



FILOGENIA

Dilvan Moreira (Baseado em material do
prof. André Carvalho)

Leitura

- Introduction to Computational Genomics:
A Case Studies Approach
 - Capítulo 7

Tópicos

3

- SARS
- Origem e Evolução da Epidemia
- Análise Filogenética
- Construção de Árvores Filogenéticas
- Algoritmo Neighbor-Joining
- Estudo de Casos

Epidemia de SARS

4

- Síndrome Respiratória Aguda Grave
 - *Severe Acute Respiratory Syndrome*
- Doença respiratória grave que atingiu o mundo em 2003
 - Causada pelo SARS coronavirus (SARS-COV)
 - Termo corona vem da coroa que aparece quando o vírus é observado em um microscópio eletrônico
 - Não é a gripe aviária



Epidemia de SARS

5

□ Coronavirus

- Patógenos que causam uma grande variedade de doenças em animais
 - Qualquer organismo capaz de causar doença infecciosa
- Podem apresentar mutações frequentes e, assim, infectar outras espécies
- Outros coronavírus foram identificados como causadores de hepatite em ratos e gastroenterite em suínos
- É o vírus mais comum em patologia veterinária

Epidemia de SARS

6

□ Fevereiro de 2003

- Hospital francês de Hanói, Vietnam, chamou a OMS com um relatório de uma infecção semelhante a gripe
 - Altamente contagioso
- Especialista e doenças infecciosas da OMS, Dr Carlo Urbani, concluiu que era um novo e incomum patógeno

Epidemia de SARS

7

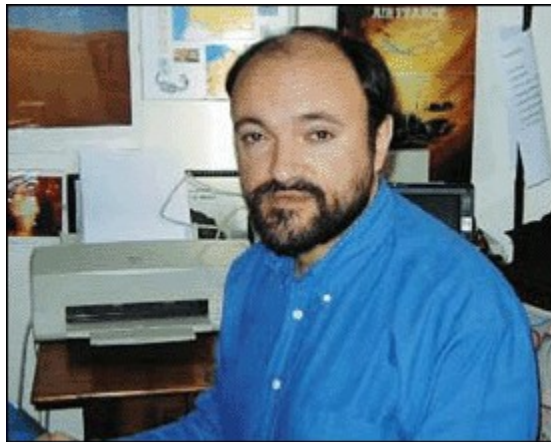
□ Fevereiro de 2003

- Durante sua estadia, Dr. Urbani coletou exames, examinou documentos do hospital e organizou quarentena de pacientes
 - Foi o primeiro a identificar e descrever a nova doença, SARS
 - Febre, tosse seca, falta de ar, piora progressiva do sistema respiratório, morte por falha do sistema respiratório
 - Em 3 semanas, Dr. Urbani mais cinco outros profissionais de saúde do hospital morreram de SARS

Epidemia de SARS

8

- Março de 2003
 - OMS emitiu um alerta global, chamando SARS de risco à saúde mundial



Epidemia de SARS

9



Hospital de Hanói
Março de 2003

Origem da Epidemia de SARS

10

- Primeiros casos ocorreram em 11/2002 na província de Guangdong, China
 - 106 pessoas adoeceram em hospital da cidade de Guangzhou
 - Resto do mundo não sabia disso
 - Médico desse hospital visitou Hong Kong em 21 de fevereiro de 2003
 - Hospedou-se no 9o andar do hotel metrópole
 - Adoeceu e morreu, com diagnóstico de pneumonia
 - Várias pessoas que estiveram no 9o andar tornaram-se transmissores da doença





Origem da Epidemia de SARS

13

- Um dos visitantes do 9o andar foi um executivo americano
 - Primeiro paciente tratado no hospital francês de Hanói
 - Infectou 80 pessoas antes de morrer
 - Outros visitantes do 9o andar levaram a doença para Canadá, Cingapura e EUA
 - Em abril de 2003, foram relatados 4300 casos com 250 mortes em 25 países

Origem da Epidemia de SARS

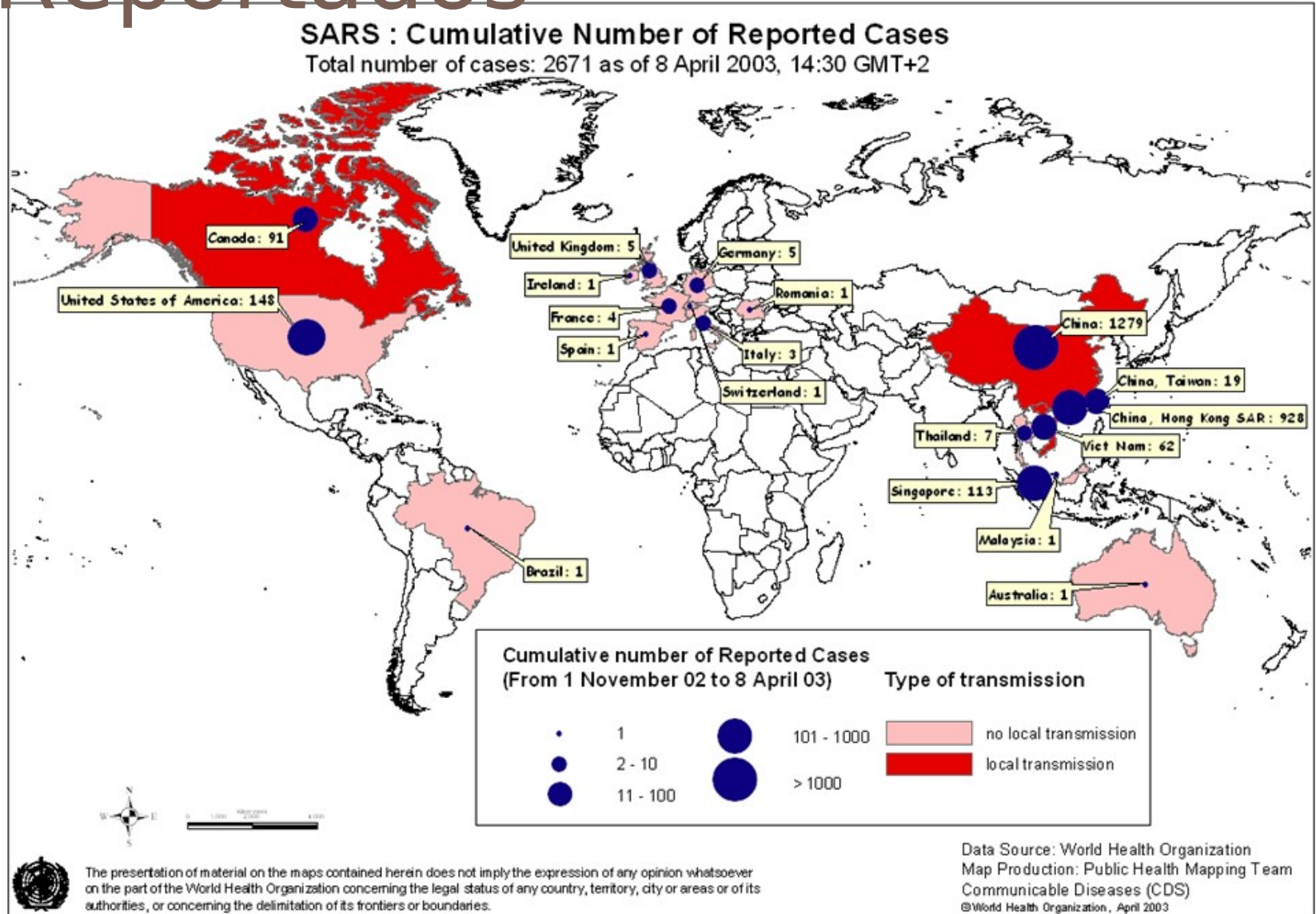
14

□ Março de 2003

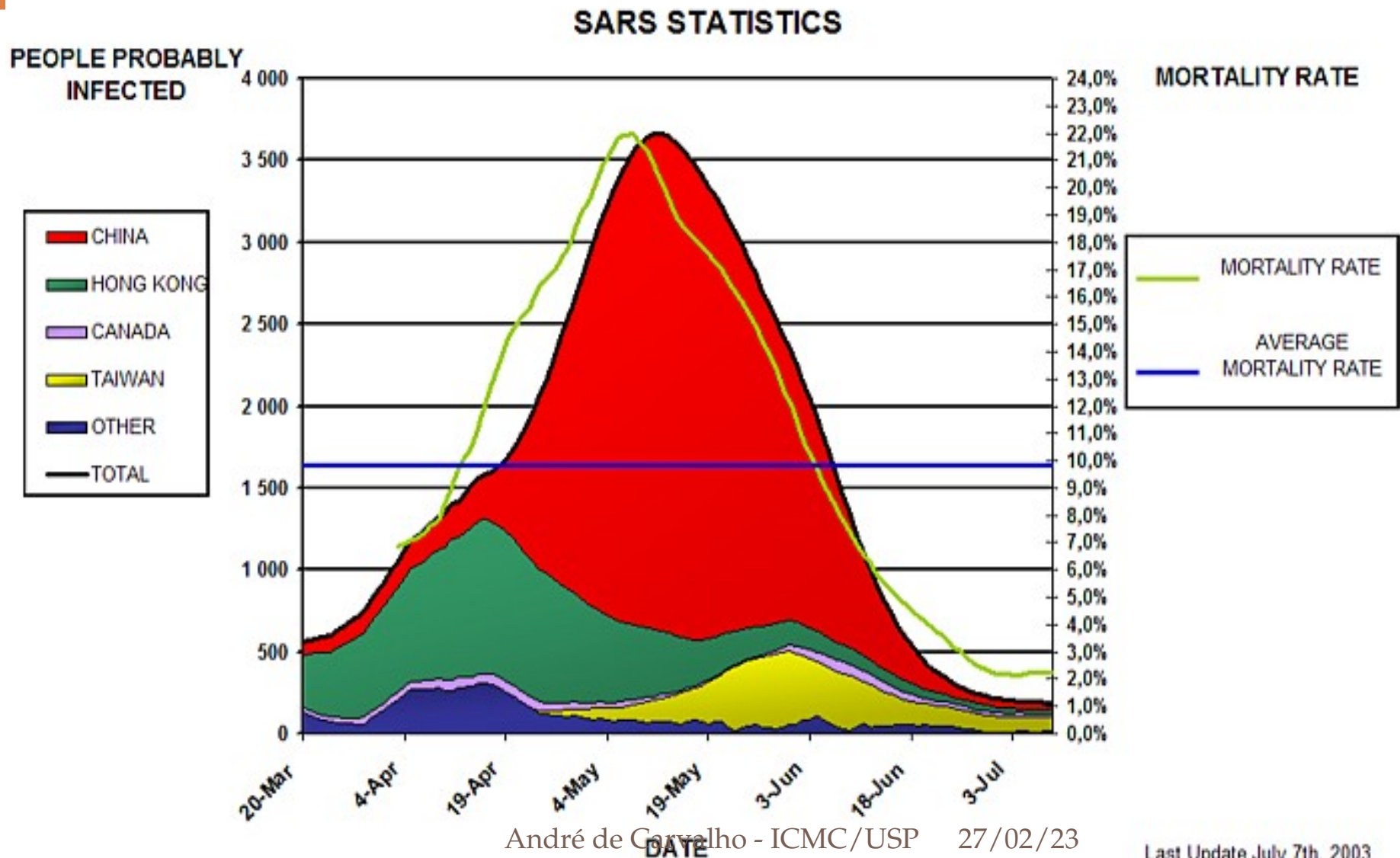
- Início do mês, OMS coordenou pesquisa internacional
- Final do mês, novo vírus que causa SARS foi identificado de forma independente em:
 - Alemanha, Canadá, EUA e Hong Kong
- SARS-CoV é um RNA viral (como HIV)
- Comum em humanos e animais, coronavírus causam ~25% de todas as infecções respiratórias superiores
 - Ex.: gripe comum

