

Bioinformática: da Biologia à Flexibilidade Molecular



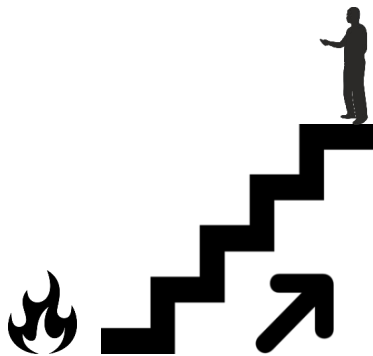
Filogenia Molecular

Alunos: Luiza Adami, Luiza Favarato e Marlonni Maurastoni

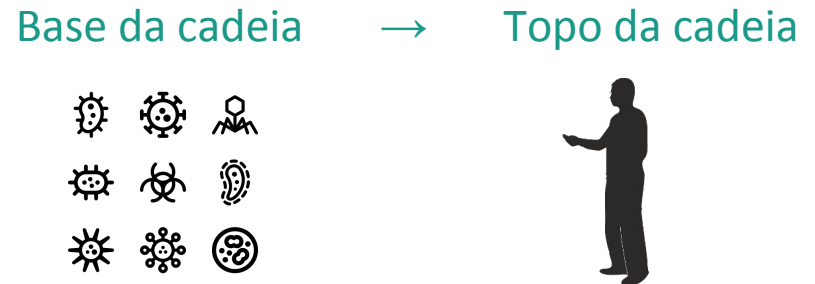
Professor: PhD Teodiano Freire Bastos Filho

5.1. Introdução

- A humanidade tem tendência de organizar e classificar o mundo à sua volta com o objetivo de facilitar o entendimento e a comunicação;
- Período pré-Darwiniano: **sistemas artificiais.**



Escala Natural de Platão



Sistema Classificatório de Lineu

5.1. Introdução

- Final do século XVIII e início do século XIX: **Sistemas Naturais de Classificação**:
 - Ordem natural dos seres vivos.
- E.G. Saint-Hillaire (1818):
 - **Homologia** – as asas de um morcego, as nadadeiras de uma baleia e os braços de um macaco;
 - Critério principal para uma classificação natural dos organismos.

5.1. Introdução

- “A Origem das Espécies” (Darwin, 1859):
 - Ancestralidade em comum.



A primeira árvore filogenética moderna (esboço de Darwin no manuscrito de A Origem das Espécies)

5.1. Introdução

- **Taxonomia Evolutiva**: incorporou o vetor tempo e adicionou uma quantificação da divergência estrutural entre os grupos;



A primeira árvore filogenética moderna (esboço de Darwin no manuscrito de A Origem das Espécies)

5.1. Introdução

- **Fenética:** buscava incluir na classificação dos organismos o máximo possível de características



A primeira árvore filogenética moderna (esboço de Darwin no manuscrito de A Origem das Espécies)

5.1. Introdução

- **Cladística** (Willi Hennig, 1950): organismos que compartilhassem características derivadas (apomórficas) poderiam ser considerados descendentes do organismo ancestral, na qual a característica em seu estado primitivo (ou plesiomórfico) passou para o estado derivado.



A primeira árvore filogenética moderna (esboço de Darwin no manuscrito de A Origem das Espécies)

5.1. Introdução

- **Sequenciamento**: proteico e genômico.
 - Sanger (1954);
 - Comparação de genes em **larga escala**:
 - Entre si ou associadas a outros dados (fósseis).
 - Precauções:
 - Genes podem sofrer diferentes **processos evolutivos** que **alteram sua estrutura e/ou função**, como mutações e rearranjos, ou ainda duplicações e perdas de função;
 - Muitas sequências do genoma não chegam à etapa de tradução, podendo conter **elementos regulatórios ou transponíveis**.

5.2. Aplicações

- Ao empregarmos as técnicas filogenéticas, o objetivo é coletar e analisar dados capazes de fornecer a melhor estimativa para chegarmos à filogenia verdadeira;
- A taxa de evolução molecular refere-se à velocidade na qual os organismos acumulam diferenças genéticas ao longo do tempo:
 - Hipótese do Relógio Molecular: as mudanças nas sequências se acumulam a uma taxa mais ou menos constante ao longo do tempo;
 - Relógios Moleculares Relaxados: desvios no comportamento temporal de diferentes linhagens moleculares.

5.2. Aplicações

- Além de desvendar as **relações evolutivas entre organismos**, a filogenética molecular pode ser aplicada para:
 - **Estimativa de tempo** às filogenias e datar surgimento de espécies;
 - Disseminação de organismos;
 - Entender grandes **eventos biológicos** que ocorreram no passado;
 - Modelos de dinâmica populacional;
 - **Reconstruir** genomas ancestrais;
 - Distribuição geográfica de indivíduos (**filogeografia**).

5.2. Aplicações

Genética

Epidemiologia

Microbiologia

Evolução

A filogenia molecular busca inferir a história evolutiva de organismos ou outras entidades biológicas (como proteínas e genes) a partir de sequências de ácidos nucleicos ou aminoácidos.

Resolução de crimes

Virologia

Zoologia

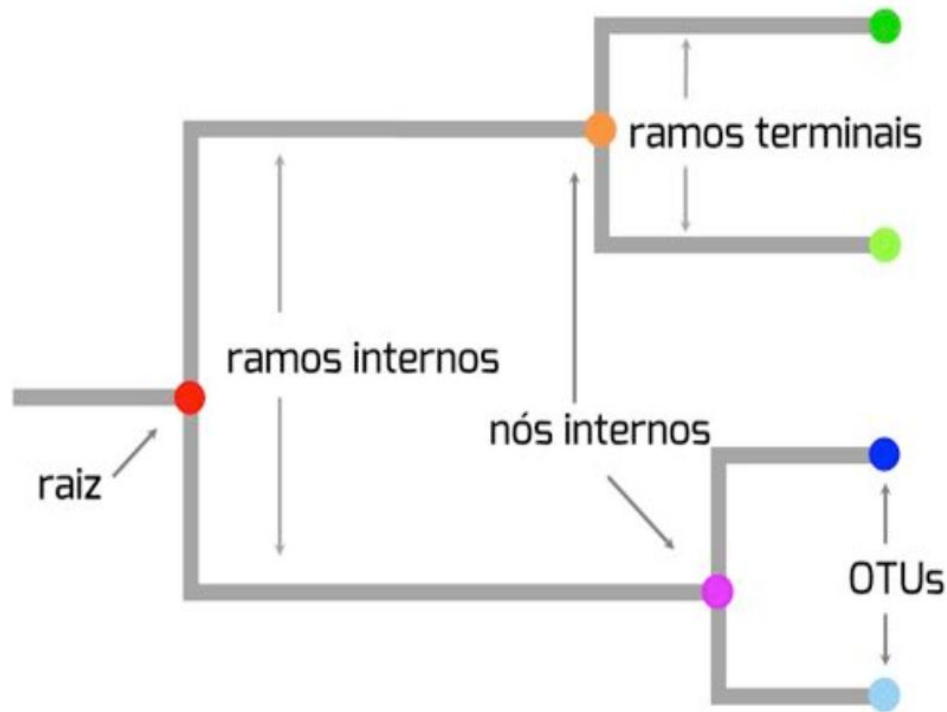
Botânica

Parasitologia

5.3. Representação de Árvores

- As hipóteses sobre a história evolutiva são o resultado dos estudos filogenéticos e se chamam **Filogenia** (árvores filogenéticas):
 - Representação **gráfica**;
 - Formadas por **nós** (pontos) ligados por **ramos** (linhas);
 - Os **nós terminais**, mais externos na filogenia, identificam os **indivíduos, genes ou proteínas** que foram amostrados

5.3. Representação de Árvores



OTUs = *operational taxonomical units* ou unidades taxonômicas operacionais.

* Ancestral comum.

Figura 1-5: Nomenclatura associada a árvores filogenéticas.

Estas notações foram desenvolvidas para utilização computacional da informação filogenética.

5.3. Representação de Árvores

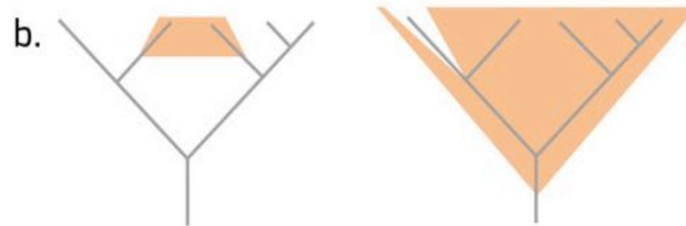
- Representação de filogenias: derivação **dicotômica** e derivação **politômica**.
 - Politomias verdadeiras;
 - Politomias falsas.
- Grupamentos segundo a sua **ancestralidade**:
 - Monofiléticos;
 - Parafiléticos;
 - Polifiléticos.

5.3. Representação de Árvores

Monofiléticos



Polifilético
(convergência
evolutiva)



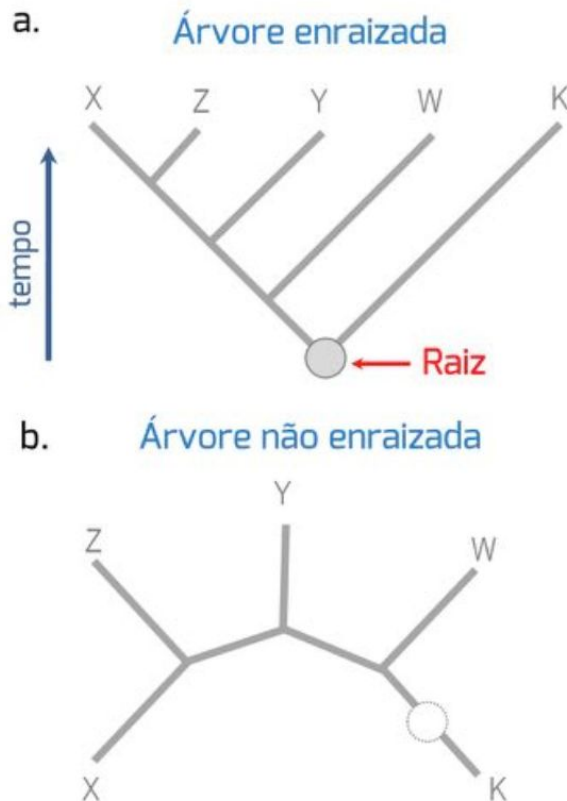
Parafilético



Monofiléticos
aninhados

Figura 3-5: (a) Exemplos de clados destacados em verde. (b) Exemplos de organizações da topologia que não caracterizam a existência de um clado, destacados em laranja. (c) Diferentes níveis de clados que podem estar embutidos em um clado de maior ordem. Observe que os clados de diferentes ordens, quando embutidos, formam clados monofiléticos.

5.3. Representação de Árvores



A identificação de uma **raiz** nas filogenias geralmente requer a **inclusão de uma ou diversas OTUs** que representem **grupos externos**. Os grupos externos devem ser **ancestrais comuns** das OTUs em estudo, já conhecidos, que indicarão caracteres presentes em organismos mais próximos aos ancestrais, provendo um **direcionamento** para a interpretação dos processos evolutivos.

