

Analise filogenética baseada em alinhamento de domínios

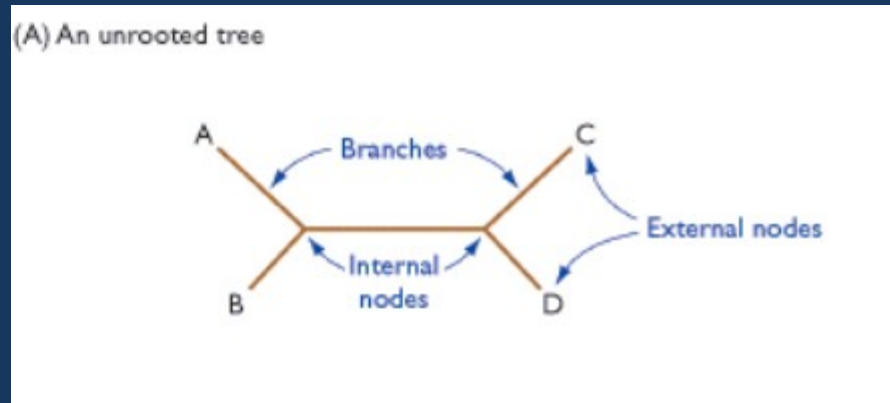
Moléculas biológicas e evolução

- Como já foi comentado anteriormente sabemos que o DNA de qualquer espécie de ser vivo sofre mutações ao longo do tempo e que parte destas mutações se fixam
- Quando duas espécies surgem a partir de um ancestral comum elas passam a acumular mutações distintas
- Deste modo a análise destas mutações permitira a inferência do processo evolutivo dos organismos comparados

Moléculas biológicas e evolução

- Quando analisamos seqüências de organismos distantemente relacionados o DNA não possui semelhança o suficiente para permitir um alinhamento confiável, devido a isso utilizamos seqüências de proteínas para análise evolutiva
- Deve-se lembrar entretanto que ao analisarmos evolução de seqüências protéicas nos temos que considerar que as mutações que ocorrem nas proteínas são meramente um reflexo das mutações do DNA e portanto qualquer modelo para estudo de evolução de proteínas deve levar em conta o código genético

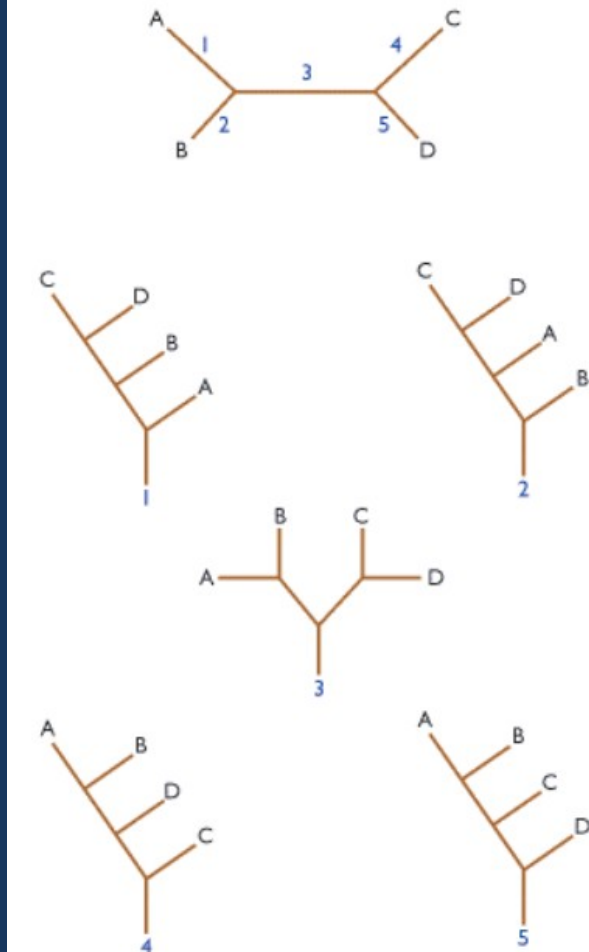
Arvore filogenética



- Uma maneira de representar o processo evolutivo de uma família de proteínas (ou genes) é através de uma árvore filogenética
- Esta árvore é composta de nós externos que representam os organismos ou seqüências estudados. Os ramos são as linhas que interconectam estes nós e os ramos internos conectam os ramos.

Arvore filogenética

(B) Rooted trees



- A árvore pode ser representada na forma não-enraizada, que apresenta a evolução mas não determina o ponto onde se encontrava o ancestral comum
- No exemplo mostrado ao lado existem 5 pontos diferentes para a raiz da árvore e portanto 5 árvores enraizadas podem ser deduzidas, cada uma com diferentes implicações evolutivas