SARS Número de Casos Reportados

15

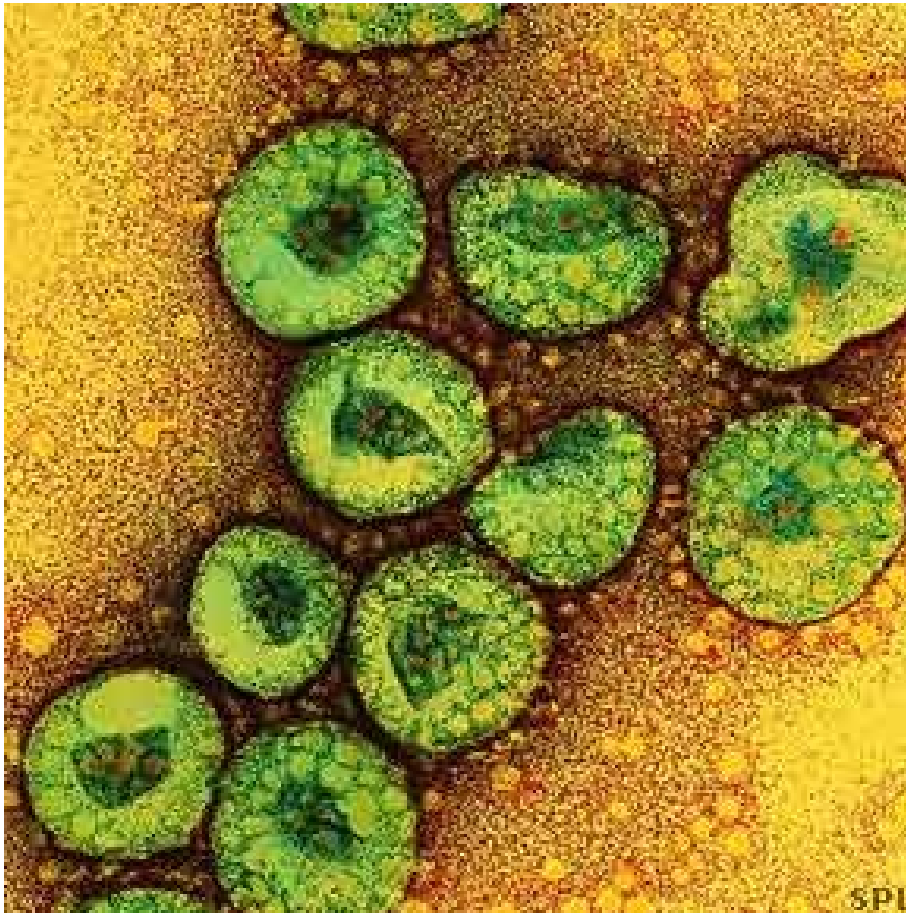


SARS Estatísticas



Coronavirus SARS

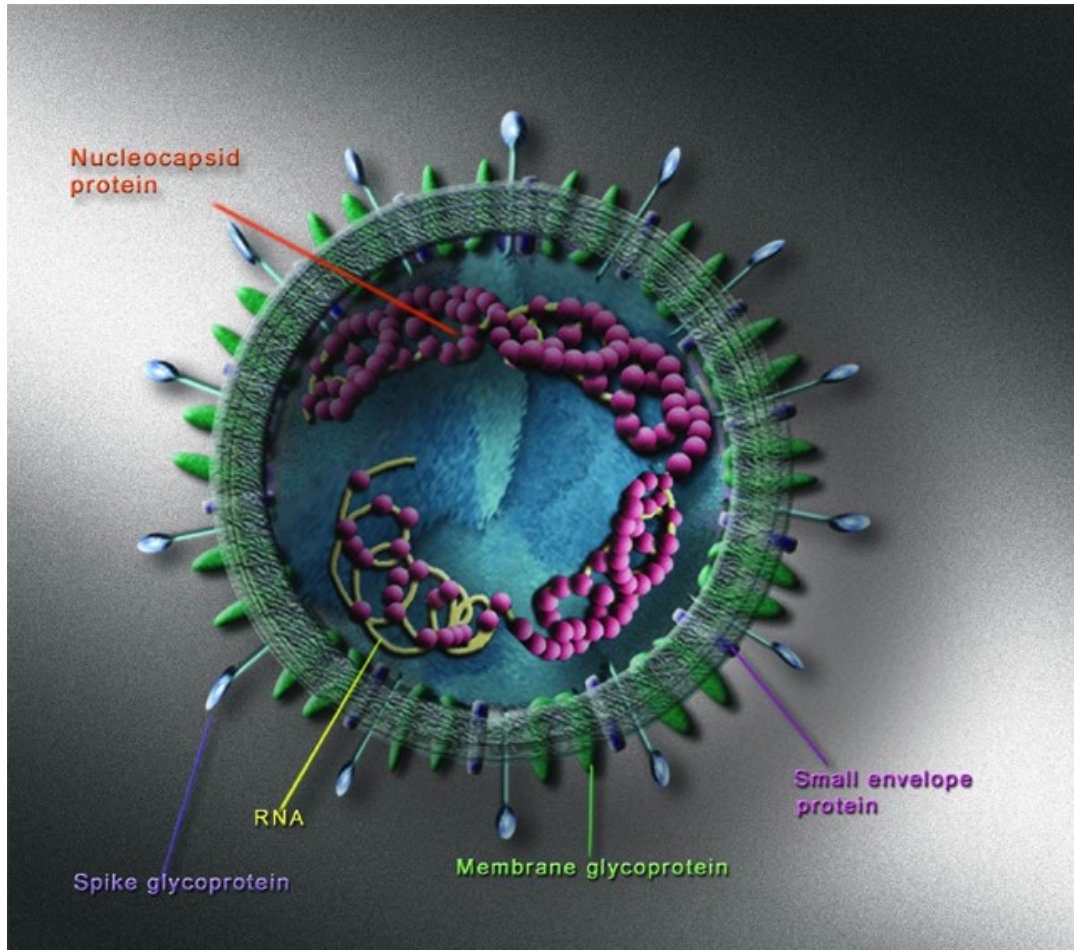
17



Fonte: BBC

Coronavirus SARS

18



www.cell-research.com

27/02/23

Origem da Epidemia de SARS

19

□ Abril de 2003

- Laboratório canadense seqüência a seqüência de RNA do vírus SARS CoV
- Análise filogenética do vírus mostrou que o coronavírus mais próximo é o da civeta
- Alimento popular em Guangdong



Origem da Epidemia de SARS

20

□ Maio de 2003

- Dois artigos da *Science* apresentaram o genoma completo do SARS CoV
 - Genoma contém 29.751 bp
 - É substancialmente diferente de todos os CoVs humanos
 - Diferentes também dos CoVs dos pássaros – nenhuma relação com gripe aviária

□ Final de 2003

- SARS se espalhou por todo o mundo

Análise Filogenética do SARS

21

- Análise filogenética pode responder questões como:
 - Que tipo de vírus causou a infecção original?
 - Qual a fonte da infecção?
 - Quando e onde o vírus cruzou fronteiras entre espécies?
 - Quais são as mutações chave que possibilitaram esse cruzamento?
 - Qual a trajetória seguida para o espalhamento do vírus?

Análise Filogenética do SARS

22

- Para responder às perguntas anteriores, vamos:
 - Examinar alguns algoritmos chave de análise filogenética
 - Aplicar esses algoritmos a dados de SARS
 - Disponíveis no GenBank e no site do livro

