

# Curso sobre Genômica no Estudo de Microrganismos – Módulo Filogenoma

Professoras UFPR  
Dra. Chirlei Glienke  
Dra. Desirrê Alexia Petters Vandresen  
GS Treinamentos e Consultoria

# Filogenômica

- Definição: análises baseadas no genoma completo para inferir filogenia
- Na prática: não necessariamente todos os genes de um determinado conjunto de espécies: Ortólogos de Cópia Única (SCOs)
- Por que?
  - uso de parálogos, genes duplicados de qualquer tipo ou pseudogene pode causar ruído filogenético nos dados
  - Inferência de árvores filogenéticas com base em múltiplos genes marcadores amostrados nos genomas de interesse

# Problemas

- A seleção de genes reduz a resolução da amostragem “filogenômica”
  - especialmente no caso de SCOs (Ortólogos de Cópia Única), clados distantemente relacionados com níveis mais baixos de homologia reduzirão drasticamente o número de genes comparados
  - grupos externos empregados para enraizar árvores
- Importante desafio na filogenômica:
  - potencial incongruência entre as histórias evolutivas de genes individuais, que pode ser generalizada em microrganismos devido à prevalência da transferência horizontal de genes

# Lembrando.....

1 gtgcccaagtg agcgaggact gca  
61 ggggaccgcga acaacatgga tag  
121 acgtttccatg gggccaagga ggt  
181 atgggcctca tataacaacag gat  
241 tgtgccactt gtgaacagat tgc  
301 accaccaatc cactaatcag gca  
361 gctatggaac agatggctgg atc  
421 cagactaggc agatgggtaca tgc  
481 ggtctgaaag atgaccttct tga  
541 atgcagcgat tcaagtgtatc ctc

# TABELA DO CÓDIGO GENÉTICO

		SEGUNDA LETRA					
		U	C	A	G		
PRIMEIRA LETRA [5']	U	UUU UUC UUA UUG	UCU UCC UCA UCG	UAU UAC UAA UAG	UGU UGC UGA UGG	U C A G	
	C	CUU CUC CUA CUG	CCU CCC CCA CCG	CAU CAC CAA CAG	CGU CGC CGA CGG	U C A G	
	A	AUU AUC AUA AUG Met (M) (iniciador)	ACU ACC ACA ACG	AAU AAC AAA AAG	AGU AGC AGA AGG	U C A G	TERCEIRA LETRA [3']
	G	GUU GUC GUA GUG	GCU GCC GCA GCG	GAU GAC GAA GAG	GGU GGC GGA GGG	U C A G	

= Códon de iniciação da cadeia polipeptídica

= Códon de término da cadeia polipeptídica



# Considerações

- Nem todas as mutações genéticas nos genes levam a mutações nas proteínas correspondentes
- Quando ocorrem mutações em aminoácidos, elas podem afetar a proteína de três maneiras:
  - podem ser prejudiciais (deletérias) para a proteína, caso em que a seleção purificadora irá eliminá-las
  - podem ser neutros, o que significa que não terão efeito na função da proteína
  - podem ser benéficos

# Considerações

- Como as mutações neutras não são penalizadas pela selecção natural, podem acumular-se ao longo do tempo
- Mutações benéficas geralmente se fixam mais rapidamente na população devido à seleção positiva

# Consequência

- Sequências de proteínas ortólogas de duas espécies que divergiram de um ancestral comum:
  - irão diferir apenas ligeiramente quando houver pouco tempo para que as mutações ocorram
  - mais diferenças terão se acumulado quando tiver passado bastante tempo

# Consequência

- Diferentes proteínas podem evoluir em taxas muito diferentes:
  - proteínas “housekeeping” evoluem lentamente - catalisam funções essenciais dentro das células (cytochrome C)
  - proteínas menos essenciais geralmente evoluem mais rapidamente

# Consequência

- Os fatores que determinam as taxas relativas de substituição de aminoácidos durante a evolução das proteínas são complexos e variam entre os táxons

# Por que analisar sequências de proteínas?

- Seleção natural:
- Embora o DNA (4 bases) contenha todas as informações necessárias para criar uma proteína, geralmente não é o próprio DNA que está sujeito à seleção natural
- As proteínas (20 aa) são os catalisadores de praticamente todas as transformações químicas na célula sobre as quais a seleção natural atua

# Por que analisar sequências de proteínas?

- Códons preferencias:
- Nem todos os organismos têm a mesma preferência de códons (leveduras, protistas e todos os animais têm preferências diferentes)
- Como consequência, a mesma sequência de proteínas em cada um desses organismos é o resultado de diferentes sequências de DNA
  - essas diferenças estão relacionadas ao viés de códons (códons preferenciais) e não apenas à evolução

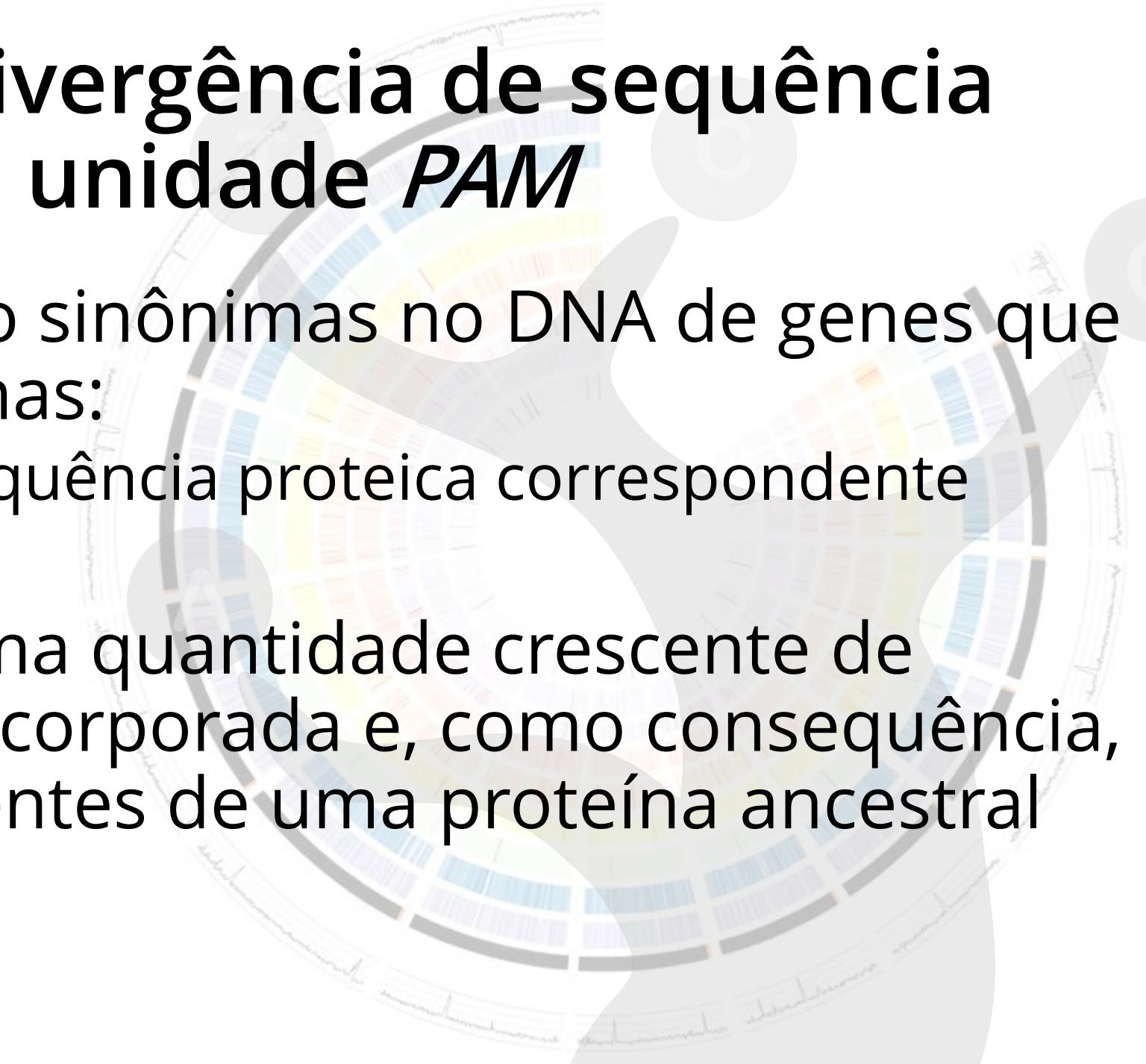
# Por que analisar sequências de proteínas?

- Código não tão universal:
- Alguns protistas usam os códons TAA e TGA para codificar a glutamina, em vez de “STOP”
- Nas mitocôndrias, o códon TGA codifica o triptofano, em vez de “STOP”

