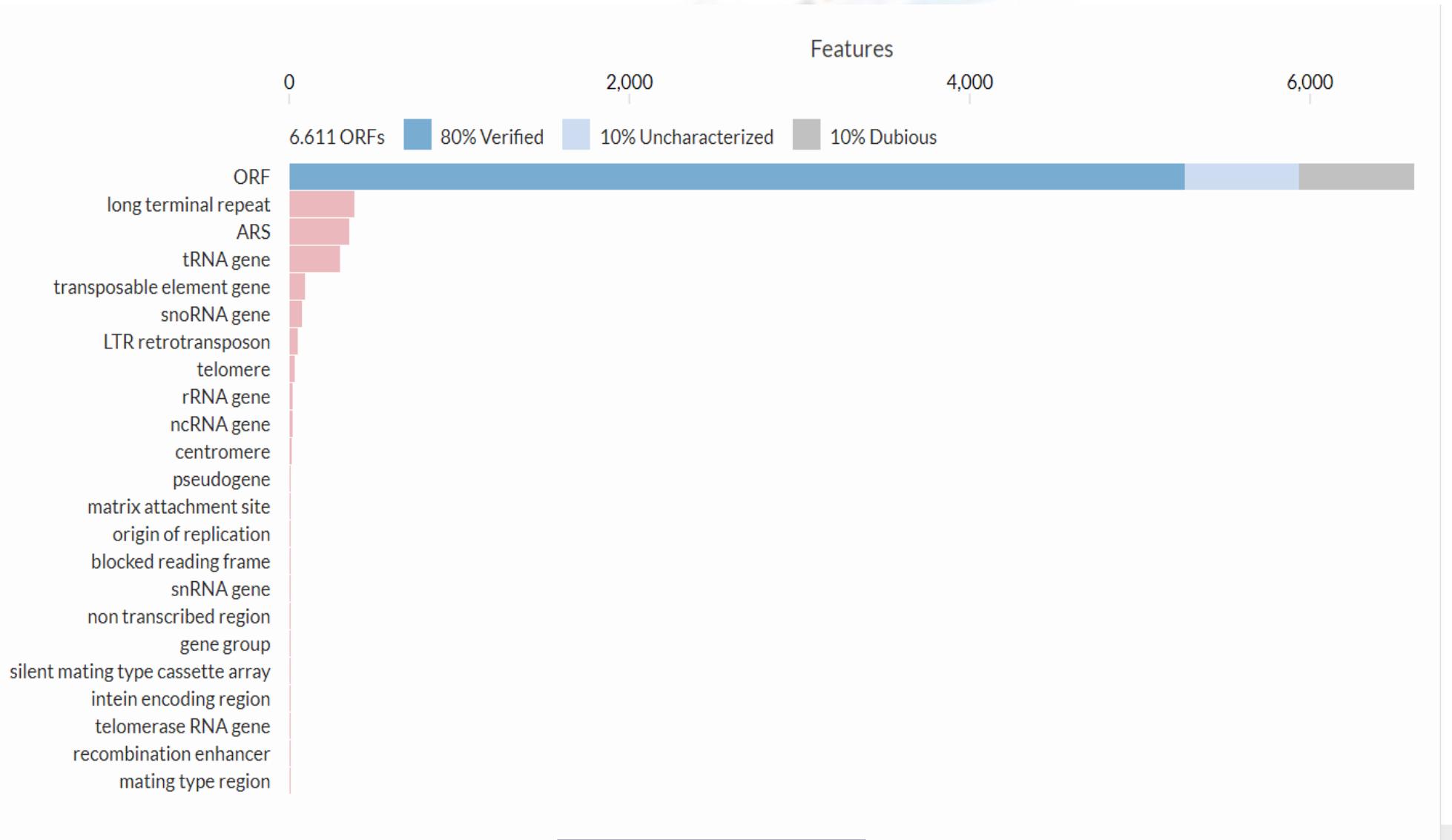
Table 1 General features of *Diaporthe* genomes

<i>Diaporthe</i> species	<i>D. caulinora</i>	<i>D. capsici</i>	<i>D. citri</i>	<i>D. destruens</i>	<i>D. longicolla</i>	<i>D. phragmitis</i>
Total genome size (bp)	57,864,239	57,558,510	63,685,968	56,108,228	64,714,568	58,328,132
Coverage (X)	270	ND	271	121	145	50
Number of contigs	10	20	34	47	985	28
Maximum contig length (bp)	14,464,108	8,755,198	12,370,252	6,293,594	1,124,325	7,711,659
N50 (bp)	8,708,519	5,171,887	5,472,022	2,479,481	204,364	3,550,333
GC content (%)	52.97	51.27	46.72	48.7	48.26	50.82
Complete BUSCO (%)	97.8	ND	98.5	97.93	98.21	97.9
Duplicated BUSCO (%)	0.6	ND	ND	3.33	ND	1
Fragmented BUSCO (%)	0.43	ND	ND	ND	ND	ND
Missing BUSCO (%)	1	ND	ND	ND	ND	1.1
Protein-coding gene number	18,385	14,425	15,921	13,754	16,606	12,393
Total gene length (bp)	23,421,216	23,205,508	26,007,773	ND	28,344,980	16,320,211
Average gene length (bp)	1690	1609	1633.5	ND	1709	1317
Protein-coding gene number (Augustus with <i>D. caulinora</i>) ^a	18,385	15,675	15,113	13,717	15,232	15,655
Predicted secretome ^a	1501	1588	1383	1298	1535	1539
Predicted effectors > 80% probability ^a	1598	1204	1088	1103	1229	1168
Cytoplasmic effectors ^a	1448	1043	938	982	1073	1009
Apoplastic effectors ^a	150	161	150	121	156	159
Genbank accession no.	BioProject PRJNA717308	WNXA000000000	JACTAD0000000000	JACAAM0000000000	JUJX0000000000	JACDXY0000000000
References	this paper	[6]	[34]	[36]	[28]	[37]

ND no data

^a data from this paper

Saccharomyces cerevisiae Genome Snapshot



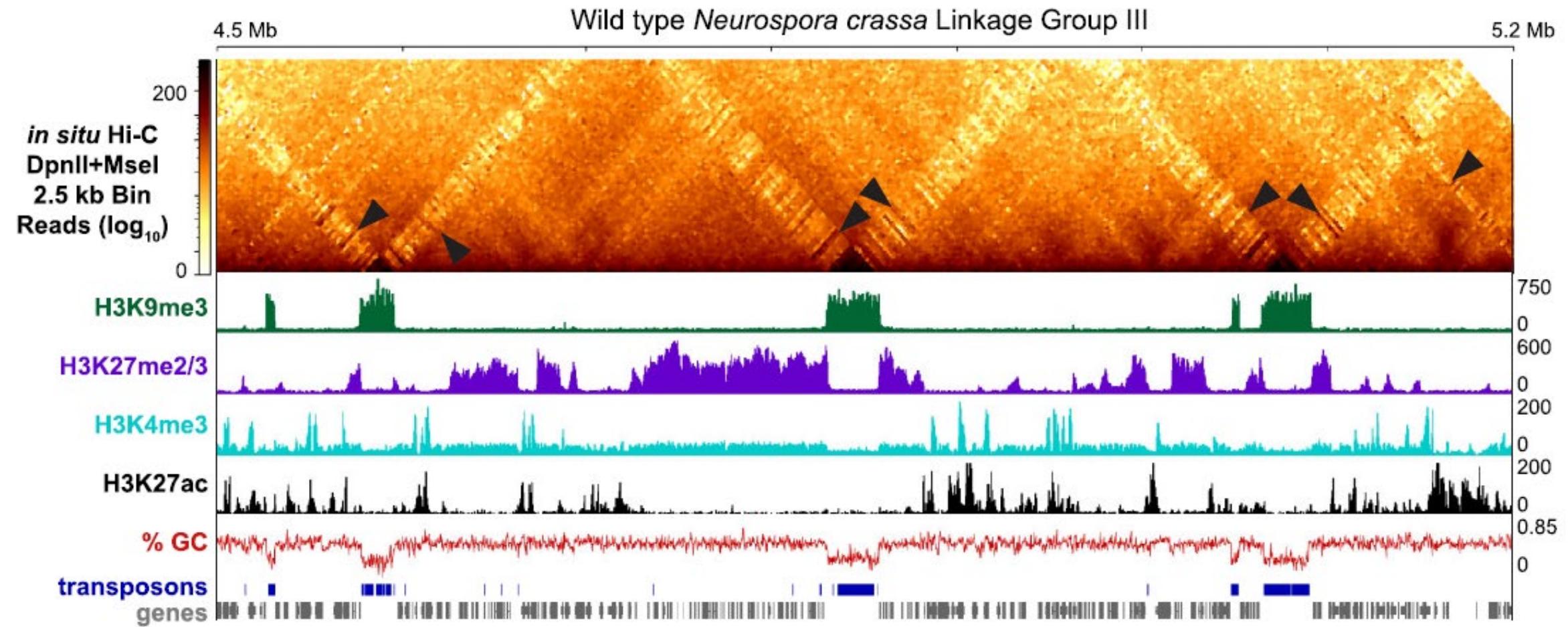
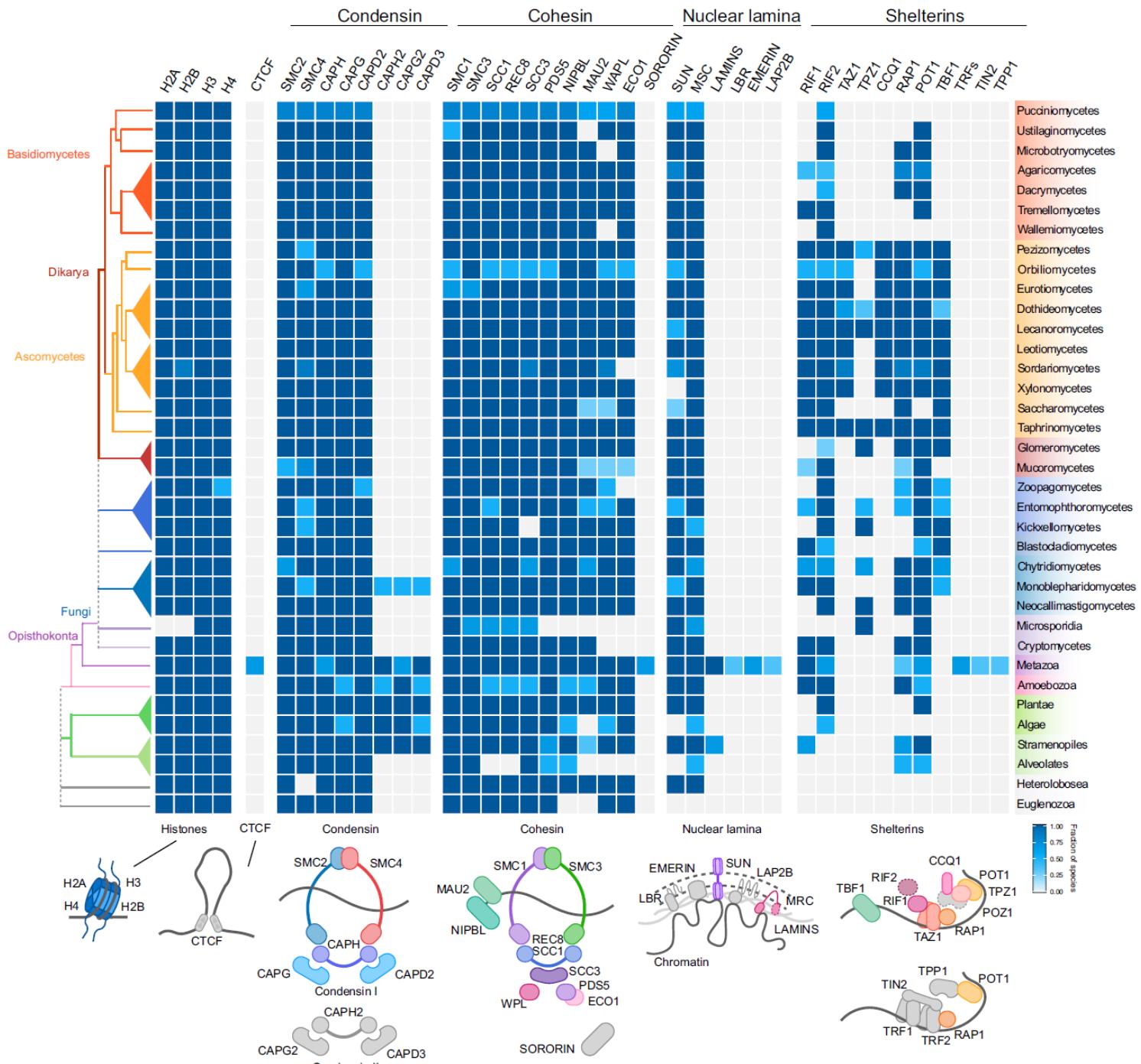