Figura 4-5: Comparação de árvores (a) enraizadas e (b) não enraizadas. No primeiro caso, é possível definir a direção das mudanças evolutivas, devido à presença do vetor tempo dado pela presença da raiz.

5.3. Representação de Árvores

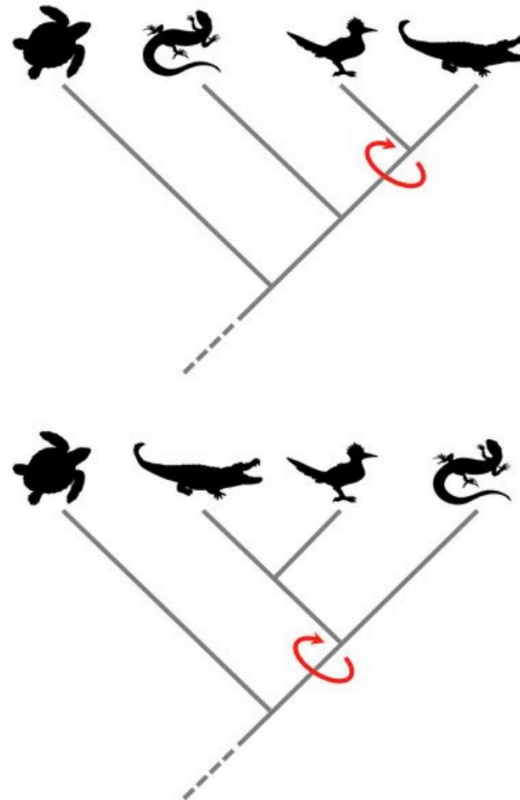


Figura 5-5: A porção terminal da árvore dos vertebrados (representada na Figura 2-5) foi reorganizada de diferentes maneiras (as setas indicam o ponto de rotação). Conforme a analogia de um móvel, todas elas representam a mesma relação evolutiva.

5.3. Representação de Árvores

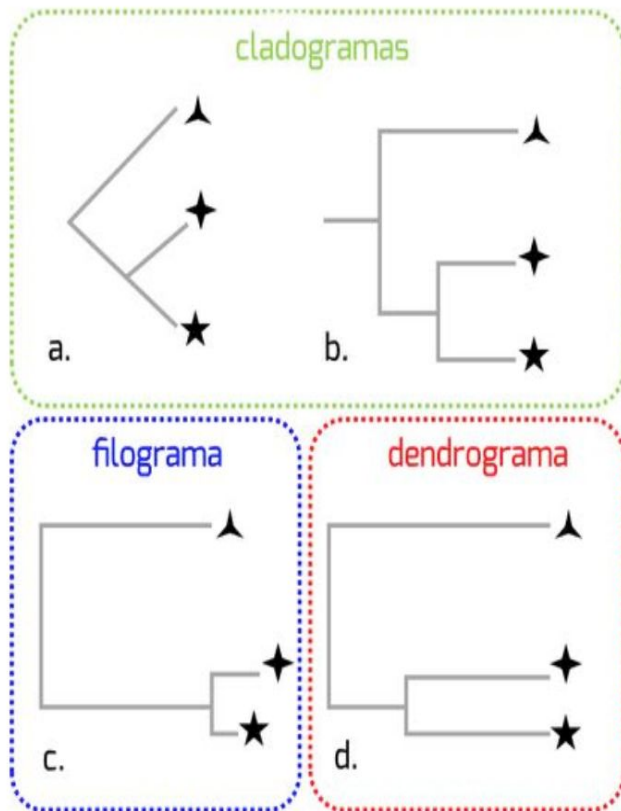
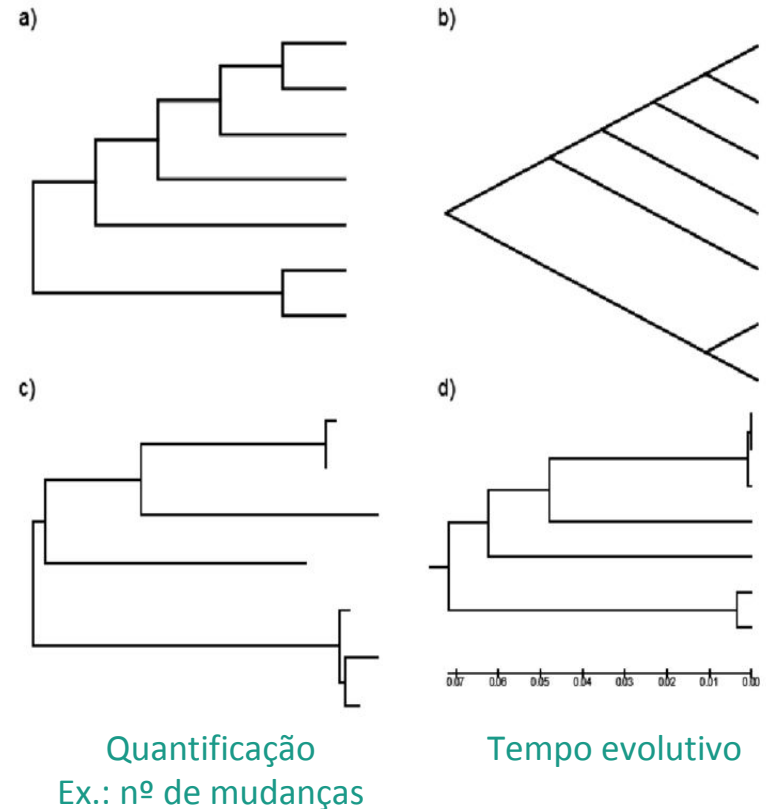


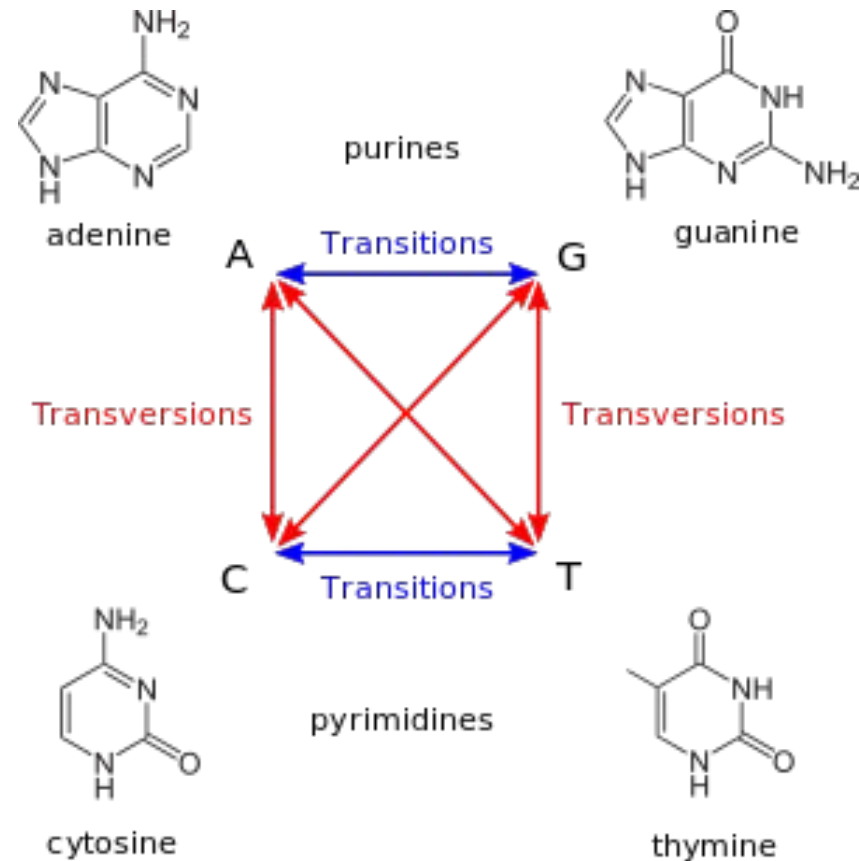
Figura 6-5: Nomenclatura de árvores filogenéticas. Observe que os cladogramas *a* e *b* são equivalentes, mas o filograma *c* e o dendrograma *d* não o são.

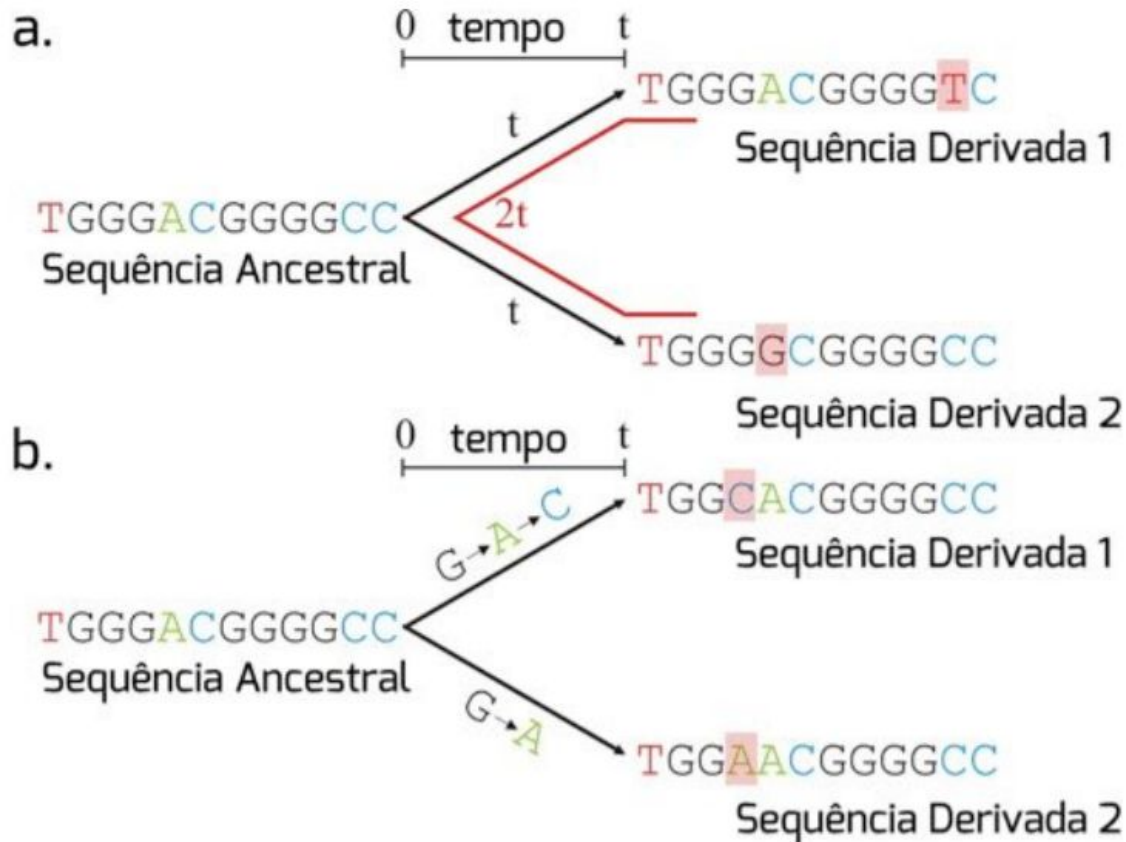


5.3. Representação de Árvores

- O tipo de dado molecular a ser empregado nas análises também deve ser levado em conta:
 - Sequências de aminoácidos;
 - Sequências de nucleotídeos.
- É preciso, assim, conhecer o caso de estudo e o tipo de pergunta que se busca responder com cada filogenia.