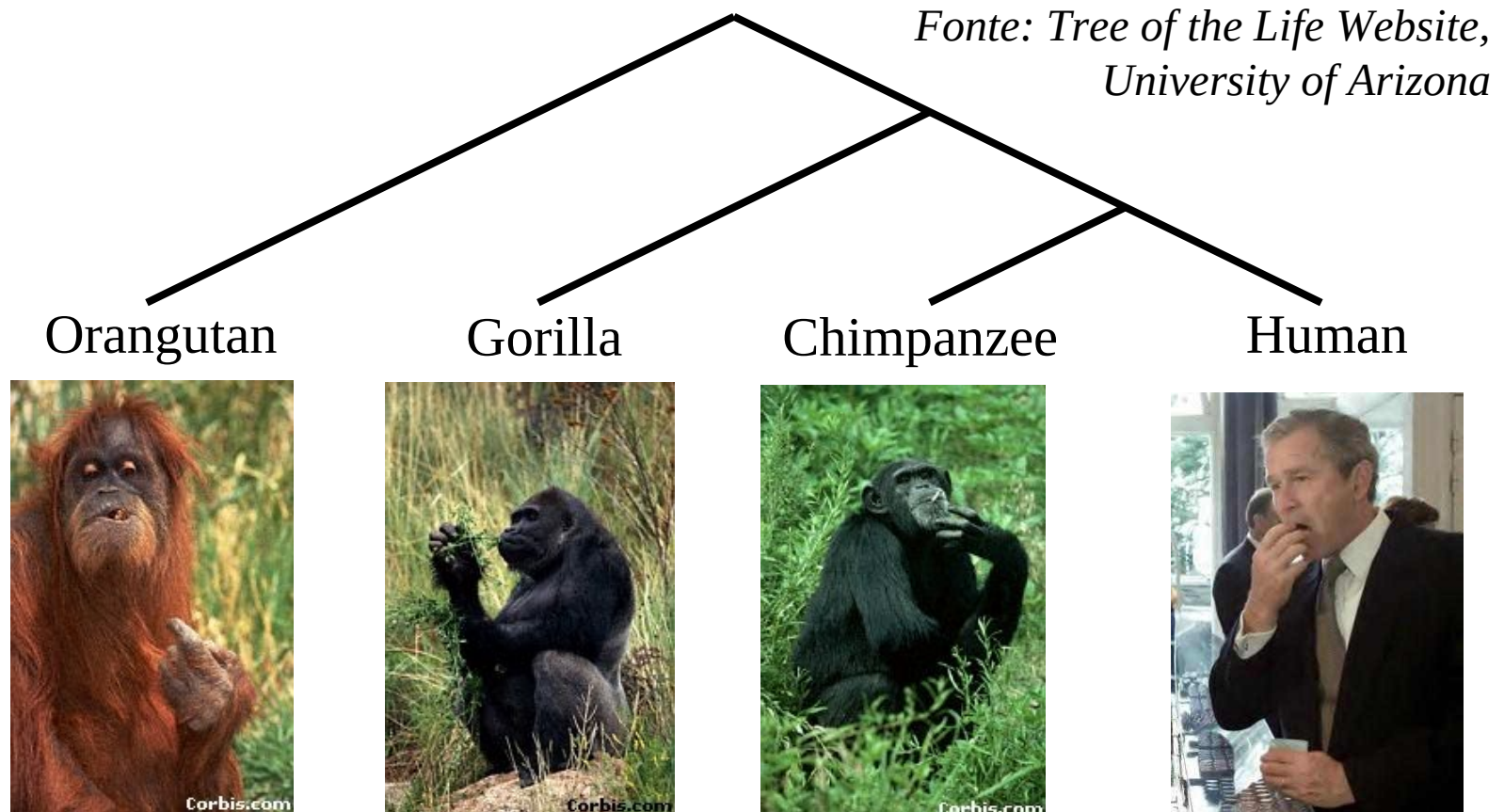
Árvores e Evolução

23

- Trajetória de avanço do SARS pode ser representada por uma árvore
 - Todos os vírus de SARS que apareceram no mundo se originaram do vírus achado na China
 - Novos ramos aparecem quando o vírus se espalha
- Tradicionalmente, história evolutiva ligando grupos de espécies tem sido representada por uma árvore
 - Única figura no livro “ On the origin of species” de Darwin

