

Reconstrução Filogenética

Rodrigo Mendes

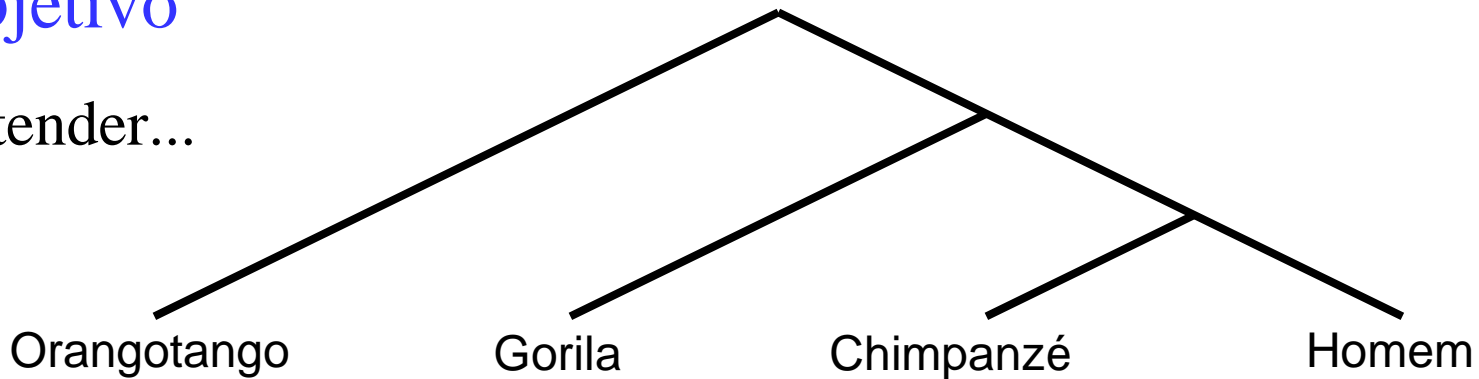
PPG-GMP ESALQ/USP

rmendes@esalq.usp.br

Seminário

Objetivo

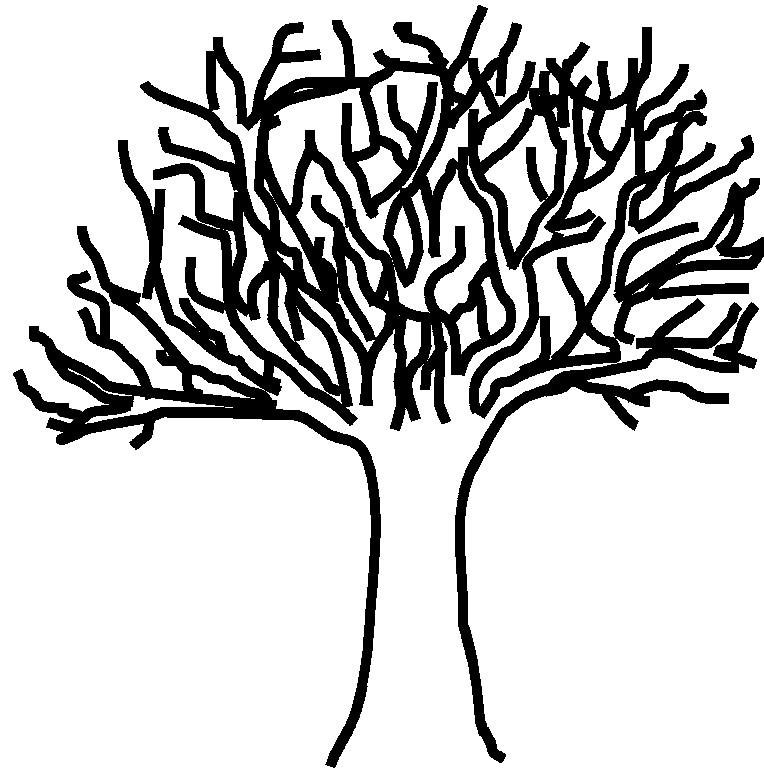
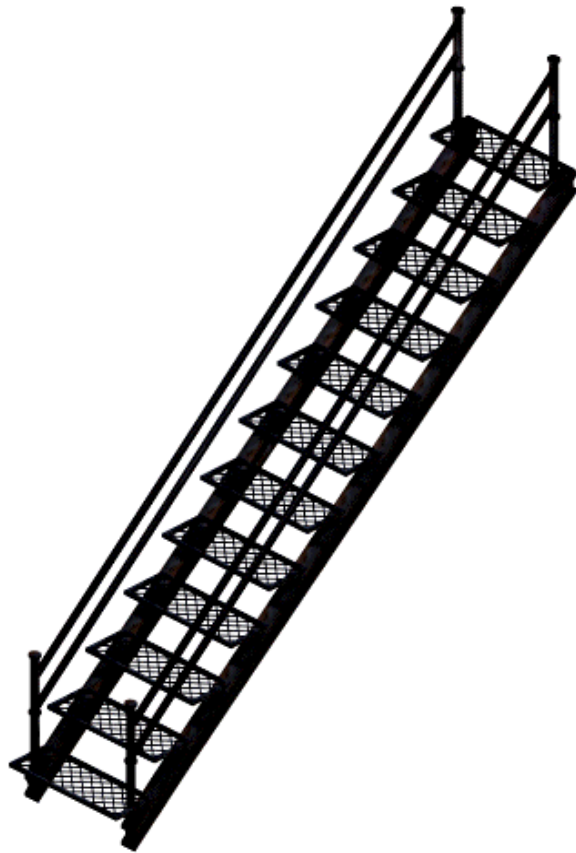
Entender...



Tree of the Life Website - University of Arizona

Evolução do Conceito

Organização do Universo dos Seres Vivos



Filogenia

Definição de árvore filogenética

Árvores filogenéticas ilustram as relações evolucionárias entre grupos de **organismos**, ou entre uma família de **ácidos nucleicos** relacionados ou seqüências de **proteínas**.

