0410107 - PRINCÍPIOS DE SISTEMÁTICA E BIOGEOGRAFIA

Instituto de Biociências - Universidade de São Paulo

Prática 5

Exercício 1

Por que topologias de consenso possuem sempre maior número de transformações do que suas topologias fundamentais?

Exercício 2

Considere o diagrama não-enraizado ilustrado na Figura 1

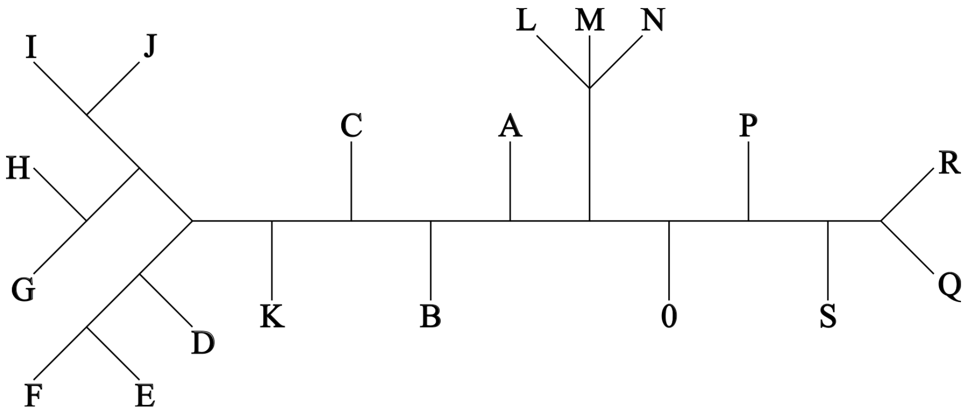


Figura 1: Diagrama não-enraizado para o Exercício 2.

i. Enraize este diagrama utilizando os terminais **J** e **I** como grupos-externos.

ii. Descreva esta topologia em um único parágrafo de forma que sua leitura reproduza a topologia em questão.

Exercício 3

O que comprimento de ramos (internos e externos) representam e como eles podem ser utilizados para inferir tempos absolutos de eventos de cladogênese?

Exercício 4

Considere o cladograma ilustrado na Figura 2.

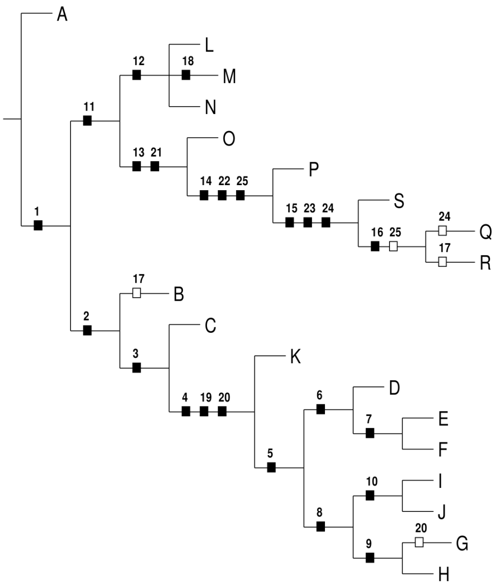


Figura 2: Cladograma para o Exercício 4. Transformações representadas por "■" indicam polarização $0 \to 1$ e aquelas representadas por "□" indicam polarização $1 \to 0$.

i. Indique onde ocorreu uma reversão que sustenta a relação de algum grupo.

ii. Indique onde ocorreu uma reversão autapomórfica.

iii. Qual caráter possui otimização ambígua e qual seria a otimização alternativa?

iv. As relações entre os terminais O, P, Q, R e S podem ser expressas, no que chamamos de notação parentética da seguinte forma: (O(P(S(Q R)))). Usando a mesma lógica, descreva da mesma forma as relações entre os terminais K, D, E, F, G, H, I e J.

Exercício 5

Considere as seguintes notações parentéticas:

```
Topologia 1:  (S(T((U(V X))(Y (W Z)))))
Topologia 2:  (S(T((U V X)((Y W)Z))))
Topologia 3:  (S(T(((U V)X)(Y W Z))))
```

i. Qual seria o consenso estrito entre estas três topologias?

ii. Qual seria o consenso semi-estrito entre as topologias 1 e 2?

