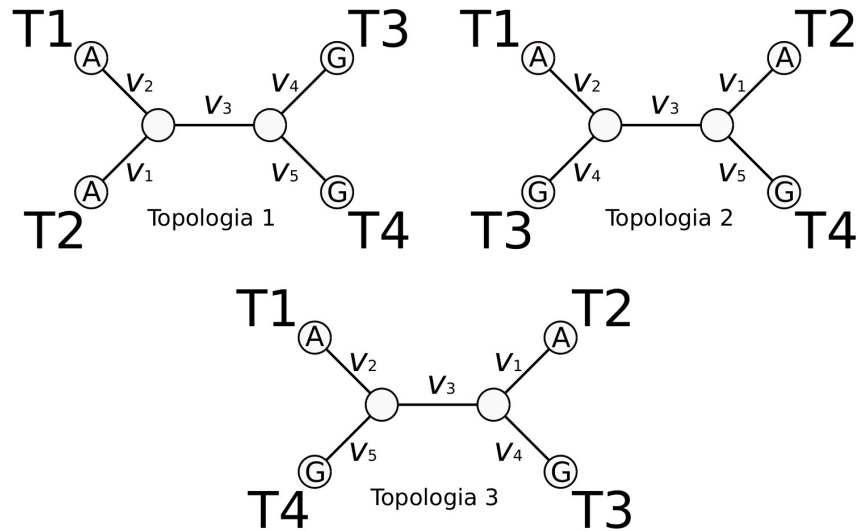


IB-USP 41.107. – PRINCÍPIOS DE SISTEMÁTICA E BIOGEOGRAFIA
Exercício 5 suplementar (ML)

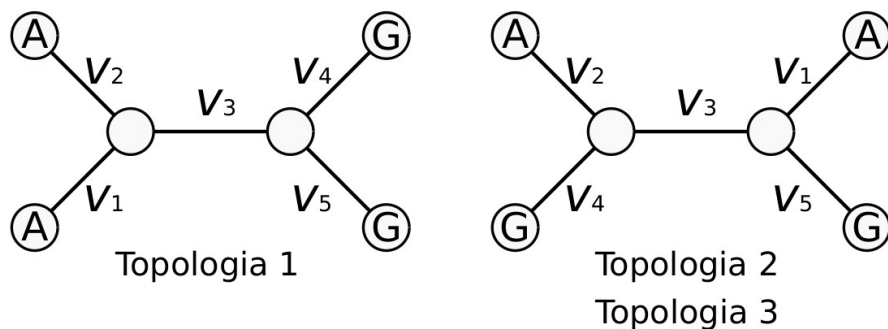
1. As topologias abaixo representam os três diagramas não-enraizados possíveis para 4 terminais. Cada par, i.e., T1 e T2, e T3 e T4, compartilham o mesmo nucleotídeo, A e G, respectivamente, em uma determinada região.



Assuma as seguintes premissas:

- i.* o comprimento de ramo V_1 sempre igual a V_2 .
- ii.* o comprimento de ramo V_4 sempre igual a V_5 .

Sendo verdadeiras essas premissas, embora tenhamos três topologias possíveis para estes quatro terminais, dada a distribuição de caracteres poderíamos considerar somente duas hipóteses, sendo que uma delas representa duas topologias distintas:

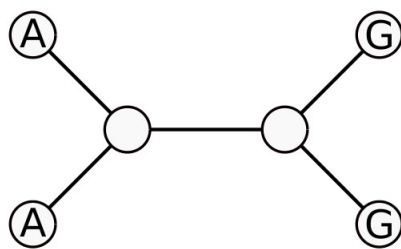


Isso explica porque nos exercícios seguintes iremos apenas comparar duas topologias.

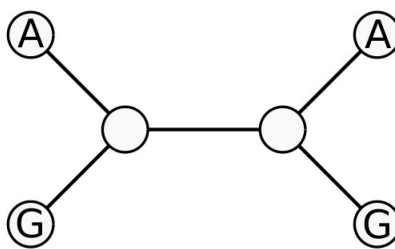
2. Considere a **Topologia 1 e 2** e ilustre (desenhe) as seguintes topologias com seus respectivos comprimento de ramos:

Topologia 1: $v_1 = v_2 = v_3 = v_4 = v_5$	Topologia 2: $v_1 = v_2 = v_3 = v_4 = v_5$
Topologia 1: $v_1 = v_2 = 0,8$, $v_3 = 0,01$ e $v_4 = v_5 = 0,8$.	Topologia 2: $v_1 = v_2 = 0,8$, $v_3 = 0,01$ e $v_4 = v_5 = 0,8$.
Topologia 1: $v_1 = v_2 = 0,05$, $v_3 = 0,1$ e $v_4 = v_5 = 0,1$.	Topologia 2: $v_1 = v_2 = 0,05$, $v_3 = 0,1$ e $v_4 = v_5 = 0,1$.