- Para construção de uma filogenia foi necessário a elaboração de modelos evolutivos que descrevem matematicamente as mudanças nos organismos ao longo do tempo.
- Matematicamente não é factível colocar os processos de **substituição, inserção, deleção e duplicação**, bem como ocorrência de **transposição** ou até mesmo de **retrotransposição** para explicar inteiramente o processo evolutivo.
- As **substituições** têm sido o principal alvo para o desenvolvimento de modelos matemáticos e compõem a base de diversos métodos de inferência filogenética,
 - relevância para explicar os mecanismos evolutivos
 - disponibilidade de modelos de probabilidade estatística





Distância p :

contagem das diferenças
dividida pelo número total de
sítios do alinhamento

8 sítios diferentes

100pb

Distância $p = 0,08$

A ocorrência de múltiplas substituições ao longo do tempo na divergência de sequências homólogas pode mascarar as verdadeiras diferenças entre as sequências (*multiple hits*).

Distância genética: o número de substituições de nucleotídeos que se acumularam nas sequências desde a divergência.

Distância genética

- O conceito de distância genética, muitas vezes, auxilia o pesquisador na construção de filogenias.
- Está diretamente relacionada com a relação evolutiva entre duas OTUs:
 - **Menor distância genética indica uma relação evolutiva mais próxima.**
 - **Maior distância sugere uma derivação evolutiva proporcionalmente maior.**

Inferência filogenética

O alinhamento, além de ser o primeiro passo garantir maior confiabilidade nas análises posteriores, é requerido por todos os métodos de inferência filogenética para construção da árvore.

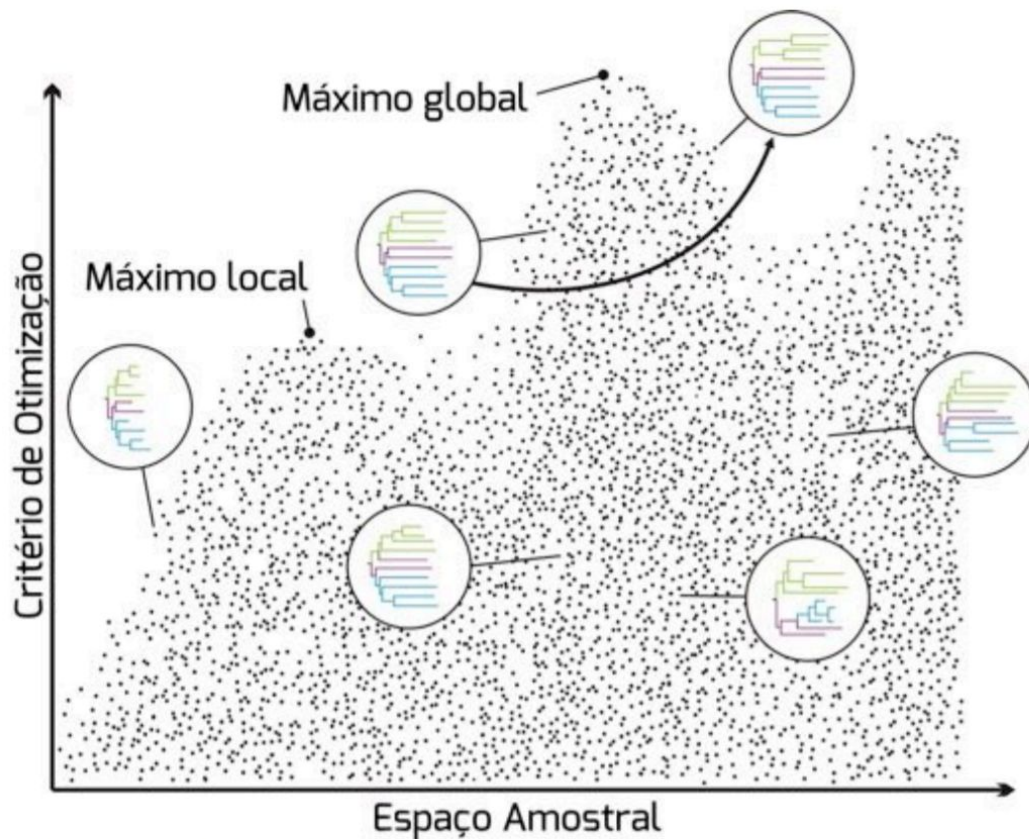
- Métodos para estimar filogenia:

Quantitativos: convertem o alinhamento em matrizes de distância par-a-par para todas as sequências incluídas. São utilizados para construção de uma filogenia inicial, que posteriormente é submetida a algum método do grupo qualitativo. Desvantagem por apresentar apenas uma filogenia inicial.

Qualitativos: atribuem um critério de otimização para escolher a melhor filogenia. Constroem várias filogenias, o critério atribui um valor a cada filogenia e as ordena segundo este valor.

Tabela 1-5: Comparação entre os tipos de métodos para inferência de filogenias.

Tipo	Método	Princípio	Programa
Métodos Quantitativos	UPGMA	Agrupa sequencialmente as OTUs com menor distância evolutiva entre si	Geneious MEGA MEGA
	Aproximação dos vizinhos	Busca a árvore com a menor soma total de ramos	Geneious HyPhy
	Máxima Parcimônia	Busca a filogenia com menor número de eventos evolutivos	PAUP MEGA Mesquite
Métodos Qualitativos	Máxima Verossimilhança	Busca a árvore com o valor de maior verossimilhança entre todas as filogenias construídas	PAUP PAML phyML MEGA
	Estatística Bayesiana	Amostra um número representativo de filogenias a partir do espaço amostral total de árvores e busca a mais provável	Mr. Bayes BEAST BAMBE



Métodos qualitativos exigem algoritmos para escolha da melhor filogenia.
Exatos e heurísticos.

Exatos

Seleção aleatória inicial de três OTUs e construção de árvore inicial. Por tentativa, um a um, novas OTUs do alinhamento, são inseridas em diferentes posições na árvore. Esse procedimento é repetido até todos os táxons serem inseridos, gerando todas as filogenias possíveis para o alinhamento.

Utilizando um critério de otimização (dado pelo método qualitativo) as filogenias são organizadas em um espaço virtual que contém todas as filogenias possíveis para o alinhamento empregado

Heurísticos

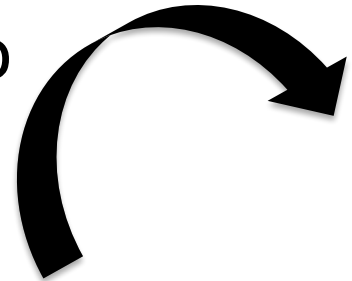
procuram pela melhor filogenia em um subconjunto de todas as filogenias possíveis
estes métodos não garantem que a filogenia correta seja encontrada, pois apenas algumas árvores do espaço amostral total serão consideradas

5.6 Abordagens Quantitativas

UPGMA – Método de agrupamento par a par usando médias aritméticas não ponderadas

- ❖ Proposto por Sneath e Sokal, 1973;
- ❖ Mais simples para reconstrução filogenética;
- ❖ Pressuposto: Relógio Molecular;
- ❖ Medidas de distância evolutiva são computadas e organizadas na forma de uma matriz;

Exemplo



5.6 Abordagens Quantitativas

UPGMA – Método de agrupamento par a par usando médias aritméticas não ponderadas

Matriz Original

	A	B	C	D	E
A	0	10	12	8	7
B	10	0	4	4	14
C	12	4	0	6	16
D	8	4	6	0	12
E	7	14	16	12	0



	A	B	C	D	E
A		10	12	8	7
B			4	4	14
C				6	16
D					12
E					

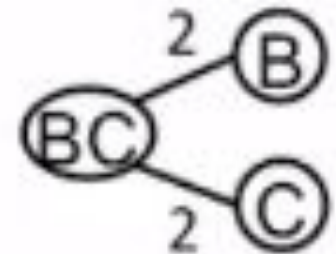
5.6 Abordagens Quantitativas

UPGMA – Método de agrupamento par a par usando médias aritméticas não ponderadas

Matriz Original

	A	B	C	D	E
A		10	12	8	7
B			4	4	14
C				6	16
D					12
E					

Árvore



5.6 Abordagens Quantitativas

UPGMA – Método de agrupamento par a par usando médias aritméticas não ponderadas

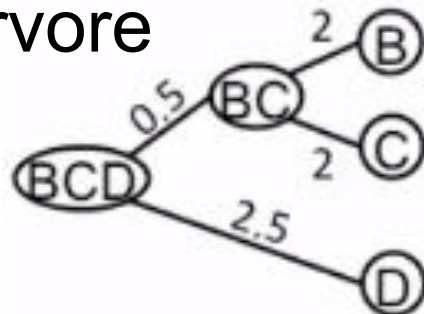
Matriz Original

	A	B	C	D	E
A		10	12	8	7
B			4	4	14
C				6	16
D					12
E					



	A	BC	D	E
A		11	8	7
BC			5	15
D				12
E				

Árvore



5.6 Abordagens Quantitativas

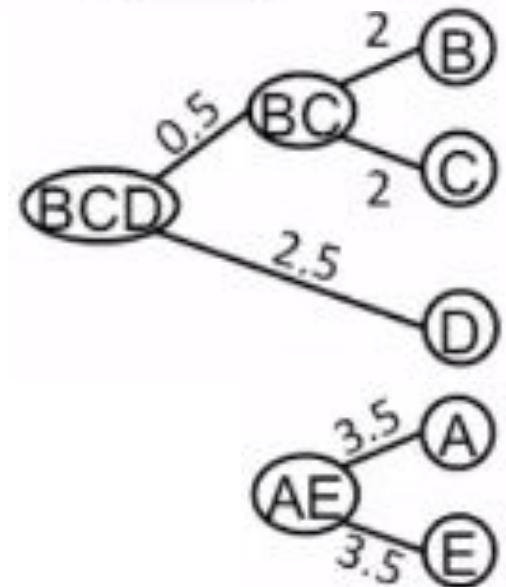
UPGMA – Método de agrupamento par a par usando médias aritméticas não ponderadas