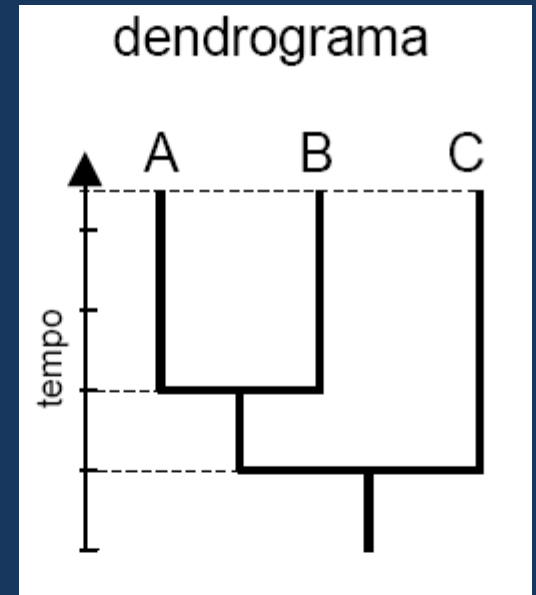
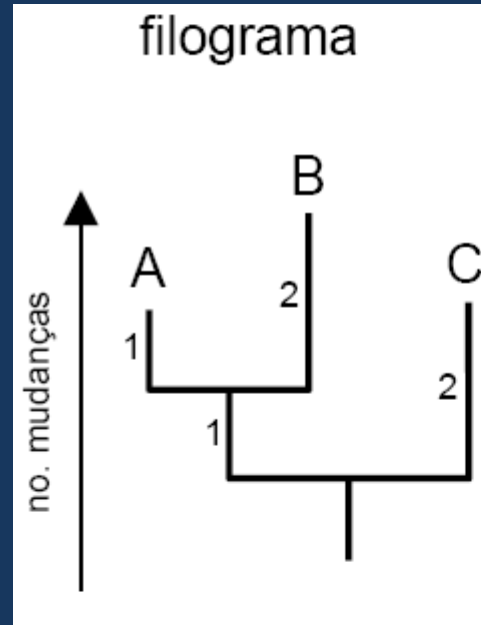
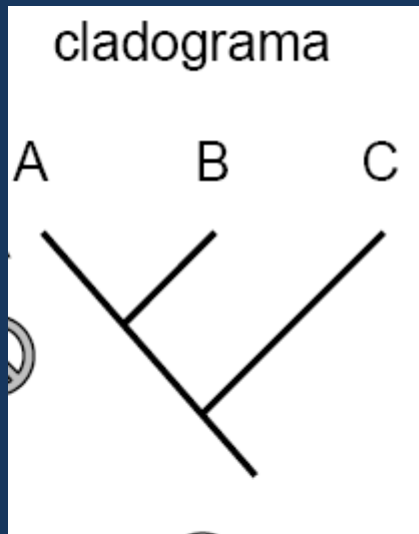
Arvore filogenética

Como saber onde ocorre a raiz de uma arvore filogenética?

Uma das abordagens possíveis é colocar a raiz no ponto médio da arvore (ponto eqüidistante dos nós terminais) entretanto este tipo de abordagem parte do pressuposto que as taxas de evolução dos genes é uniforme o que muitas vezes não é verdade

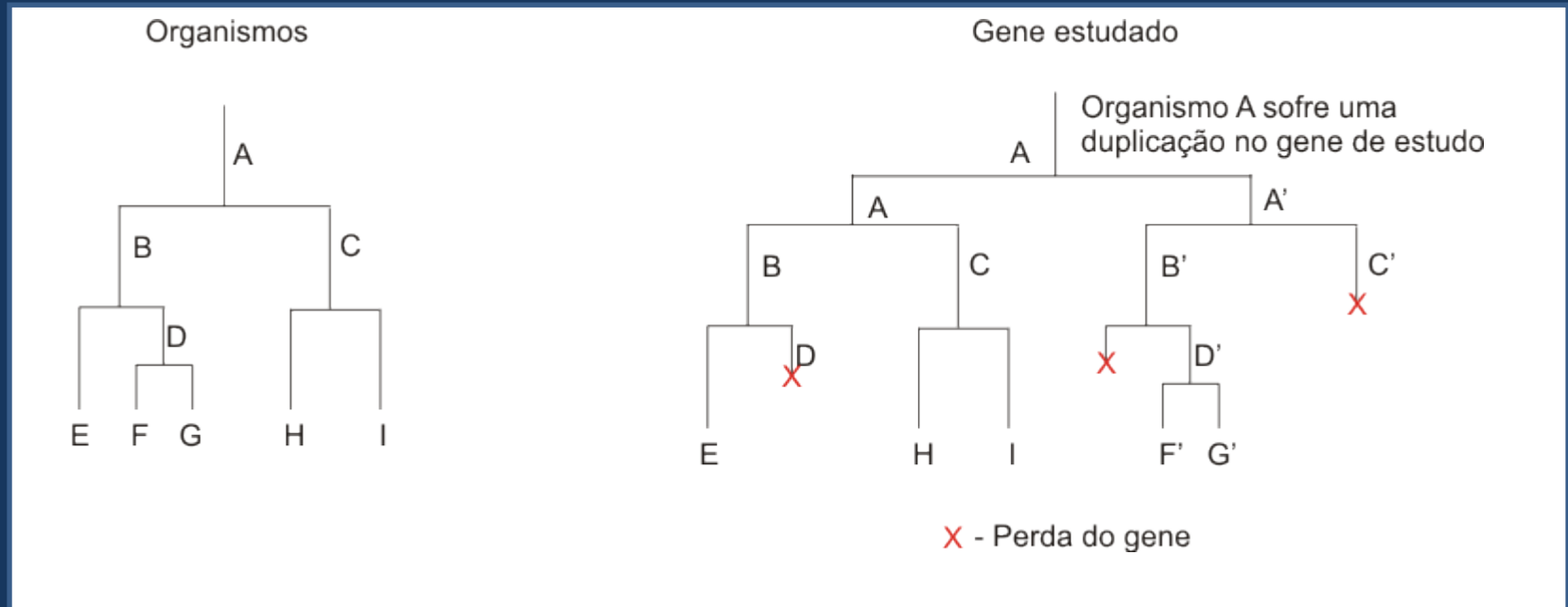
Uma segunda abordagem muito utilizada é a escolha de um “outgroup” , uma proteína que conhecidamente é mais distante do resto do grupo e a arvore passa a ser enraizada no ramo que liga esta proteína aos outros grupos