Exercício 6

Considere o seguinte modelo de substituição de nucleotídeos,

$$JC = \begin{pmatrix} P_{(AA)} & P_{(AC)} & P_{(AG)} & P_{(AT)} \\ P_{(CA)} & P_{(CC)} & P_{(CG)} & P_{(CT)} \\ P_{(GA)} & P_{(GC)} & P_{(GG)} & P_{(GT)} \\ P_{(TA)} & P_{(TC)} & P_{(TG)} & P_{(TT)} \end{pmatrix} \alpha$$

na qual $P_{(AA)}$ é a probabilidade de não-mudança do estado A, $P_{(AC)}$ é a probabilidade da transformação $A \rightarrow C$ ocorrer, $P_{(AG)}$ é a probabilidade da transformação $A \rightarrow G$ ocorrer, ... *ad nauseam*!

Este modelo, uma simplificação do modelo de Jukes–Cantor (1969)[1] – o primeiro modelo de substituição publicado –, refere-se às probabilidades de substituição ($L_{(i,j)}$) multiplicada pela taxa (α) pela qual qualquer substituição irá ocorrer. Neste exercício, esta taxa de substituição será expressa pelo comprimento dos ramos "v". As Tabelas 1 e 2, abaixo, contém as probalilidades de tranformação de acordo com os repectivos valores de v.

	A	**C**	**G**	**T**
A	0,9064	0,0312	0,0312	0,0312
C	0,0312	0,9064	0,0312	0,0312
G	0,0312	0,0312	0,9064	0,0312
T	0,0312	0,0312	0,0312	0,9064

Tabela 1: Matriz de probabilidades de tranformação para $v = 0.1$.

	A	**C**	**G**	**T**
A	0,9516	0,0161	0,0161	0,0161
C	0,0161	0,9516	0,0161	0,0161
G	0,0161	0,0161	0,9516	0,0161
T	0,0161	0,0161	0,0161	0,9516

Tabela 2: Matriz de probabilidades de tranformação para $v = 0.05$.

[1] Jukes, T.H. & C.R. Cantor. 1969. Evolution of protein molecules. *In*: H.N. Munro (ed.) Mammalian protein metabolism. Academic Press, NY. 21-132 pp.

Com base nestas informações, você deverá calcular a verossimilhança máxima (*Maximum Likelihood*, **ML**) de uma das topologias abaixo. Esse exercício deverá ser feito em dupla, de modo que cada um faça o cálculo para umas das topologias abaixo e vocês possam comparar os resultados. Lembre-se se algumas regras para o cálculo de **ML**:

i. Para cada reconstrução, as probabilidades de substituição $L_{(i,j)}$ no comprimento dos ramos são multiplicadas. Por exemplo, a probabilidade da Reconstrução 1 da Topologia 1 ($P_{R1_{T1}}$) é $P_{(AA|v=0.05)} * P_{(AA|v=0.05)} * P_{(AA|v=0.1)} * P_{(AG|v=0.1)} * P_{(AG|v=0.1)}$, ou seja, $0{,}9516 * 0{,}9516 * 0{,}9064 * 0{,}0312 * 0{,}0312 = 0.000798984$ (veja Tabelas 1 e 2, e a Reconstrução 1 da Topologia 1). Observe que existem componentes probabilísticos idênticos – o que minimiza o número de cálculos necessários.

ii. A **ML** de uma topologia é dada pela soma das **ML**s de cada reconstrução. Por exemplo, a verossimilhança máxima da Topologia 1 (*i.e.*, L_{T1} é a soma das probabilidades das 16 possíveis reconstruções, ou seja, $L_{T1} = P_{R1_{T1}} + P_{R2_{T1}} + P_{R3_{T1}} ... P_{R16_{T1}}$.