Matriz Original

	A	B	C	D	E
A		10	12	8	7
B			4	4	14
C				6	16
D					12
E					



	A	BCD	E
A		10	7
BCD			14
E			



Árvore

5.6 Abordagens Quantitativas

UPGMA – Método de agrupamento par a par usando médias aritméticas não ponderadas

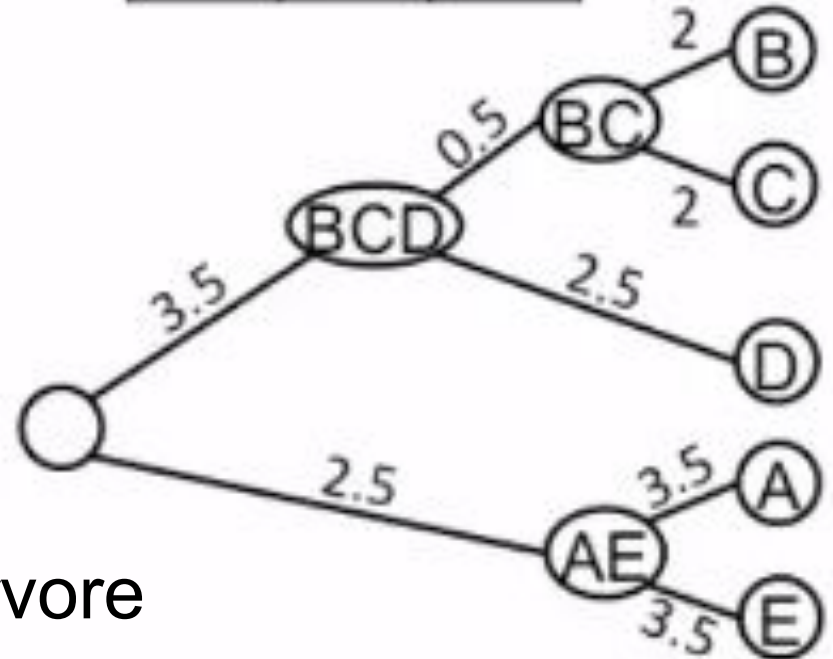
Matriz Original

	A	B	C	D	E
A		10	12	8	7
B			4	4	14
C				6	16
D					12
E					

$$\frac{10+12+8+14+16+12}{6} = 12$$



	AE	BCD
AE		12
BCD		



Árvore

5.6 Abordagens Quantitativas

UPGMA – Método de agrupamento par a par usando médias aritméticas não ponderadas

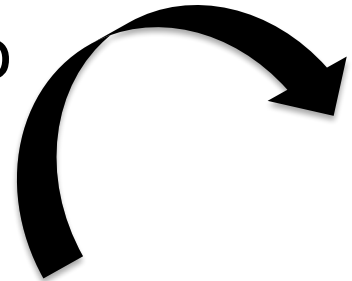
“Por se basear na hipótese do relógio molecular, o UPGMA pode levar à obtenção de topologias falsas quando tal hipótese não for satisfeita pelos dados. Sabe-se que o método é muito sensível a variações nas taxas evolutivas entre linhagens, fato este que levou a proposição de métodos onde as variações são ajustadas para a obtenção de sequências que satisfaçam o relógio molecular. Apesar disso, devido ao surgimento de métodos mais robustos e mais eficientes em lidar com dados não uniformes, o UPGMA encontra-se praticamente abandonado como alternativa para reconstrução filogenética.”

5.6 Abordagens Quantitativas

NJ - Aproximação de Vizinhos

- ❖ Proposto por Saitou e Nei, 1987;
- ❖ Aceleração dos algoritmos de evolução mínima;
- ❖ Anteriormente: Todas as árvores precisavam ser construídas para se verificar a de menor soma;
- ❖ Similar à UPGMA, mas se baseia na distância entre dois em relação aos outros.
- ❖ Diminui a taxa de erro do UPGMA.

Exemplo



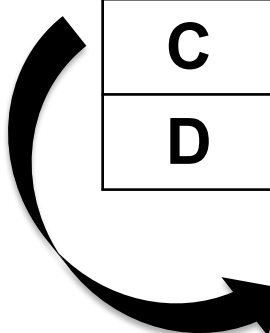
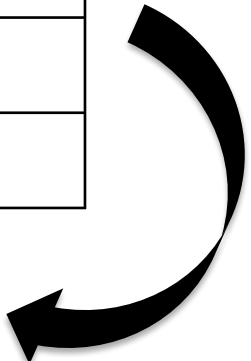
5.6 Abordagens Quantitativas

NJ - Aproximação de Vizinhos

<i>d</i>	A	B	C	D
A	0	8	4	6
B	8	0	8	8
C	4	8	0	6
D	6	8	6	0

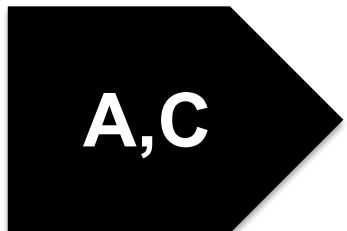
Σ

<i>u</i>	
A	18
B	24
C	18
D	20


$$Q_{(i,j)} = (r - 2)d(C_i, C_j) - u(C_i) - u(C_j)$$




Q	A	B	C	D
A	0	-26	-28	-26
B	-26	0	-26	-28
C	-28	-26	0	-26
D	-26	-28	-26	0



A,C

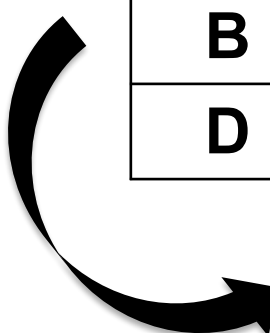
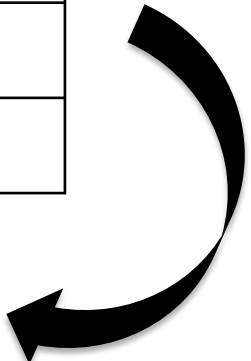
5.6 Abordagens Quantitativas

NJ - Aproximação de Vizinhos

d	A,C	B	D
A,C	0	6	4
B	6	0	8
D	4	8	0

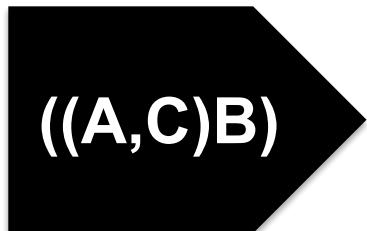
 Σ

u	
A,C	10
B	14
D	12


$$Q_{(i,j)} = (r - 2)d(C_i, C_j) - u(C_i) - u(C_j)$$




Q	A,C	B	D
A,C	0	-18	-18
B	-18	0	-18
D	-18	-18	0



((A,C)B)

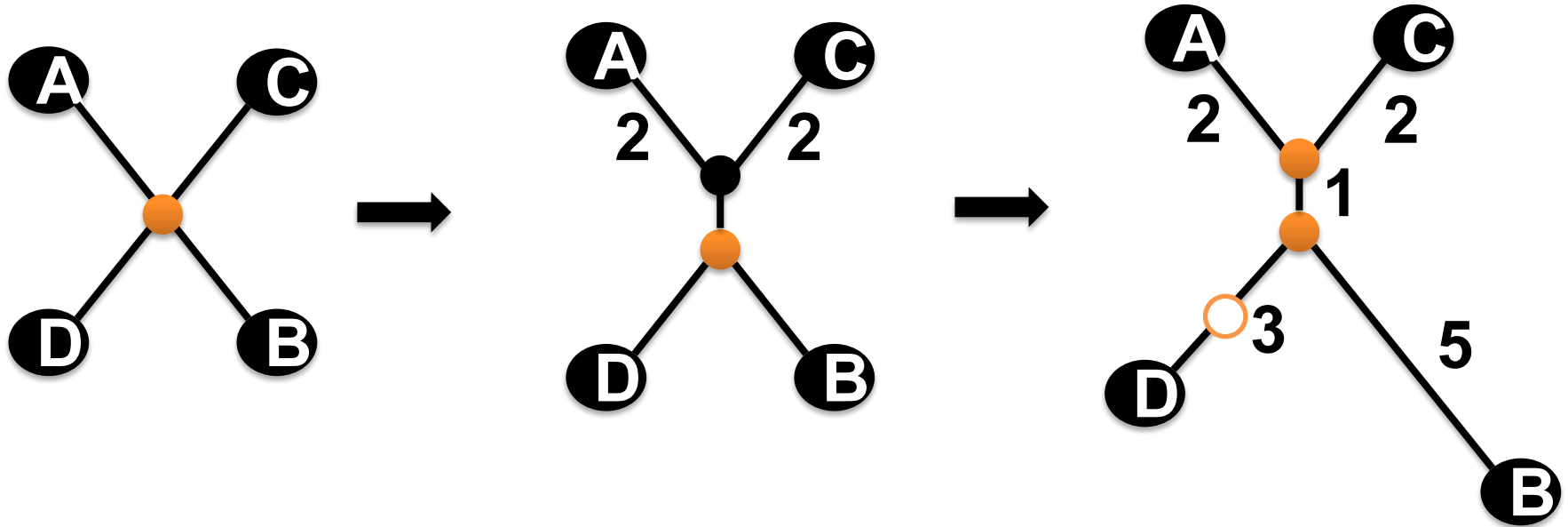
5.6 Abordagens Quantitativas

NJ - Aproximação de Vizinhos

d	A,B,C	D
A,B,C	0	3
D	3	0

 Σ

u	
A,B,C	3
D	3



Parcimônia

- Atualmente a máxima parcimônia foi substituída por outros métodos, como máxima verossimilhança e inferência Bayesiana devido, principalmente, às **simplificações nos processos evolutivos** assumidas pelo método e, sobretudo, nas limitações de seu uso.
- Dentro de um conjunto de filogenias, aquela filogenia que apresentar o menor número de eventos evolutivos (substituições) deve ser a mais provável para explicar os dados do alinhamento.
- O método calculará as probabilidades de mudanças dos nucleotídeos nos ramos da árvore. A parcimônia considera cada sítio do alinhamento individualmente e calcula as probabilidades de ocorrência dos quatro nucleotídeos nos táxons ancestrais

Devido ao caráter probabilístico do método, é necessário que certas pressuposições sejam estabelecidas para especificar o custo de substituição dos nucleotídeos.

Matriz de custo igual: assume que as substituições de nucleotídeos tem custo 1, enquanto que a não alteração não é penalizada.

Matriz de custo desigual: proposto para especificar as transições e as transversões, leva em consideração a diferença na probabilidade de mudança entre purinas e pirimidina. Considera a maior probabilidade de ocorrência de transições em relação às transversões.