# Por que analisar sequências de proteínas?

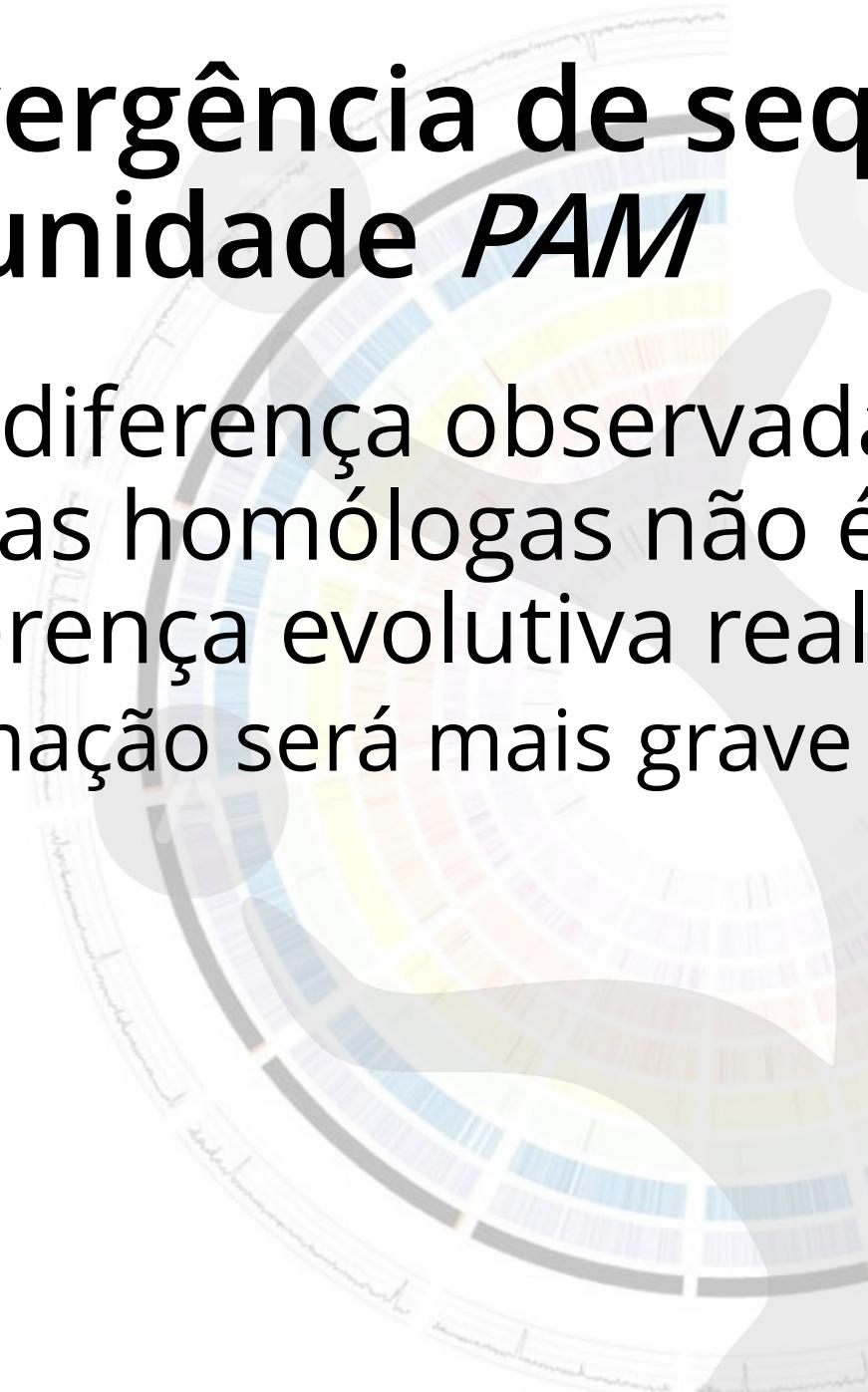
- Substituições múltiplas ou saturação de substituição
  - DNA: 4 bases (64 possíveis códons)
  - Proteína: 20 aminoácidos
- Organismos muito distantes: uso de sequencias de aminoácidos (traduzidas a partir de sequencia de nucleotídeos)
- Organismos muitos próximos: uso de sequencia de nucleotídeos

# Natureza da divergência de sequência em proteínas - unidade *PAM*



- Substituições não sinônimas no DNA de genes que codificam proteínas:
  - Alterações na sequência proteica correspondente
- Com o tempo, uma quantidade crescente de mutações será incorporada e, como consequência, os dois descendentes de uma proteína ancestral irão divergir

# Natureza da divergência de sequência em proteínas - unidade *PAM*



- A percentagem de diferença observada entre duas sequências proteicas homólogas não é proporcional à diferença evolutiva real
  - o nível de subestimação será mais grave em distâncias evolutivas maiores

# Valor *PAM*

- Número de mutações pontuais aceitas por 100 aminoácidos
- Introduzido por Dayhoff em 1978
- Medida que é proporcional à verdadeira distância evolutiva entre duas proteínas

# Relação entre a porcentagem observada de distância entre duas sequências e seu valor *PAM*

- Duas proteínas homólogas que tiveram um ancestral comum e têm uma distância *PAM* de 250-300 (80%-85% de diferença):
  - não podem ser distinguidas de forma confiável de duas proteínas escolhidas aleatoriamente de comprimento semelhante e alinhadas

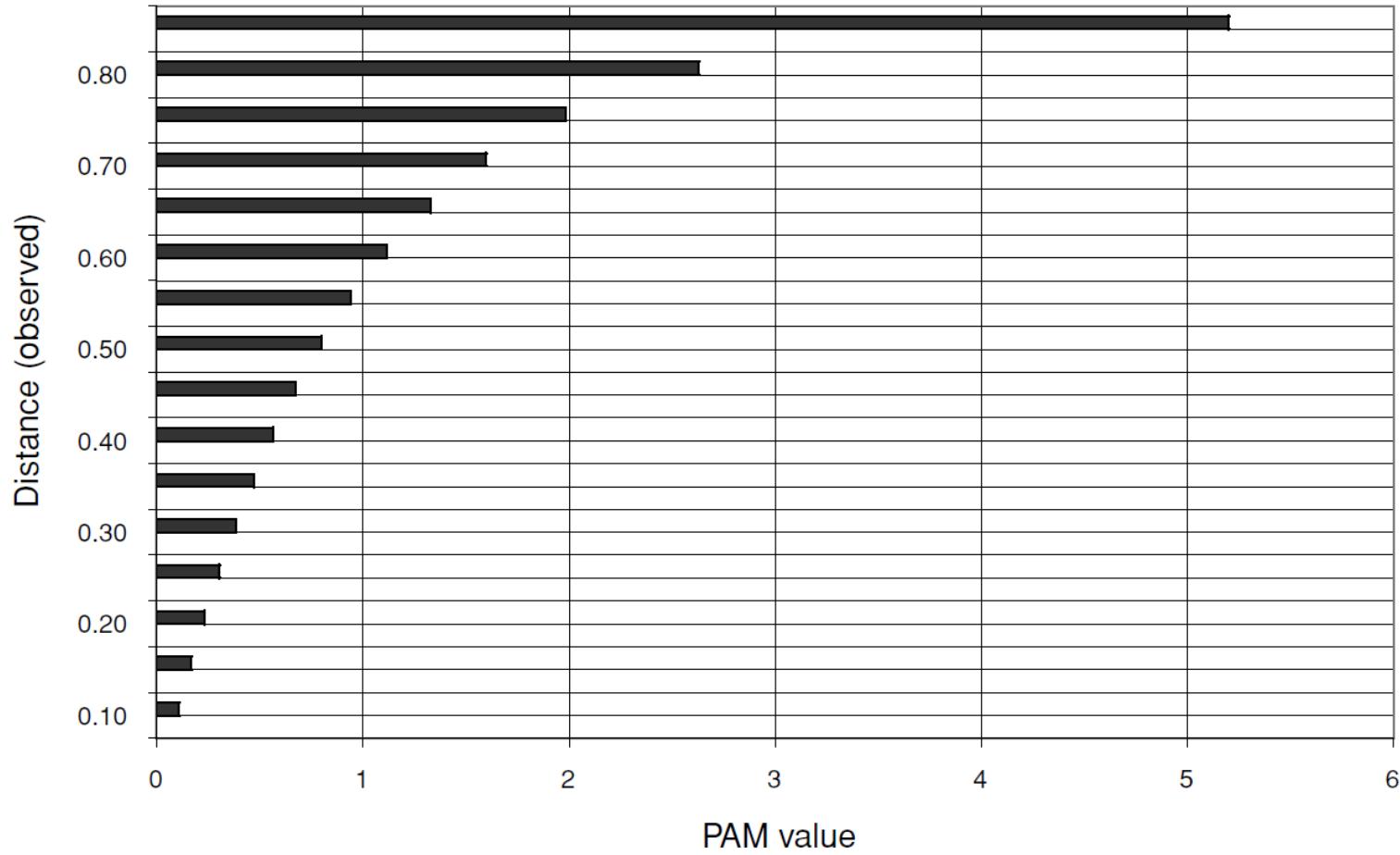
# Relação entre a porcentagem observada de distância entre duas sequências e seu valor *PAM*

- Portanto, as análises filogenéticas utilizando sequências de proteínas devem geralmente ser limitadas a proteínas que diferem em menos de 80% a 85%

# Relação entre a porcentagem observada de distância entre duas sequências e seu valor *PAM*

- “zona crepuscular”:

Quando o valor da distância *PAM* entre duas proteínas distivamente relacionadas se aproxima do valor 250, torna-se difícil dizer se as duas proteínas são homólogas ou se são duas proteínas escolhidas aleatoriamente que podem ser alinhadas por acaso



# Considerações importantes

- Proteínas com funções não essenciais à sobrevivência do organismo podem, em geral, evoluir a uma taxa relativamente elevada
  - a informação evolutiva pode ser rapidamente apagada e tais proteínas só podem ser utilizadas para estudos filogenéticos de organismos intimamente relacionados

# Considerações importantes

- As proteínas de manutenção (genes *housekeeping*), como as histonas, as enzimas da maquinaria de replicação, as enzimas do metabolismo central e as proteínas do citoesqueleto, evoluem lentamente e incorporam entre 1 a 10 mutações por 100 resíduos e por 100 milhões de anos
  - leva um tempo considerável até que estas proteínas tenham incorporado tantas substituições que apaguem toda a informação evolutiva

# Considerações importantes

- Devido a esta lenta taxa de evolução, as proteínas de genes *housekeeping* são excelentes marcadores para estudos filogenéticos de organismos distamente relacionados

Type of protein	Rate of Change (PAMs/100 myrs)	Theoretical Lookback Time (myrs)
Pseudogenes	400	45
Fibrinopeptides	90	200
Ig lambda chain C region	27	670
Somatotropin	25	800
Ribonucleases	21	850
Haemoglobin alpha chain	12	1500
Acid proteases	8	2300
Cytochrome c	4	5000
Adenylate kinase	3.2	6000
Glyceraldehyde-P dehydrogenase	2	9000
Glutamate dehydrogenase	0.9	18000

Useful lookback time = 360 PAMs

(Adapted from Table 1 in Dayhoff, 1978)

# Alinhamento de sequências de duas proteínas

- Proteínas homólogas quase nunca são idênticas, mas apenas semelhantes
- Além das substituições, haverá inserções e deleções em uma sequência em relação à outra