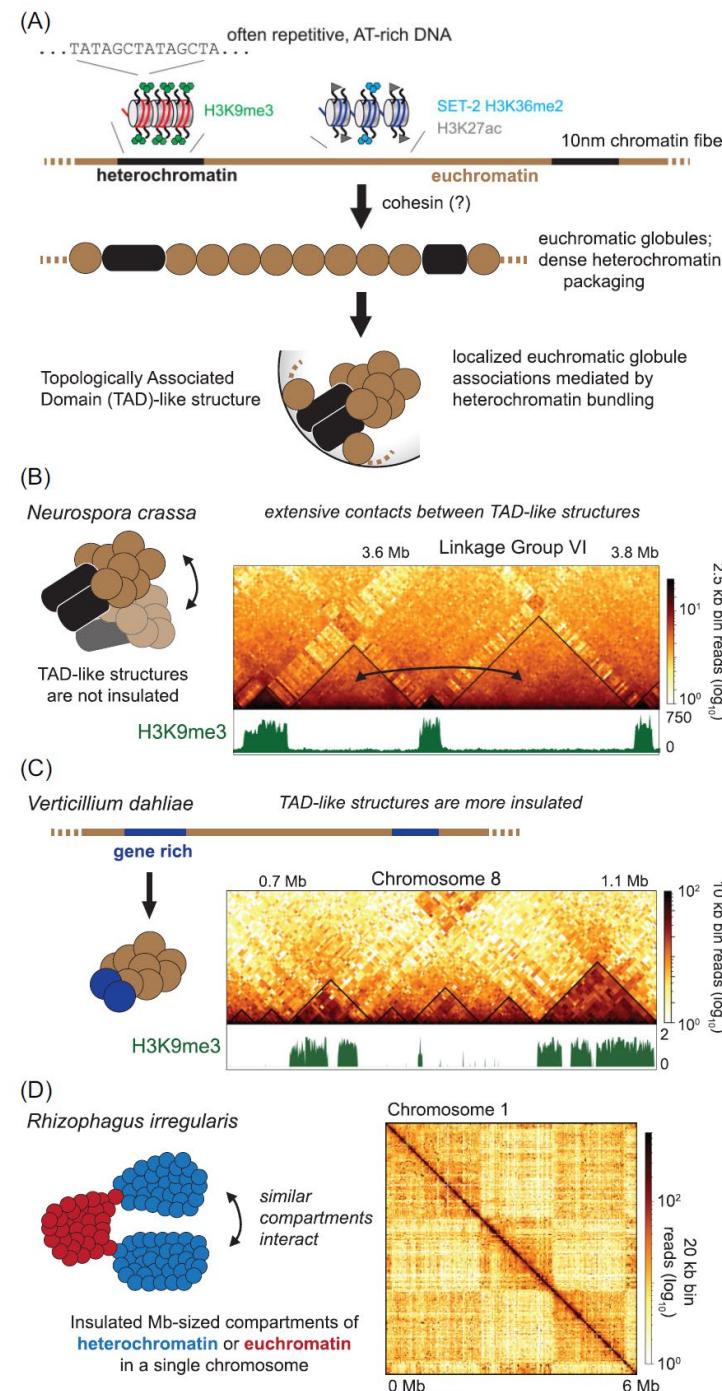
Figure 1. Chromatin profile and evidence for cross-compartment (heterochromatic–euchromatic) contacts in *N. crassa*. Histone post-translational modifications [both activating (H3K4me3 and H3K27ac) and repressive (H3K9me3 and H3K27me2/3)] demarcate the chromatin of *N. crassa*. Regions of the *N. crassa* genome that contain numerous transposon relicts are often rich in adenine/thymine base pairs (AT-rich) and devoid of genes.



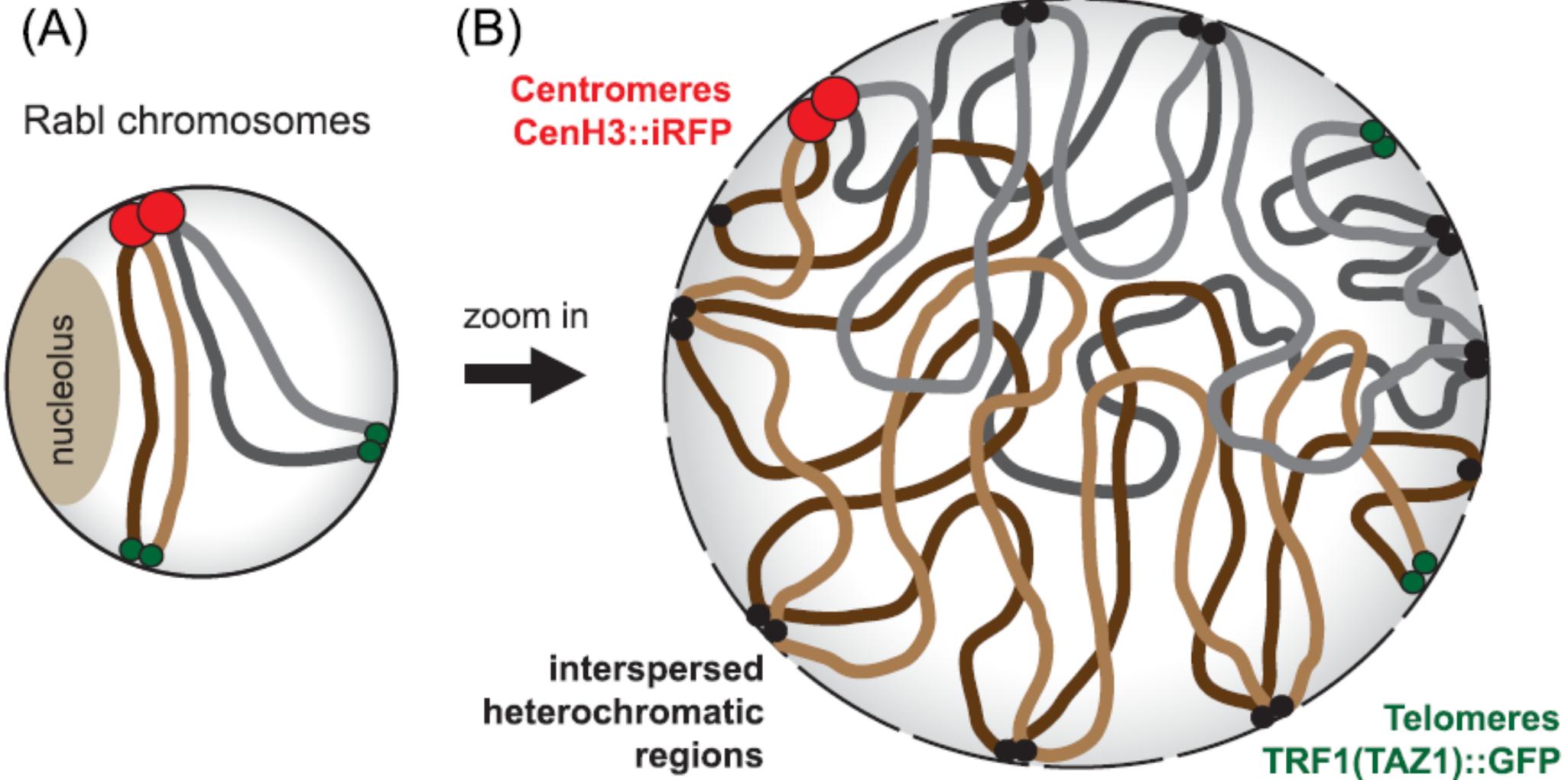
Gene orthologs encoding most of the proteins involved in genome organization are conserved in fungi