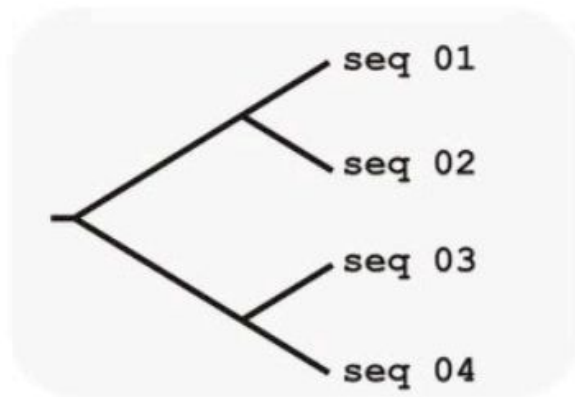
a.

$$\text{Matriz de custo igual} = \begin{matrix} & \begin{matrix} A & C & G & T \end{matrix} \\ \begin{matrix} A \\ C \\ G \\ T \end{matrix} & \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \end{matrix}$$

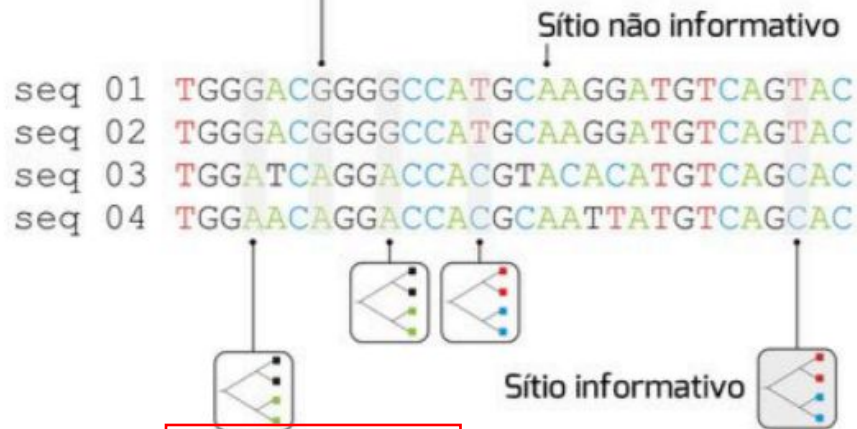
b.

$$\text{Matriz de custo desigual} = \begin{matrix} & \begin{matrix} A & C & G & T \end{matrix} \\ \begin{matrix} A \\ C \\ G \\ T \end{matrix} & \begin{bmatrix} 0 & 4 & 1 & 4 \\ 4 & 0 & 4 & 1 \\ 1 & 4 & 0 & 4 \\ 4 & 1 & 4 & 0 \end{bmatrix} \end{matrix}$$

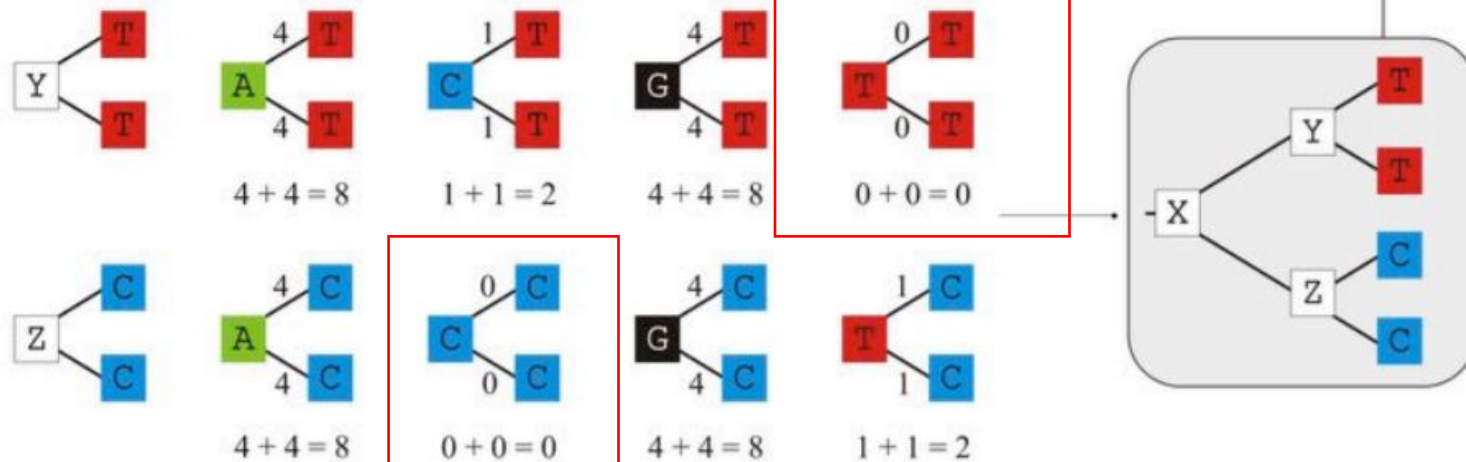
a.



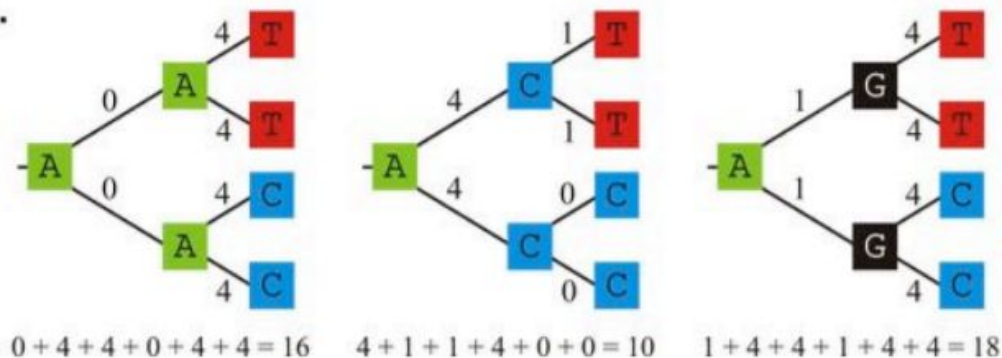
b.



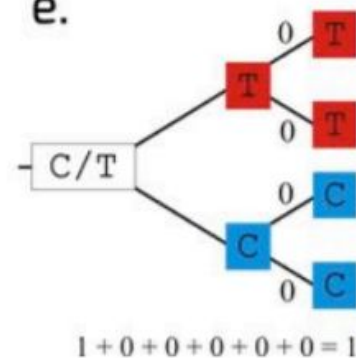
c.



d.



e.



Desvantagens

Falham ao estimar a relação evolutiva entre um grande número de táxons, especialmente se diferentes linhagens possuem **taxas evolutivas variáveis ou taxas evolutivas muito rápidas**. Nestes casos, é comum que o método agrupe incorretamente os táxons com maiores taxas de evolução, levando à inferência da filogenia errada (atração de ramos longos)

Por não ter um modelo de substituição especificado, o método de parcimônia é incapaz de considerar **mutações reversas ou múltiplas substituições**.

5.7 Abordagens Qualitativas

Máxima Verossimilhança

- ❖ Alia informações do alinhamento a um modelo estatístico capaz de lidar com a mudança de um nucleotídeo de maneira mais completa;
- ❖ Amplamente utilizado, devido à qualidade da abordagem estatística;

“Devido ao uso de diferentes algoritmos, na prática, um mesmo conjunto de sequências submetido a diferentes programas para inferência filogenética por máxima verossimilhança dificilmente resultará na mesma árvore.”

5.7 Abordagens Qualitativas

Máxima Verossimilhança

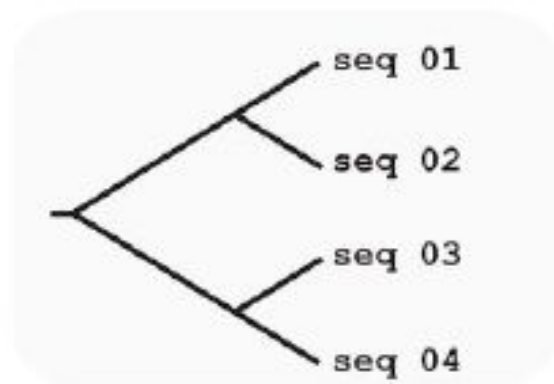
- ❖ Se tornou popular na área da filogenética a partir de 1981, com o desenvolvimento de um algoritmo para estimar filogenias baseadas no alinhamento de nucleotídeos;
- ❖ Diversos programas utilizam esse método, como o PAUP, MEGA, PHYLIP, fastDNAmI, IQPNNI e METAPIGA
- ❖ Infere a história evolutiva mais consistente;

Verossimilhança de uma árvore

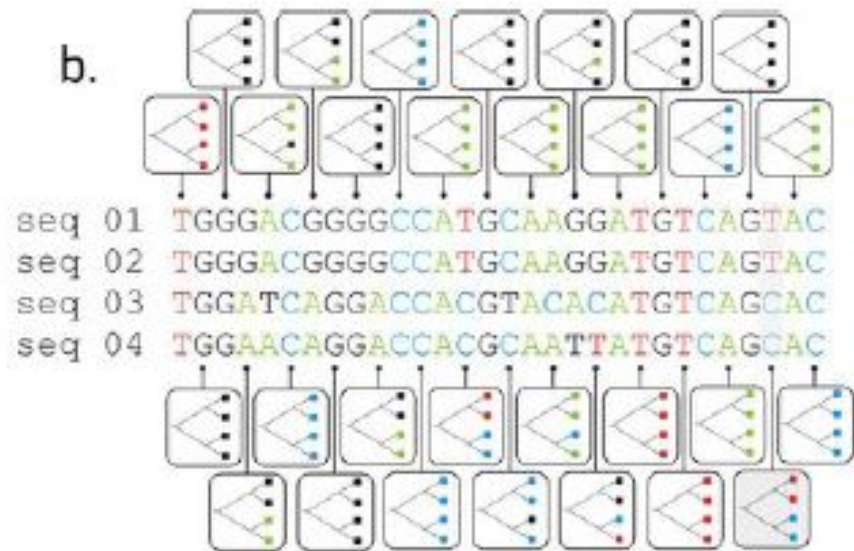
α

Probabilidade de explicar os dados do alinhamento

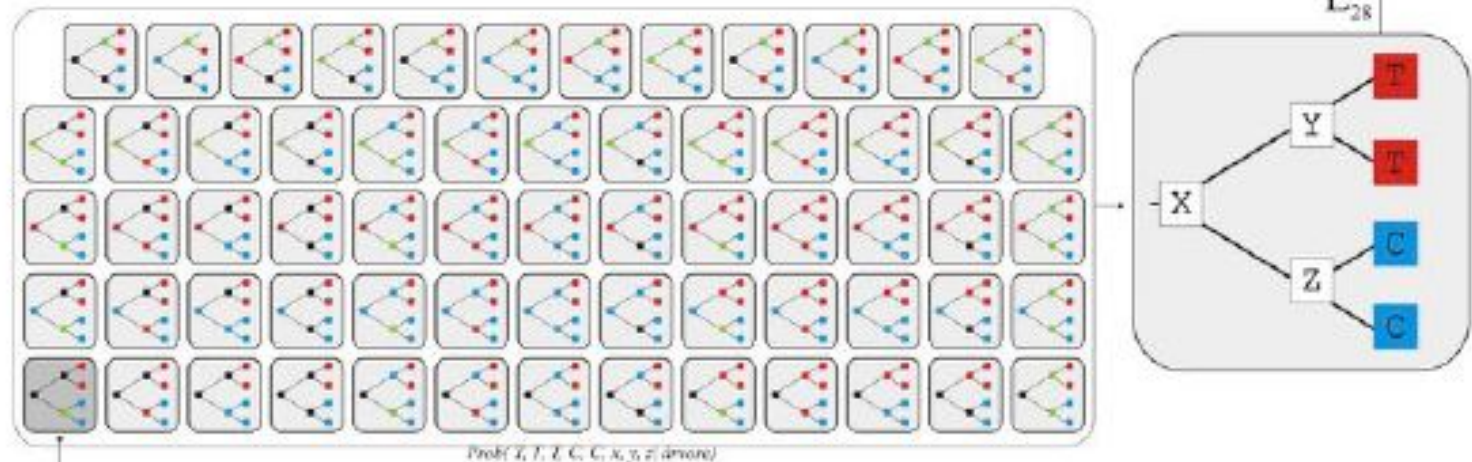
a.