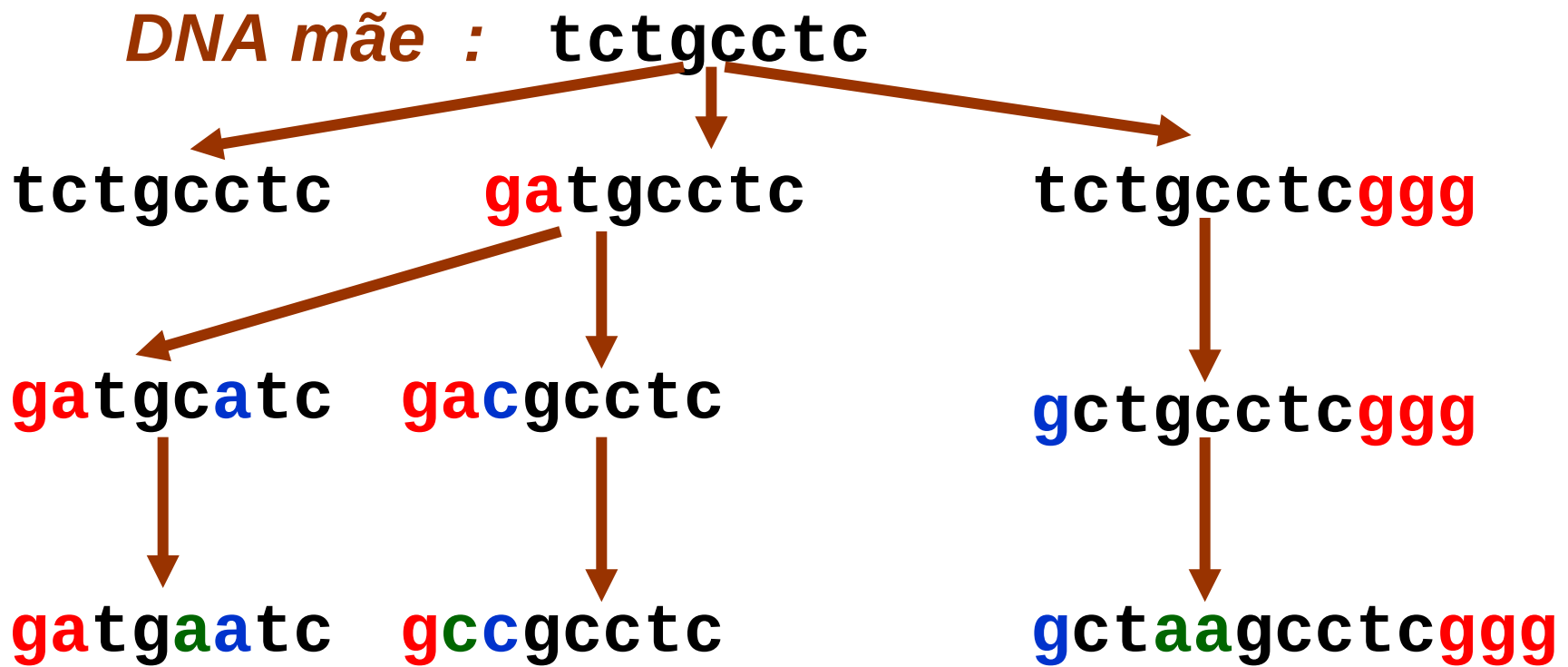
Árvores Filogenéticas

24



Árvores Filogenéticas

25



Espécies atuais

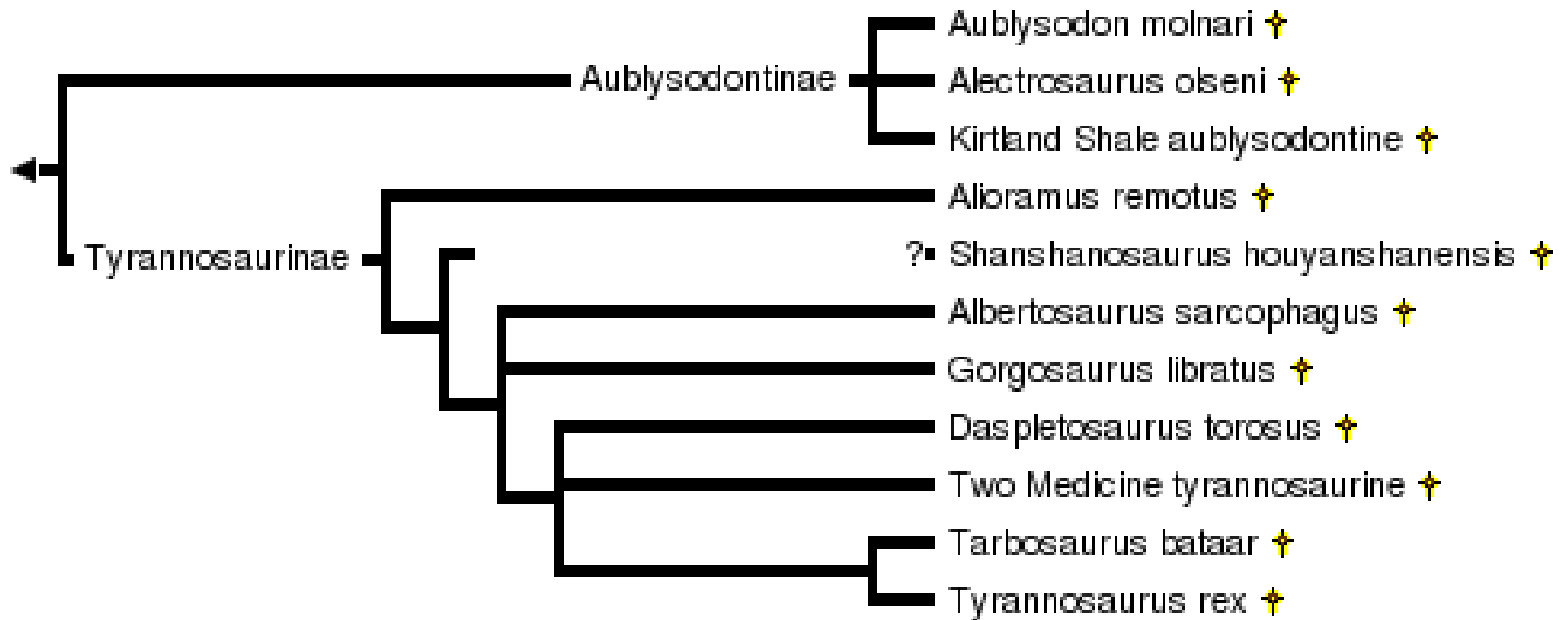
Filogenia

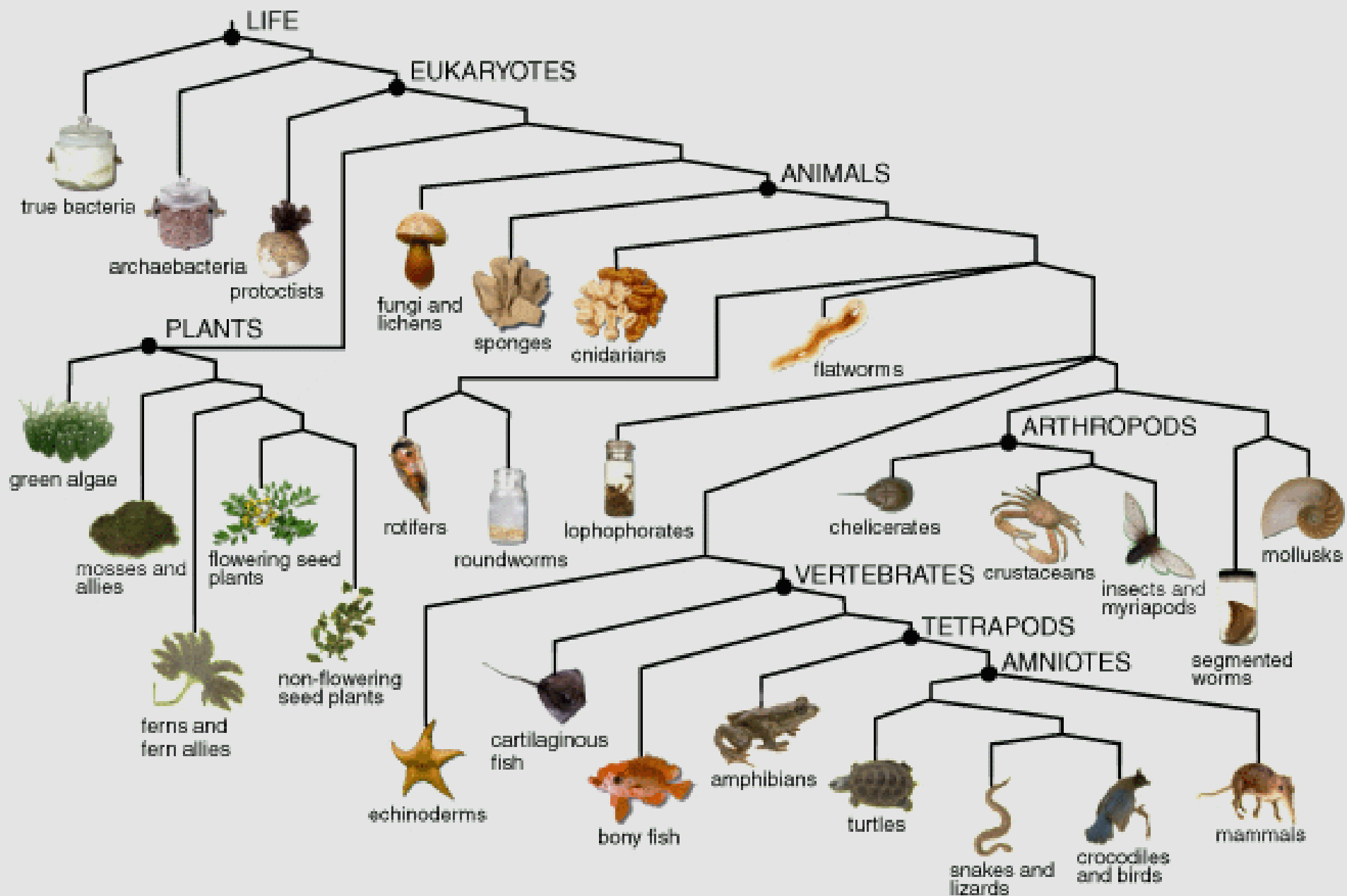
26

- Estudo da relação evolutiva entre vários grupos de organismos
 - Espécies
 - Populações
 - Etc.
- Representada por diagrama em forma de árvore (árvore filogenética)
 - Análise cladística
 - Geralmente baseada em dados morfológicos

Cladísticas

27





Árvores Filogenéticas

29

- Mostram os relacionamentos evolutivos entre várias espécies ou indivíduos
 - Que se acredita terem um ancestral em comum
- Forma de cardiograma
 - Cada nó com descendentes representa o ancestral comum mais recente deles
 - Tamanho das arestas corresponde a estimativas de tempo

Árvores Filogenéticas

30

- Cada nó é denominado uma unidade taxonômica (taxon, plural taxa)
- Nós internos são unidades taxonômicas hipotéticas
 - Não podem ser diretamente observados
- Relacionamentos mais complexos podem assumir a forma de redes

Estruturas de Árvores Filogenéticas

31

- Árvores apresentam duas ou mais taxa
 - Espécie ou indivíduo
- Nós externos representam taxa existentes
 - Nós internos apresentam seus ancestrais (em geral extintos)
- Árvores podem ser de:
 - Bifurcação
 - Cada nó interno tem no máximo 2 filhos
 - Multifurcação
 - Cada nó interno pode ter mais que 2 filhos
- Árvores podem ter ou não raiz

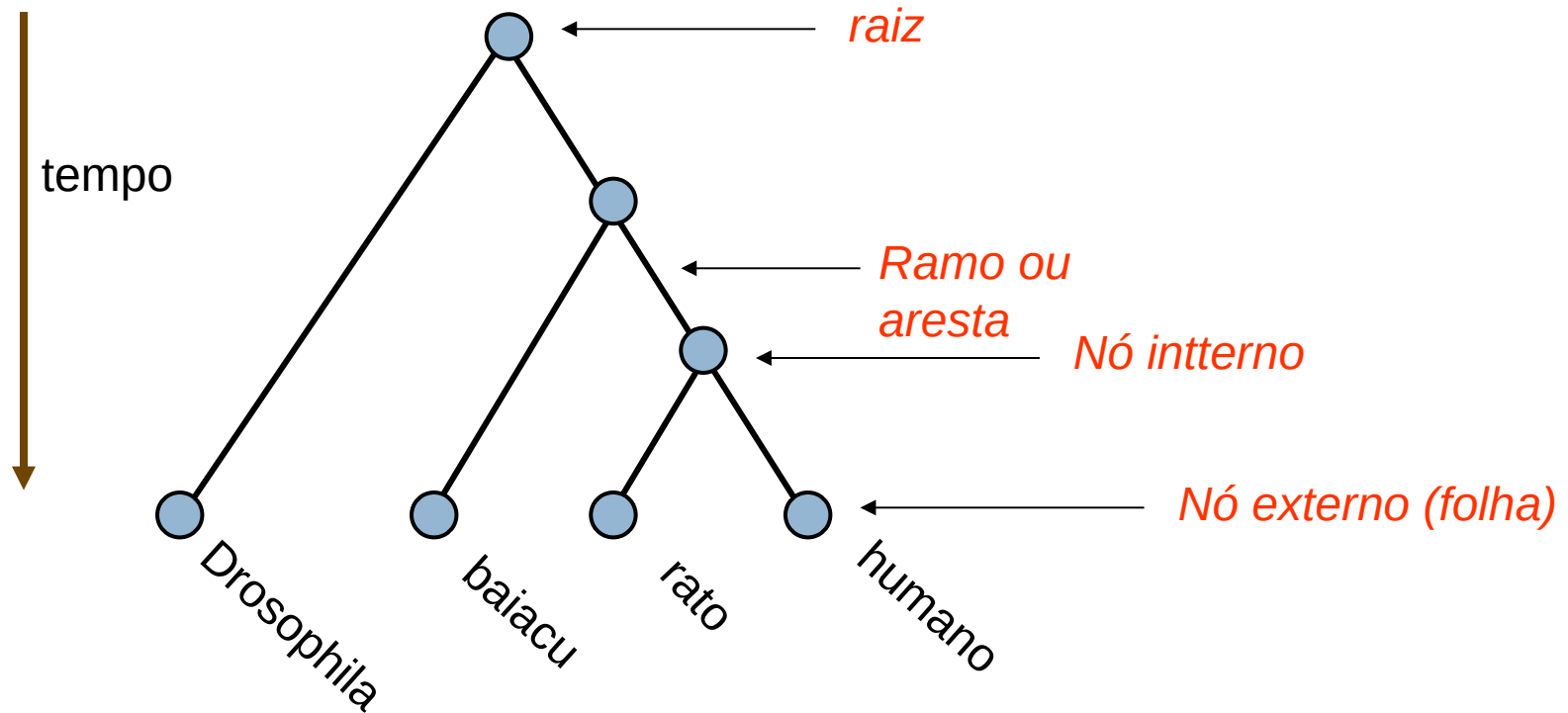
Árvores Filogenéticas com Raiz

32

- Define-se um nó interno especial, chamado raiz
 - Ancestral comum a todos os outros nós
 - Todos os caminhos evolutivos levam à raiz
 - Ramos são orientados da raiz aos nós externos