FEMS Microbiology Reviews, 2023, 47, 1–22 DOI: [10.1093/femsre/fuad021](https://doi.org/10.1093/femsre/fuad021)

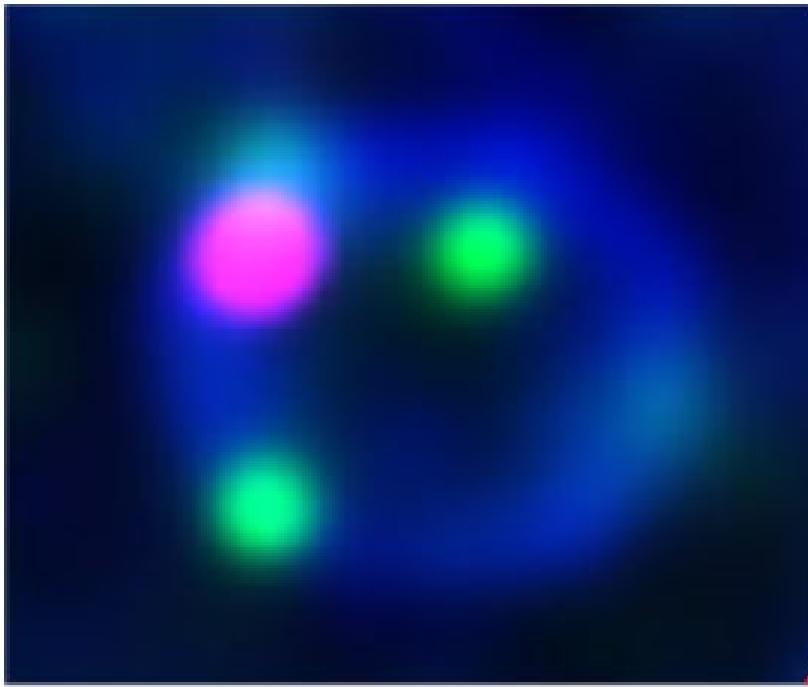
- FEMS Microbiology Reviews , 2023, 47



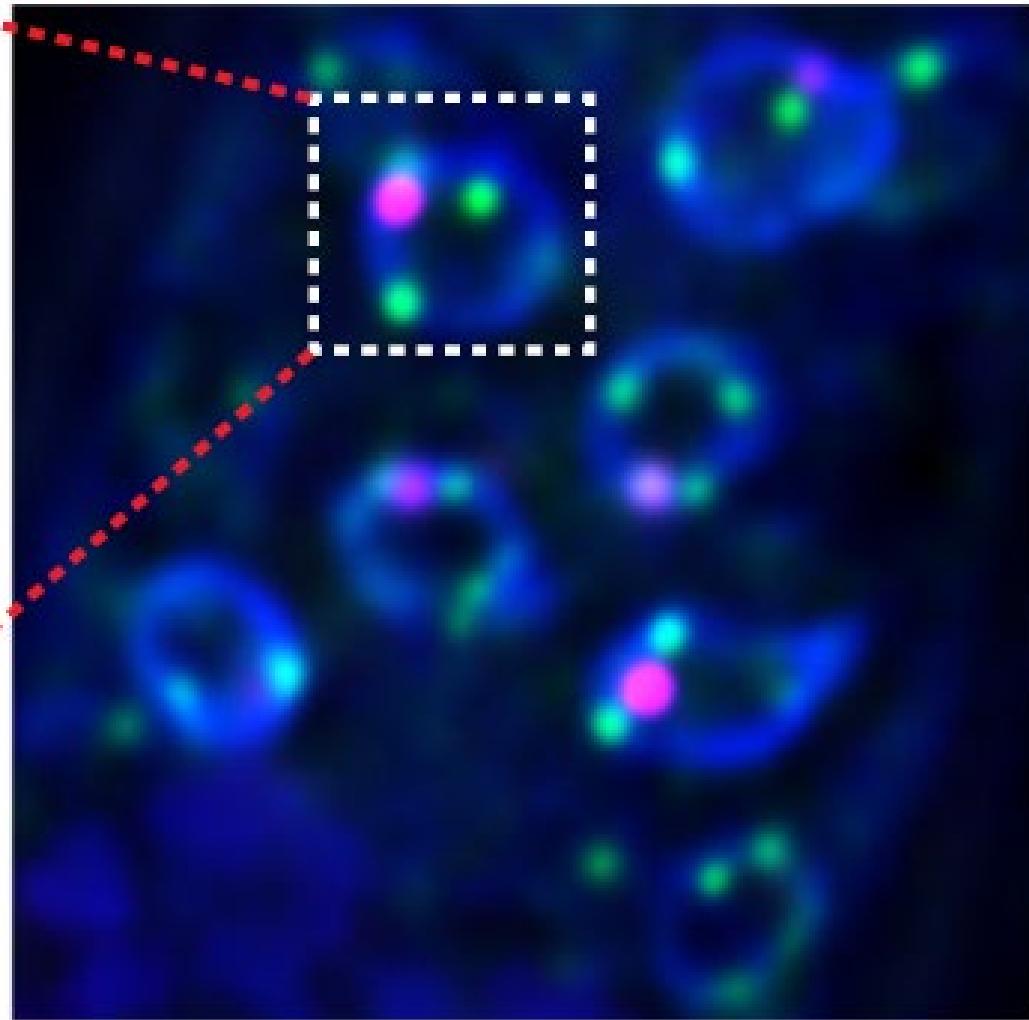
A formação da “conformação do cromossomo Rabl” em núcleos fúngicos (Carl Rabl)



Rabl Cromossomo em fungos (Carl Rabl)



Centromeres
Telomeres
Nuclear Membrane



Genoma 3D de *Penicillium oxalicum*

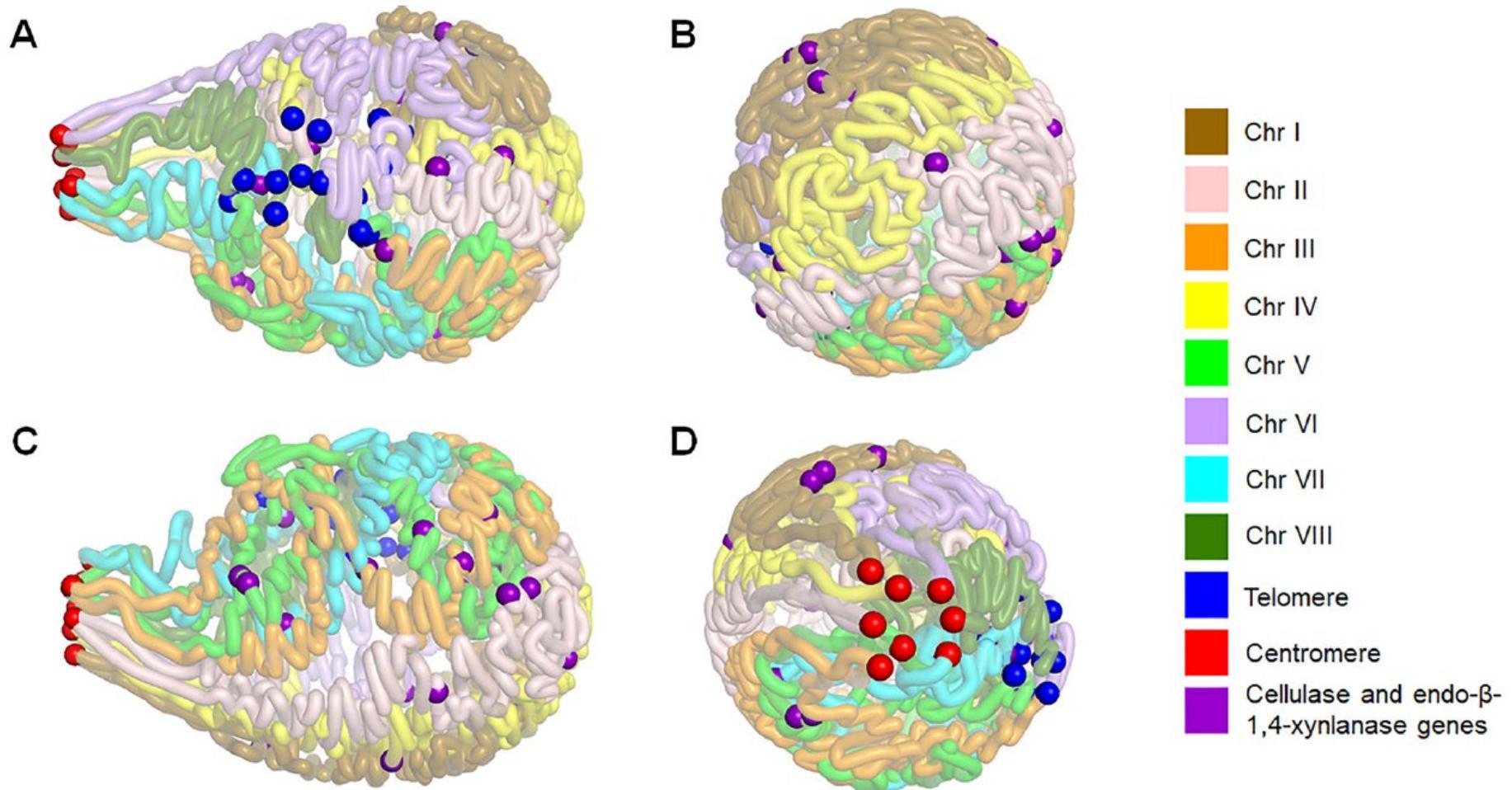
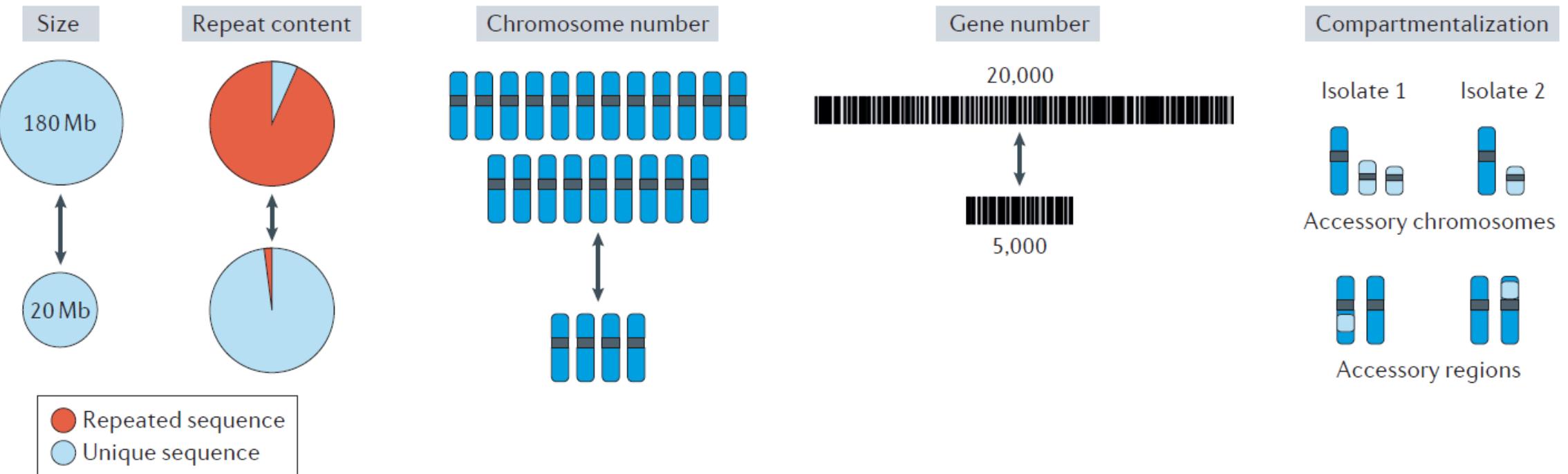