# Alinhamento de sequências de duas proteínas

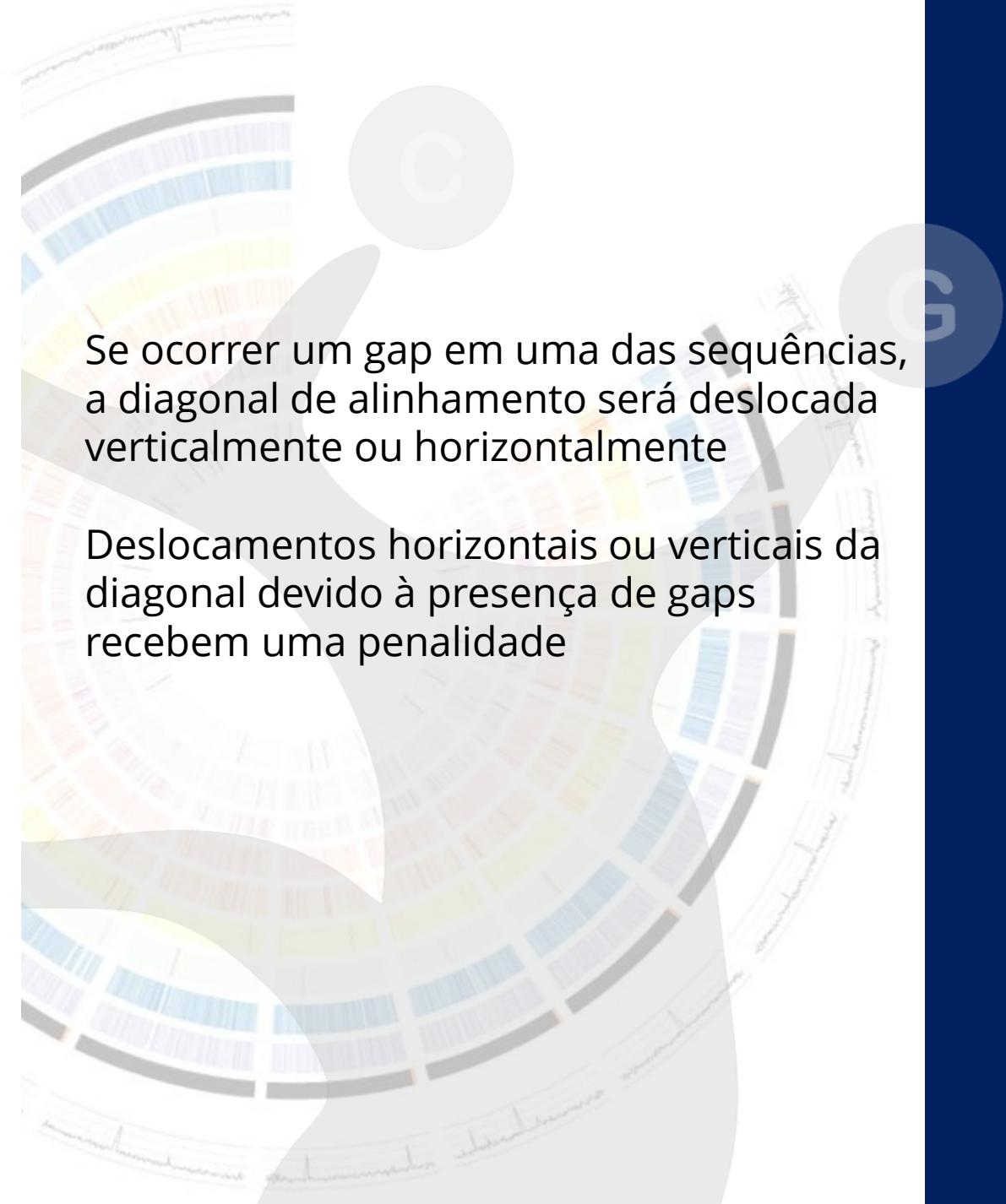
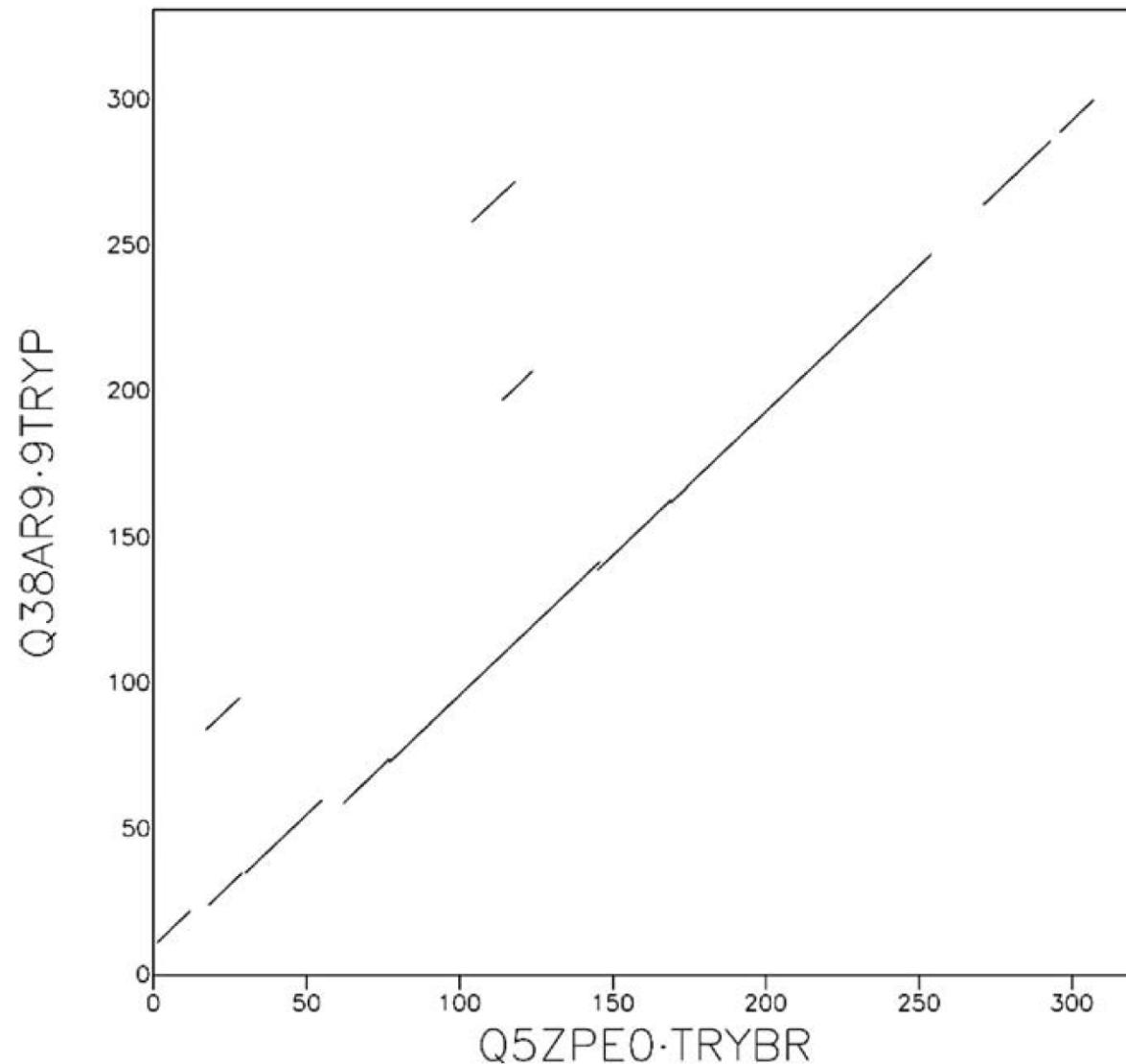
- As análises filogenéticas devem ser realizadas apenas em resíduos homólogos (ou seja, aqueles resíduos em cada uma das sequências que se originam de um resíduo ancestral comum)
- Portanto, é essencial que duas ou mais sequências estejam devidamente alinhadas

# Alinhamento de sequências de duas proteínas

- Uso de procedimento semelhante ao método de matriz de pontos
- As duas sequências a serem alinhadas são escritas como colunas e linhas de uma matriz

## Dotmatcher: Q5ZPE0·TRYBR vs Q38AR9·9TRYF

(windowsize = 10, threshold = 23.00 24/02/07)



# Alinhamento de sequências de duas proteínas

- A escolha do valor de cada pontuação positiva em relação à penalidade de gap influencia fortemente a qualidade do alinhamento resultante
  - Um valor de penalidade de gap muito baixo levará a uma situação em que os resíduos em regiões altamente diferentes não estarão alinhados entre si, mas sim com gaps
  - Uma penalidade de gap muito alta levará a alinhamento de resíduos ou regiões não homólogas

# Alinhamento de sequências de duas proteínas

- Matrizes de pesos diferentes:
  - desenvolvidas para programas de alinhamento de sequências para refletir algumas regras de mutação teóricas ou observadas

# Alinhamento de sequências de duas proteínas

- As matrizes de peso empírico são geralmente baseadas nas taxas de substituição observadas de aminoácidos em grandes coleções de sequências homólogas em bancos de dados como o SwissProt
  - A escolha da matriz de pesos fica ao critério científico do usuário e dependerá do conjunto de dados que está sendo analisado

# Alinhamento de sequências de duas proteínas

- Matrizes de probabilidades logarítmicas:
  - PAM250 (Dayhoff, 1978)
  - BLOSUM62 (Henikoff & Henikoff, 1992)
- Matriz Jones – Taylor – Thornton (JTT) (Jones et al., 1992) é uma atualização da matriz PAM
- Matriz Gonnet (Gonnet et al., 1992)
- Matriz Whelan e Goldman (WAG): é uma melhoria na matriz JTT baseada em métodos de verossimilhança (Whelan & Goldman, 2001)
- Matriz de substituição proteica (VT) (Muller e Vingron, 2000): baseada em um método que aproxima a máxima verossimilhança

# Ajuste manual de um alinhamento de proteínas

- Alinhamento de múltiplas sequências produzido automaticamente muitas vezes precisa de ajuste manual para melhorar sua qualidade
  - especialmente na posição dos gaps

# Ajuste manual de um alinhamento de proteínas

- Necessita do conhecimento disponível sobre um tipo específico de proteína
  - Informações sobre resíduos de sítios ativos e elementos de estrutura secundária, como hélices α, fitas β e alças

# Ajuste manual de um alinhamento de proteínas

- Conhecimento das propriedades físico-químicas dos 20 aminoácidos e ter em mente uma série de regras práticas para a mutabilidade dos vários aminoácidos

**Table 9.2** Single and 3-letter codes for amino acids. All proteins are polymers of the 20 naturally occurring L-amino acids

Alanine	Ala	A
Cysteine	Cys	C
Aspartic Acid	Asp	D
Glutamic Acid	Glu	E
Phenylalanine	Phe	F
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Lysine	Lys	K
Leucine	Leu	L
Methionine	Met	M
Asparagine	Asn	N
Proline	Pro	P
Glutamine	Gln	Q
Arginine	Arg	R
Serine	Ser	S
Threonine	Thr	T
Valine	Val	V
Tryptophan	Trp	W
Tyrosine	Tyr	Y

In addition, B may be used for Asx (aspartate or asparagine) and X for Glx (glutamate or glutamine). J, O and U are not used. The one-letter code is invariably used when comparing and aligning sequences of proteins. Most are easily remembered by their initial letters. Note that cysteine and methionine are the only two sulphur-containing amino acids.