Árvores Filogenéticas com Raiz

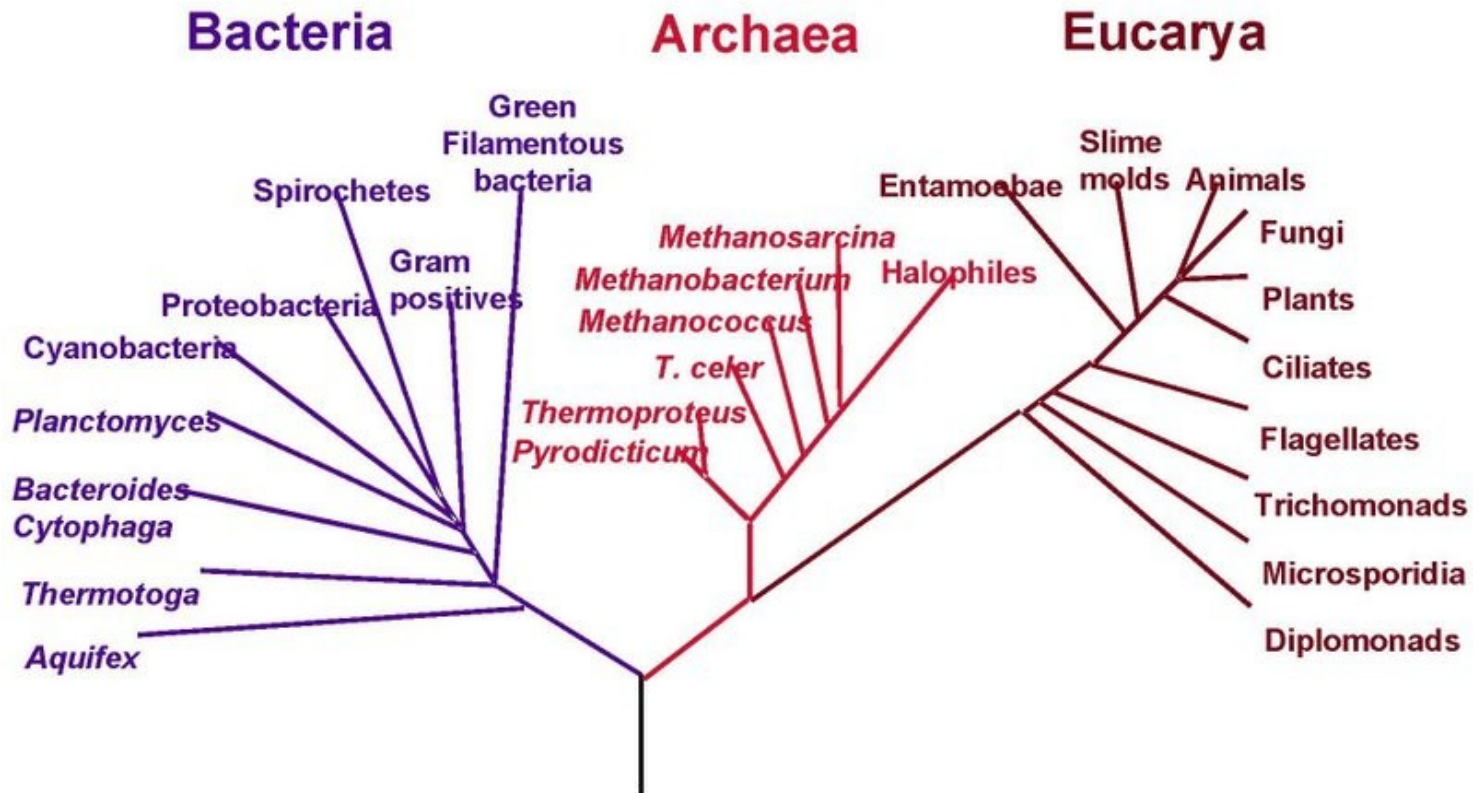
33



Árvores Filogenéticas com Raiz

34

Phylogenetic Tree of Life



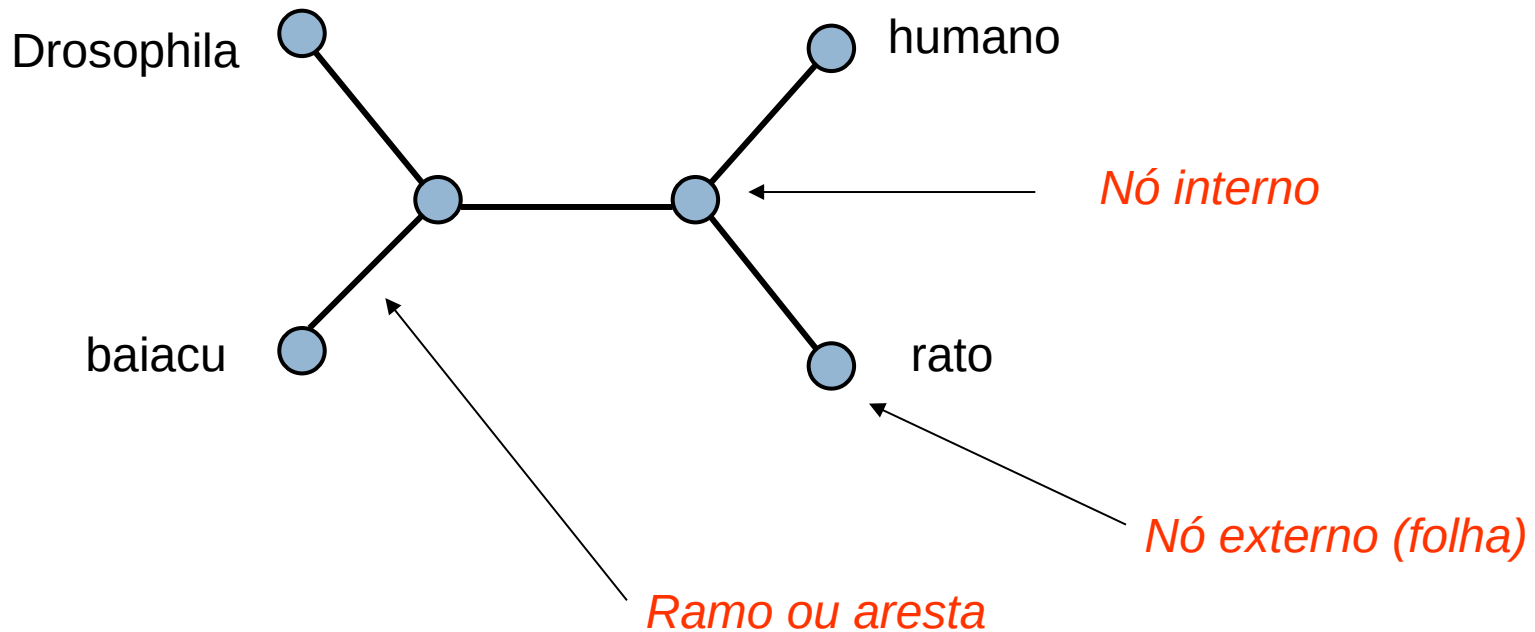
Árvores Filogenéticas sem Raiz

35

- Ramos não possuem orientação
 - Mostram relacionamento topológico entre taxa, sem identificar um ancestral comum
- Existem métodos para definir uma raiz para uma árvore sem raiz
 - Escolhem uma aresta para colocar nó raiz
 - Requer informação biológica externa ou pelo menos suposição de onde por a raiz

Árvores Filogenéticas sem Raiz

36



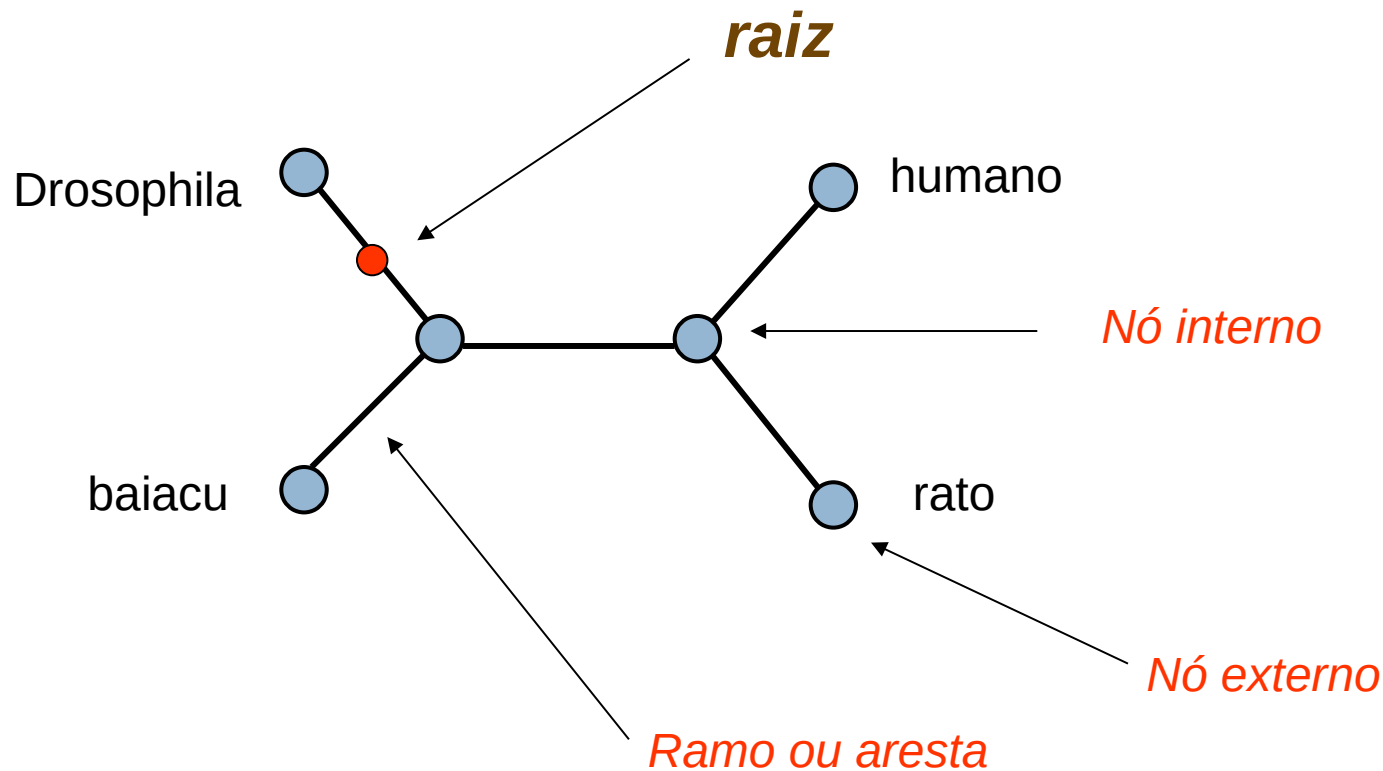
Árvores Filogenéticas sem Raiz

37

- Raiz é geralmente definida incluindo uma ou mais taxa no conjunto de dados
 - Que se sabe serem resultados de divisão mais antiga
 - Relação mais distante a cada uma das taxa
 - Essa taxon (ou taxa) externa é chamada de *outgroup*
 - Ramo da árvore onde o *outgroup* se junta as demais taxa é considerado conter o nó raiz

Árvores Filogenéticas sem Raiz

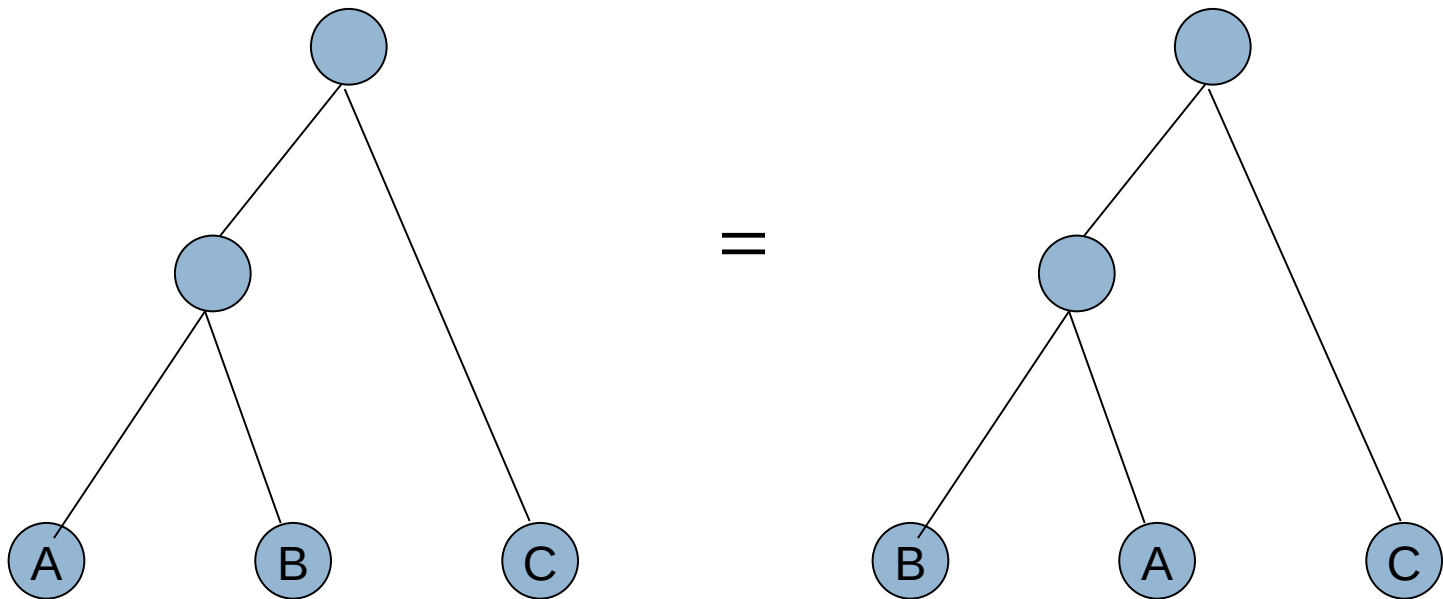
38



Estruturas de Árvores Filogenéticas

39

- Rotação dos ramos de um nó interno não altera relacionamento entre taxa



Invariante à rotação

Número de Possíveis Árvores

40

- Reconstrução de árvores filogenéticas a partir de sequências de DNA
 - Complicado pelo grande número de possíveis árvores