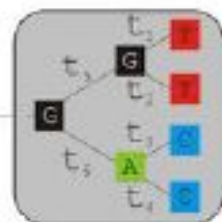
b.



c.



d.



$$\text{Prob}(T, T, T, C, C, x, y, z | \text{árvore}) =$$

$$\pi_G \cdot P_{GG} \cdot (t_5) \cdot P_{GT} \cdot (t_1) \cdot P_{GT} \cdot (t_2) \cdot P_{GA} \cdot (t_6) \cdot P_{AC} \cdot (t_3) \cdot P_{AC} \cdot (t_4)$$

Probabilidade de ocorrência dos nucleotídeos (A, C, G, T) no nó mais ancestral é dada pela frequência estacionária.

Probabilidade de mudança do nó X para o nó Y multiplicando pelo tamanho do ramo t_i e considerando variações em parâmetros do modelo de substituição.

5.7 Abordagens Qualitativas

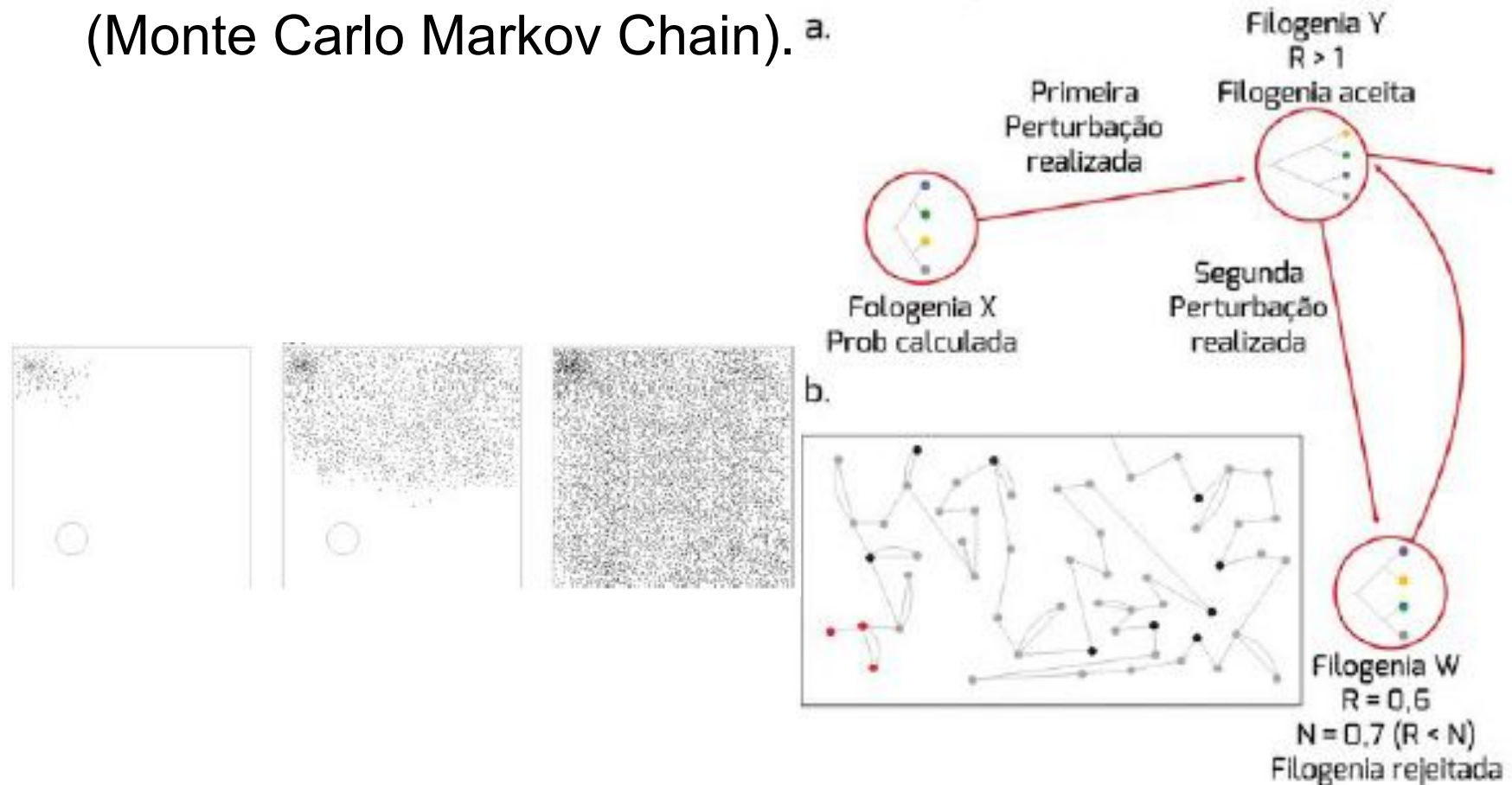
Análises Bayesianas

- ❖ A inferência Bayesiana engloba o método de máxima verosimilhança e informações a respeito da filogenia, do alinhamento ou dos táxons (probabilidade anterior);
- ❖ Ao final das análises, é possível verificar a probabilidade de cada filogenia (probabilidade posterior);

5.7 Abordagens Qualitativas

Análises Bayesianas

- ❖ Perturbações aleatórias são realizadas nas filogenias geradas, para descartar ou manter determinada filogenia (Monte Carlo Markov Chain).



5.7 Abordagens Qualitativas

Análises Bayesianas

Através da Análise Bayesiana é possível:

- ❖ Fazer inferência filogenética;
- ❖ Refinar as análises;
- ❖ Datar filogenias e ramos específicos;
- ❖ Inferir caracteres ancestrais;
- ❖ Analisar a dinâmica populacional sob uma ótica evolutiva.

Confiabilidade

- Reconstrução fidedigna das relações evolutivas.
- **Precisão:** número de filogenias que podem ser excluídas, a partir do conjunto total de filogenias, por não serem “verdadeiras”. **Quanto maior o número de filogenias excluídas neste processo, mais preciso é o método.**
 - **Reavaliar:** gene, tamanho das sequências, número de indivíduos e o grupo externo.
- Como é possível saber se a amostragem foi suficiente e a filogenia é confiável?
 - Reamostragem – Usar novas amostras e a filogenia resultante ser a mesma
 - Algoritmos: *bootstrap* – método de reamostragem

O Bootstrap

A inferência destas filogenias será realizada pelo método de construção especificado pelo usuário, seja aproximação de vizinhos, máxima parcimônia ou máxima verossimilhança

Ao final, o algoritmo analisará os clados e automaticamente verificará a presença de determinados agrupamentos em todas as filogenias construídas. **Se, por exemplo, encontramos as sequências 1 e 2 formando um clado em 70% das filogenias construídas, atribuiremos a confiabilidade de 70 ao clado formado por estas duas sequências**

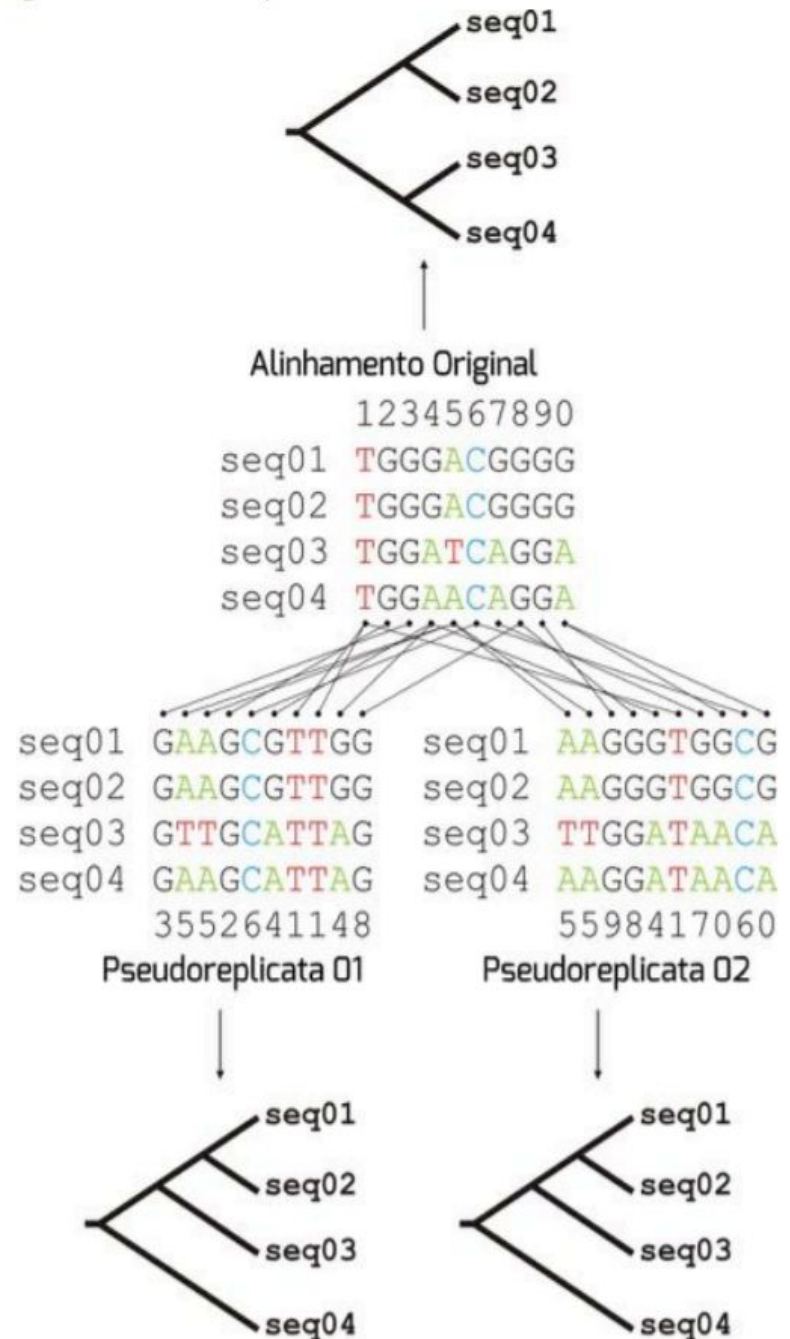
majority-rule consensus tree: Neste método, o algoritmo tabulará todos os clados formados em todas as replicatas geradas. Aqueles clados que mais aparecerem servirão para montar a filogenia consenso.

A confiabilidade em filogenias Bayesianas já é construída durante o processo.