FIG 5 3D model of *P. oxalicum* genome under Avicel induction. (A) Front view. (B) Right flank view. (C) Back view. (D) Left flank view. Red, blue, and purple balls represent the predicted centrosomes, telomeres, and cellulase and endo- β -1,4-xylanase genes, respectively.

Genomas de fungos fitopatogênicos

Variação em tamanho de genoma, sequências repetitivas, número de cromossomos, número de genes e compatimentalização

a Fungal plant pathogen genomes are highly diverse



Diversidade de TEs entre genomas de fungos

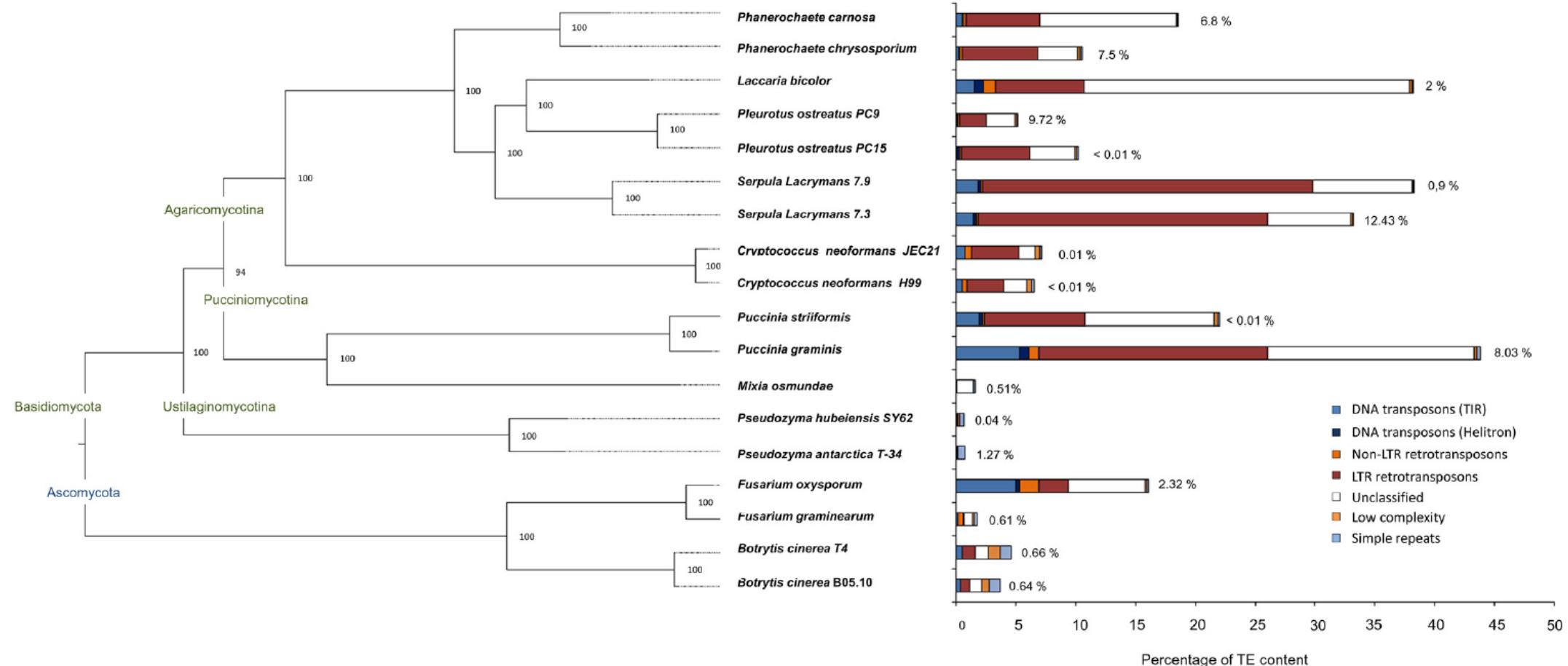
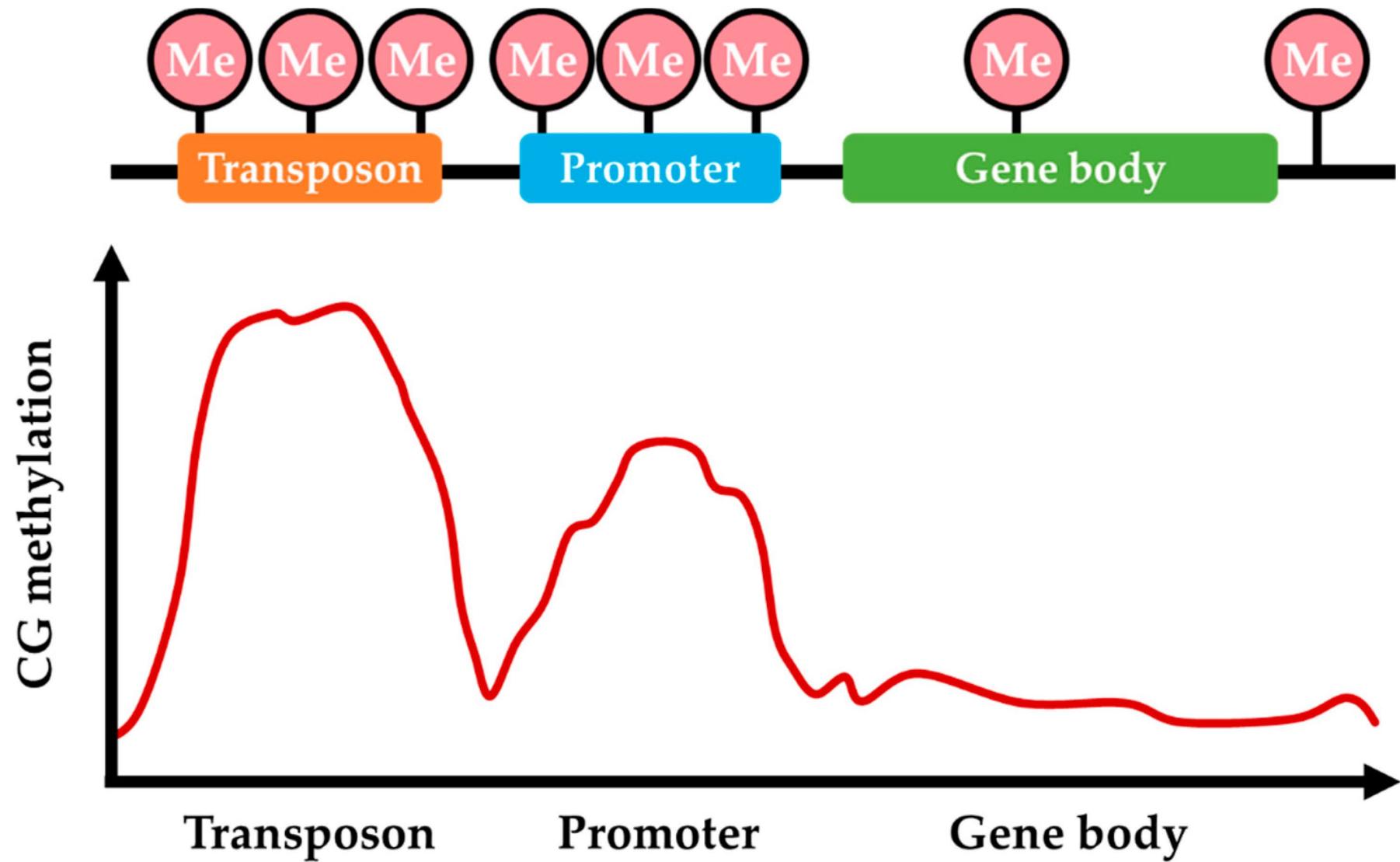