# Ajuste manual de um alinhamento de proteínas

- Numa proteína dobrada, os resíduos D, R, E, N e K são preferencialmente mutados para resíduos de propriedades semelhantes
- Por serem polares ou carregados, são encontrados principalmente na superfície da proteína dobrada
- Desempenham um papel menos importante no enovelamento de proteínas:
  - sofrem mutações com bastante facilidade

# Ajuste manual de um alinhamento de proteínas

- Os resíduos hidrofóbicos (F, A, M, I, L, Y, V e W):
  - são substituídos preferencialmente por outros hidrofóbicos
- Estão posicionados principalmente internamente e determinam o enovelamento da proteína
  - sofrem mutações lentamente

# Ajuste manual de um alinhamento de proteínas

- Os resíduos C, H, Q, S e T são geralmente indiferentes e podem ser substituídos por qualquer outro tipo de resíduo
- Os resíduos (D, R, E, N, K, C, H, Q, S e T), quando conservados ao longo do alinhamento, são muito provavelmente resíduos que estão envolvidos no sítio ativo de uma enzima
  - Portanto o alinhamento múltiplo deve ser ajustado de forma a manter a homologia posicional destes resíduos

# Modelos evolutivos

- A utilização de um ou outro modelo de evolução pode mudar os resultados de uma análise filogenética
- Portanto, torna-se crucial decidir qual é o modelo mais adequado

# Modelos evolutivos

- Cuidado com dados heterogêneos: um único modelo de evolução pode não se ajustar bem a todos os dados
  - por exemplo: união de genes diferentes para a análise filogenética, ou regiões codificantes e não codificantes
  - diferentes regiões dos genomas estão sujeitas a diferentes pressões seletivas e restrições evolutivas
- análise combinada em que cada partição de dados (por exemplo, genes diferentes) tem seu próprio modelo

# Modelos evolutivos

- Em geral, modelos mais complexos ajustam-se melhor aos dados do que modelos mais simples porque têm mais parâmetros
- Alternativa: selecionar um modelo de evolução mais complexo - com mais parâmetros

# Modelos evolutivos

- Problema: quando se utilizam modelos complexos, é necessário estimar um grande número de parâmetros, o que tem várias desvantagens
  - a análise torna-se computacionalmente difícil e requer uma grande quantidade de tempo
  - como mais parâmetros precisam ser estimados a partir dos mesmos dados, mais erros são incluídos em cada estimativa

# Modelos evolutivos

- Solução:
  - escolher um modelo que seja complexo suficiente para explicar os dados, mas não tão complicado que exija tempo impraticável para os cálculos ou grandes conjuntos de dados para obter estimativas precisas

# Modelos evolutivos

- O modelo de evolução mais adequado para um determinado conjunto de dados pode ser selecionado usando técnicas estatísticas

# Modelos evolutivos

- Diferentes abordagens para selecionar o modelo de evolução mais adequado dentro de uma coleção de modelos candidatos:
  - testes hierárquicos de razão de verossimilhança (hLRTs)
  - critérios de informação (ex. Akaike Information Criterion – AIC)
  - abordagens Bayesianas (Bayesian Information Criterion – BIC, aplicando Markov chain Monte Carlo - MCMC)
  - abordagens baseadas em desempenho (decision theoretic framework (DT) ou referencial teórico de decisão
    - ranqueamento

# Estimativa de modelos

- Programas como ModelTest e ProtTest

## Conjunto de modelos candidatos avaliados em ModelTest. Para cada modelo, foram avaliados: modelo, modelo+I, modelo+G e modelo +I+G

I = proporção de sítios invariáveis (1 parâmetro)  
 G = parâmetro de forma da distribuição gama (1 parâmetro)

Model	Free parameters	Base frequencies	Substitution code <sup>a</sup>	Reference
JC	0	equal	000000	(Jukes & Cantor, 1969)
F81	3	unequal	000000	(Felsenstein, 1981)
K80	1	equal	010010	(Kimura, 1980)
HKY	4	unequal	010010	(Hasegawa <i>et al.</i> , 1985)
TrNef	2	equal	010020	(Tamura & Nei, 1993)
TrN	5	unequal	010020	(Tamura & Nei, 1993)
K81	2	equal	012210	(Kimura, 1981)
K81uf	5	unequal	012210	(Kimura, 1981)
TIMef	3	equal	012230	(Posada, 2003)
TIM	6	unequal	012230	(Posada, 2003)
TVMef	4	equal	012314	(Posada, 2003)
TVM	7	unequal	012314	(Posada, 2003)
SYM	5	equal	012345	(Zharkikh, 1994)
GTR	8	unequal	012345	(Rodríguez <i>et al.</i> , 1990)

Numbers refer to how many distinct parameters are used to describe the relative rates  $A \leftrightarrow C$ ,  $A \leftrightarrow G$ ,  $A \leftrightarrow T$ ,  $C \leftrightarrow G$ ,  $C \leftrightarrow T$ , and  $G \leftrightarrow T$ . For example, 010020 implies three different parameters:  $0 \approx A \leftrightarrow C = A \leftrightarrow T = C \leftrightarrow G = G \leftrightarrow T$ ,  $1 \approx A \leftrightarrow G$ , and  $2 \approx C \leftrightarrow T$ .

**Conjunto de modelos candidatos avaliados no ProtTest. Para cada modelo (0 parâmetros livres) foram avaliados: modelo, modelo+F, modelo+I, modelo+G, modelo+F+I, modelo+F+G,modelo+I+G e modelo+F+I+G**

Model	Data type	Reference
JTT	nuclear	(Jones <i>et al.</i> , 1992)
MtREV	mitochondrial	(Adachi & Hasegawa, 1996)
MtMam	mammal mitochondria	(Cao <i>et al.</i> , 1998)
MtArt	arthropod mitochondria	(Abascal <i>et al.</i> , 2007)
Dayhoff	nuclear	(Dayhoff <i>et al.</i> , 1978)
WAG	nuclear	(Whelan & Goldman, 2001)
RtREV	retroviral polymerase	(Dimmic <i>et al.</i> , 2002)
CpREV	chloroplastic	(Adachi <i>et al.</i> , 2000)
Blosum62	nuclear	(Henikoff & Henikoff, 1992)
VT	nuclear	(Muller & Vingron, 2000)

F = frequências de aminoácidos (19 parâmetros)

I = proporção de sítios invariáveis (1 parâmetro)

G = parâmetro de forma da distribuição gama (1 parâmetro)

# Alinhamento de aminoácidos

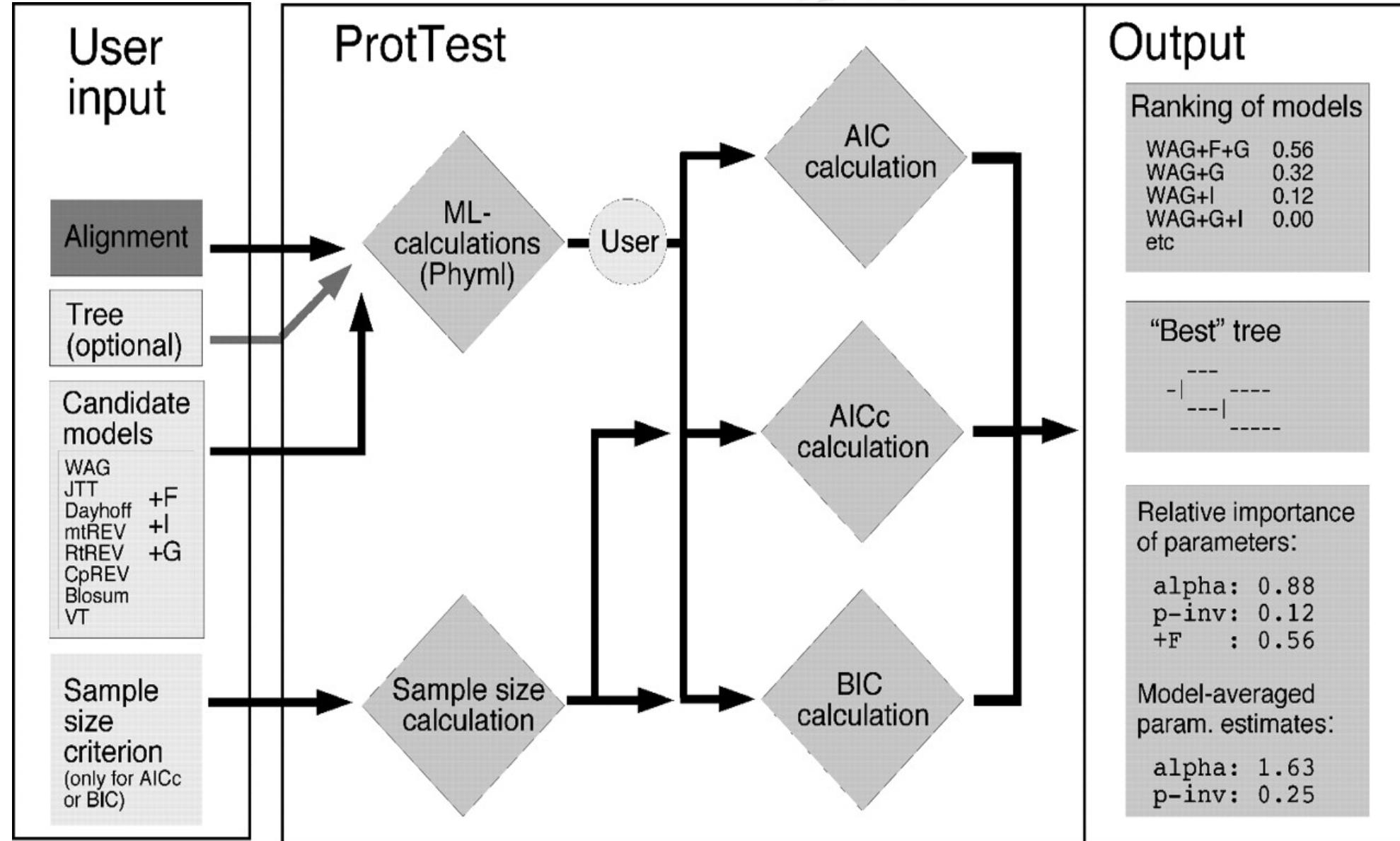
mosca	GAKKVIISAP	SAD.APM..F	VCGVNLDAYK	PDMKVVSNAS	CTTNCLAPLA
humano	GAKRVIISAP	SAD.APM..F	VMGVNHEKYD	NSLKIISNAS	CTTNCLAPLA
planta	GAKKVIISAP	SAD.APM..F	VVGVNEHTYQ	PNMDIVSNAS	CTTNCLAPLA
bactéria	GAKKVVMTGP	SKDNTPM..F	VKGANFDKY.	AGQDIVSNAS	CTTNCLAPLA
levedura	GAKKVVITAP	SS.TAPM..F	VMGVNEEKYT	SDLKIVSNAS	CTTNCLAPLA
árquea	GADKVLISAP	PKGDEPVKQL	VYGVNHDEYD	GE.DVVSNAS	CTTNSITPVA



mosca	VINDNFEIV	EGLMTTVHAT	TATQKTVVDGP	SGKLWRDGRG	AAQNIIPAST
humano	VIHDNFGIV	EGLMTTVHAI	TATQKTVVDGP	SGKLWRDGRG	ALQNIIPAST
planta	VVHEEFGIL	EGLMTTVHAT	TATQKTVVDGP	SMKDWRGGRG	ASQNIIPSST
bactéria	VINDNFGII	EGLMTTVHAT	TATQKTVVDGP	SHKDWRGGRG	ASQNIIPSST
levedura	VINDAFGIE	EGLMTTVHSL	TATQKTVVDGP	SHKDWRGGRT	ASGNIIPSST
árquea	VLDEEFGIN	AGQLTTVHAY	TGSQNLMDGP	NGKP.RRRRA	AAENIIPST