A Árvore da Vida

O início



Charles Darwin
1859

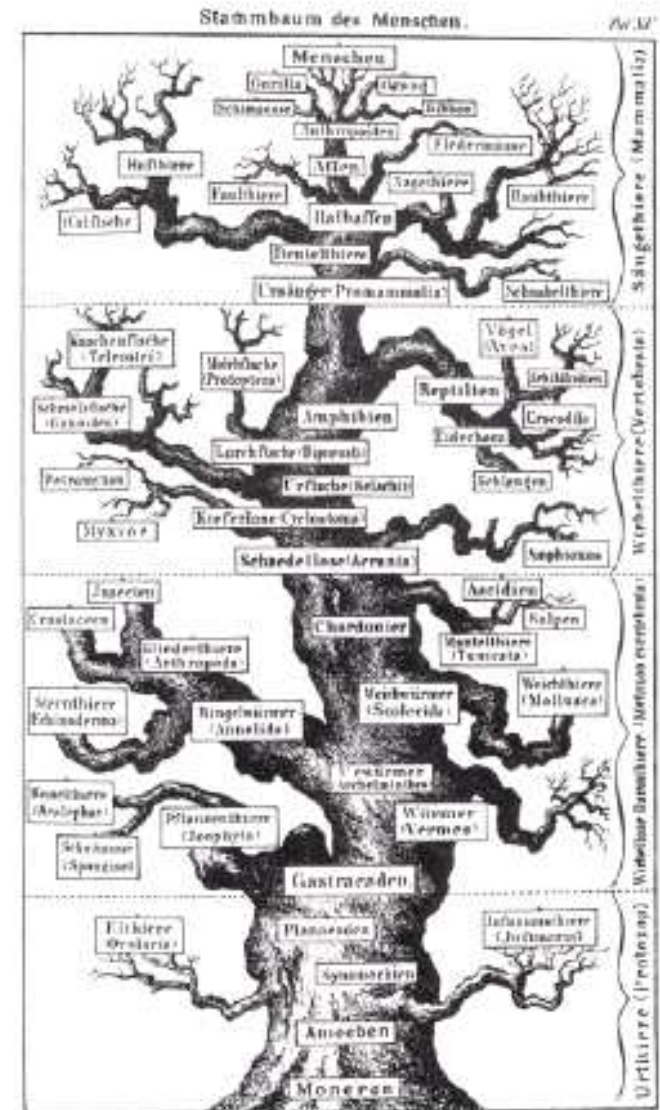
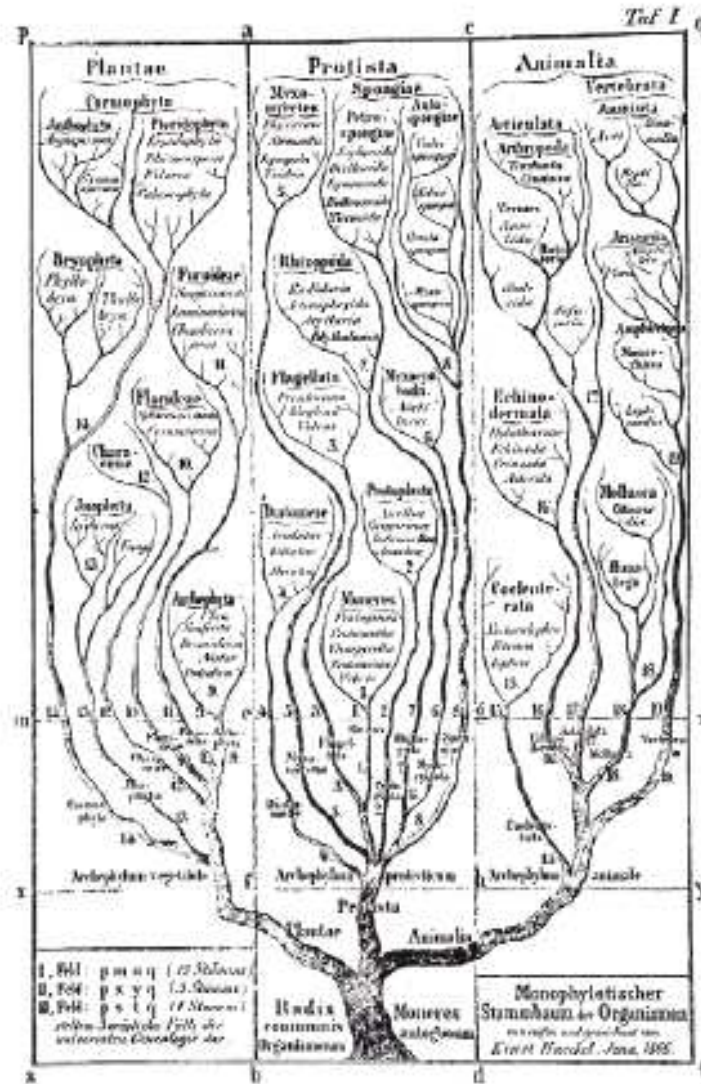
*"The affinities of all the beings of the same class have sometimes been represented by a great tree... As buds give rise by growth to fresh buds, and these if vigorous, branch out and overtop on all sides many a feebler branch, so by generation I believe it has been with the great **Tree of Life**, which fills with its dead and broken branches the crust of the earth, and covers the surface with its ever branching and beautiful ramifications."*

A Árvore da Vida

O início

Filogenias
“universais”