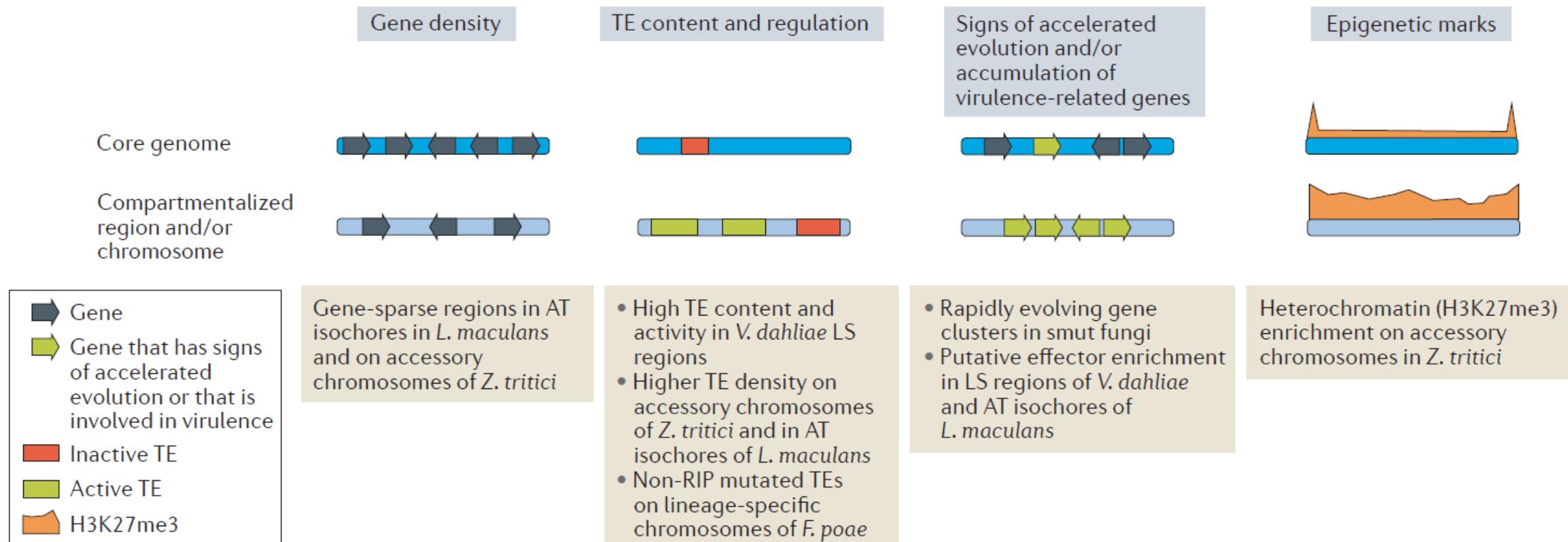
Fig 7. Phylogeny and repeat content of eighteen fungal species. Maximum-likelihood phylogeny inferred with RAxML based on 551 genes and 100 bootstraps. Percentages of assembly gaps are shown near to each bar. Dashed lines are used to align each branch to the tip.

Silenciamento de transposons



Compartimentos do genoma x core genoma

b Characteristics of genome compartments



Compartimentalização de genomas - Efetores

- Podem ser encontrados:
 - dispersos no genoma – associação com elementos repetitivos
 - organizados em clusters no genoma em regiões específicas - compartimentalização
- Compartimentalização do genoma – implicações:
 - Em nível funcional: co-regulação
 - Sequencias regulatórias compartilhadas
 - Regulação epigenética
 - Em nível evolutivo
 - Clusters podem ser co-transferidos horizontalmente como uma unidade de função
 - Co-segregação de alelos
 - Alta taxa de mutação

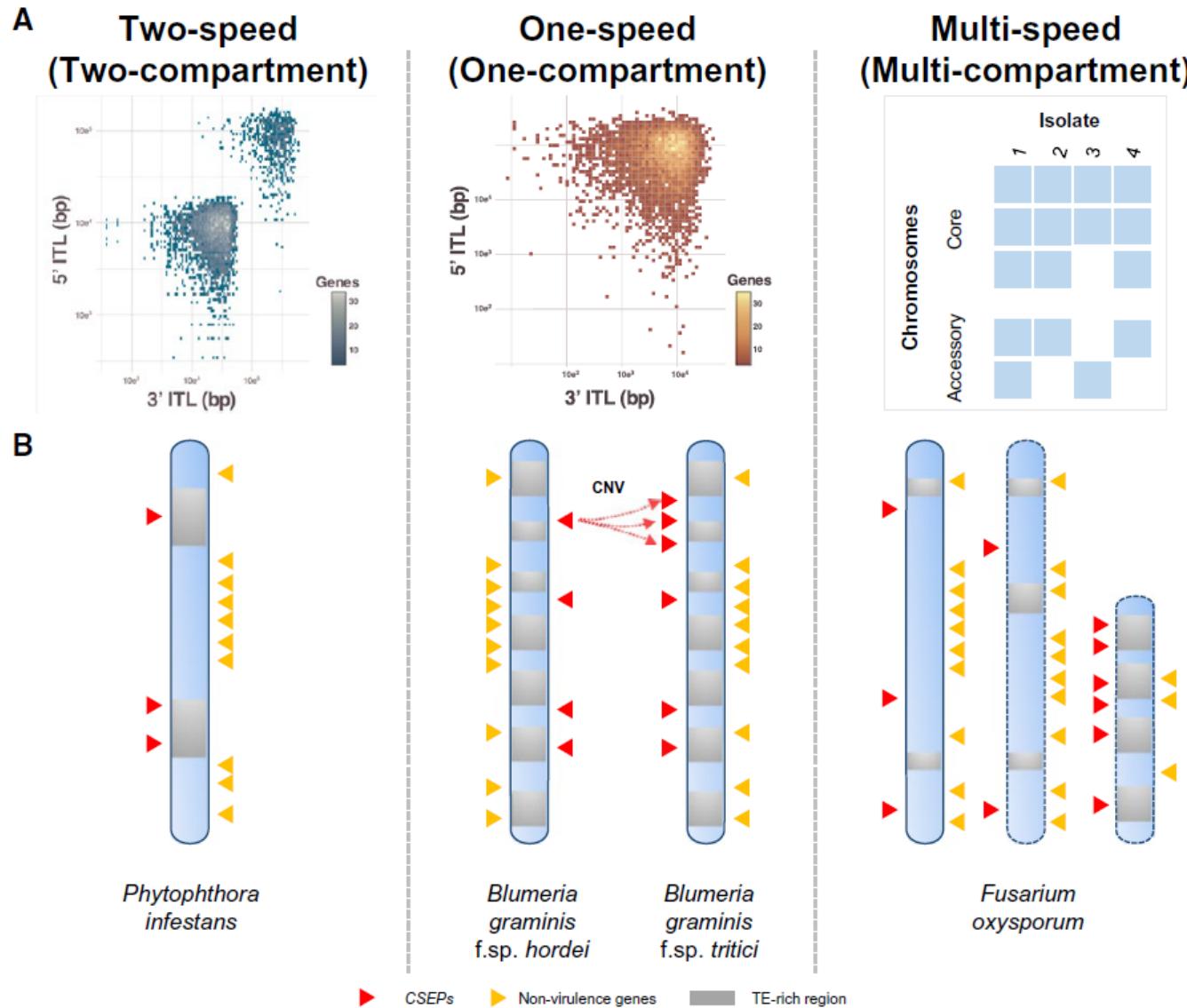
Conceito two-speed

- Housekeeping genes (HG)
 - Conservados - seleção
- Não HG
 - Exemplos:
 - efetores
 - metabolismo secundário
 - Rápida evolução - Isolados da mesma espécie:
 - Apenas moderada sintenia (colinearidade de genes)
 - Frequentemente polimorfismo gênico e cromossômico

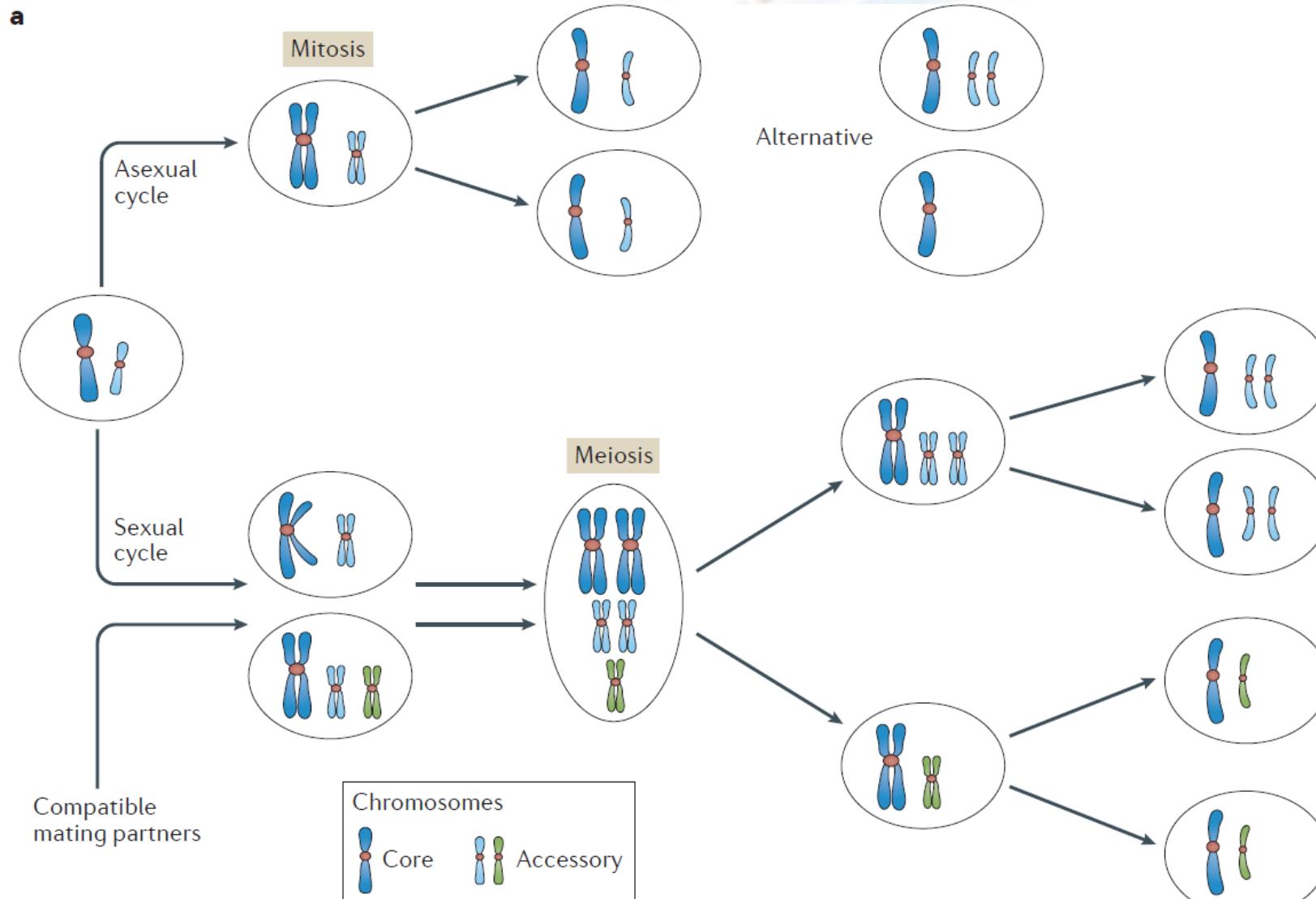
Conceito Genomas two-speed (Doing et al., 2015)