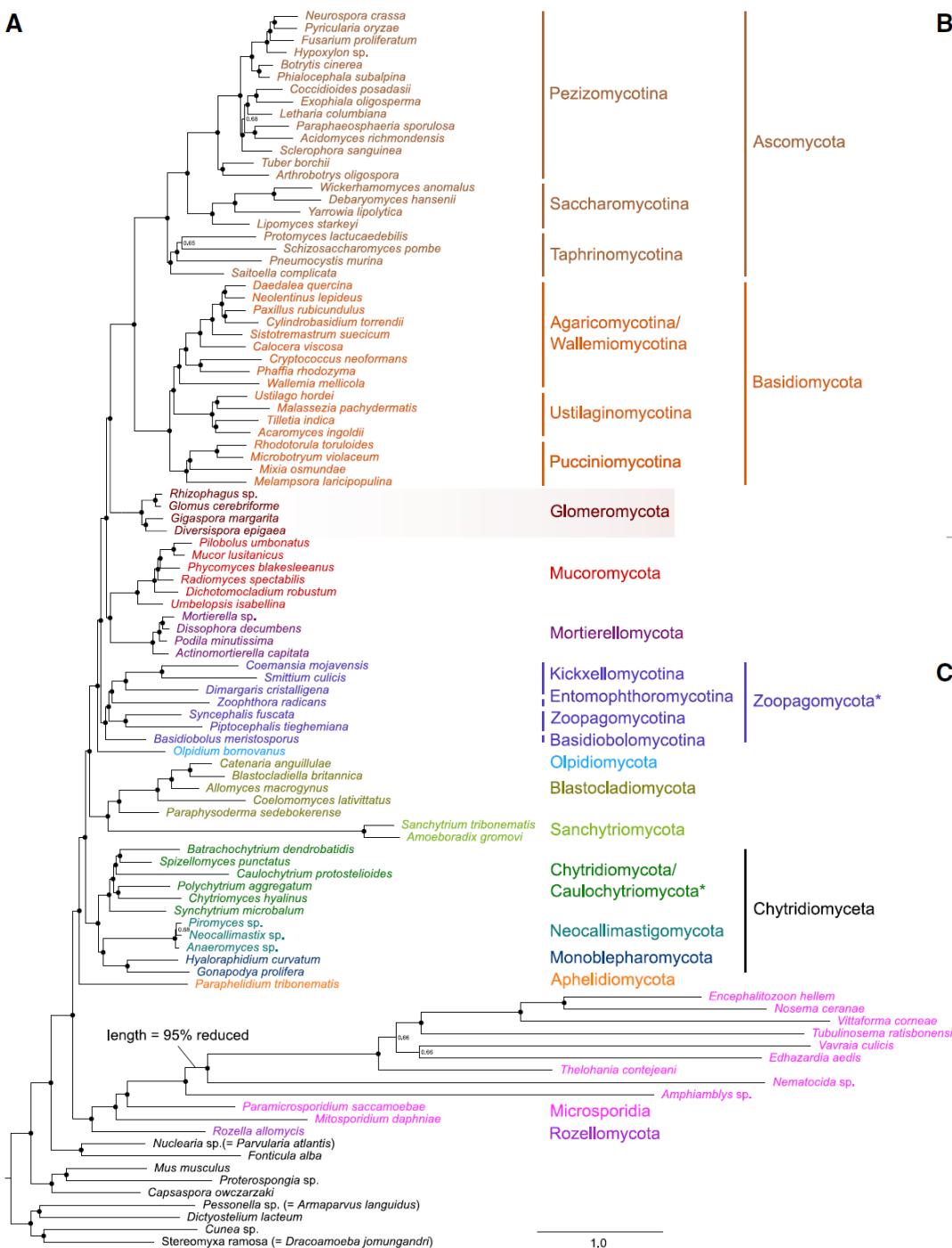
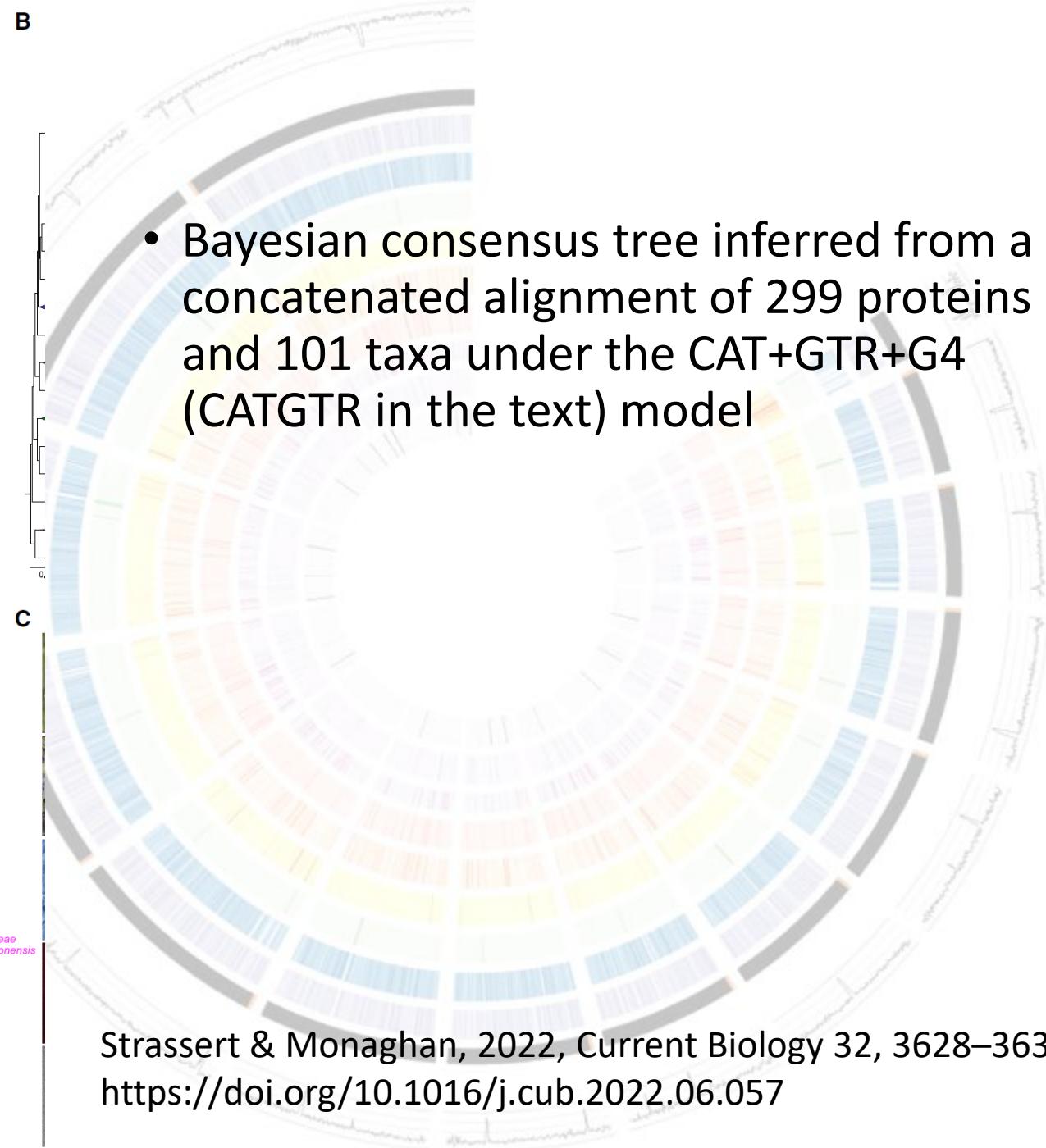
mosca	GAAKAVGKVI	PALNGKLTGM	AFRVPTPNVS	VVDLTVRLGK	<b>GASYDEIKAK</b>
humano	GAAKAVGKVI	PELNGKLTGM	AFRVPTANVS	VVDLTCRLEK	PAKYDDIKKV
planta	GAAKAVGKVL	PELNGKLTGM	AFRVPTSNVS	VVDLTCRLEK	<b>GASYEDVKAA</b>
bactéria	GAAKAVGKVL	PELNGKLTGM	AFRVPTPNVS	VVDLTVRLEK	<b>AATYEQIKAA</b>
levedura	GAAKAVGKVL	PELQGKLTGM	AFRVPTVDVS	VVDLTVKLNK	ETTYDEIKKV
árquea	GAAQAATEVL	PELEGKLDGM	AIRVPVPNGS	ITEFVVDLDD	DVTESDVNA





# EXEMPLOS

- Uso de filogenômica para resolver a diversificação inicial da árvore da vida fúngica
- Conjunto de dados de 299 proteínas com 661 táxons que representam todas as linhagens fúngicas reconhecidas existentes