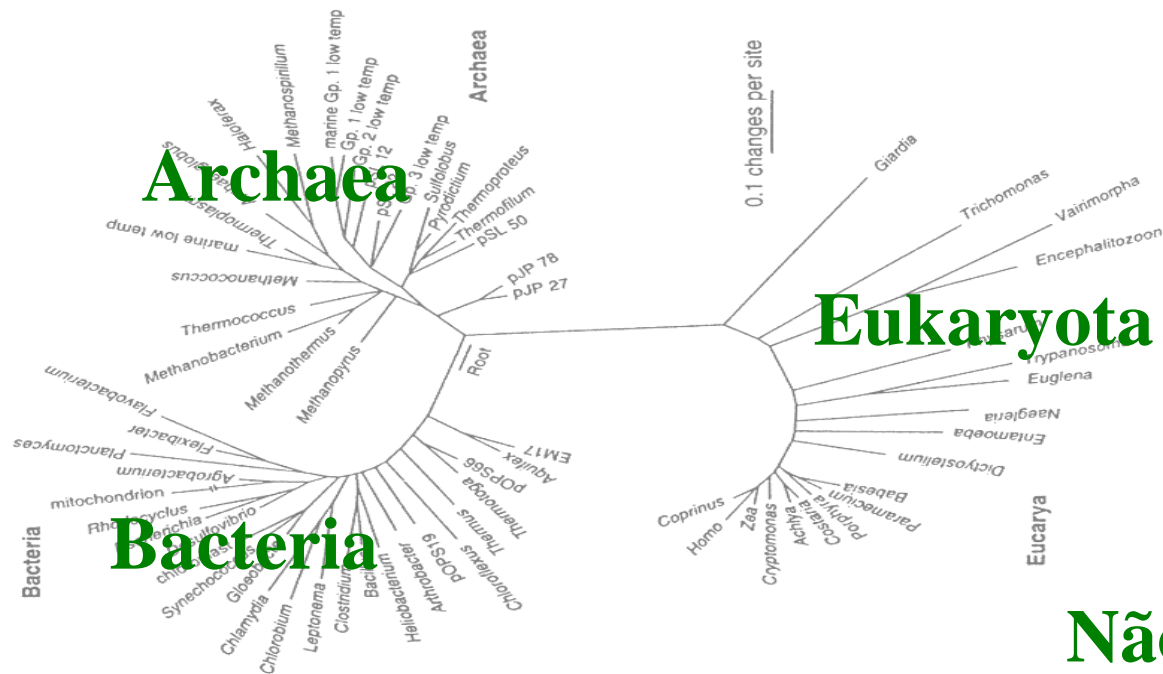
Ernest Haeckel
1860



Classificação Atual

National Center for Biotechnology Information

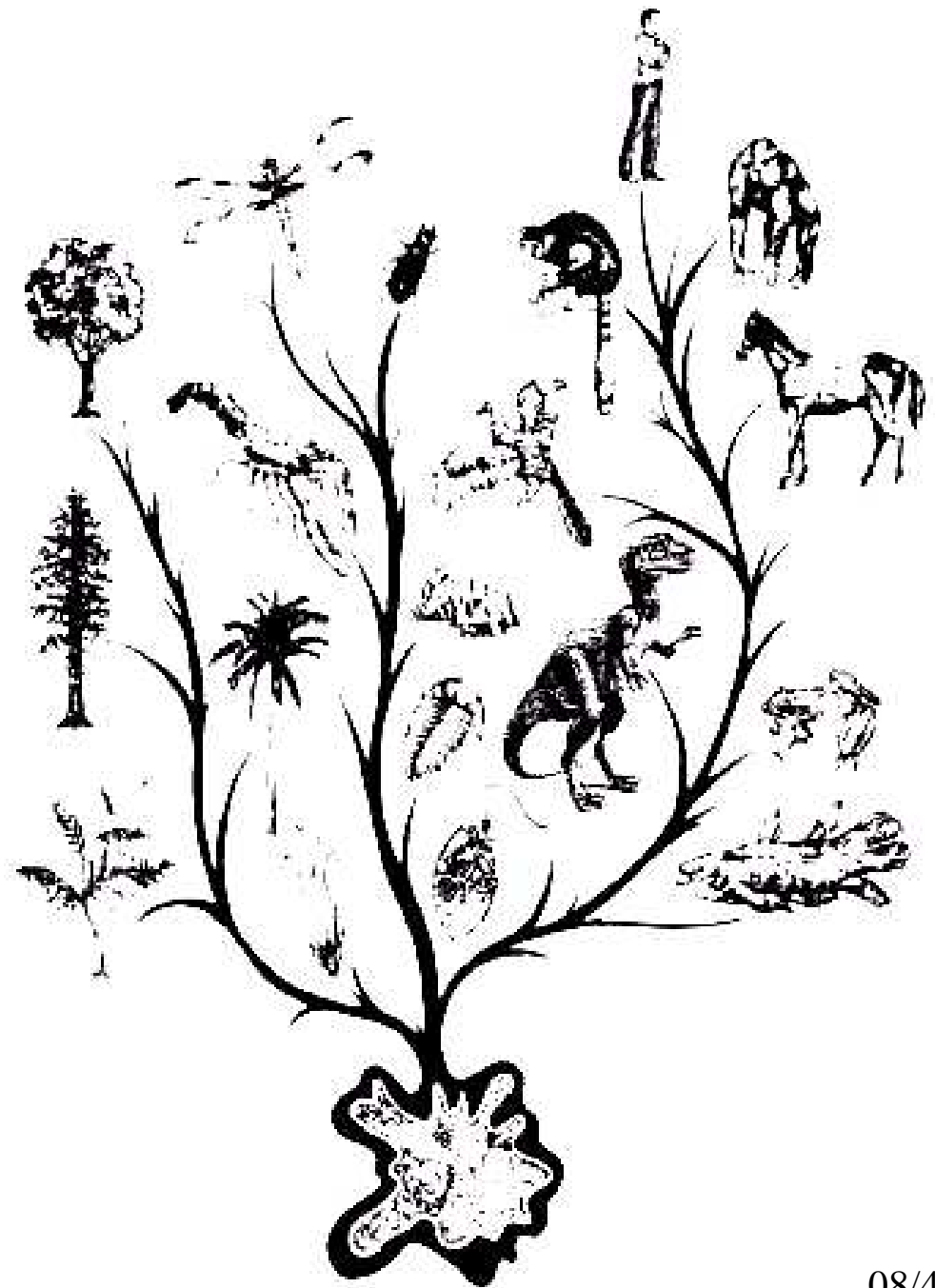
Nível máximo de classificação do banco de dados taxonômico mantido pelo NCBI/GenBank.



Reconstrução

Pressupostos

- Ancestralidade comum
- Os organismos trazem em si parte da história evolutiva
- Igualdades significam origem comum



Reconstrução Filogenética

Dados utilizados

- **Tradicionalmente:** características morfológicas (formato da asa, número de pernas, disposição de olhos).

Sistemática Evolutiva (baseado nas diferenças)

- **Atualmente:** principalmente dados moleculares (seqüências de DNA e proteínas - marcadores moleculares).

Taxonomia Numérica ou Sistemática Fenética (baseado nas igualdades)

Filogenia Molecular

Consequência do uso... (ex. 1)

João Marcelo Pereira Alves

Caracterização e Filogenia Moleculares de *Acanthamoeba*

Tese de Doutorado | ICB | USP | 2001

“Entretanto, apesar dos grandes avanços no conhecimento da organização genética e filogenia no gênero *Acanthamoeba*, surgidos com as técnicas de análise molecular, a determinação das espécies continuou baseada nos critérios antigos, embora os novos dados indicassem total inadequação da nomenclatura vigente.”

pg 128

Filogenia Molecular

Consequências do uso... (ex. 2)



Available online at www.sciencedirect.com



Molecular Phylogenetics and Evolution 32 (2004) 83–100

MOLECULAR
PHYLOGENETICS
AND
EVOLUTION

www.elsevier.com/locate/ympev

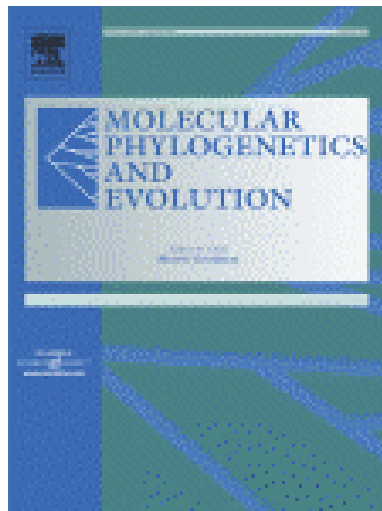
A phylogeny of four mitochondrial gene regions suggests a revised taxonomy for Asian pitvipers (*Trimeresurus* and *Ovophis*)[☆]

Anita Malhotra* and Roger S. Thorpe

School of Biological Sciences, University of Wales Bangor, Bangor, Gwynedd LL57 2UW, UK

Received 1 August 2003; revised 5 February 2004

Available online 12 April 2004



BioCode

<http://www.rom.on.ca/biodiversity/biocode/biocode.html>

Draft BioCode (1997): the prospective international rules for the scientific names of organisms

Prepared and edited by

W. Greuter¹, D. L. Hawksworth², J. McNeill³, M. A. Mayo⁴, A. Minelli⁵, P. H. A. Sneath⁶, B. J. Tindall⁷, P. Trehane⁸ & P. Tubbs⁹ (the IUBS/TUMS International Committee for Bionomenclature)

Fourth draft, revised at a meeting of the Committee at Egham, U.K., 21-25 April 1997, by

W. Greuter, D. L. Hawksworth, J. McNeill, A. Minelli, B. J. Tindall, P. Trehane & P. Tubbs

Estudo Filogenético

Aplicações

- **Árvore da Vida:** analisar as mudanças que tem ocorrido na evolução dos diferentes organismos.
- Relações filogenéticas entre **genes** podem ajudar a prever quais deles apresentam funções similares (detecção de ortólogos).
- Detectar mudanças em espécies de evolução rápida (**vírus**).