Tipos de arvores



- Cladograma- demonstra apenas a topologia da arvore
- Filograma- comprimento do ramos são proporcionais ao numero de mudanças aferidas
- Dendrograma- Comprimento do ramos reflete o tempo transcorrido (ramos com diferentes taxas de mutação)

Evolução de genes



Nem sempre é possível associar a filogenia detectada para um gene com a filogenia dos organismos. Fenômenos de duplicação, deleção e recombinação podem alterar significativamente a filogenia obtida.

Tipos de métodos para construção de árvores filogenéticas

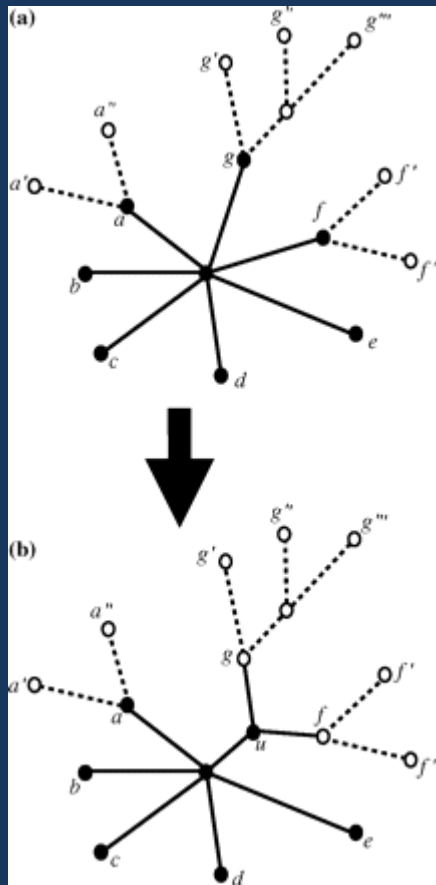
- Máxima parcimônia- Este método busca a árvore que envolva o menor número de eventos de mutação
- Máxima probabilidade- Método semelhante a máxima parcimônia, mas utiliza modelos evolucionários que computam diferentes chances de mutação dependendo da base que sofreu mutação
- Métodos de distancia- Utiliza comparações par a par entre distâncias (número de mutações) das diferentes seqüências estudadas para construir uma árvore que reproduza esta distância de modo mais próximo possível.

Construção de árvores filogenéticas

- Numero de árvores possíveis de acordo com o numero de espécies (ou proteínas) diferentes analisadas

Species	Number of trees
1	1
2	1
3	3
4	15
5	105
6	945
7	10,395
8	135,135
9	2,027,025
10	34,459,425
11	654,729,075
12	13,749,310,575
13	316,234,143,225
14	7,905,853,580,625
15	213,458,046,676,875
16	6,190,283,353,629,375
17	191,898,783,962,510,625
18	6,332,659,870,762,850,625
19	221,643,095,476,699,771,875
20	8,200,794,532,637,891,559,375
30	4.9518×10^{38}
40	1.00985×10^{57}
50	2.75292×10^{76}

Método de Neighbor joining



A partir da distancia entre os diversos pontos (genes ou proteínas) de uma arvore filogenética são buscados vizinhos, que são um par de seqüências que possuem um parâmetro Q mínimo

Este par de vizinhos darão origem a um nó interno que passará a representar este par de seqüências

Após isso o parâmetro Q será recalculado considerando o nó interno

Calculo do parâmetro Q

	A	B	C	D	E
A	—	22	39	39	41
B	—	—	41	41	43
C	—	—	—	18	20
D	—	—	—	—	10
E	—	—	—	—	—

Primeiro calcula-se um fator u , cujo o valor é igual a somatória das distancias daquele no terminal em relação aos outros pontos dividido por (numero de pontos-2)

- $u_a = (22 + 39 + 39 + 41) / (5 - 2) = 47$
- $u_b = (22 + 41 + 41 + 43) / (5 - 2) = 49$
- $u_c = (39 + 41 + 18 + 20) / (5 - 2) = 39.3$
- $u_d = (39 + 41 + 18 + 10) / (5 - 2) = 36$
- $u_e = (41 + 43 + 20 + 10) / (5 - 2) = 38$