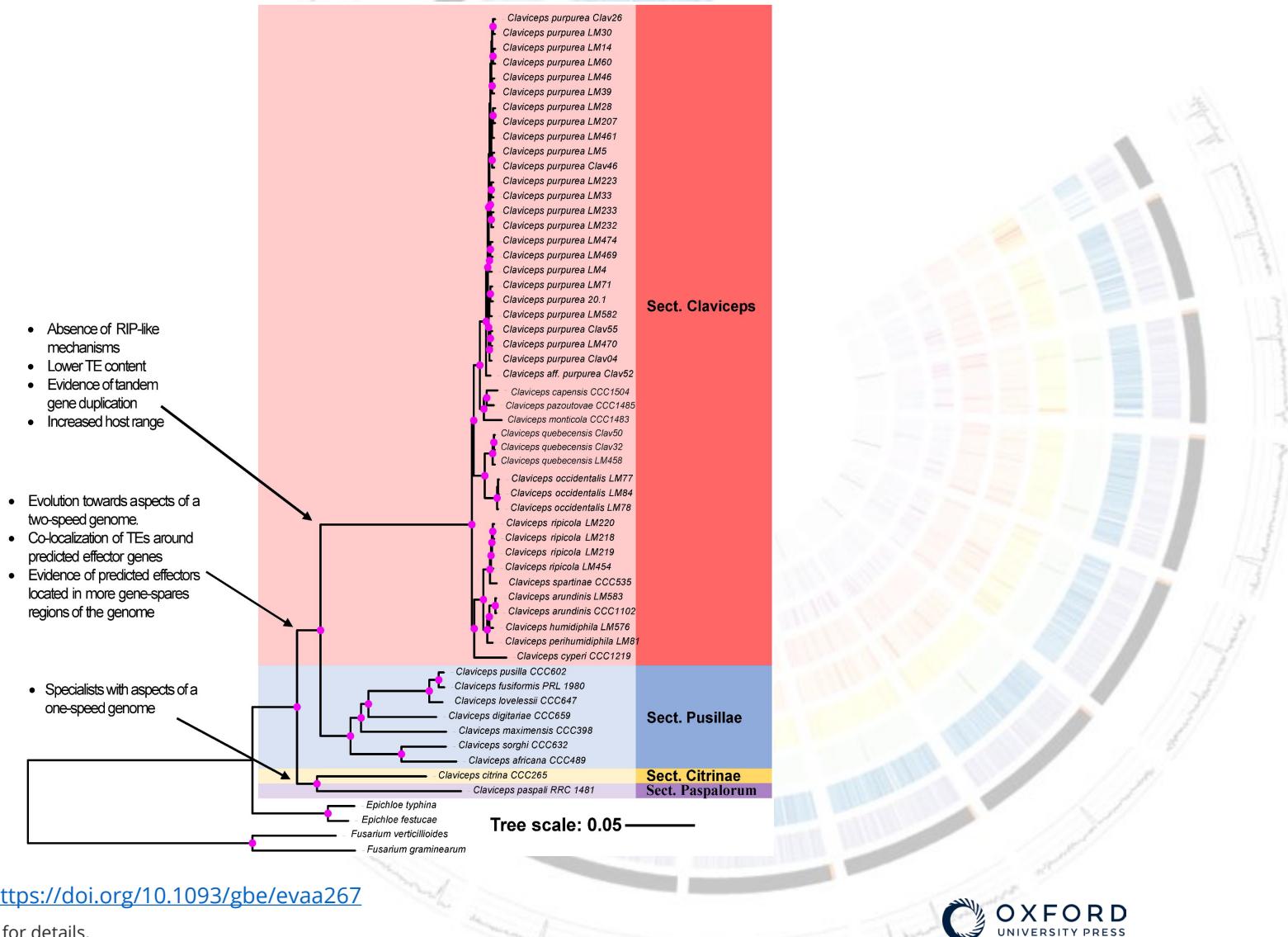
**A****B**

Strassert & Monaghan, 2022, Current Biology 32, 3628–3635  
<https://doi.org/10.1016/j.cub.2022.06.057>



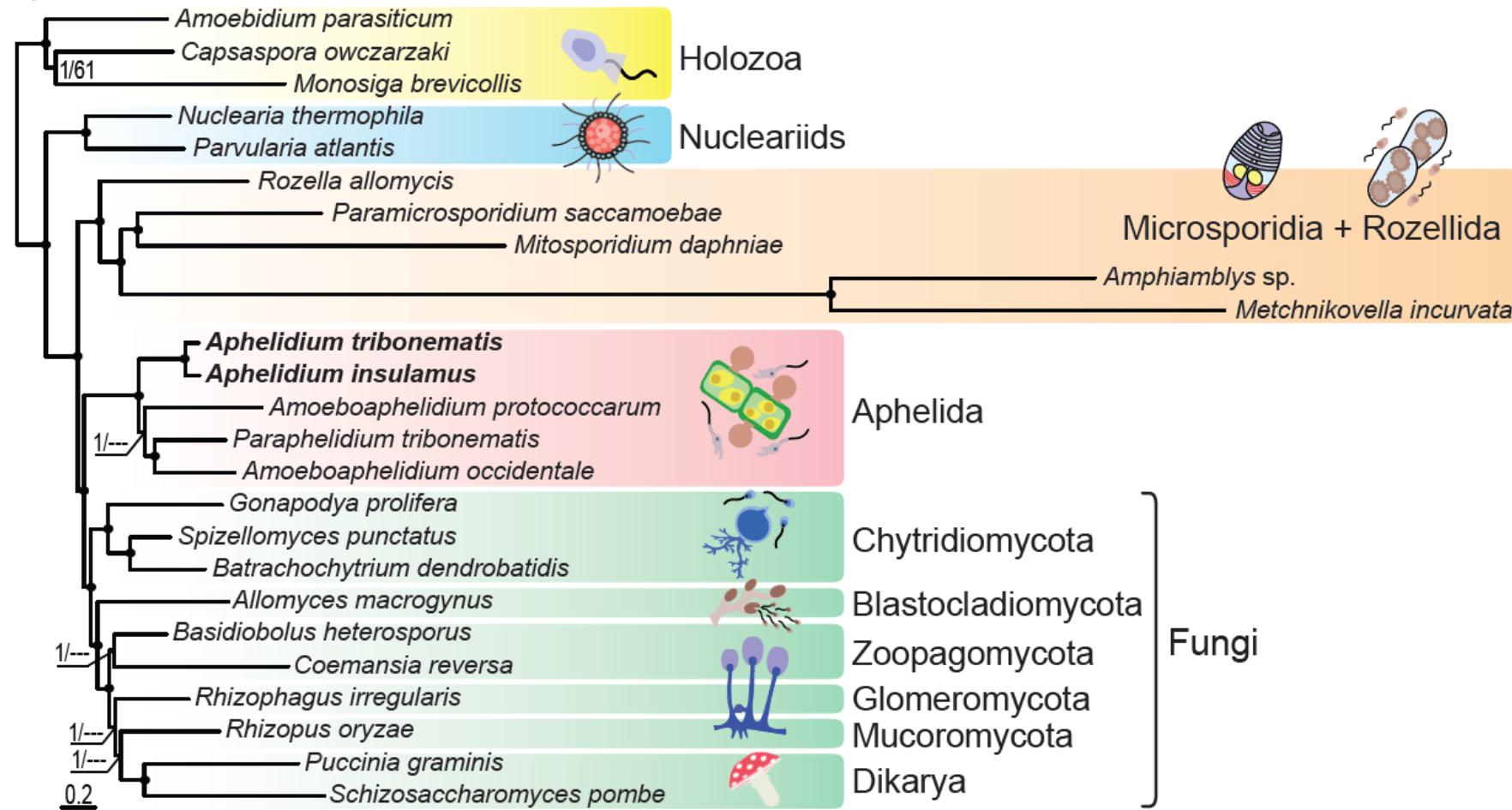
# Reconstrução filogenética (baseada em ML) do gênero *Claviceps* usando sequências de aminoácidos de 2.002 ortólogos de cópia única com 1.000 réplicas de bootstrap

Pink dots at branches represent bootstrap values  $\geq 95$   
Arrows and descriptions indicate potential changes in genomic architecture between *Claviceps* sections identified in this study



# Filogenômica apoia a monofilia de afelideos e fungos

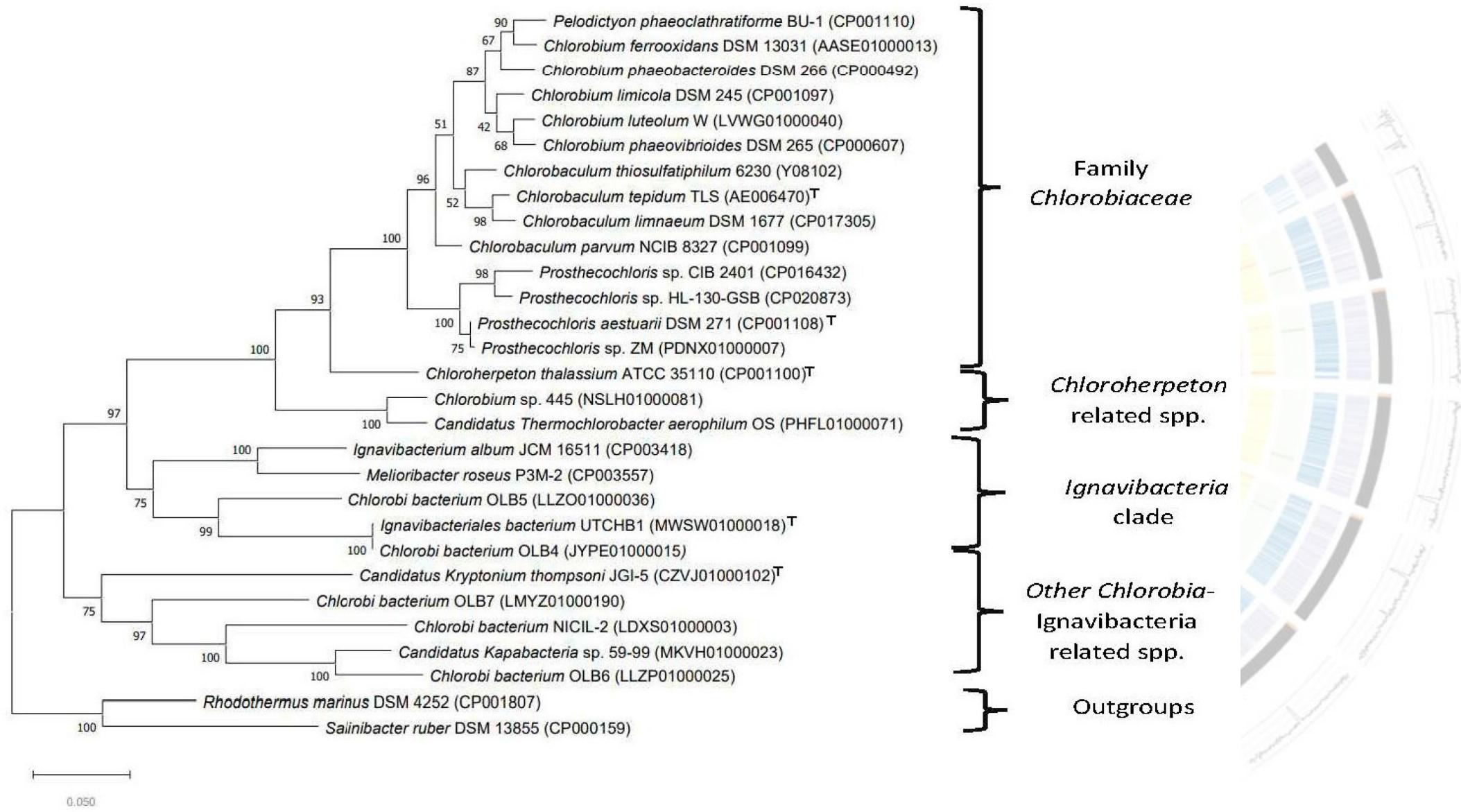
a)