Árvores Filogenéticas

Terminologia (1)

- **OTUs:** *operational taxonomic units*, objetos de comparações.
- **Nós internos:** unidades ancestrais, hipoteticamente.
- **Raiz:** ancestral comum de todas OTUs do estudo. O caminho entre a raiz e os nós definem o caminho evolutivo da OTU.

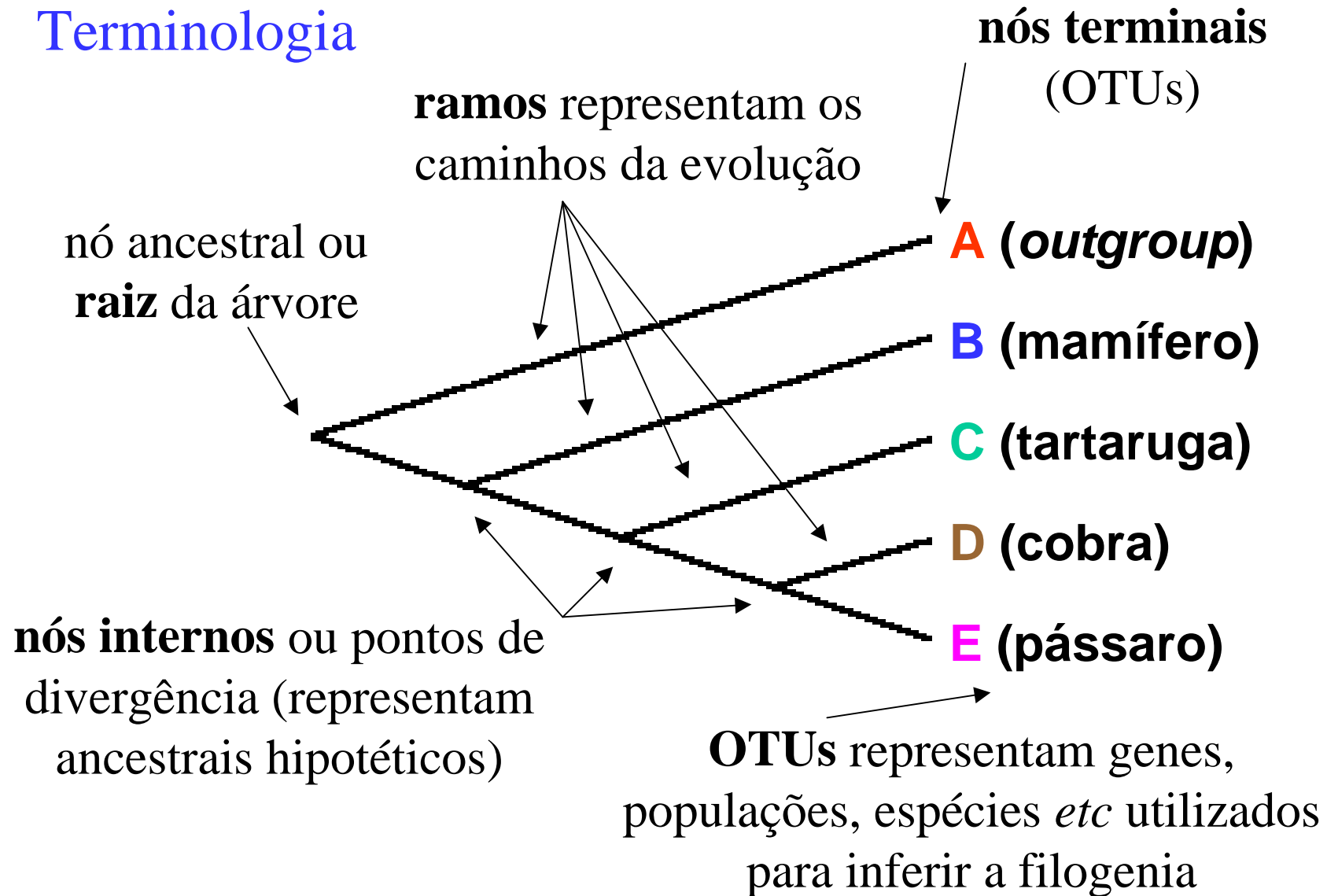
Árvores Filogenéticas

Terminologia (2)

- **Não-enraizada:** relações entre OTUs sem representar o caminho evolutivo. Obs: pode ser enraizada na presença de um *outgroup* (grupo externo o qual acredita-se que tenha sido ramificado anteriormente).
- **Topologia:** padrão de enraizamento da árvore.
- **Comprimento do ramo:** quantidade de diferenças que ocorrem ao longo do ramo.