Calculo do parâmetro Q

	A	B	C	D	E
A	—	22	39	39	41
B	—	—	41	41	43
C	—	—	—	18	20
D	—	—	—	—	10
E	—	—	—	—	—

Para cada par de nós terminais i, j calcula-se Q definido por: $D_{i,j} - u_i - u_j$ e selecciona-se aquele com menor valor

$$AB \rightarrow 22 - 47 - 49 = -74$$

$$AC \rightarrow 39 - 47 - 39.3 = -47.3$$

$$AD \rightarrow 39 - 47 - 36 = -44$$

$$AE \rightarrow 41 - 47 - 38 = -44$$

$$BC \rightarrow 41 - 49 - 39.3 = -47.3$$

$$BD \rightarrow 41 - 49 - 36 = -44$$

$$BE \rightarrow 43 - 49 - 38 = -44$$

$$CD \rightarrow 18 - 39.3 - 36 = -57.3$$

$$CE \rightarrow 20 - 39.3 - 38 = -57.3$$

$$DE \rightarrow 10 - 36 - 38 = -64$$

Calculo do novo nó

- A distancia dos ponto A e B ao novo nó (AB) será calculada como:

$$v_a = \frac{1}{2} D_{AB} + \frac{1}{2} (u_a - u_b)$$

$$v_a = 11 + \frac{1}{2} (47 - 49) = 10$$

$$v_b = \frac{1}{2} D_{AB} + \frac{1}{2} (u_b - u_a)$$

$$v_b = 11 + \frac{1}{2} (49 - 47) = 12$$

Calculo do novo nó

A	(AB)	C	D	E
(AB)	—	29	29	31
C	—	—	18	20
D	—	—	—	10
E	—	—	—	—

- Os nós terminais A e B são apagados e em seu lugar será inserido um nó (AB) que será considerado como um novo ponto
- As distancias deste nó para os outros nós terminais (k) será calculado seguindo a seguinte fórmula:

$$D_{(AB),k} = (D_{ak} + D_{bk} - D_{ab})/2$$

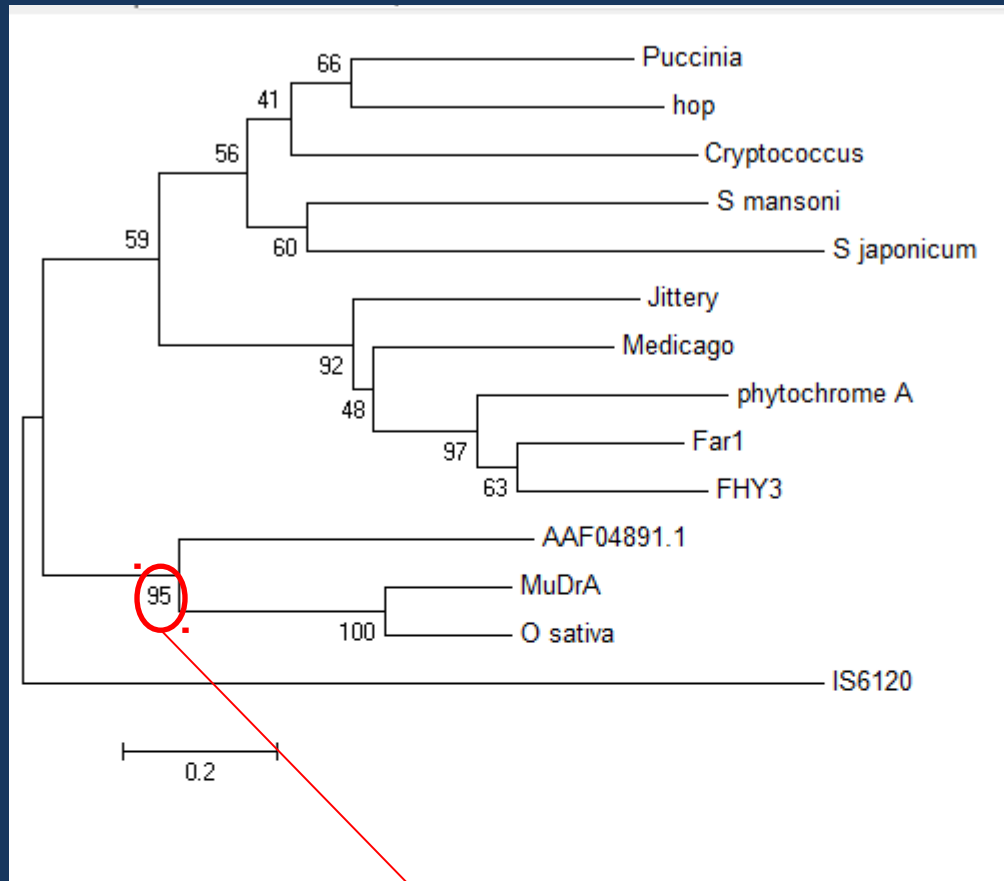
Calculo do bootstrap



O calculo de bootstrap para arvores filogenéticas permite com que tenhamos um parâmetro que reflita a robustez da análise filogenética produzida. O bootstrap é gerado a partir da criação de vários conjuntos de seqüências nas quais são escolhidas randomicamente colunas do alinhamento múltiplo. São geradas novas análises para cada um dos novos conjuntos gerados.