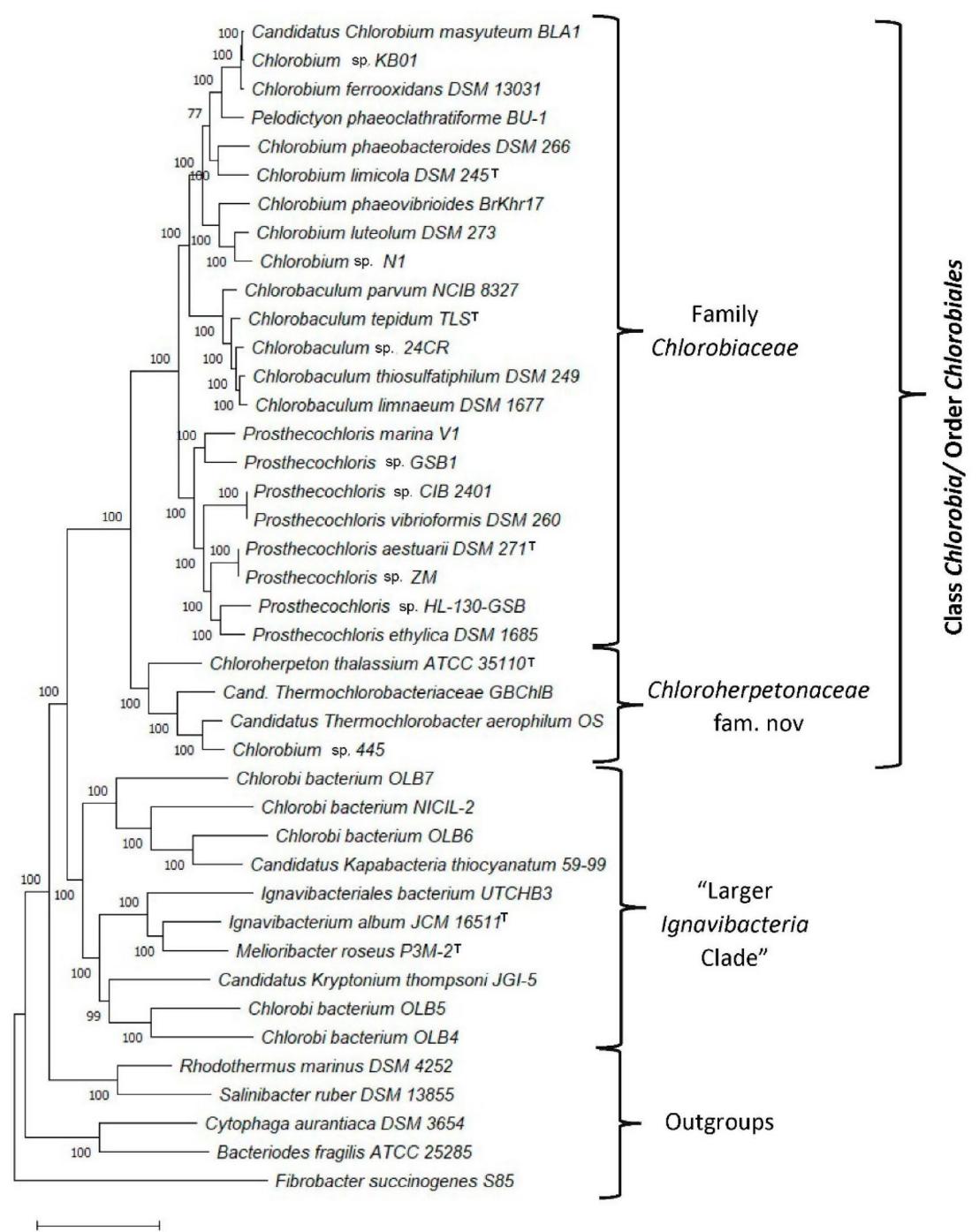


Os números nas ramificações indicam probabilidades posteriores bayesianas e valores de bootstrap de ML; valores de bootstrap <50% são indicados por ---. Ramos com valores de suporte maiores ou iguais a 0,99 probabilidade posterior de BI e 99% de bootstrap ML são indicados por pontos pretos;

Análise filogenômica de Holomycota.  
Árvore filogenômica de inferência bayesiana (BI) baseada em 175 proteínas conservadas de Lax et al. (2018). A árvore foi reconstruída usando 25 espécies e 59.889 posições de aminoácidos com o modelo CAT-GTR e o modelo LG+F+R6+C60 usando a aproximação PMSF para máxima verossimilhança (ML).

# Maximum-likelihood phylogenetic tree based on 16S rRNA gene sequences for the type strains of all *Chlorobi* species

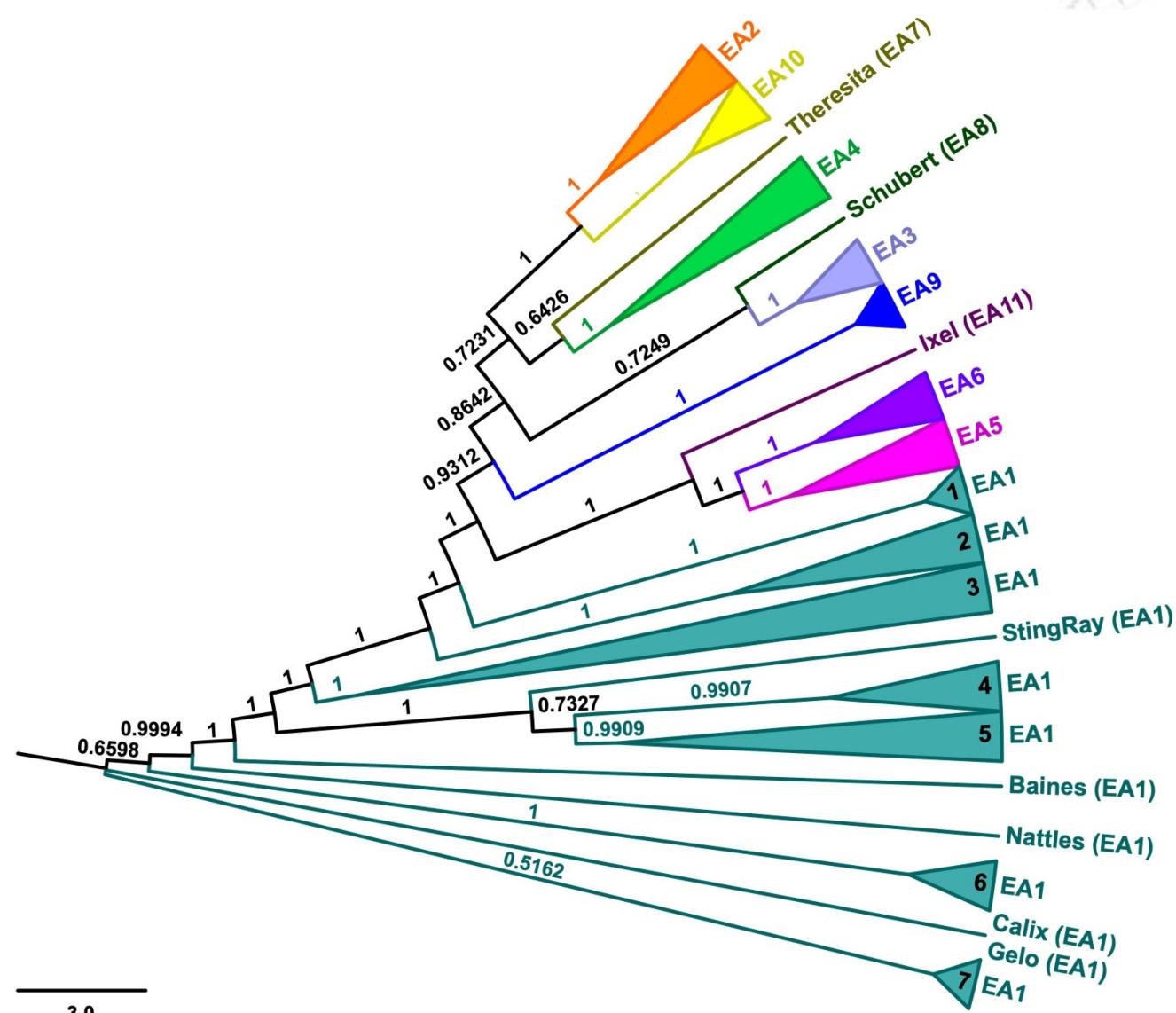




As relações evolutivas entre espécies/cepas de *Chlorobia* e *Ignavibacteria* foram examinadas usando análises filogenômicas e comparativas de sequências genômicas

- A bootstrapped maximum-likelihood tree for different genome-sequenced *Chlorobia*/*Ignavibacteria* and related species/strains based on concatenated sequences for 282 conserved proteins that are core proteins for these species