- Presença de compartmentalização física do genoma
- Diferentes taxas de evolução nos dois compartimentos
 - Taxa de mutação por unidade de tempo
- Genomas de fungos fitopatogênicos:
 - Compartimentalizados
 - Arquitetura de genoma mosaico – regiões:
 - densas em genes (genes housekeeping conservados) e pobres em sequencias repetitivas
 - Esparsas em genes e ricas em regiões repetitivas, contendo genes associados a virulência de rápida evolução
 - Presença de AT-rich isochores e cromossomos acessórios

Variações do conceito

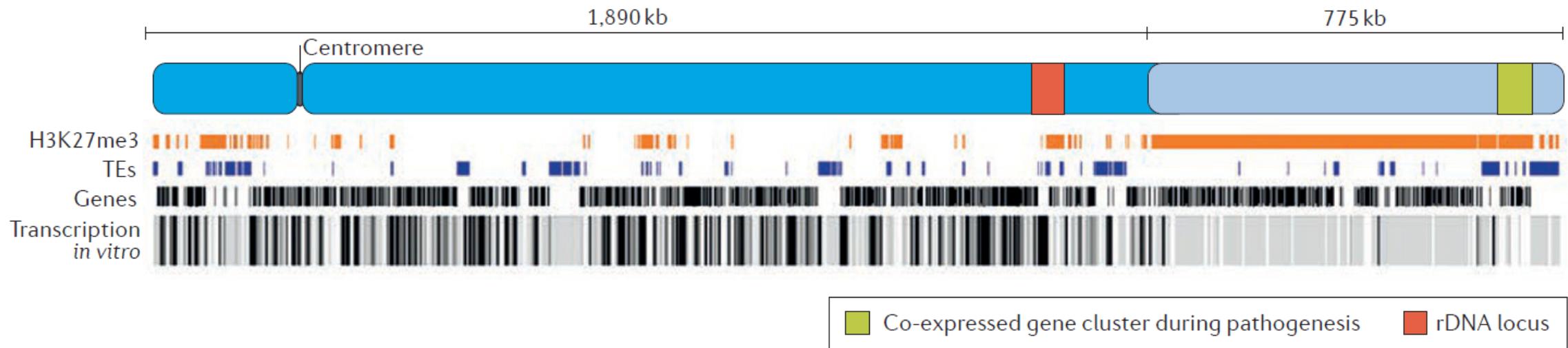


Dinâmica de cromossomos acessórios



Exemplo de região acessória e de core genoma no mesmo cromossomo

c Chromosome 7 of *Z. tritici* has characteristics of core and accessory chromosomes