- Possíveis árvores sem raiz ($n \geq 3$)

$$\frac{(2n - 5)!}{2^{n-3}(n - 3)!}$$

- Possíveis árvores com raiz ($n \geq 2$)

$$\frac{(2n - 3)!}{2^{n-2}(n - 2)!}$$

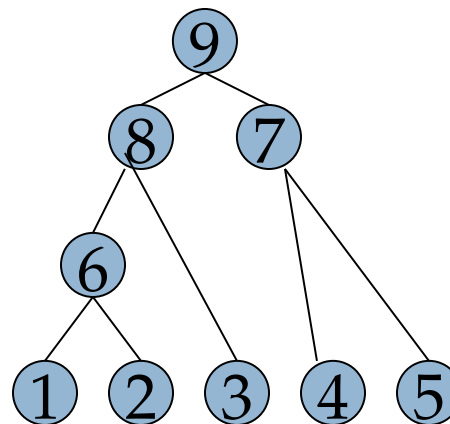
n: número de taxa

Representação de Árvores

41

- Existem várias formas não gráficas de representar uma árvore

9	8	7
8	3	6
7	4	5
8	1	2
5	-	-
4	-	-
3	-	-
2	-	-
1	-	-



Formato padrão popular:

Newick: (((1,2),3),((4,5),(6,7)))

Inferindo Árvores

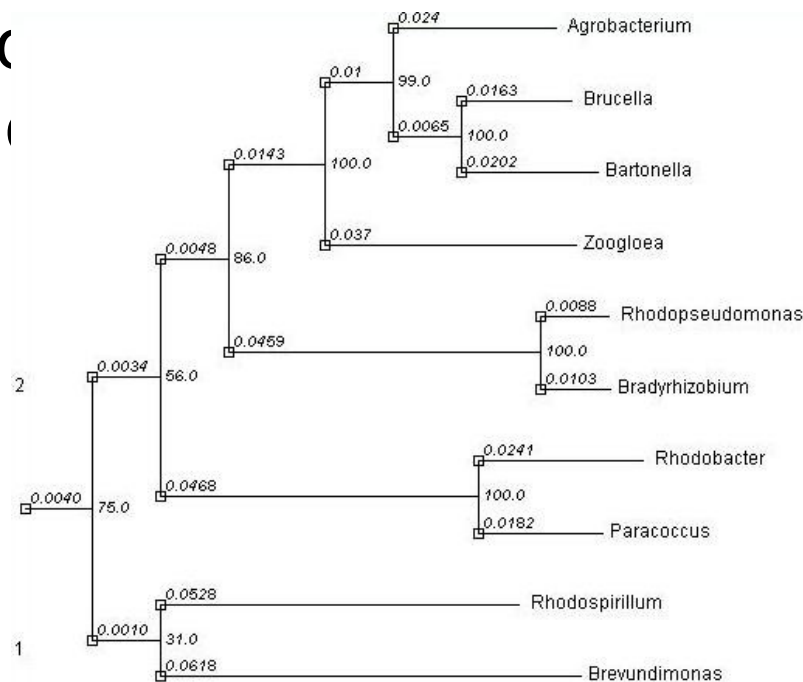
42

- Até pouco tempo, relacionamento entre taxa era inferidos por características morfológicas
- Atualmente são utilizadas sequências de DNA
 - Tecnologia de sequenciamento
 - Mutações deixam uma trilha
 - Árvores podem ser inferidas a partir da similaridade entre sequências homólogas

Inferindo Árvores

43

- Ramos da árvore podem ter tamanhos diferentes
 - Quanto maior o tamanho, maior a incerteza nas inferências,



Inferindo Árvores

44

- Dadas seqüências homólogas de um grupo de taxa
 - Existem vários métodos de reconstruir seus relacionamentos filogenéticos
- Métodos podem ser divididos em dois grupos:
 - Os que ordenam todas as possíveis árvore por meio de algum critério para encontrar a melhor
 - Os que constroem a árvore diretamente a partir dos dados (sem definir uma função de avaliação)

Inferindo Árvores

45

- Ordenando as possíveis árvores
 - Critérios em geral procuram a árvore com o menor número de mutações
 - Por causa do enorme número de possíveis árvores:
 - Pode demorar muito para encontrar a melhor árvore
 - Quando usa aproximações para acelerar busca, pode não encontrar melhor árvore

Inferindo Árvores

46

- Métodos que constroem a partir dos dados
 - Árvore é construída por métodos filogenéticos que usam algoritmos e estatística
 - Frequentemente baseados na computação da distância entre pares de taxa
 - Muito populares por serem geralmente rápidos

Inferindo Árvores

47

- Métodos que constroem a partir dos dados
 - Método baseado em distância mais popular é o algoritmo neighbor-joining (NJ)
 - Apesar de não ser necessariamente tão bem comportado estatisticamente quanto outros métodos
 - Robusto e acurado
 - Garantido de inferir a árvore verdadeira se as distâncias usadas refletem a distância real entre seqüências
 - Resultado não garantido por outros métodos estatisticamente mais sofisticados

Métodos Baseados em Distância

48

- Dadas n taxa
- Construir matriz de distâncias entre taxa
 - Se os ramos da árvore têm um tamanho específico, distância entre dois nós quaisquer pode ser facilmente computada
 - Tamanho total do caminho único unindo eles
 - Permite especificar distância entre nós folha da árvore usando distância aditiva

Métodos Baseados em Distância

49

□ Distância aditiva

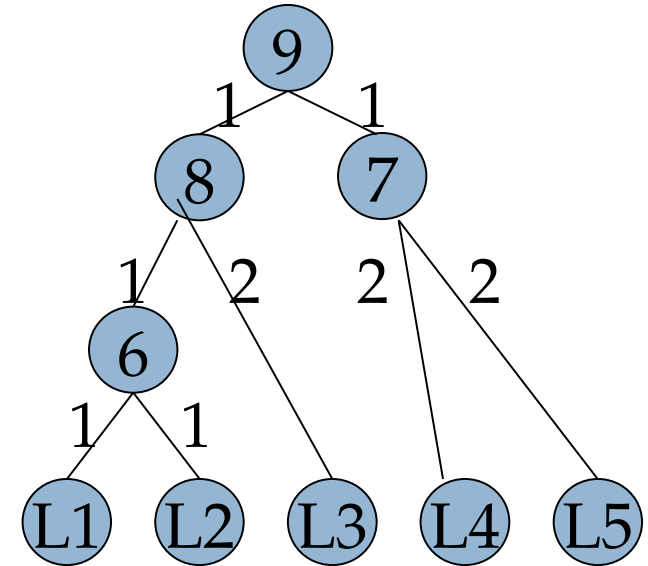
- Biologicamente, aditividade é uma propriedade importante para uma matriz de distâncias
- Número de substituições separando duas taxa de seu último ancestral comum forma uma distância aditiva
 - Distância sobre o caminho do nó i ao nó j
 - Modelo de Jukes-Cantor é freqüentemente usado como modelo de substituição

Métodos Baseados em Distância

50

Matriz de Distâncias

	L1	L2	L3	L4	L5
L1	0	2	4	6	6
L2	2	0	4	6	6
L3	4	4	0	6	6
L4	6	6	6	0	4
L5	6	6	6	4	0



Algoritmo Neighbor-Joining

51

- Desenvolvido em 1987 por Naruya Saitou e Masatoshi Nei
 - Mesmo Nei do método Nei-Gojobori
- Algoritmo guloso (*greedy*)
 - Começa com uma filogenia em forma de estrela
 - Todas as taxa estão conectadas diretamente a um único nó raiz
 - Iterativamente combina pares de nós

Algoritmo Neighbor-Joining

52

- Chave para sucesso do algoritmo:
 - Critério que define como são selecionados os nós a serem combinados a cada iteração
 - Identifica nós que são topologicamente vizinhos na árvore
- Taxa selecionadas são combinadas em um táxon
 - Uma nova matriz de distâncias é então calculada
- Processo repetido até que todas as taxa sejam combinadas
- Árvore gerada é sem raiz

Algoritmo Neighbor-Joining

53

- Cálculo do tamanho dos ramos
 - Supor 3 taxa em uma árvore sem raiz

Fórmula 3-pontos:

$$L_x + L_y = d_{AB}$$

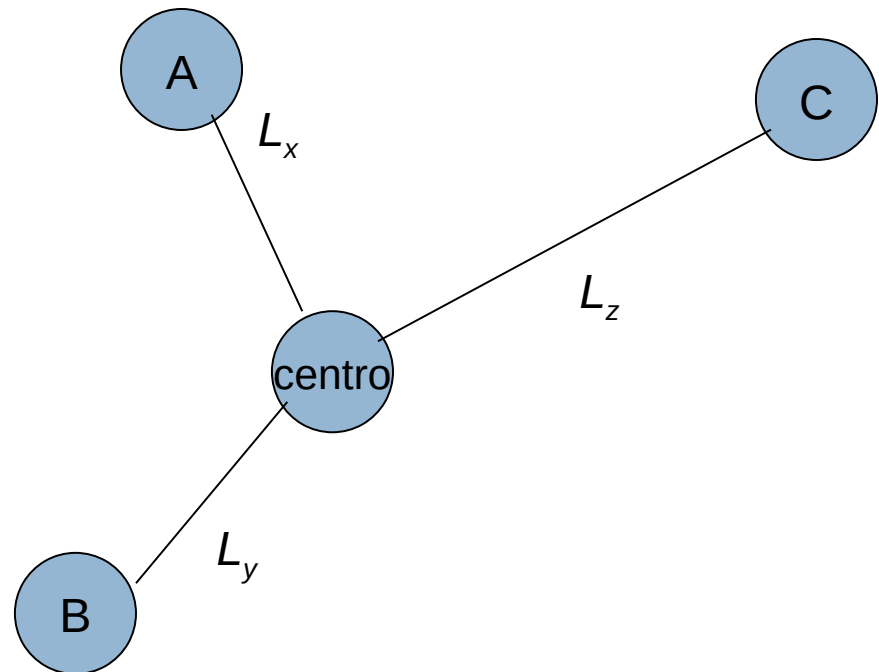
$$L_x + L_z = d_{AC}$$

$$L_y + L_z = d_{BC}$$

$$L_x = (d_{AB} + d_{AC} - d_{BC})/2$$

$$L_y = (d_{AB} + d_{BC} - d_{AC})/2$$

$$L_z = (d_{AC} + d_{BC} - d_{AB})/2$$



Algoritmo Neighbor-Joining

54

- Usa condição de 4-pontos para selecionar nós vizinhos a serem combinados
 - Supor que 1 e 2 são vizinhos