Feito os cálculos responda:

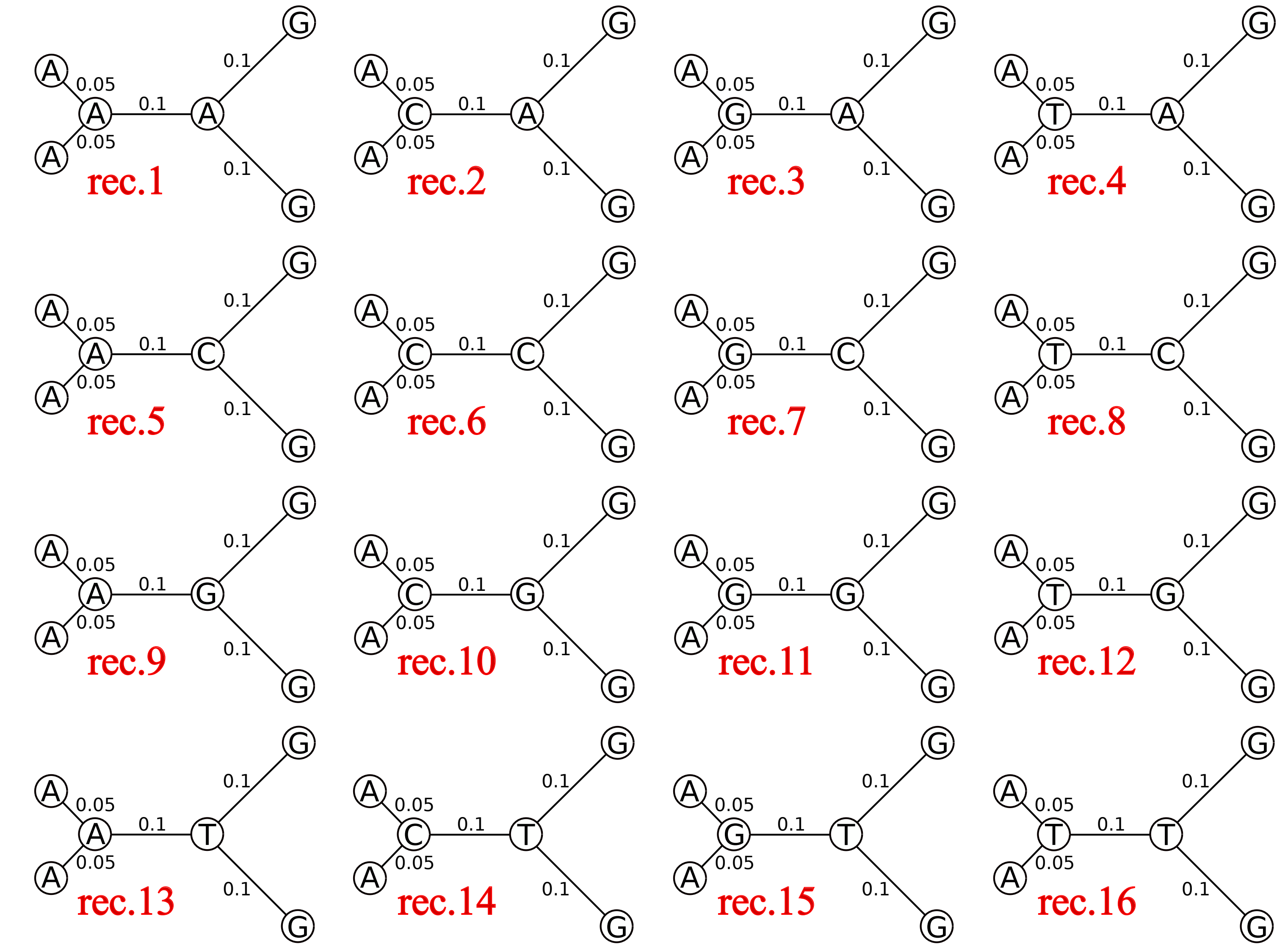
a. Qual das topologias seria escolhida por este critério, *i.e.*, verossimilhança máxima (ou *Maximum Likelihood*)? Justifique.

b. Se você estivesse utilizado como critério de seleção a parcimônia, você teria obtido o mesmo resultado? Justifique.