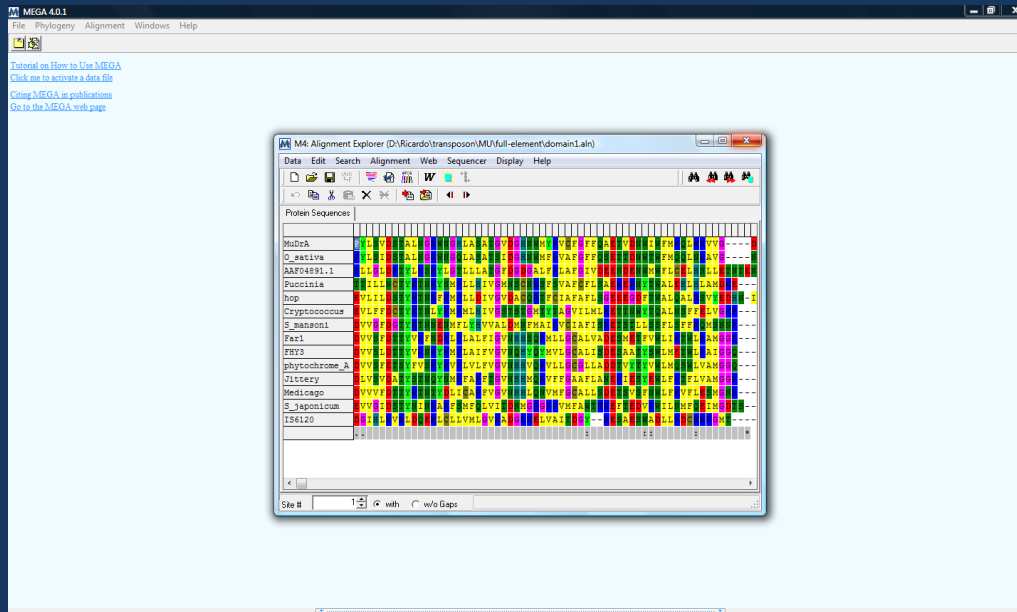
Calculo do bootstrap

- Valores nos nós interno representam valores do bootstrap – numero de ocorrências deste nó interno em 100 replicas



Valor de bootstrap indica que em 95/100 amostragens as seqüências AAF04891.1, MuDrA e O sativa estavam reunidas em um ramo contendo apenas as três seqüências

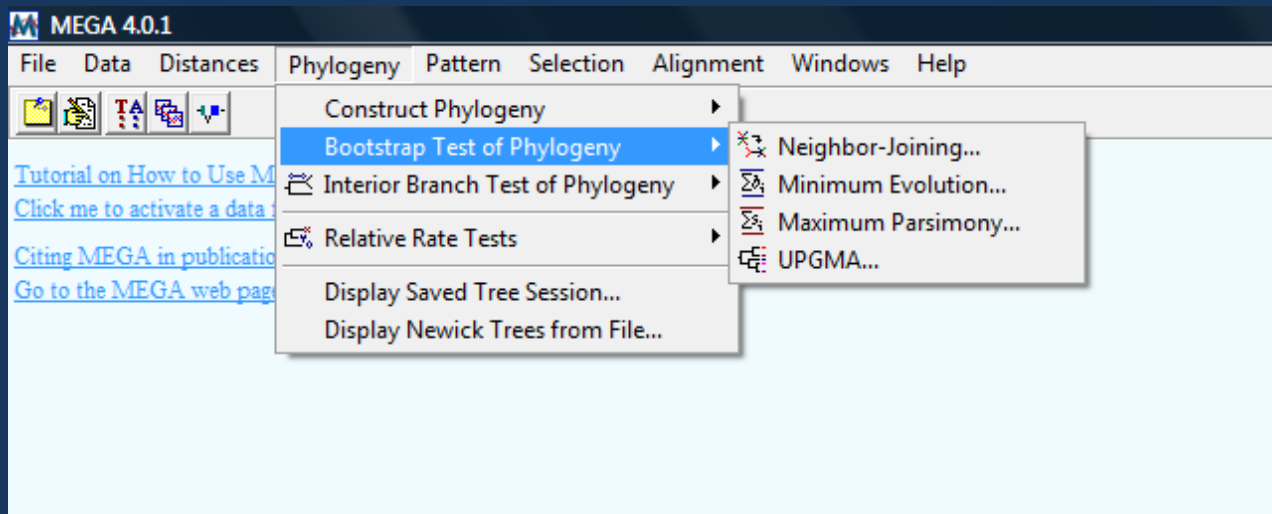
MEGA



O programa MEGA Realiza uma serie de analises evolutivas baseadas em alinhamentos múltiplos de seqüências. O programa no entanto não realiza o alinhamento múltiplo e é necessário importar alinhamentos realizados com outros programas