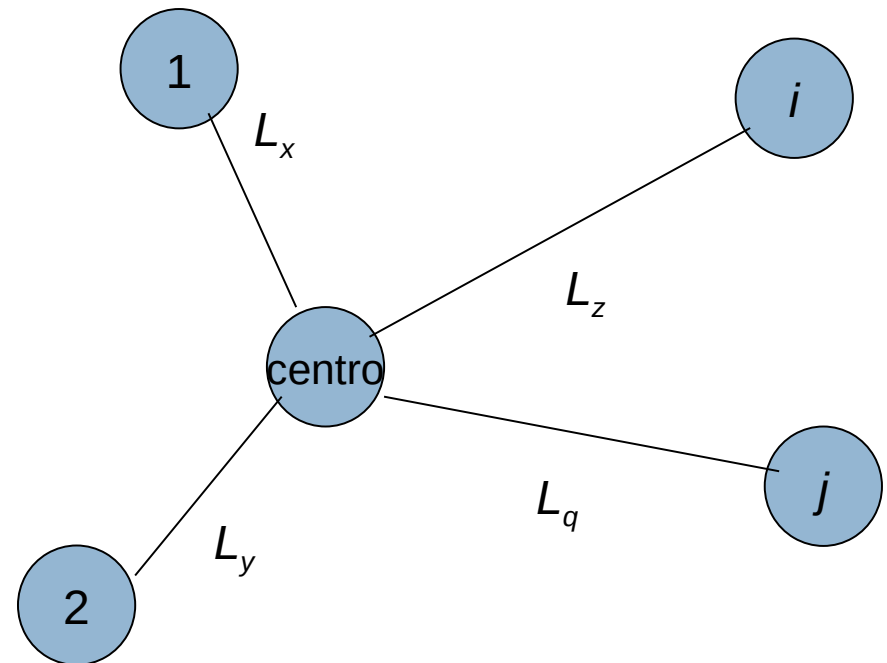
$$d(1,2) + d(i,j) < d(i,1) + d(2,j)$$

$$R_i = \sum_j d(i,j)$$

$$M(i,j) = (n-2)d(i,j) - R_i - R_j$$

$$M(i,j) < M(i,k) \text{ para todo } k \neq j$$

Grau de vizinhancidade



Algoritmo Neighbor-Joining

55

Entrada: matriz $n \times n$ de distancias D e um outgroup

Output: árvore filogenética T com raiz

- 1:** Atualizar tabela M usando D e escolher menor valor de M para selecionar duas taxa para combinar
- 2:** Combinar duas taxa t_i e t_j em um novo nó V e usar fórmula de 3-pontos para atualizar a matriz de distâncias D' onde t_i e t_j são substituídos por V .
- 3:** Calcular tamanho dos ramos de t_k a V usando a fórmula de 3-pontos, $T(V,1) = t_i$ e $T(V,2) = t_j$, $TD(t_i) = L(t_i, V)$ e $TD(t_j) = L(t_j, V)$
- 4:** Matriz de distâncias D' agora tem $n - 1$ taxa.
se existe mais 2 taxa ir para 1:
se existem 2 taxa, combina-las por um ramo de tamanho $d(t_i, t_j)$.
- 5:** Definir o nó raiz como o ramo ligando o outgroup ao resto da árvore

UPGMA

56

- Algoritmo NJ pode ser reduzido a um método mais simples, UPGMA, quando $M = D$
 - Unweighted Pair Group Method with Arithmetic Averages
 - Um dos primeiros métodos baseados em distância
 - Distância do nó externo para a raiz é a mesma para todos os nós externos
 - Ultrametricidade
 - Quase nunca é válida para sequências de DNA
 - Pode levar a inferência da árvore incorreta

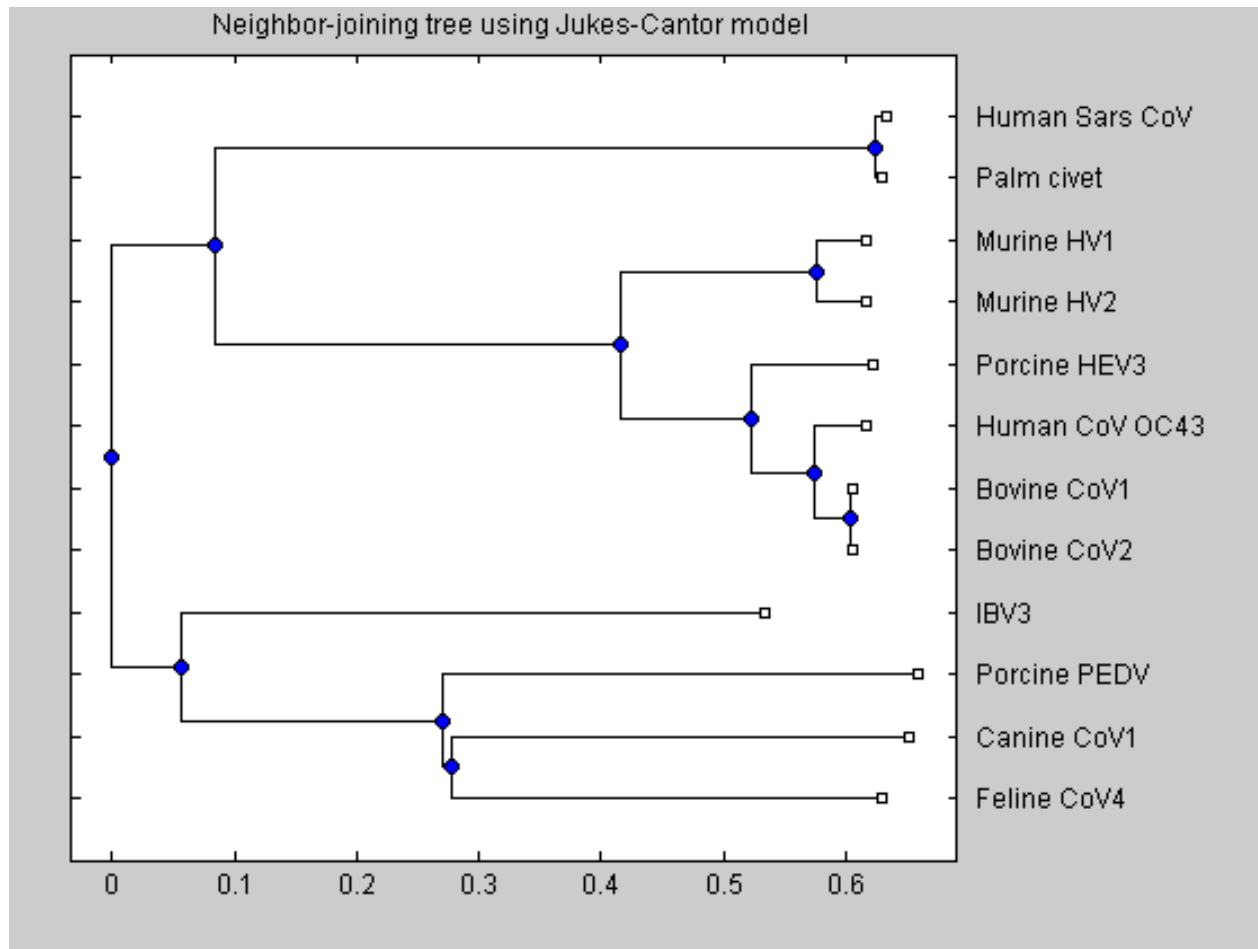
Estudo de Casos

57

- Análise filogenética da epidemia de SARS
 - Genoma do SARS-CoV tem 6 genes
 - Hospedeiro: civeta
 - Árvore epidemiológica
 - Data de origem
 - Área de origem