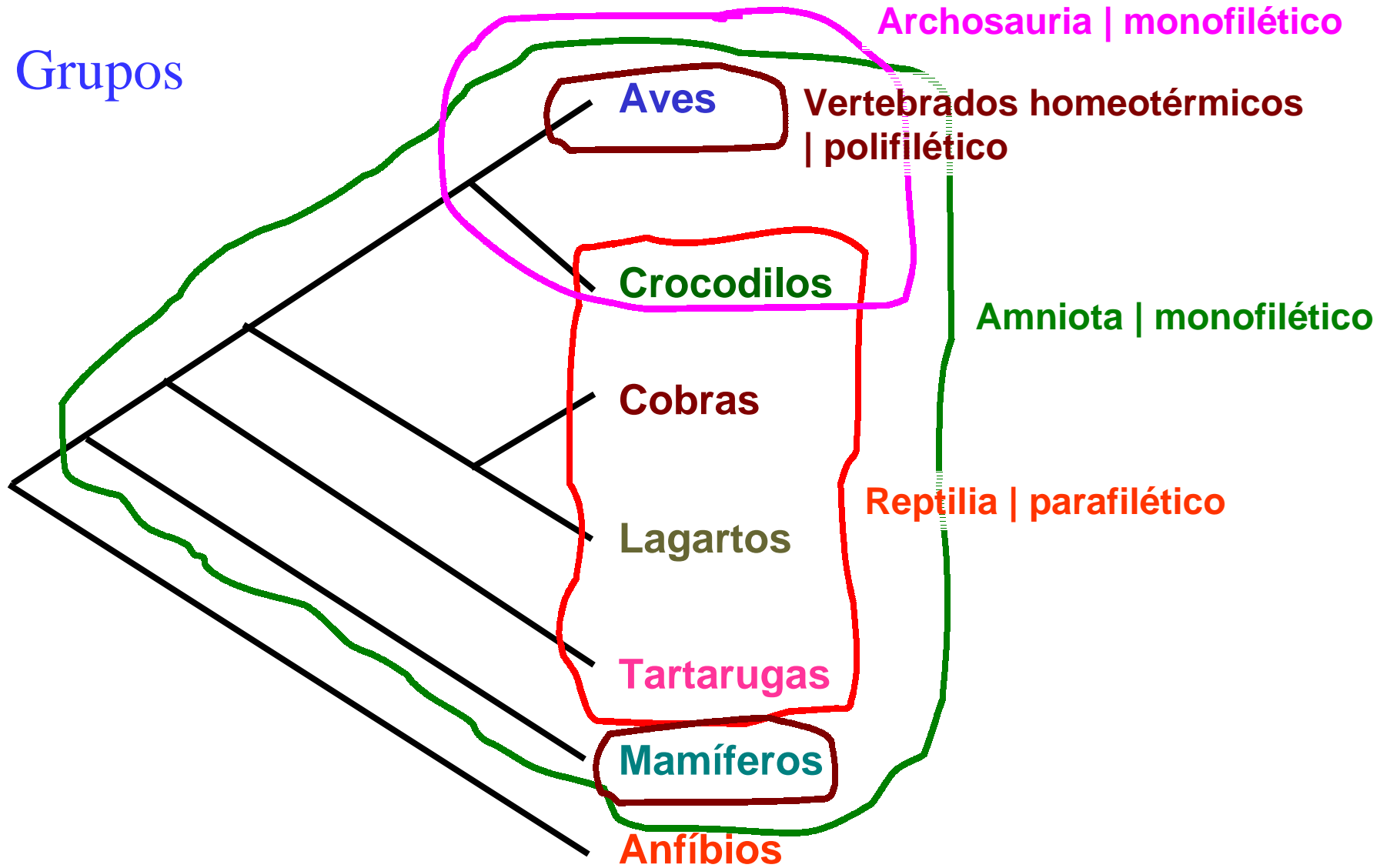
Árvores Filogenéticas

Terminologia



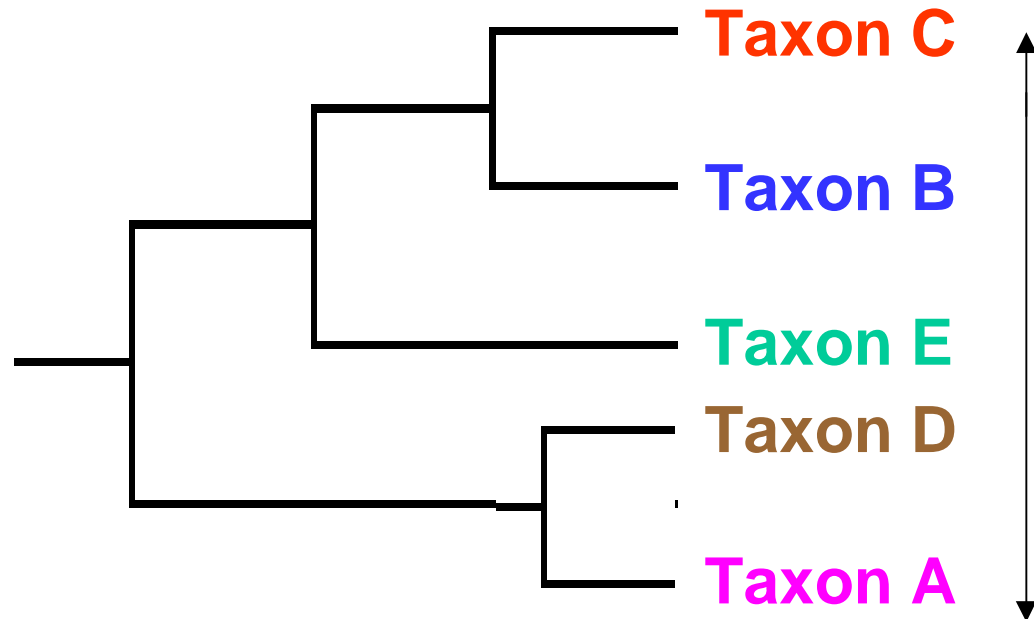
Árvores Filogenéticas

Grupos

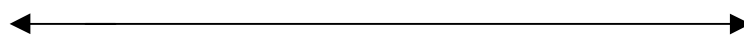


Árvore

Entendimento



a **distância** e a **ordem** (em baixo ou em cima) entre as OTUs não tem significado



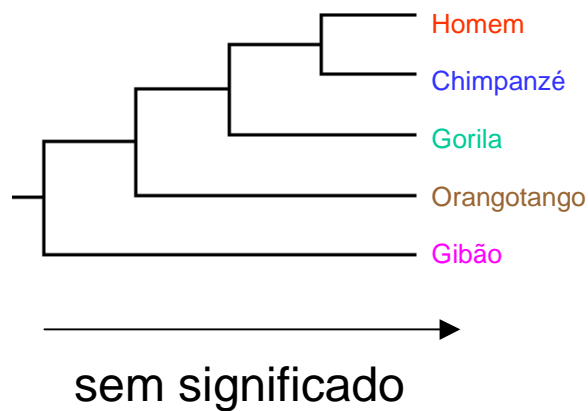
cladogramas: sem escala | **filograma:** proporcional à distância genética ou quantidade de alterações | **árvores ultramétricas:** proporcional ao tempo