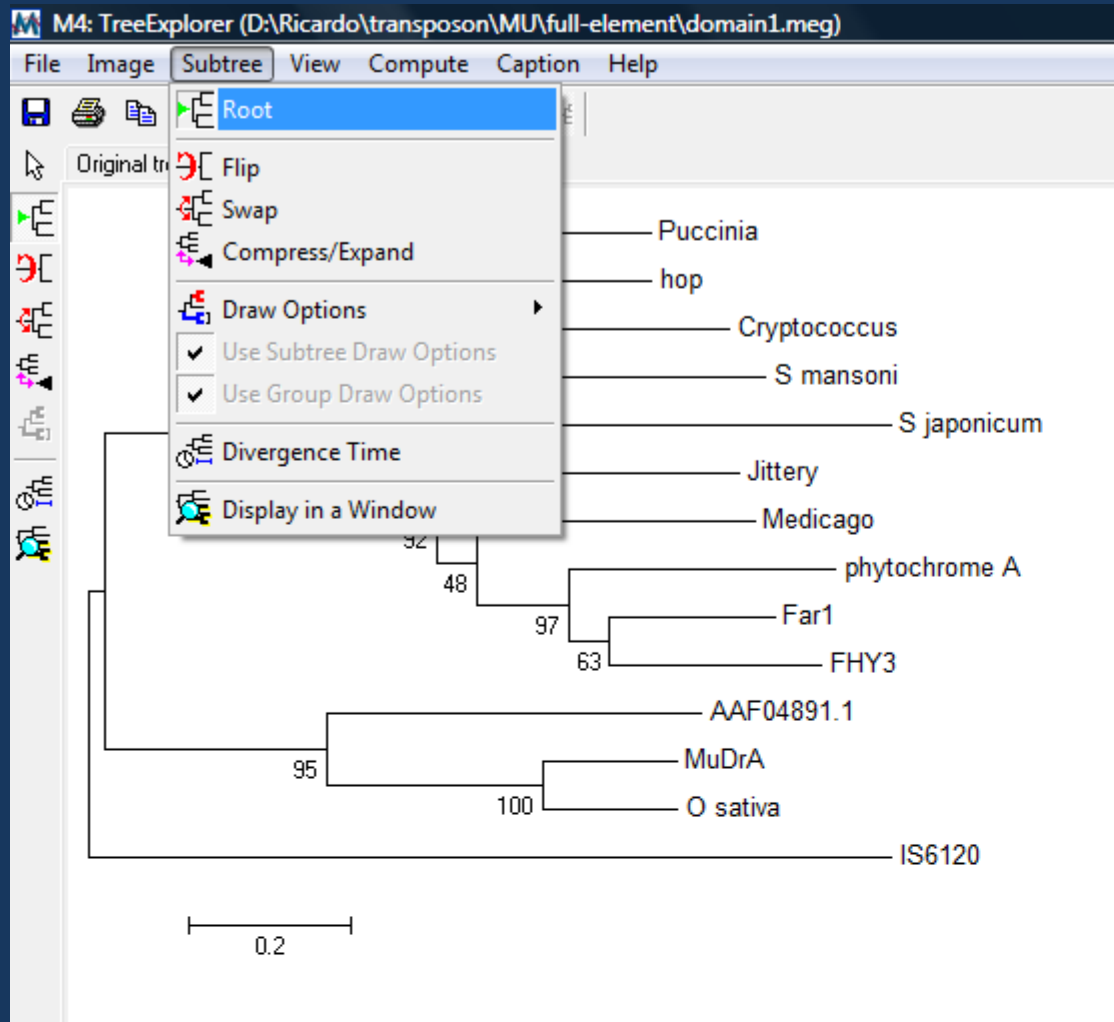
É possível importar seqüências aln resultantes do alinhamento com o clustal

MEGA



Programa permite análise filogenética de alinhamentos por diferentes métodos

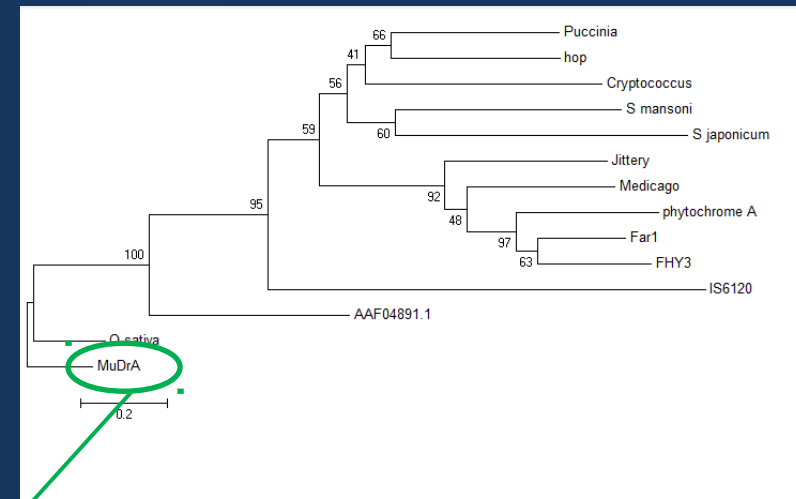
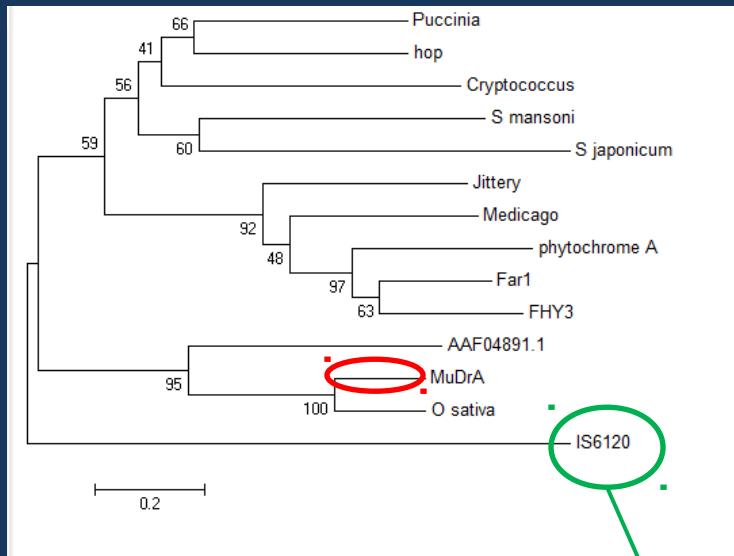
MEGA