# *Microbacterium* Cluster EA Bacteriophages: Phylogenomic Relationships and Host Range Predictions

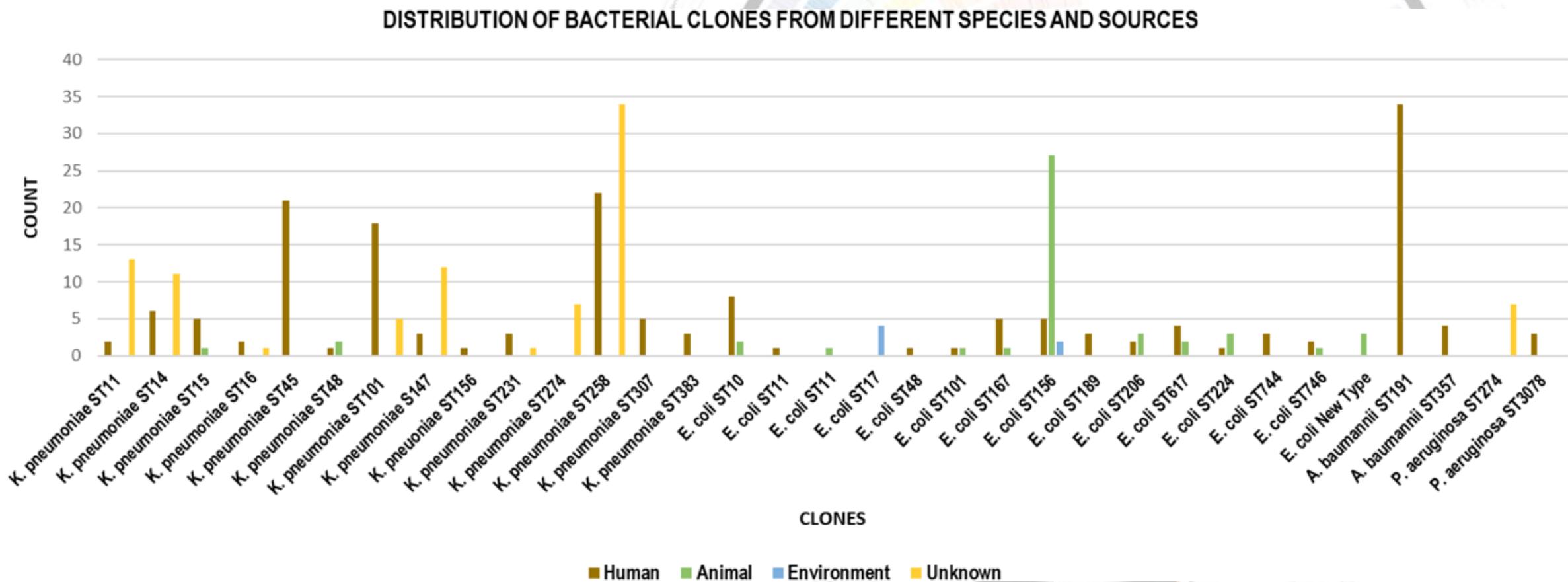


- Árvore Neighbor-joining gerada usando um alinhamento de múltiplas sequências de 125 genomas de bacteriófagos EA agrupados com 10.000 réplicas de bootstrap.
- As cores destacam a associação de bacteriófagos nos subgrupos EA1 – EA11. Os números nas ramificações indicam valores de bootstrap.

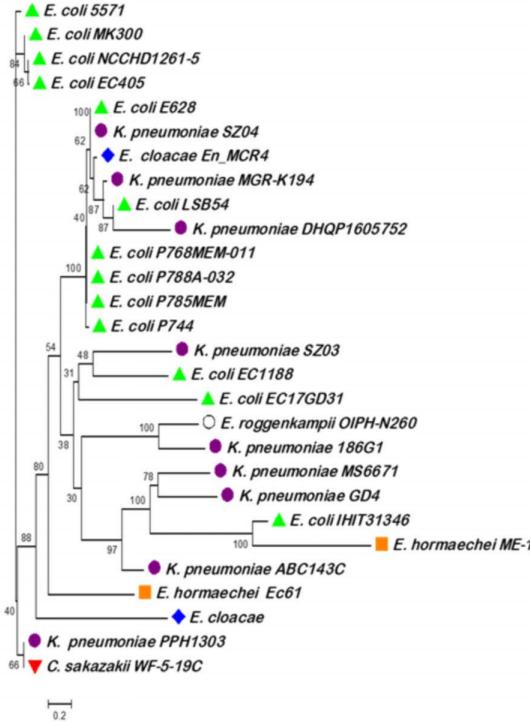
# **Phylogenomics, epigenomics, virulome and mobilome of Gram-negative bacteria co-resistant to carbapenems and polymyxins: a One Health systematic review and meta-analyses**

Um total de 877 isolados, 32 genomas e 44 sequências  
plasmídicas foram analisados

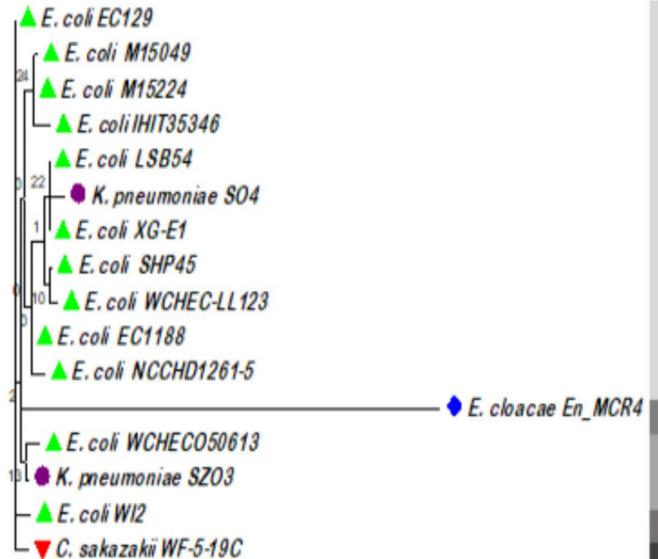
# Distribuição de tipos de sequências bacterianas com base no país, espécie e origem da amostra



A

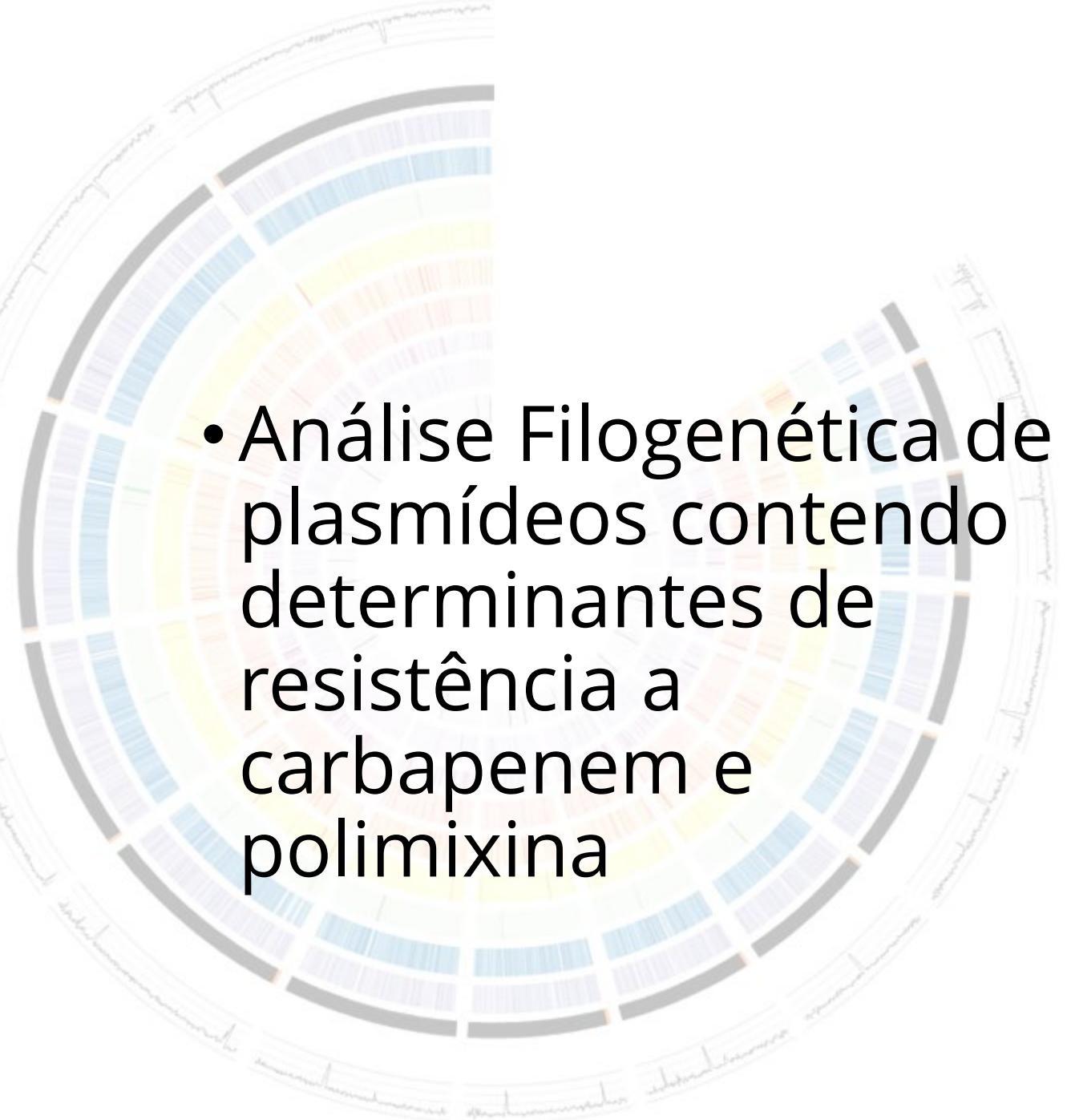


B



Clade	Clone	Plasmid	Source
I	ST31	IncC	H
	-	-	H
	ST132	-	H
	ST167	IncX3	H
	ST7153	IncX3	H
	ST25	IncX3	H
	ST84	IncX3	H
II	ST2179	IncX3	H
	ST15	A/C	H
	ST871	IncX3	A
	ST7510	IncX3	A
	ST7512	IncX3	A
	ST1602	IncX3	A
	ST25	IncX3	H
	ST3204	IncFII	H
	ST457	A/C	H
III	ST1059	IncFIB	H
	ST147	IncF	H
	ST11	IncF	H
	ST641	IncX3	H
	ST542	IncHI2	H
	ST147	IncF	H
IV	ST451	IncQ1	H
V	-	-	H
VI	-	IncB/O	A

Clade	Clone	Plasmid	Source
	ST167	IncHI2	H
	-	IncI	H
	-	IncI	H
	ST641	IncX4	H
	ST2179	IncHI2	H
I	ST25	IncX4	H
	ST466	IncHI2	A
	ST617	-	H
	ST206	IncHI2	A
	ST3204	IncI2	H
	ST132	-	H
II	ST84	CeEI	H
	-	-	H
III	ST25	IncX4	H
IV	ST1288	IncX4	H
V	-	IncI	H



- Análise Filogenética de plasmídeos contendo determinantes de resistência a carbapenem e polimixina