Método bayesiano superestima a confiabilidade

Bootstrap >70: confiante

Métodos bayesianos >90: confiante



5.9. Interpretação de filogenias

- A interpretação das implicações evolutivas associadas a um, ou a um conjunto de táxons, está diretamente relacionada à disposição dos ramos internos e externos de uma árvore.
- Raiz:
 - Enraizamento por grupo externo - utiliza organismos sabidamente relacionados ao grupo em evidência;
 - Filogenias sem raiz - é necessário ter cautela nas interpretações, pois este tipo de diagrama apenas revela a relação entre os táxons.

5.9. Interpretação de filogenias

- Avaliar os ramos:
 - Podem ter significados diferentes.
- Confiabilidade:
 - Conclusões evolutivas devem ser sustentadas em árvore confiáveis;
 - A medida de confiabilidade dos ramos deve ser denotada;
 - Ramos com maiores valores de confiabilidade gerarão conclusões mais confiáveis.

5.9. Interpretação de filogenias

- O **padrão de organização** dos ramos de uma filogenia denota o **padrão de ancestralidade**.

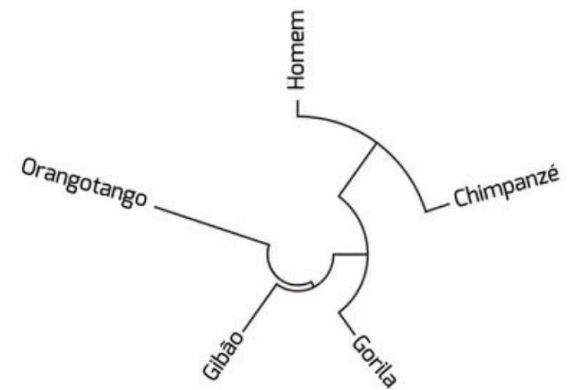
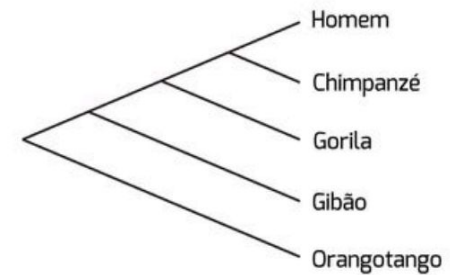
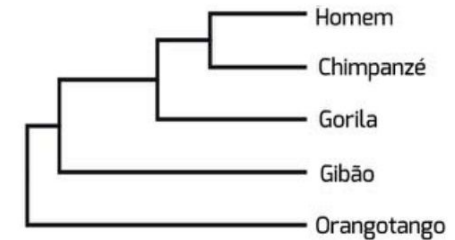


Figura 20-5: Diferentes representações da filogenia dos primatas.

5.9. Interpretação de filogenias

- Objetivo do estudo filogenético:
 - Responder uma determinada questão.
- Quando possível, é importante **reconstruir** a filogenia utilizando **diferentes métodos** de inferência e compará-las entre si:
 - **Incluir outras ferramentas** para complementar e sustentar a interpretação de uma filogenia.
 - Fósseis, por exemplo.

5.9. Interpretação de filogenias

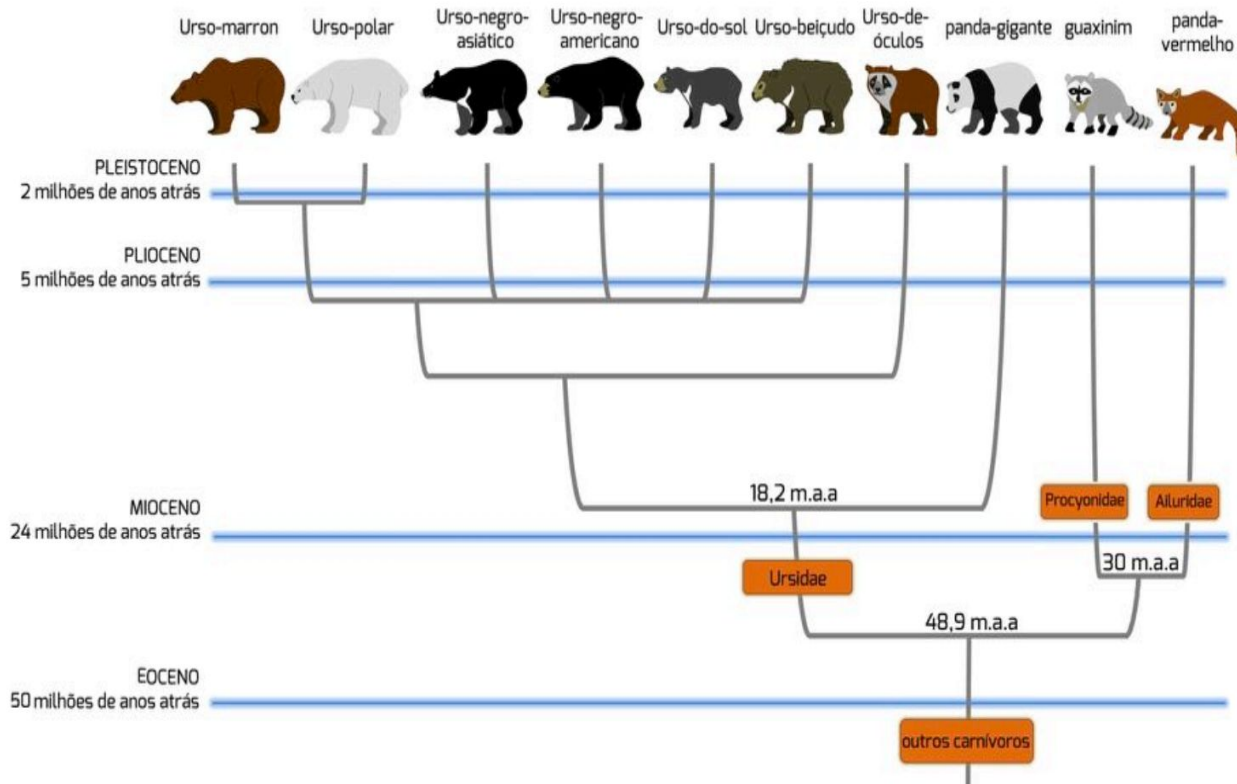


Figura 21-5: Posição filogenética do panda-gigante, baseada na combinação de diferentes tipos de dados. Baseado em BININDA-EMONDS, Olaf R.P. *Phylogenetic position of the giant panda*. Em: LINDBURG, D.G. & Baragona, K. *Giant pandas: Biology and conservation*. Berkeley: University of California Press, 2004; e em EIZIRIK, Eduardo e colaboradores: *Pattern and timing of diversification of the mammalian order Carnivora inferred from multiple nuclear gene sequences*. *Mol Phylogenet Evol*, 56, 49, 2010.