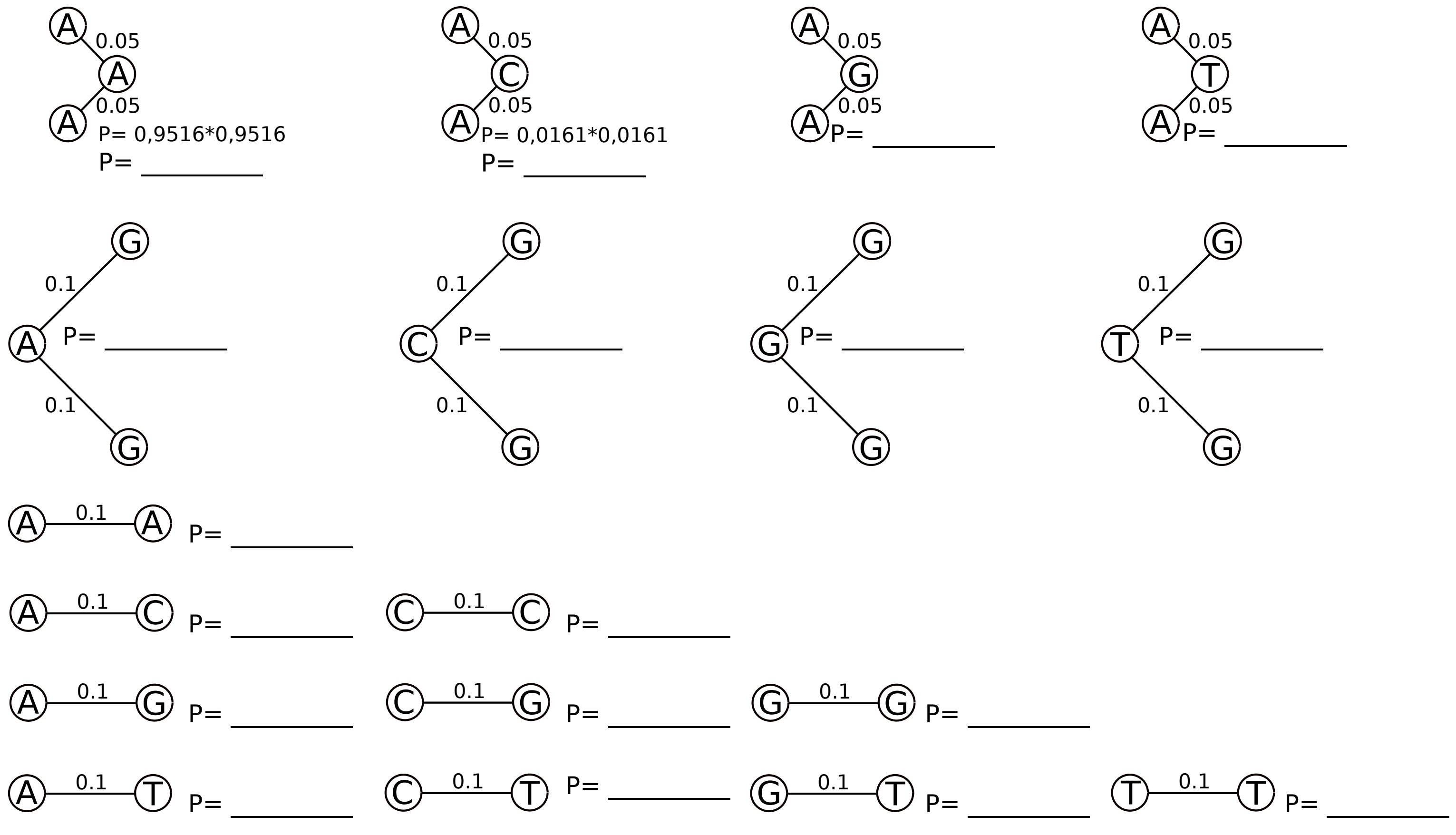
c. Para cada uma das topologias (*i.e.*, T1 e T2) qual(ais) reconstrução(ções) seria(m) selecionadas pelo critério de verossimilhança máxima?

d. Você selecionaria as mesmas reconstruções acima se o seu critério de seleção fosse a parcimômia? Justifique.