$((E, (C, B)), (D, A))$

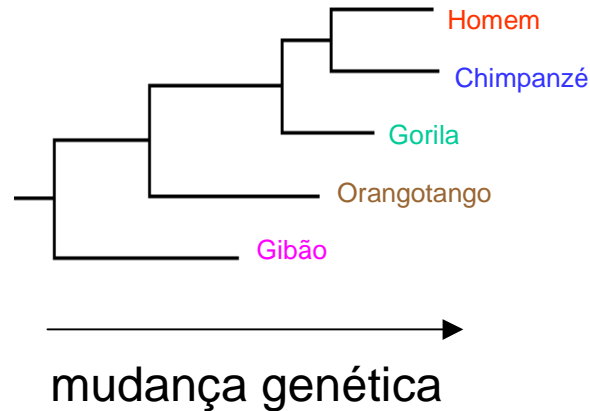
Árvores

Tipos e Inferências

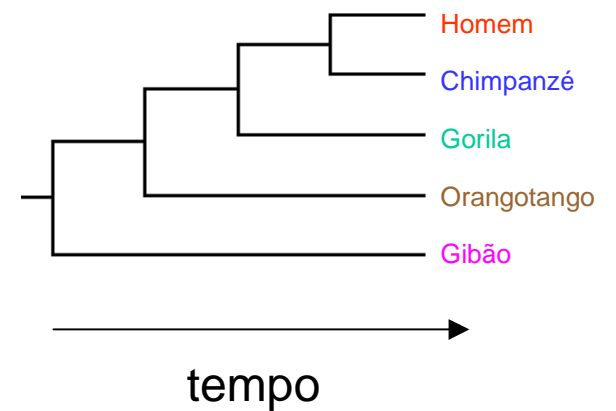
Cladograma



Filograma



Árvore Ultramétrica

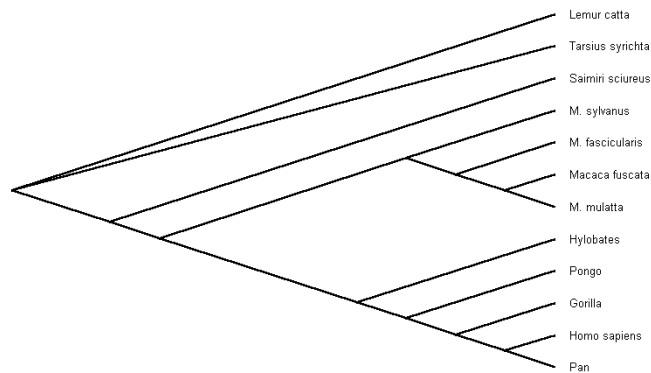


Obs: os 3 tipos apresentam a mesma ordem dos ramos (relações genéticas) entre os OTUs.

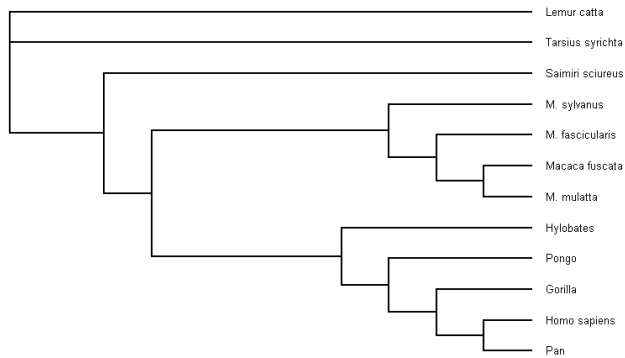
Árvores

Topologias

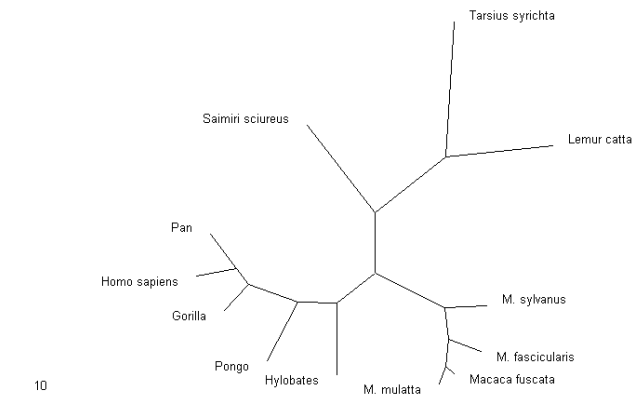
- Cladograma



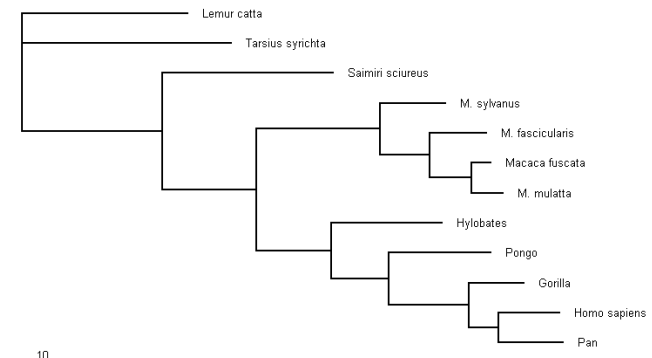
- Cladograma retangular



- Árvore radial



- Filograma



Possibilidades

Número de árvores enraizadas

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

nº de OTUs (n)	árvores enraizadas
2	1
5	105
10	34 459 425

Reconstrução Filogenética

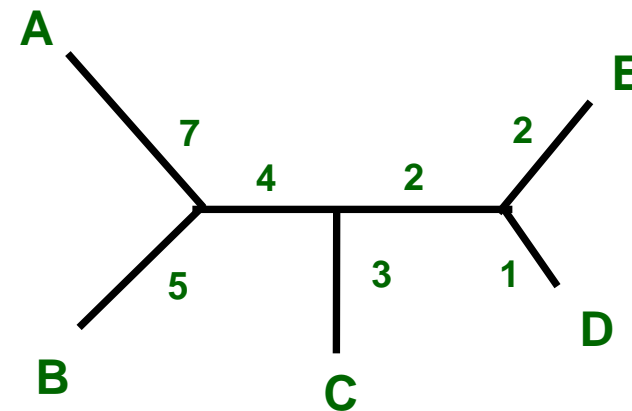
Métodos

- **Métodos de Distância:** as distâncias são computadas para todas OTUs e é construída uma árvore na qual a distância entre as OTUs refletem as distâncias evolutivas.
- **Máxima Parcimônia:** seleciona a árvore que minimiza o número de mudanças requeridas para explicar os dados.
- **Máxima Verossimilhança:** utilizando um modelo de seqüência evolutiva, encontra a árvore mais provável que reflete os dados analisados.

Reconstrução Filogenética

Método de Distância

	A	B	C	D	E
A	0				
B	12	0			
C	14	12	0		
D	14	12	6	0	
E	15	13	7	3	0



Nem sempre existe a árvore correspondente exatamente à matriz!

Reconstrução Filogenética

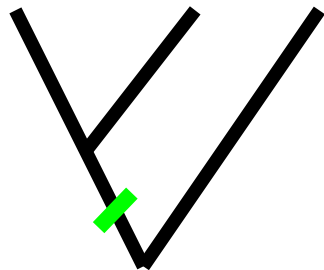
Máxima Parcimônia

- **Método:** minimiza o número de mudanças necessárias para explicar os dados.
- **Computação:** conta o número de passos de uma determinada árvore (Fitch) e busca a “melhor” entre tantas possíveis (heurístico).
- **Problema:** assume que a taxa de mudança em todos os ramos são iguais.

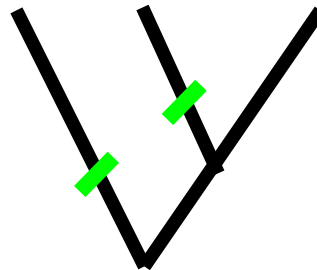
Reconstrução Filogenética

Máxima Parcimônia

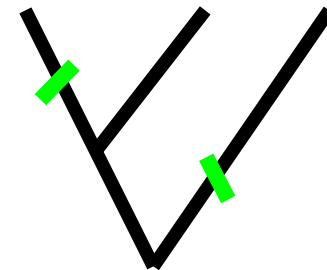
rato lagarto peixe



rato lagarto peixe



rato peixe lagarto



mais parcimonioso

◆ surgimento do pulmão

Reconstrução Filogenética

Máxima Verossimilhança

- **Princípios de independência:**

- (i) a evolução dos diferentes caracteres (bases) é independente

- (ii) seqüências após divergirem evoluem independentemente

- **Vantagem:** menor variância em relação aos outros métodos, já que é menos afetado por erros de amostragem, e mais robusto frente às violações do modelo evolutivo.