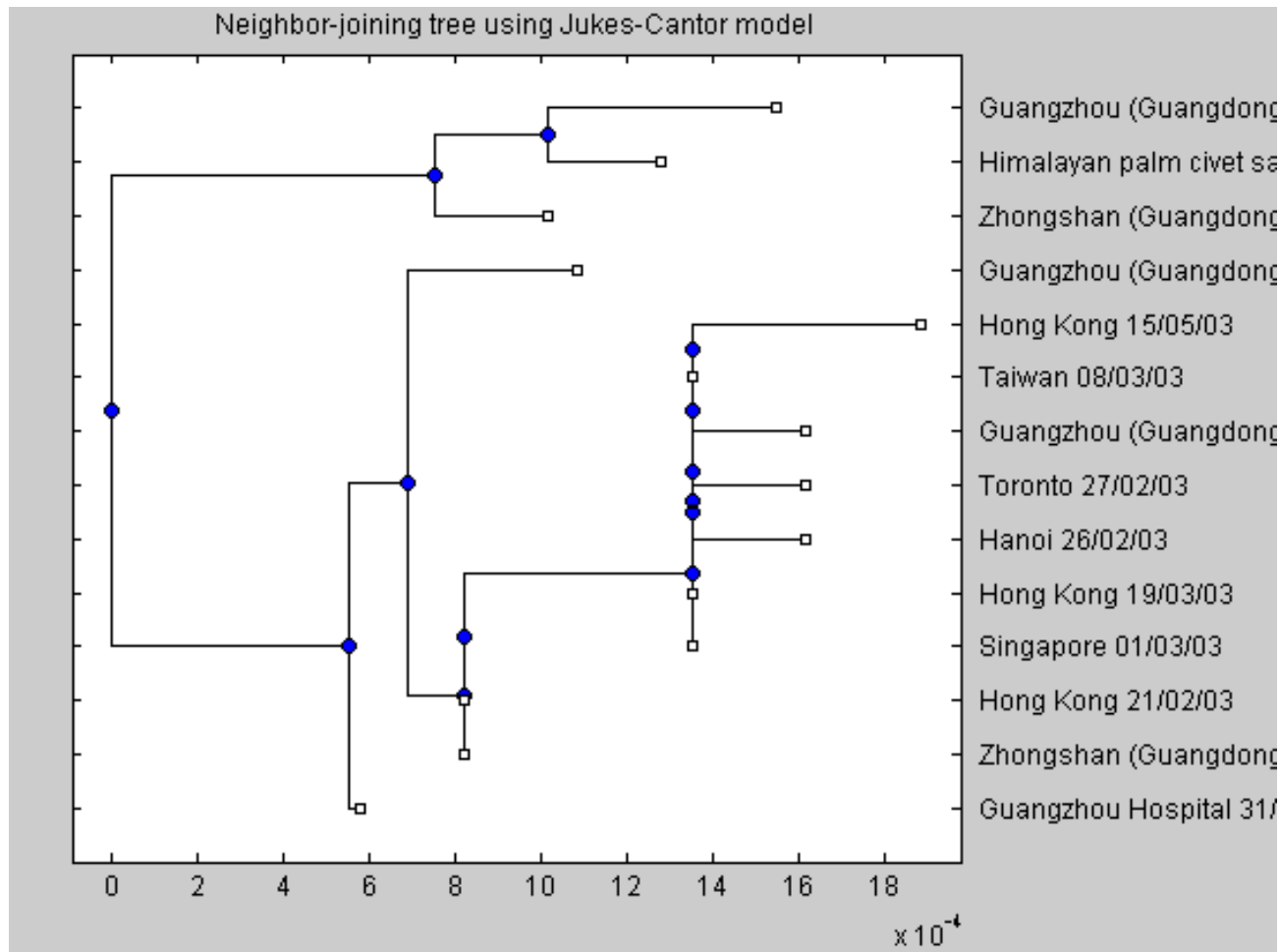
Identificação do Hospedeiro

58



Árvore Epidemiológica

59

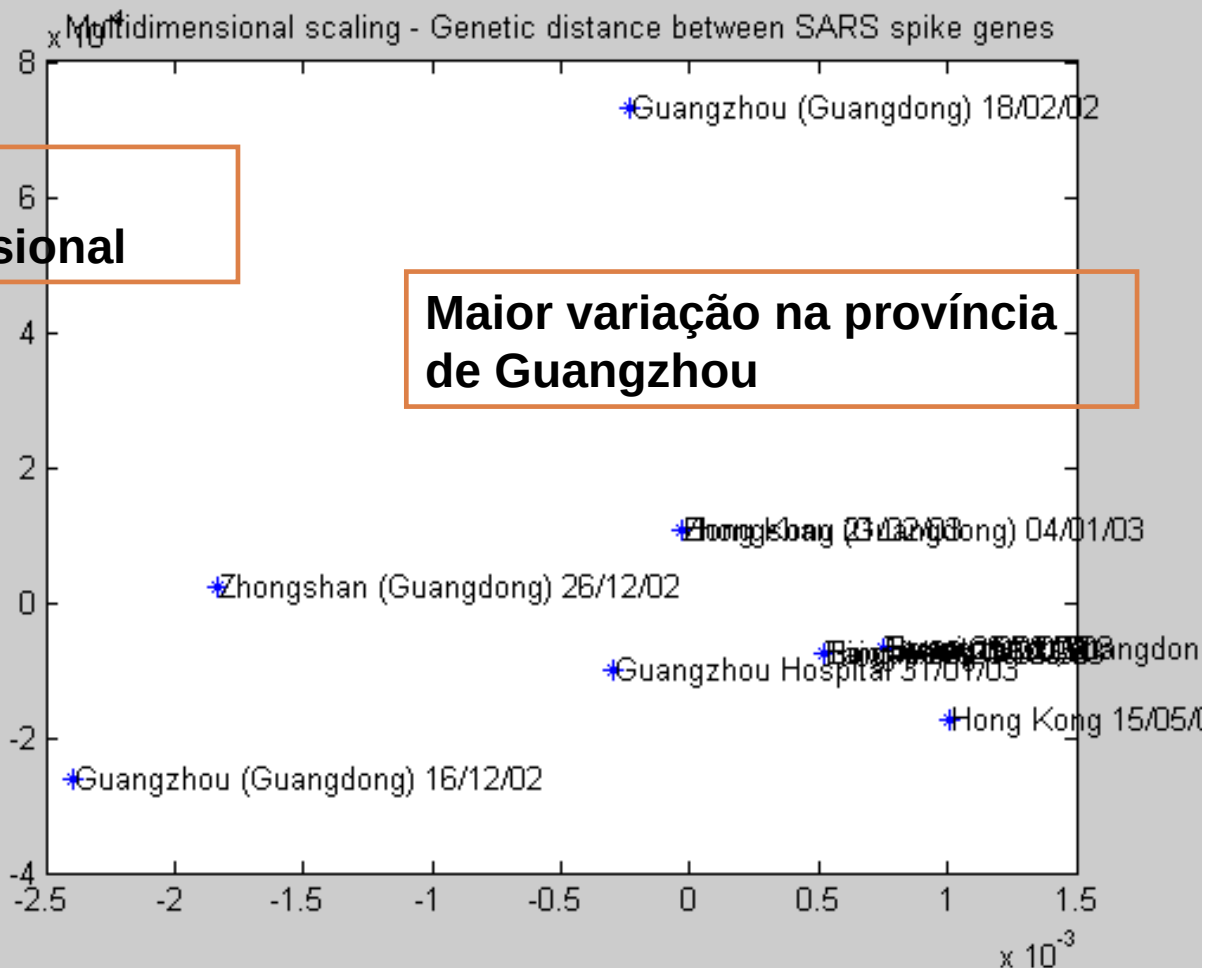


Área de Origem

60

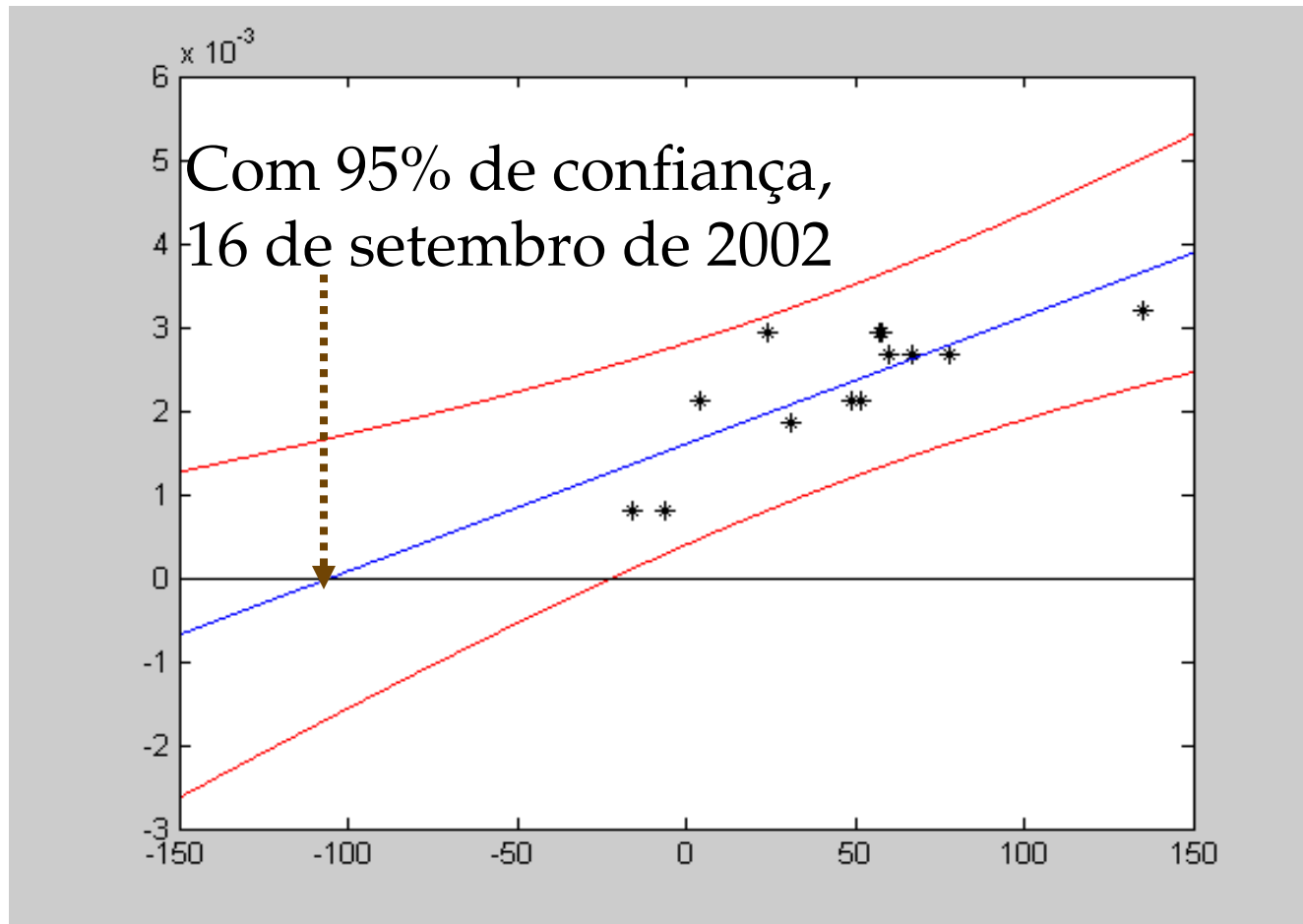
**Escala
multidimensional**

**Maior variação na província
de Guangzhou**



Data de Origem

61



Formato Newick

62

Formato de matriz:

1	2	3
2	A	B
3	C	D

Formato Newick:

`((A, B), (C, D));`



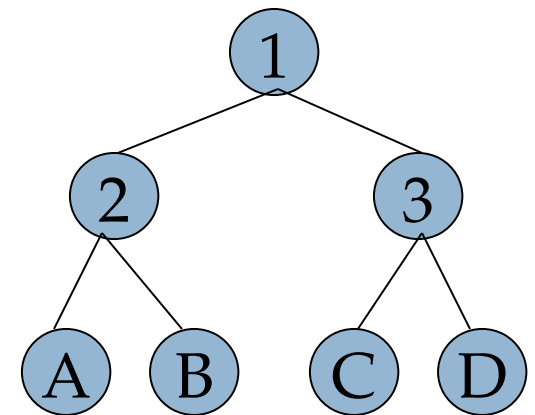
Convenção

Tamanho dos ramos pode ser representado:

`((A:1.0, B:1.0):2, (C:1, D:1):2);`

Nomes podem ser atribuídos aos nós internos:

`((A, B)2, (C, D)3);`



Conclusão

63

- SARS
- Origem e Evolução da Epidemia
- Análise Filogenética
- Construção de Árvores Filogenéticas
- Algoritmo Neighbor-Joining
- Estudo de Casos

Perguntas?