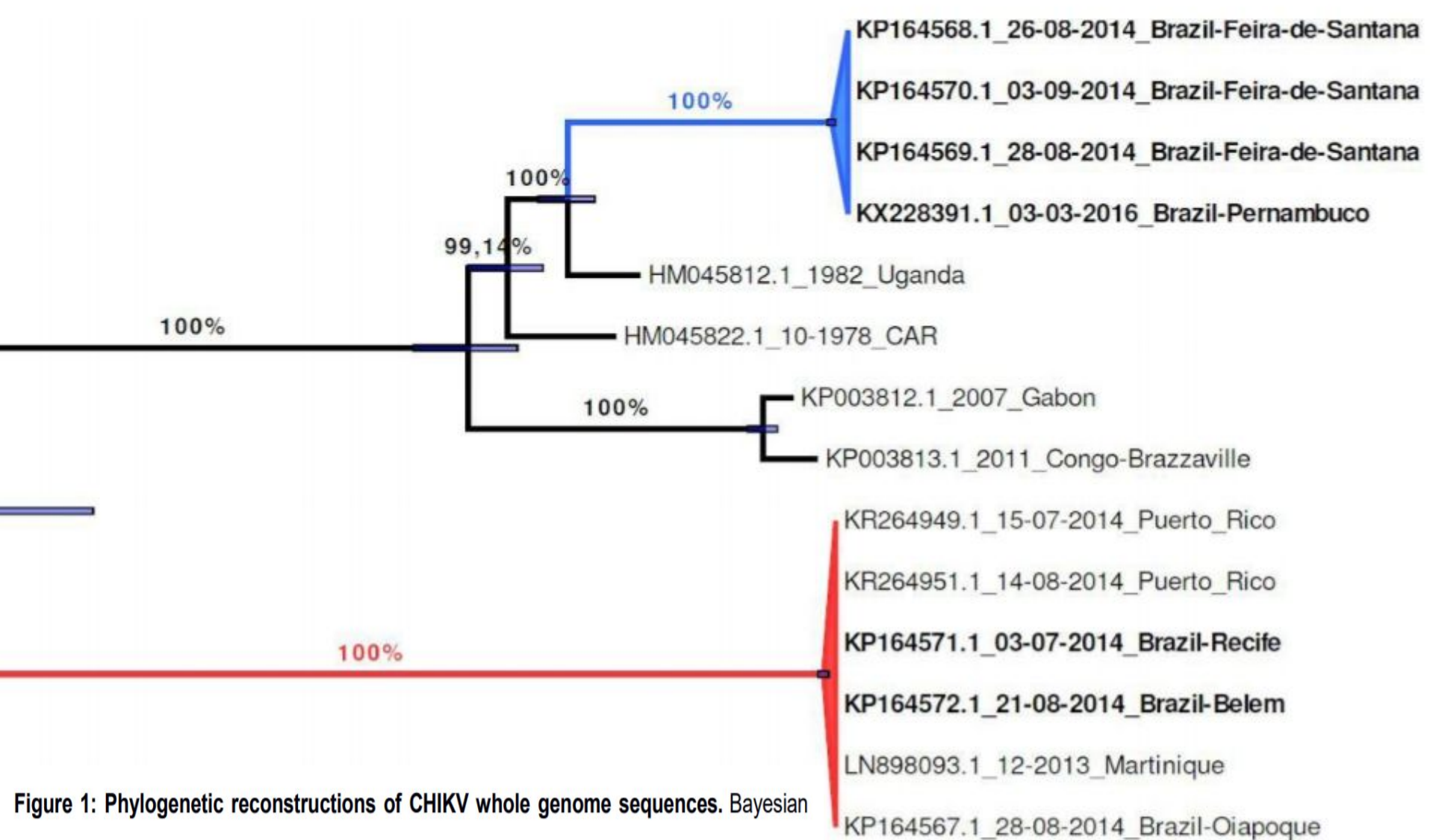
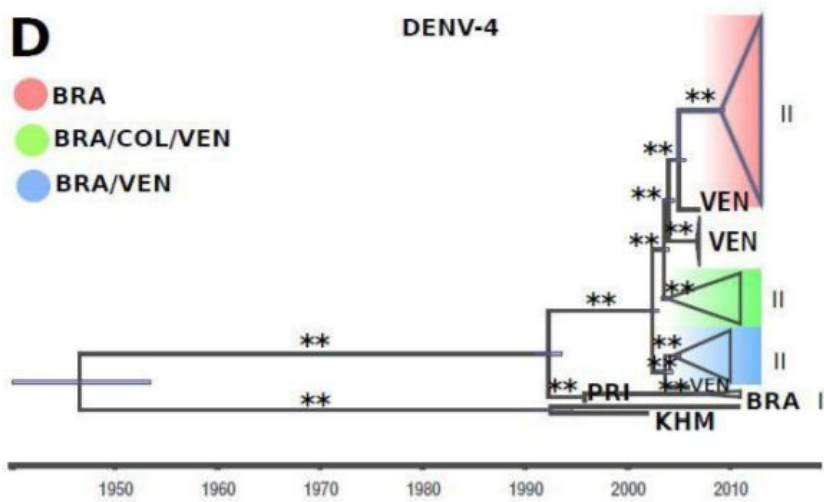
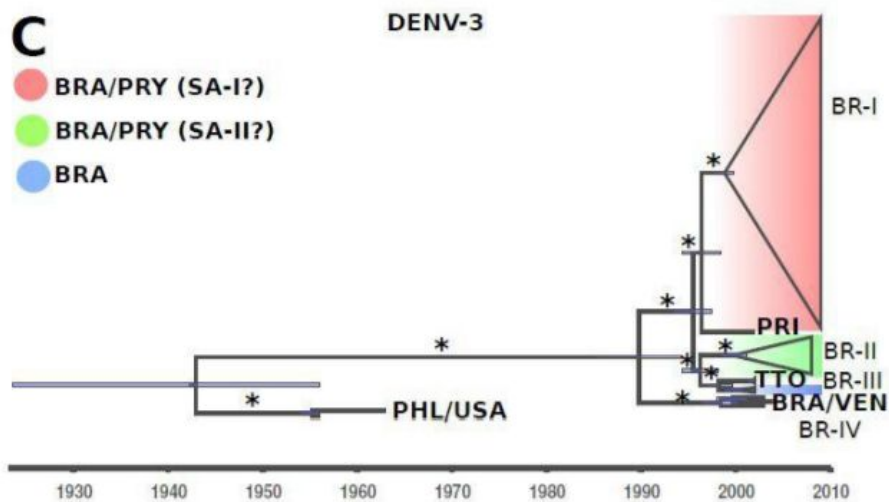
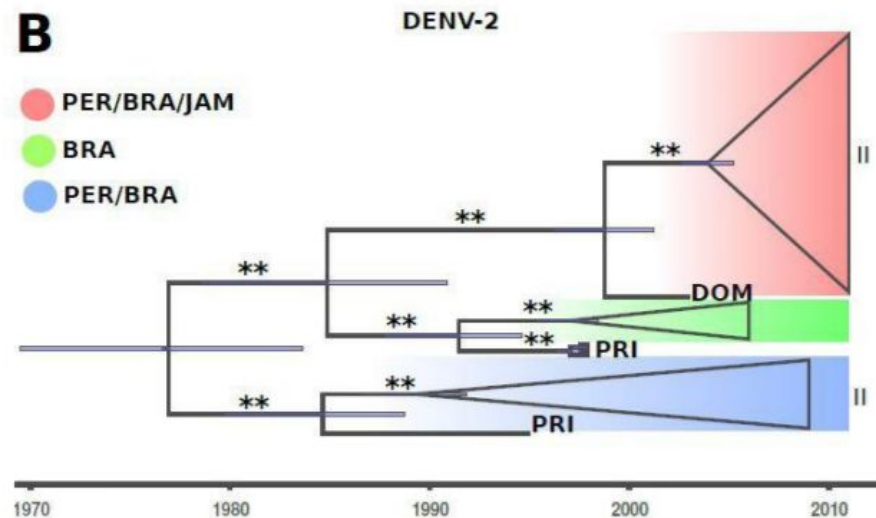
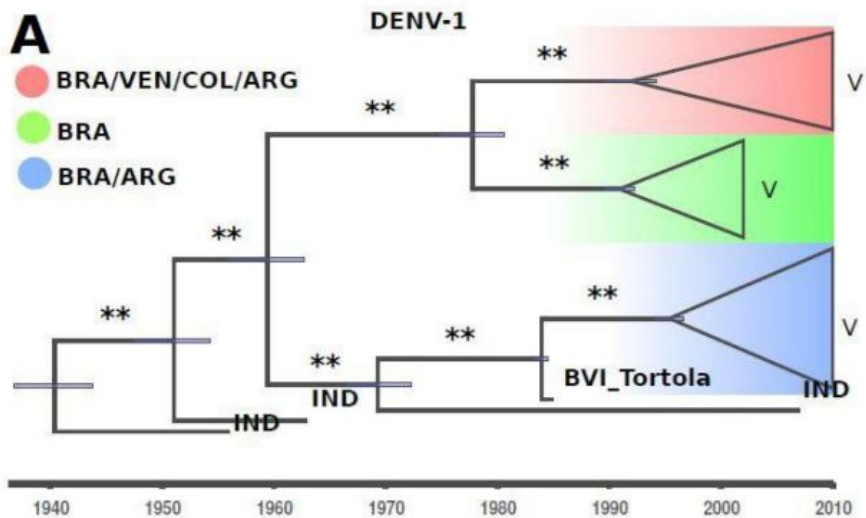


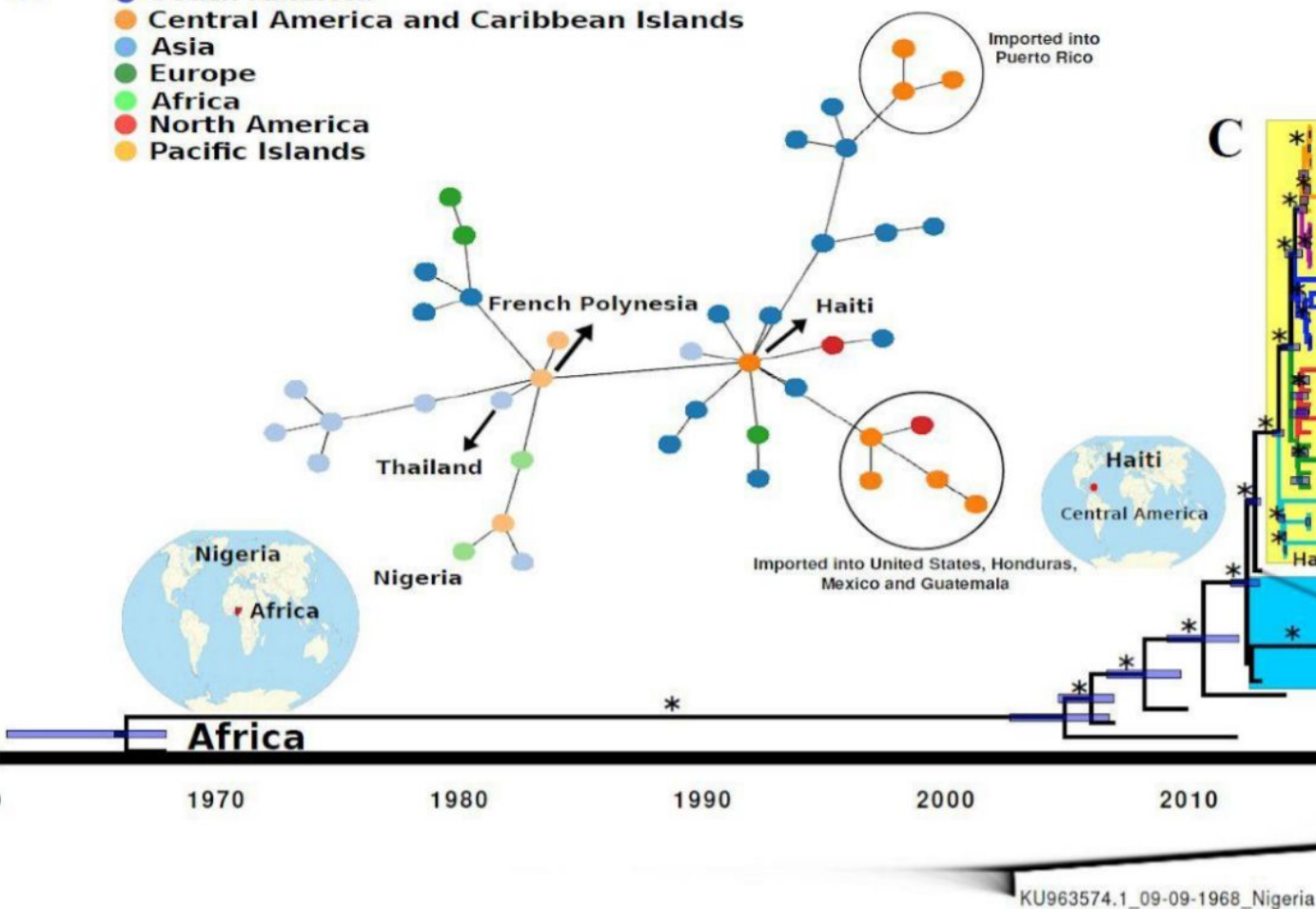
Figure 1: Phylogenetic reconstructions of CHIKV whole genome sequences. Bayesian MCC tree enforcing a strict molecular clock. The values along the tree branches represent the posterior distribution. The horizontal bars show phylogenetic uncertainty. The branches in blue and red represent Central/East African and Asian CHIKV genotypes, respectively.

Abbreviation: CAR = Central African Republic.



A

- South America
- Central America and Caribbean Islands
- Asia
- Europe
- Africa
- North America
- Pacific Islands



B

KU365780_2015_Brazil
 KU365777_2015_Brazil
 KU707826_2015_07_01_Brazil
 KU365779_2015_Brazil
 KX601168_2015_12_01_Puerto_Rico
 KX087101_2015_12_Puerto_Rico
 KU501215_2015_12_01_Puerto_Rico
 KX377337_2015_12_Puerto_Rico
 KU937936_2016_02_11_Suriname
 KU758877_2015_12_French_Guiana
 KU365778_2015_Brazil
 KX262887_2016_01_06_Honduras
 KU870645_2016_02_02_USA
 KU501217_2015_11_01_Guatemala
 KU501216_2015_12_01_Guatemala
 KX247632_2015_11_Mexico
 KU926309_2016_01_14_Brazil
 KU853013_2016_02_01_Italy
 KU853012_2016_02_01_Italy
 KU729217_2015_Brazil
 KU926310_2016_01_29_Brazil
 KU527068_2015_Brazil
 KX280026_2015_Brazil
 KU991811_2016_03_06_Italy
 KU497555_2015_11_30_Brazil
 KU729218_2015_Brazil
 KX520666_2015_08_Brazil
 KX051563_2016_02_05_USA
 KU744693_2016_02_06_China
 KX197192_2015_Brazil
 KU321639_2015_03_Brazil
 KU509998_2014_12_12_Haiti
 KX369547_2013_10_25_French_Polynesia
 KJ776791_2013_11_28_French_Polynesia
 KX117076_2016_02_17_China
 KU955589_2016_02_16_China
 KX253996_2016_02_16_China
 KU866423_2016_China
 KU820899_2016_02_17_China
 KU963796_2016_China
 KX185891_2016_02_17_China
 KU681081_2014_07_19_Thailand
 KU955593_2010_Cambodia
 EU545988_2007_06_Micronesia
 KU681082_2012_05_09_Philippines

C