- **Desvantagem:** consome grande tempo para o cálculo.

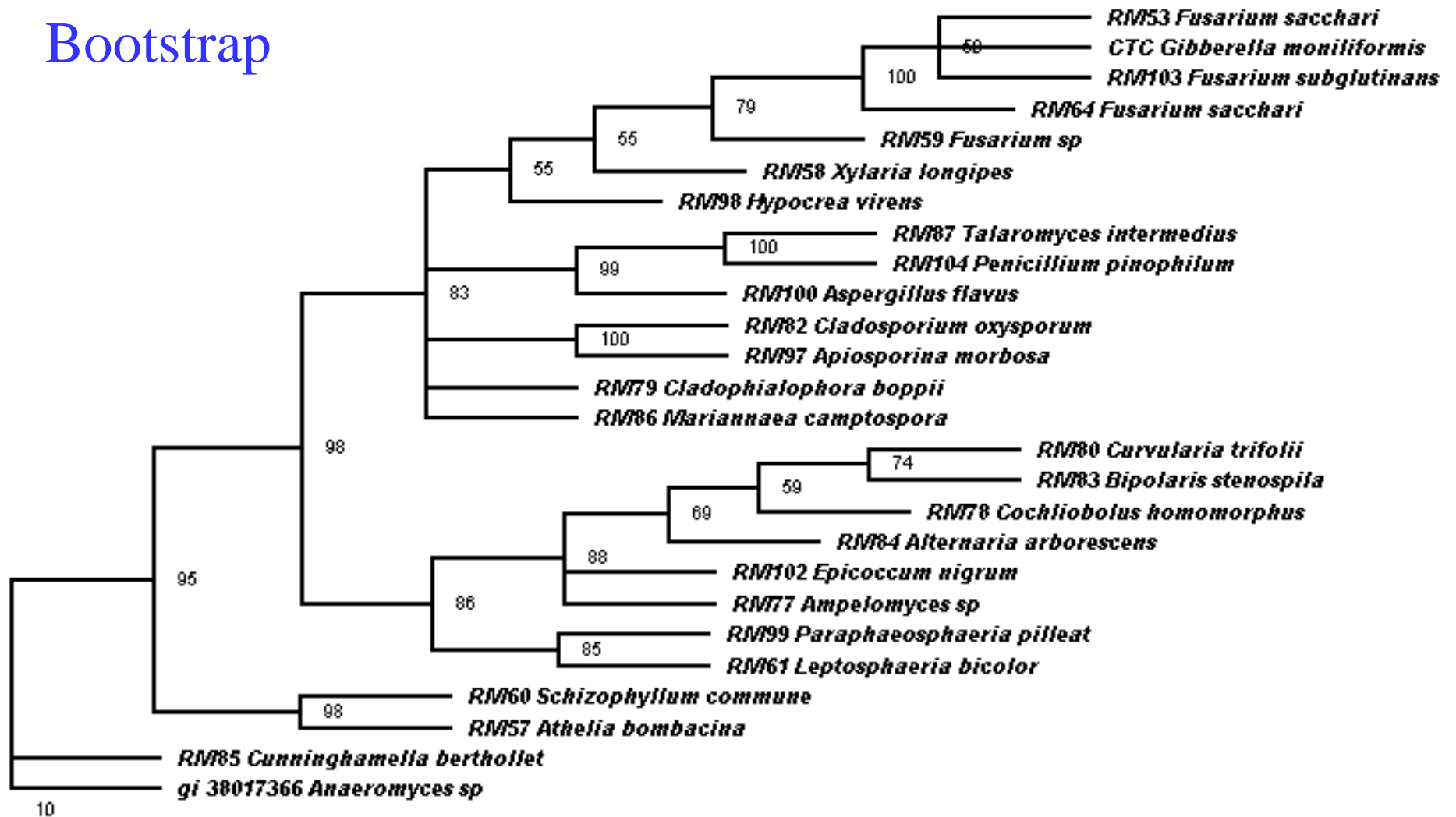
Reconstrução Filogenética

Comparações dos métodos

- **Parcimônia** e **Verossimilhança** apresentam modelos evolutivos.
- **Distância** útil para a comparação de genomas inteiros, presença ou ausência de genes, sem inferência evolutiva.
- **Parcimônia** ou **Verossimilhança**? Atualmente a maioria das pessoas acreditam que os métodos baseados na verossimilhança são melhores, embora sejam mais trabalhosos.
- Os pacotes mais comuns utilizados contêm *softwares* para os **3 métodos**. A utilização de mais de um, pode aumentar a confiabilidade da árvore.

Confiança da Árvore

Bootstrap



Reamostragem com reposição pseudoaleatória dos dados.

Homologia

“Similaridade pode ser fato, mas homologia deve permanecer como hipótese” (Russo, 2001)

- **Homologia:** ancestral comum
- **Genes Parálogos:** duplicação gênica (mesma espécie)
- **Genes Ortólogos:** divergência gênica (espécies diferentes)
- **Xenologia:** transferência horizontal
- **Plerologia:** conversão gênica
- **Estruturas homólogas:** braços de hominídeos e asas de aves
- **Estruturas análogas:** asas de aves e asas de insetos

Evolução simulada

Diferenças observadas no alinhamento

seq. ancestral	mutações nas seqüências descendentes			alinhamento
A	A	A		A A
C	C	C - A	subst. simples	C - A
T	T	T		T T
G	G	G		G G
A	A - C - T	A	subst. múltiplas	T - A
A	A	A		A A
C	C - G	C - A	subst. coincidentes	G - A
G	G	G		G G
T	T - A	T - A	subst. paralelas	A A
A	A	A		A A
A	A - C - T	A - T	subst. convergente	T T
C	C	C		C C
G	G	G		G G
C	C	C - T - C	reversão	C C

Reconstrução Filogenética

Softwares

- **PHYLIP:** *phylogenetic inference package*

Felsenstein

www.evolution.genetics.washington.edu/phylip.html

- **PAUP:** *phylogenetic analysis using parsimony*

Swofford

www.paup.csit.fsu.edu

Reconstrução Filogenética

Exemplo

GeneBee

http://www.genebee.msu.su/services/phtree_reduced.html



Reconstrução Filogenética

Na prática

1. Obter as **seqüências** incluindo um *outgroup* (arquivo texto)
2. **Formatar** as seqüências para o alinhamento (FASTA)
3. **Alinhar** utilizando o ClustalX
4. **Editar** as seqüências alinhadas utilizando o BioEdit
5. **Construir a árvore** utilizando o ClustalX
6. **Interpretar** a árvore

Uma história real

Lafayette, Louisiana EUA, 1998

“...meu ex-amante injetou sangue contaminado com HIV em mim...”