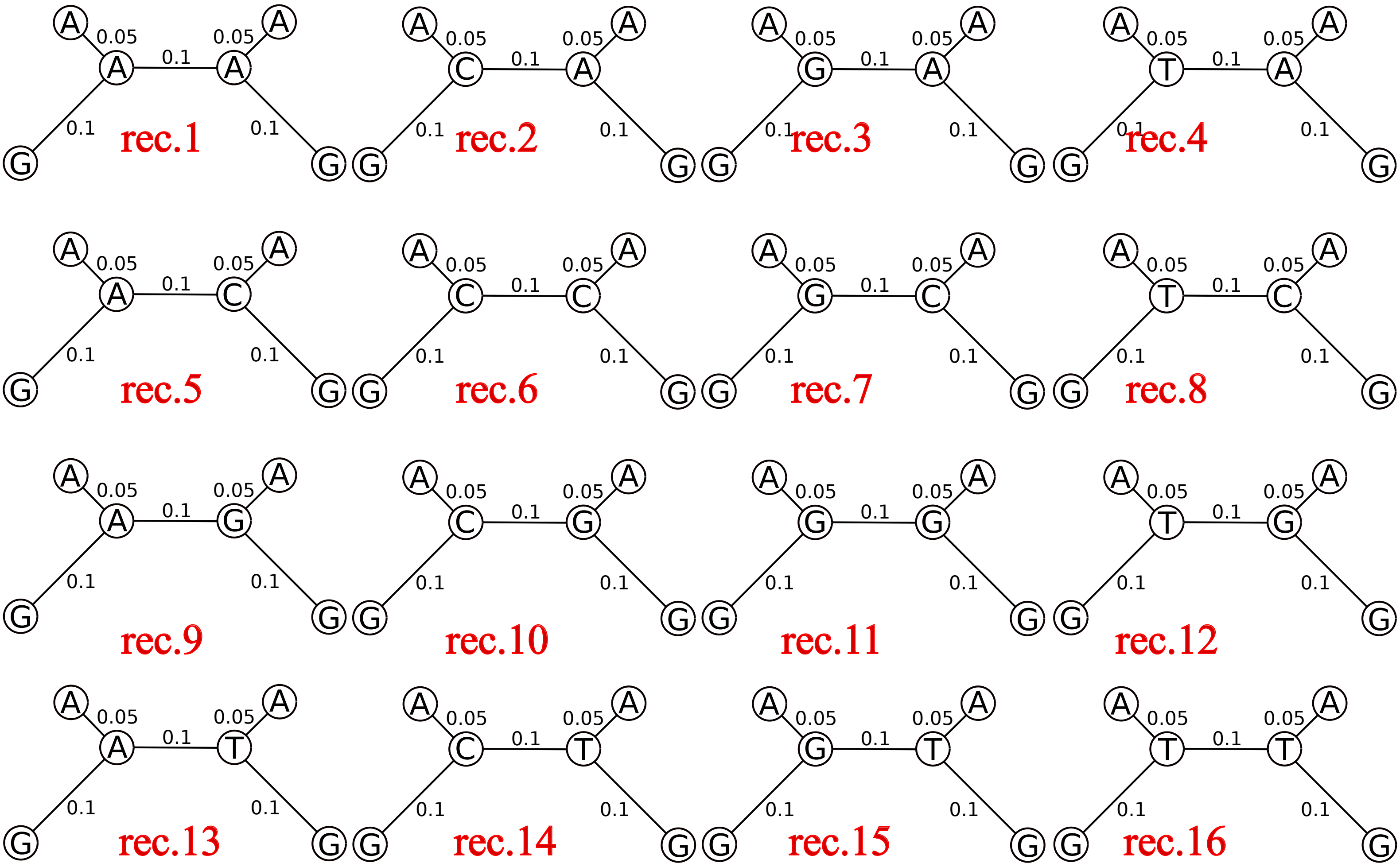
Reconstruções (T1):



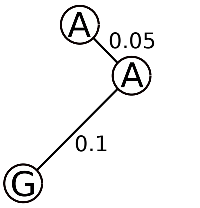
Componentes probabilísticos:



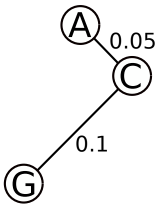
Reconstruções (T2):



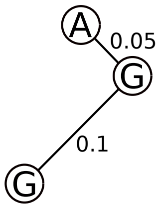
Componentes probabilísticos:



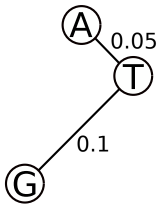
P= 0,9516*0,0312
P= __________



P= 0,0161*0,0312
P= __________



P= __________



P= __________