Variação no tamanho dos genomas

Genoma humano: $3,2 \times 10^9$
Escherichia coli: $4,6 \times 10^6$

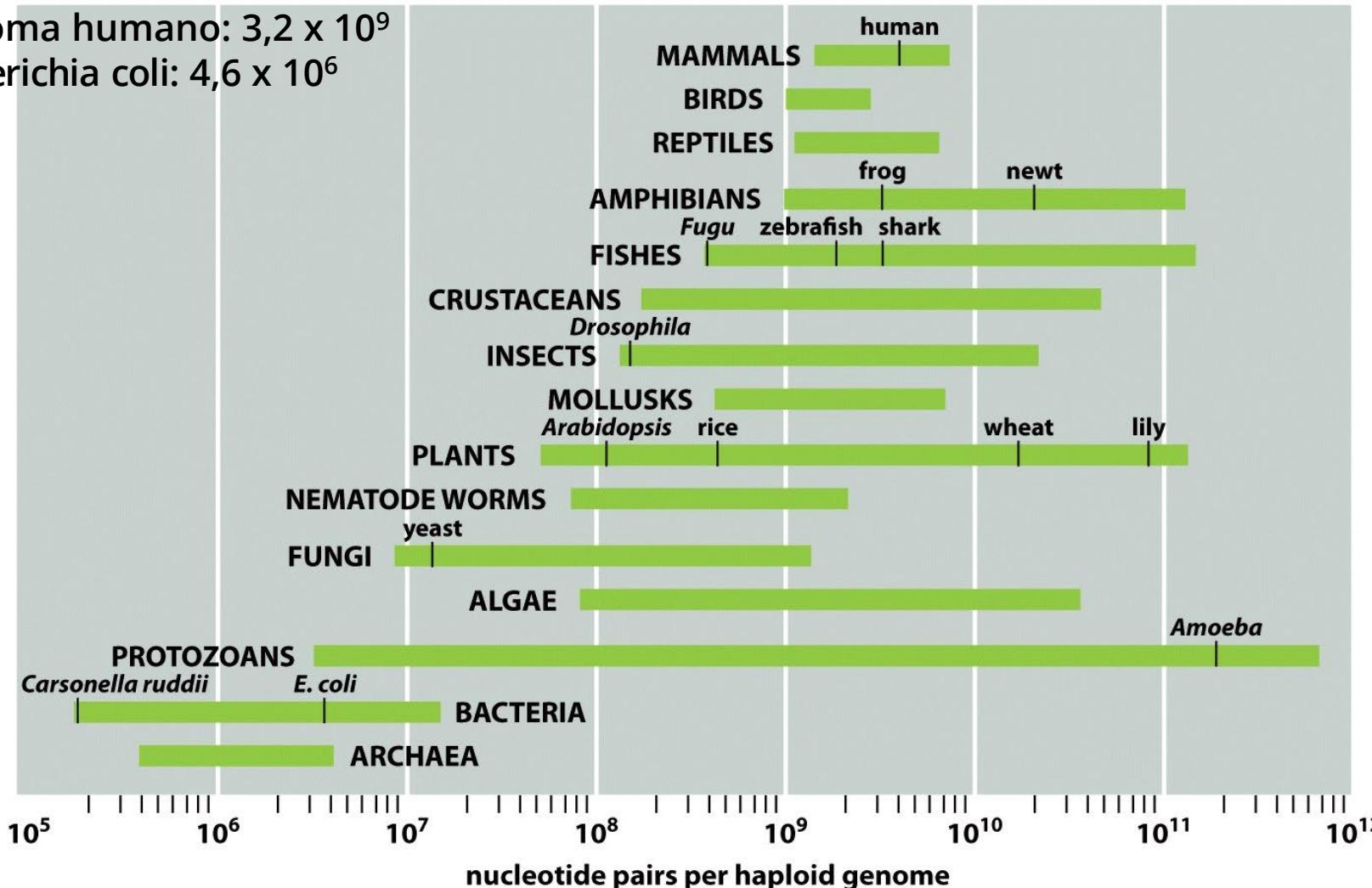
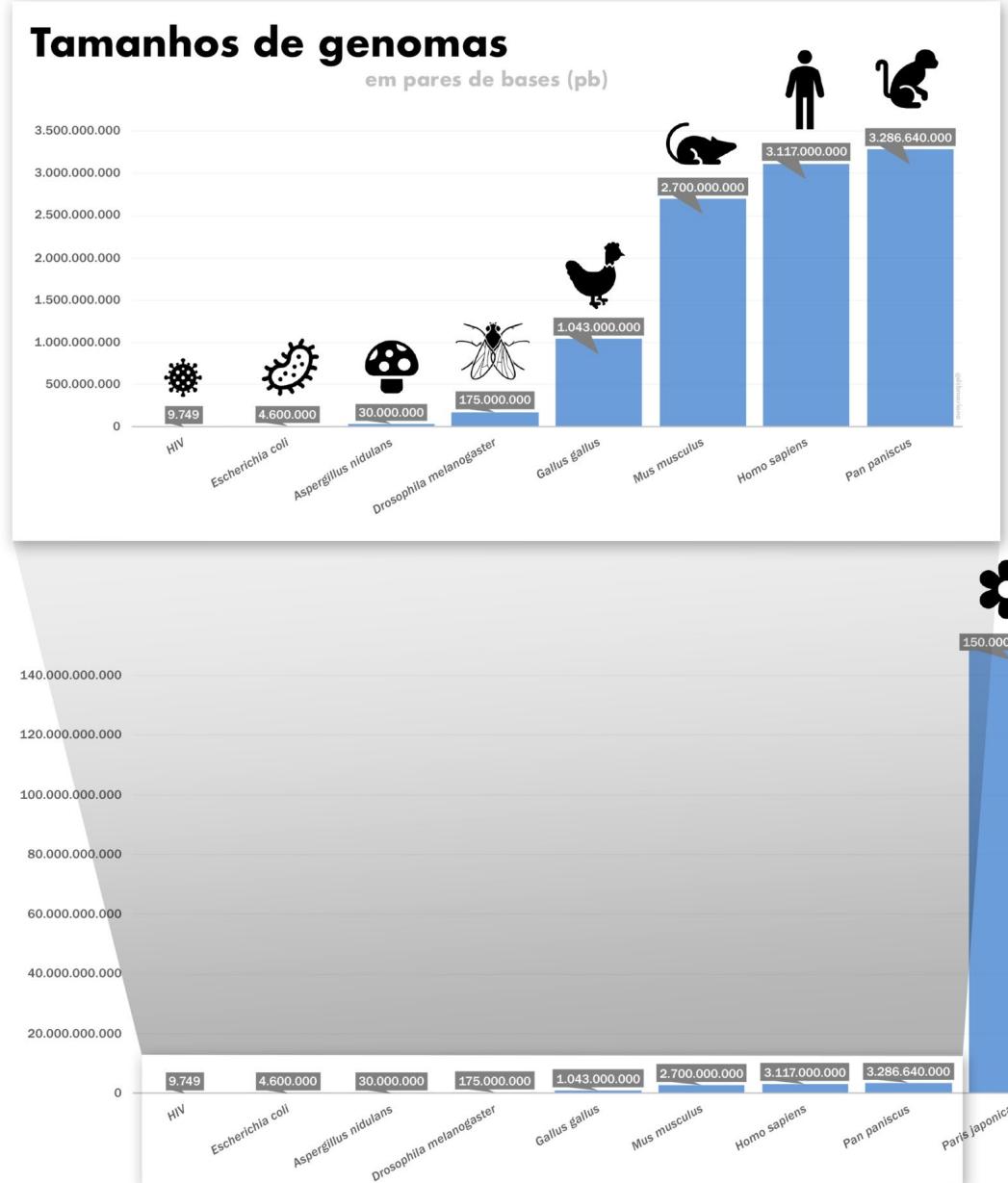
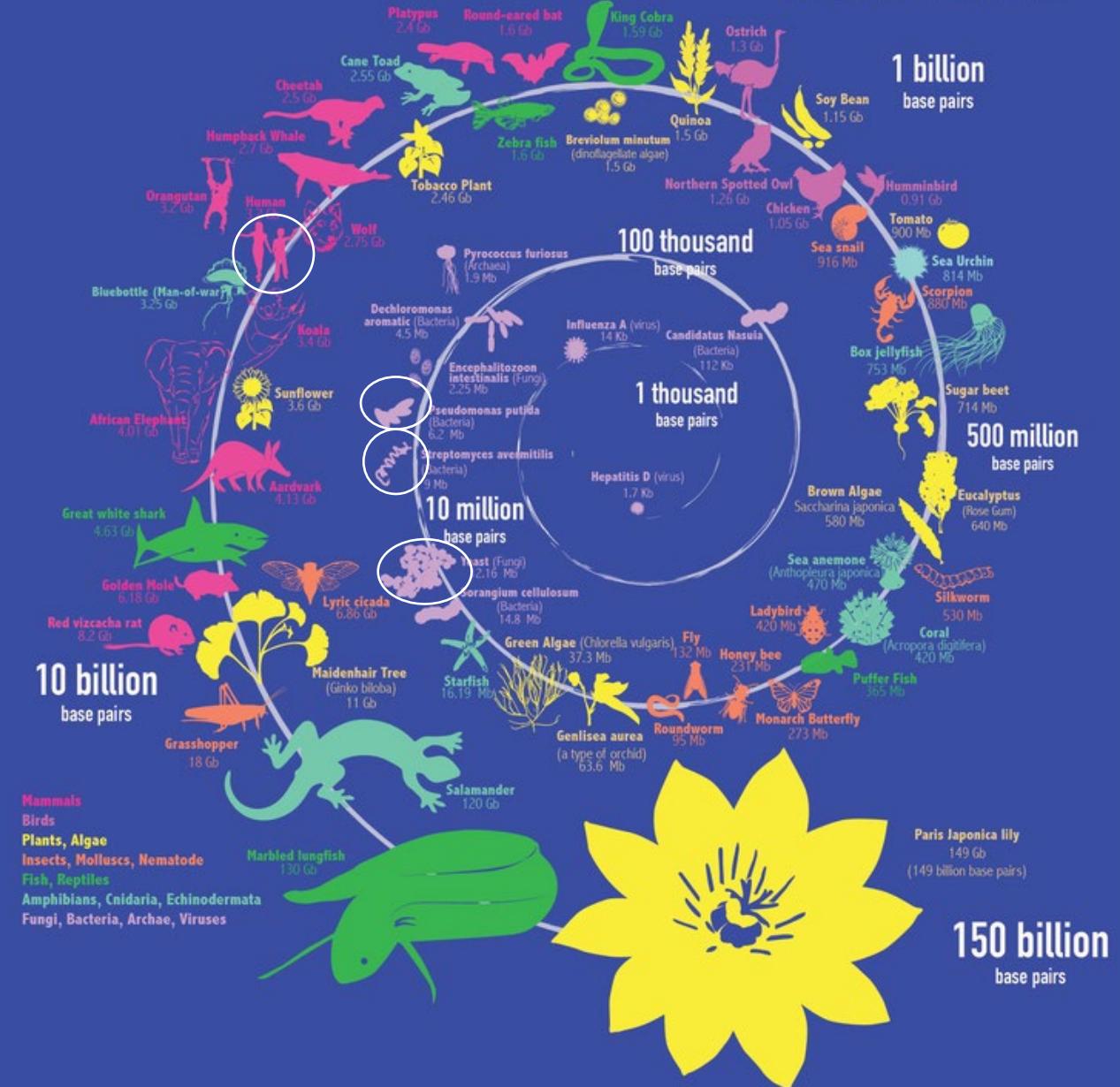


Figure 1-41 Essential Cell Biology 3/e (© Garland Science 2010)

Variação no tamanho dos genomas



Genome size (A species comparison)



Procaríotos vs. Eucaríotos

- Características do genoma de **procaríotos**
 - Genoma pequeno (0,5 a 10Mb)
 - Um cromossomo circular e haploide
 - Contém plasmídeos
 - Gene com estrutura simples
 - Sem introns
 - Alta densidade de região codificante (~90%)
 - Alguns genes são muito pequenos (<60pb)
 - Um mesmo promotor pode regular vários genes (operons)