- Existe a possibilidade de escolhermos o ramo a partir do qual a árvore ser enraizada com a opção “root”

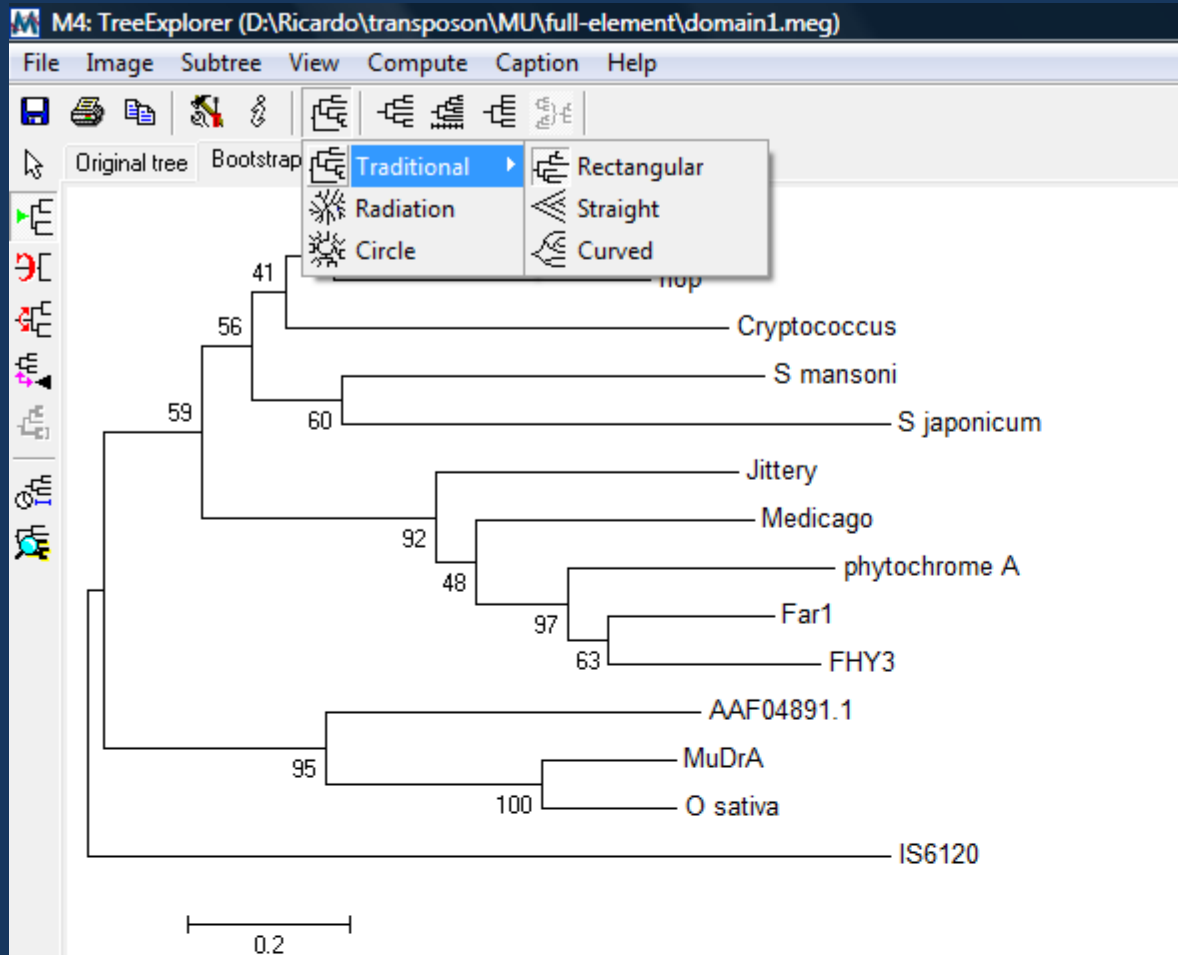
MEGA

Seleção do ramo circulado em vermelho como raiz



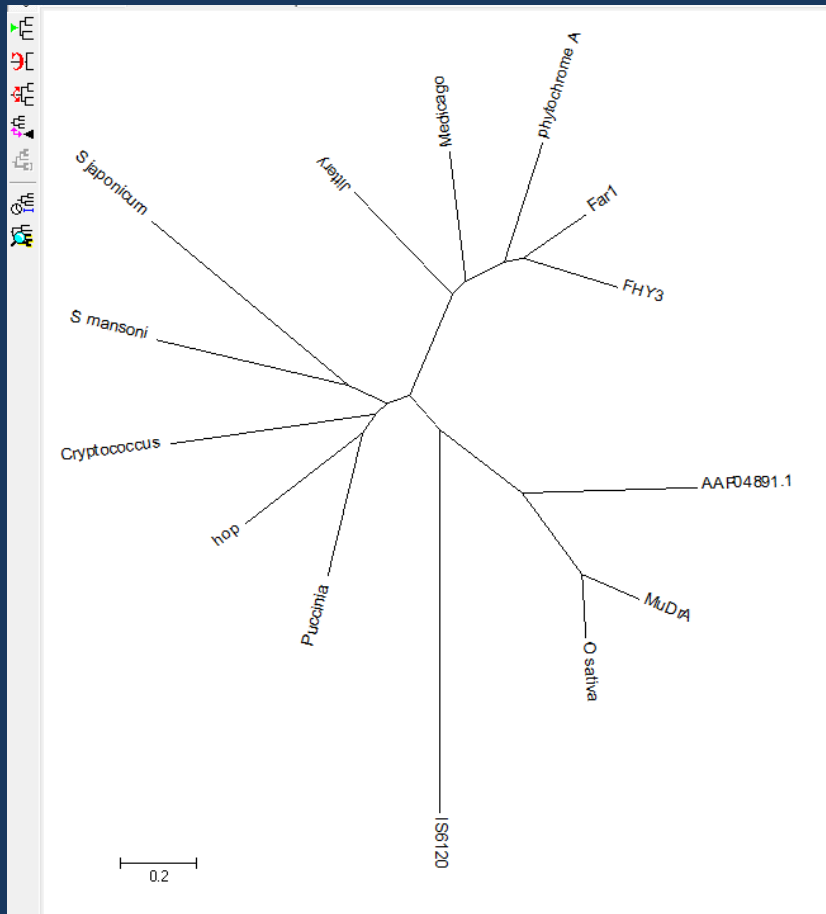
“Outgroup”

MEGA