Fato: não havia provas ou testemunhas, apenas um registro, no hospital, de retirada de sangue de um paciente com AIDS no mesmo dia em que seu amante (médico do hospital) lhe aplicou um “medicamento”.

Solução:

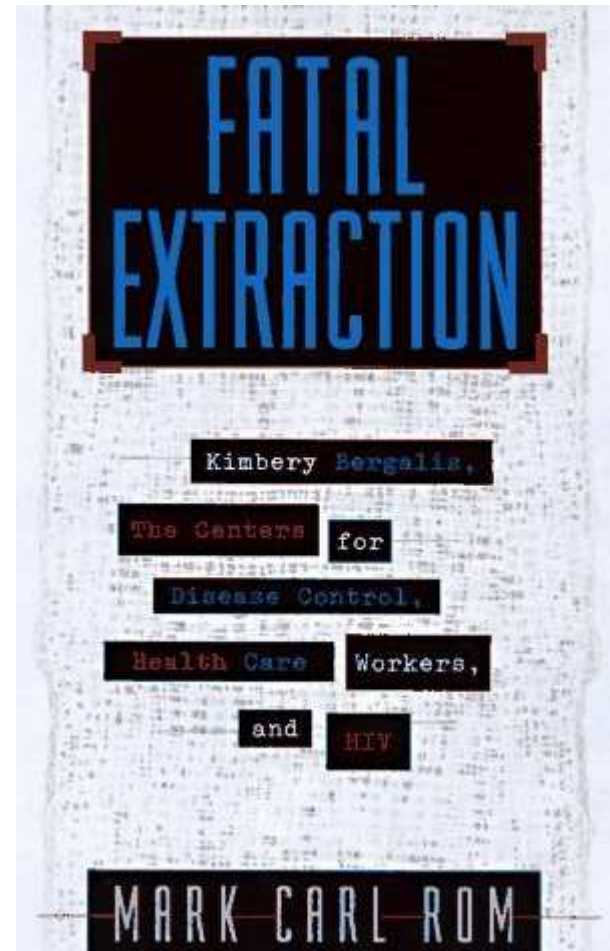
FILOGENIA

Outra história real

Flórida EUA, 1990

“Fatal Extraction: The Story Behind the Florida Dentist Accused of Infecting His Patients With HIV and Poisoning Public Health”

Mark Carl Rom



Outra história real

Flórida EUA, 1990

“...fui contaminado com HIV e nunca usei drogas ou recebi transfusão de sangue, e não tenho um comportamento sexual impróprio...”

Fato: após uma paciente acusar seu dentista (soro positivo) por contaminá-la com o vírus do HIV, foi realizado um *screening* de 1100 pacientes do dentista e foram encontrados 8 pacientes contaminados.

Solução:

A MESMA

Desvendando o mistério - Simulação

Na prática

1. Obter as **seqüências** incluindo um *outgroup* (arquivo texto)
2. **Formatar** as seqüências para o alinhamento (FASTA)
3. **Alinhar** utilizando o ClustalX
4. **Editar** as seqüências alinhadas utilizando o BioEdit
5. **Construir a árvore** utilizando o ClustalX
6. **Interpretar** a árvore

Desvendando o mistério - Simulação

Sequências (1)

```
LOCUS      Dentista.Seq      585 bp
DEFINITION Dentista.Seq, 585 bases, 9450 checksum.
ORIGIN
      1  TTTGNGNCCT  TCCAACAAAA  ATATGAGGGT  GTGGTTTGCT  GGCAACAGCG
     51  TCCGCCCCAA  GTATTTTTCA  CCCATGTCTT  TTGCGCACTT  TTTGTTTCCT
    101  GGGCGAGTTC  GCTCGCCACC  AGGACCCAAC  CATAAACCTT  TTTTTATGCA
    151  GTTGCAATCA  GCGTCAGTAT  AATAATTCAA  TTTATTAAAA  CTTTCAACAA
    201  CGGATCTCTT  GGTTCCTGGC  TCGATGAAGA  ACGCAGCGAA  ATGCGATACG
    251  TAGTGTGAAT  TGCAGAATTC  AGTGAATCAT  CGAATCTTTG  AACGCACATT
    301  GCGCCCTTTG  GTATTCCAAA  GGGCATGCCT  GTTCGAGCGT  CATTTGTACC
    351  CTCAAGCTTT  GCTTGGTGTT  GGGCGTCTTT  TTGTCTCTCC  CCTTGTTGGG
    401  GGAGACTCGC  CTTAAAACGA  TTGGCAGCCC  GACCTACTGG  TTTTCGGAGC
    451  GCAGCACAAA  TTTGCGCCTT  CCAATCCACG  GGGCGGCATC  CAGCAAGCCT
    501  TTGTTTTCTA  TAACAAATCC  ACATTTTGAC  CTCGGATCAN  GTAGGGATAC
    551  CCGCTGAACT  TAAGCATATC  AATAAGCGGA  GGAAN
```

//

Desvendando o mistério - Simulação

Formatar (FASTA) (2)

>Dentista

```
TTTGCGCCCTTCCAACAAAATATGAGGGTGTGGTTTGCTGGCAACAGCG  
TCCGCCCAAGTATTTTTCACCCATGTCTTTTGCGCACTTTTTGTTTCCT  
GGGCGAGTTCGCTCGCCACCAGGACCCAACATAAACCTTTTTTTATGCA  
GTTGCAATCAGCGTCAGTATAATAATTCAATTTATTAAACTTTCAACAA  
CGGATCTCTTGGTTCTGGCATCGATGAAGAACGCAGCGAAATGCGATACG  
TAGTGTGAATTGCAGAATTCAGTGAATCATCGAATCTTTGAACGCACATT  
GCGCCCTTTGGTATTCCAAAGGGCATGCCTGTTTCGAGCGTCATTTGTACC  
CTCAAGCTTTGCTTGGTGTGGGGCGTCTTTTTGTCTCTCCCCTTGTTGGG  
GGAGACTCGCCTTAAAACGATTGGCAGCCCGACCTACTGGTTTTTCGGAGC  
GCAGCACAAATTTGCGCCTTCCAATCCACGGGGCGGCATCCAGCAAGCCT  
TTGTTTTCTATAACAAATCCACATTTTGACCTCGGATCACGTAGGGATAC  
CCGCTGAACTTAAGCATATCAATAAGCGGAGGAACCTCAG
```

Desvendando o mistério - Simulação

Sequências (1)

Formatar (FASTA) (2)

Alinhar (ClustalX) (3)

Editar (BioEdit) (4)

Construir a árvore (ClustalX) (5)

Interpretar (6)

dentista.txt

Demonstração...

Obrigado pela Atenção!

Rodrigo Mendes

PPG-GMP ESALQ/USP

rmendes@esalq.usp.br

Apresentação disponível em <http://rodrigomendes.tripod.com>