Unidades organizacionais

Motivos

Cerca de dois a algumas dezenas de pares de bases

Podem estar dispersas ou agrupadas

Em muitos casos são unidades formadoras de repetição

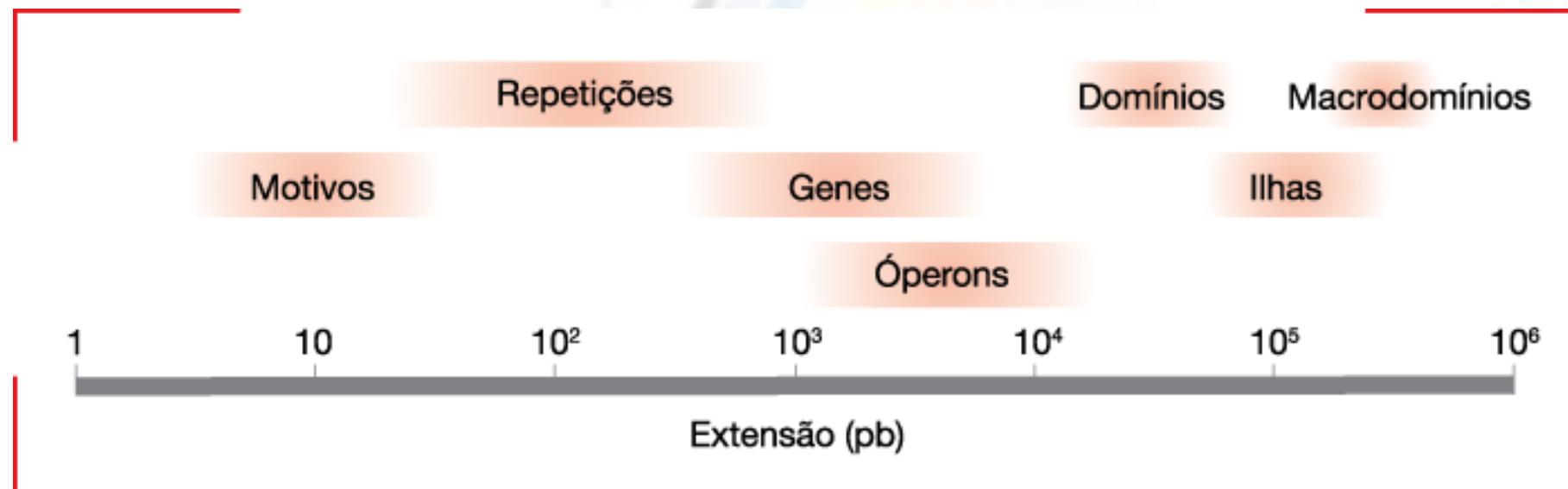


Figura 4.5

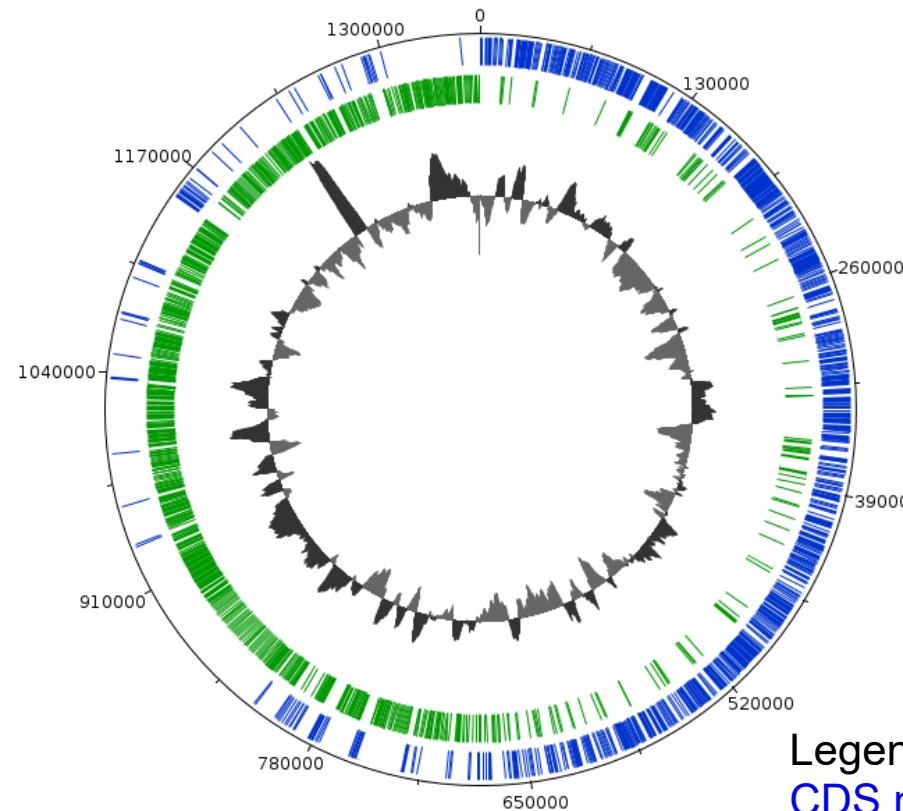
Principais unidades organizacionais de genomas procarióticos. As barras em indicam as extensões aproximadas típicas para cada tipo de unidade organizacional, conforme a escala exponencial indicada na barra inferior.

Conteúdo GC

Mycoplasma penetrans HF-2

1.358.633pb

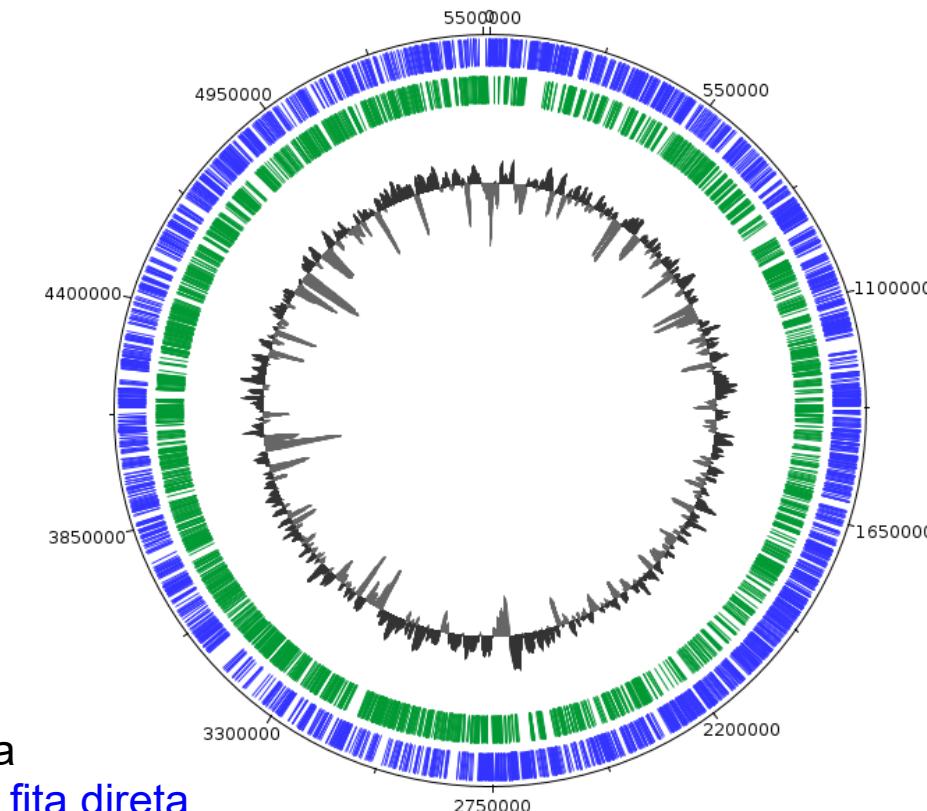
GC 25,7%



Herbaspirillum seropedicae SmR1

5.513.887pb

GC 63,4%



Legenda

CDS na fita direta

CDS na fita reversa

GC%

Ilhas genômicas

- Segmentos que diferem em algumas características do restante do genoma
 - Conteúdo GC
 - Padrões de sintenia
 - Orientação gênica
- Geralmente associadas a elementos móveis e eventos de transferência gênica horizontal
- Contém genes não essenciais
- Confere alguma vantagem ao organismo
 - Ex., ilhas de patogenicidade
 - Conferem capacidade de infectar e causar doença em um espécie hospedeira

Ilhas genômicas

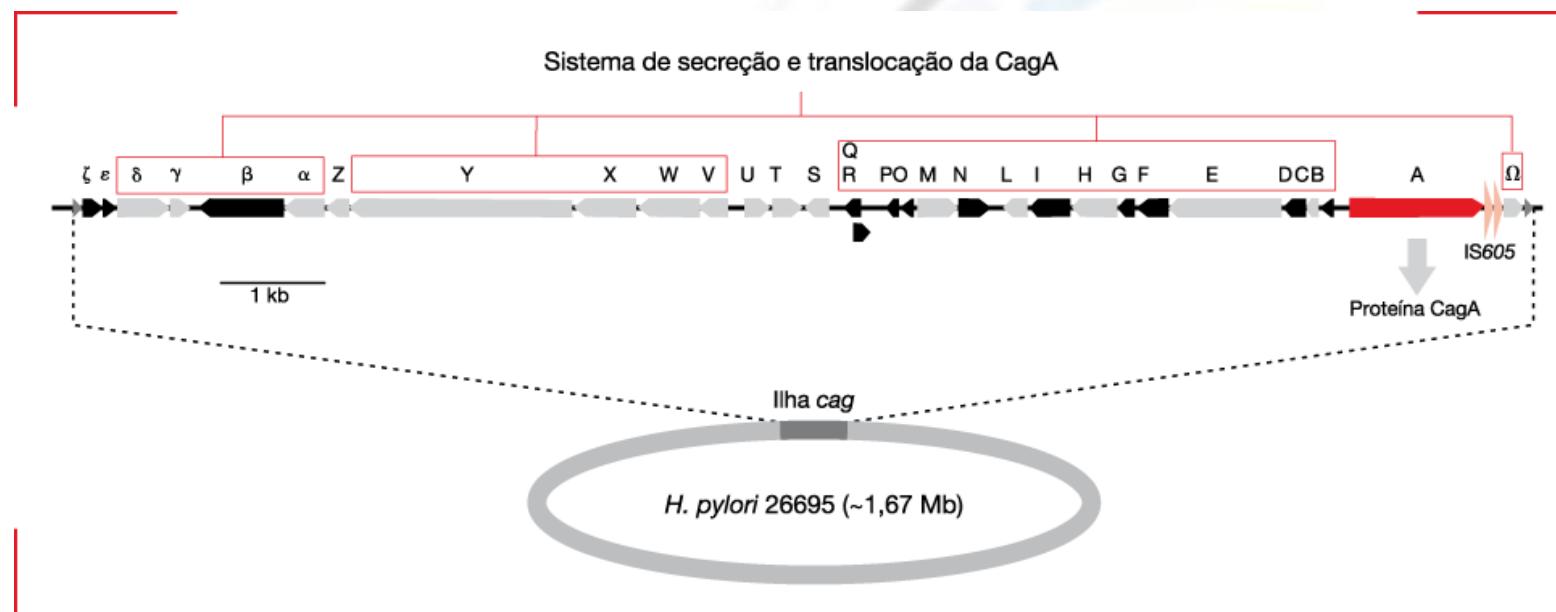


Figura 4.11

Ilha de patogenicidade *cag* de *H. pylori*. A configuração da ilha *cag* no cromossomo da linhagem 26695 de *H. pylori* está representada. Os genes codificadores de proteínas e suas respectivas orientações estão representados por setas em ■, ■ ou ■, e a letra (grega ou romana) que designa cada um deles está representada abaixo de cada seta. O gene *cagA* (em ■) codifica a proteína secretada CagA. A maioria dos demais genes (nomes assinalados por retângulos em ■) codifica proteínas essenciais para a secreção e a translocação de CagA para células do hospedeiro, sendo que as proteínas codificadas pelos genes *cag*, *cag*, *cagY*, *cagX*, *cagV*, *cagE* e *cagC* são ortólogas de proteínas do sistema de secreção do tipo IV do patógeno de plantas *A. tumefaciens*. Dezesseis genes da ilha (setas em ■) codificam produtos que levam células da mucosa gástrica a produzirem interleucina 8 (IL-8, uma citocina), o que modula a resposta imune do hospedeiro de modo favorável à bactéria. As pontas de seta menores (em ■, em cada uma das extremidades da ilha) representam as duas cópias da repetição direta de 31 pb e as pontas de seta maiores (em ■) representam as duas cópias da IS605; as pontas de seta também indicam as orientações de cada um desses elementos de sequência.

Procaríotos vs. Eucaríotos

- Características do genoma de **eucaríotos (fungos)**
 - Genoma grande (fungos: ~12Mb a 180Mb)
 - Múltiplos cromossomas lineares, mas geralmente haploides
 - Gene com estrutura complexa
 - Contém ítrons
 - Alguns ítrons são muito grandes (~300kb)
 - *Splicing* alternativo (um gene, várias proteínas)
 - Baixa densidade de região codificadora (<5% em humano)
 - Um promotor regula um único gene