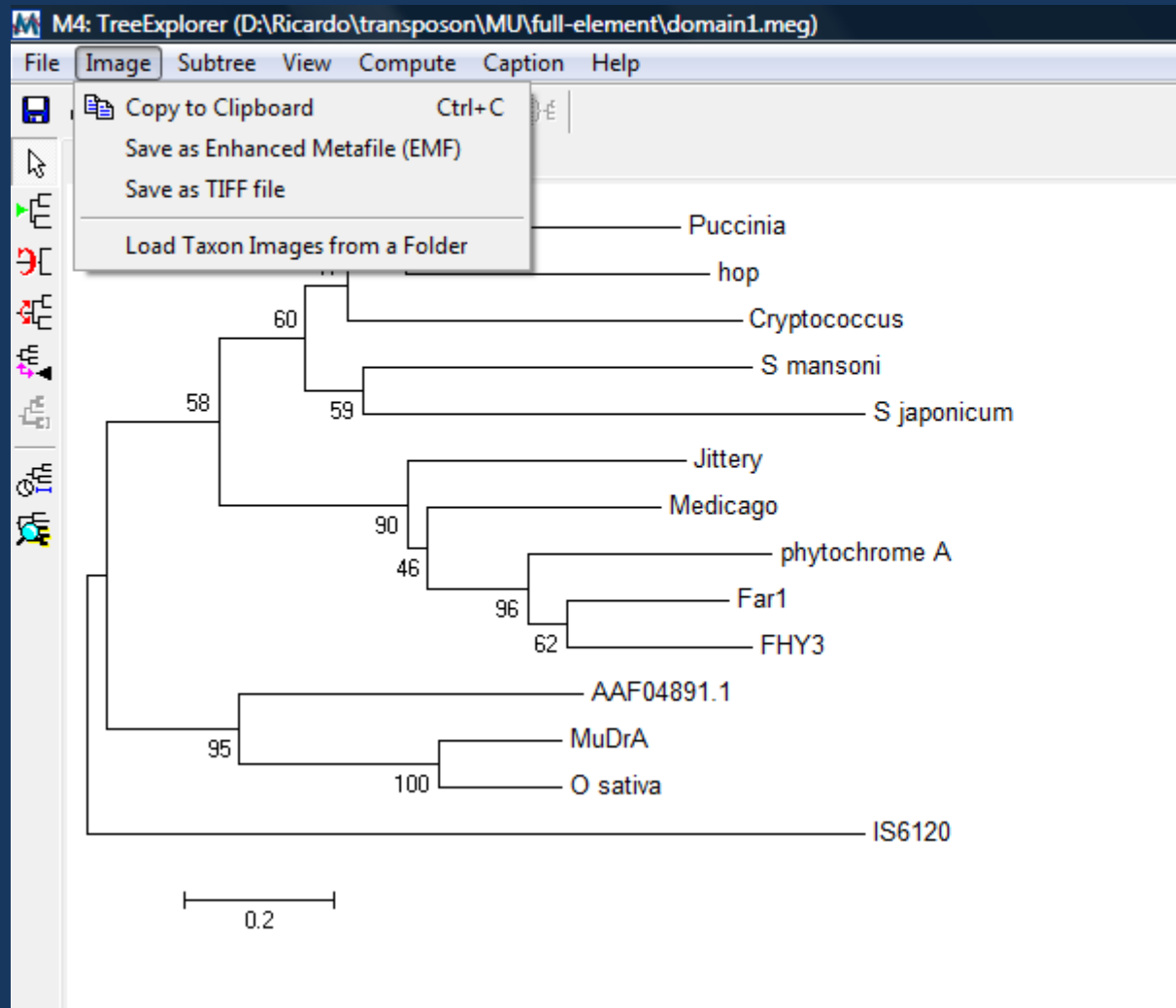
- Possibilidade de escolher entre diversos tipos de representação de árvores

MEGA



- apresentação de árvore em formato radial
- Bom modo de apresentação para árvores sem raiz

MEGA



- É possível exportar a árvore em diversos formatos