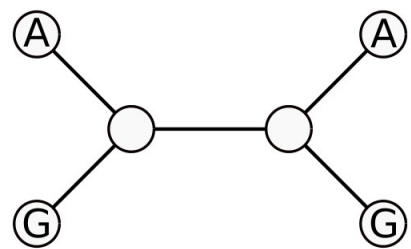
3. De acordo com os diagramas abaixo e os estados representados nos ramos terminais, atribua ao nós os estados (i.e., A, C, G ou T) que resultariam nas otimizações mais parcimoniosas para cada um destes diagramas.



Topologia 1



Topologia 2a



Topologia 2b

4. Considere a Topologia 1 do item 1 deste documento. Se $v_1 = v_2 = 0,95$, $v_3 = 0,87$ e $v_4 = v_5 = 0,80$ e assumirmos o modelo de substituição de Jukes-Cantor (1969), as probabilidades de transformação seriam as seguintes:

Probabilidades para v_1 e v_2				
	A	C	G	T
A	0.5081	0.1640	0.1640	0.1640
C	0.1640	0.5081	0.1640	0.1640
G	0.1640	0.1640	0.5081	0.1640
T	0.1640	0.1640	0.1640	0.5081

Probabilidades para v_4 e v_5				
	A	C	G	T
A	0.4613	0.1640	0.1640	0.1640
C	0.1640	0.4613	0.1640	0.1640
G	0.1640	0.1640	0.4613	0.1640
T	0.1640	0.1640	0.1640	0.5081

Probabilidades para v_3				
	A	C	G	T
A	0.48511	0.17163	0.17163	0.17163
C	0.17163	0.48511	0.17163	0.17163
G	0.17163	0.17163	0.48511	0.17163
T	0.17163	0.17163	0.17163	0.48511

- i. Qual seria a probabilidade desta topologia se fosse atribuído o estado “A” para o nó compartilhado por T1 e T2 e “G” para o nó de T3 e T4?

5. Junto com estes exercícios você baixou uma planilha eletrônica. Esta planilha deverá ser utilizada para fazer os próximos exercícios. Ela deverá economizar muito de seu tempo, pois como já deve saber, o cálculo manual das probabilidades utilizadas para selecionar topologias pelo critério de Máxima Verossimilhança é MUITO chato! Abra este arquivo em Excel ou OpenOffice, vamos examinar seu conteúdo para que você entenda como as planilhas estão organizadas. Este documento contém três planilhas internas (i.e., *sheets*), com o seguinte conteúdo:

i. A primeira, chamada “probabilidades vs -ln”, é simplesmente um gráfico ilustrando a relação entre probabilidades, que variam de 0 a 1, e o logaritmo natural (ou neperiano) negativo destas probabilidades. Este varia de 0 a ∞ . Quanto maior a probabilidade, menor é seu -ln.

ii. A segunda, chamada “Topologia 1”, calcula as probabilidades de todas as reconstruções possíveis da Topologia 1 do item 1. Esta é a planilha que deverá ser utilizada para os cálculos que serão solicitados a seguir. Algumas informações importantes:

a. No canto esquerdo superior desta planilha há uma pequena tabela com os comprimentos de ramos. Como existe uma relação de igualdade assumida entre estes comprimentos de ramos, **somente aqueles que estão marcados em vermelho deverão ser editados**. Os demais estão ligados a estes valores por *links* internos da planilha.

b. Ao inserir os valores de comprimento de ramos, automaticamente será feita a

atualização das tabelas abaixo na qual se encontram as probabilidades de transformações para cada estado de acordo com o comprimento do ramo. Adicionalmente, o cálculo de probabilidades para cada reconstrução (i.e., de 1 a 16, veja mapa de reconstruções em anexo) é executado e o valor final da probabilidade da topologia é compilado.

c. Finalmente, há uma tabela comparativa exibindo as probabilidades e seus logaritmos naturais (ou neperiano) negativos entre Topologias 1 e 2(3).

ii. A terceira planilha interna, chamada “Topologia 2 e 3”, calcula as probabilidades da Topologia 2 (e 3) da mesma forma que na planilha anterior. No entanto, os valores de comprimento de ramos estão ligados àqueles da planilha anterior. Portanto, você **não deverá editar nenhum número nesta planilha** e apenas irá utilizá-la para examinar os resultados.

A seguir você deverá utilizar os mesmos conjuntos de comprimento de ramos que você ilustrou no item 2. Note que estes comprimentos de ramos fazem parte de seu modelo, lembrando que o que você está calculando é a máxima verossimilhança (L) de seu modelo dada sua observação. Dito de outra forma, você está calculando a probabilidade de obter seus dados uma vez que seu modelo seja verdadeiro. Seu modelo, por sua vez, está decomposto em um modelo substituição + a topologia. Uma vez que você usará o mesmo modelo de substituição (i.e., JC69) seus modelos dependem apenas das topologias, incluindo seus comprimento de ramos. Assim teremos:

Modelo 1: $v_1 = v_2 = v_3 = v_4 = v_5 = 1$.

Modelo 2: $v_1 = v_2 = 0,8$, $v_3 = 0,01$ e $v_4 = v_5 = 0,8$.

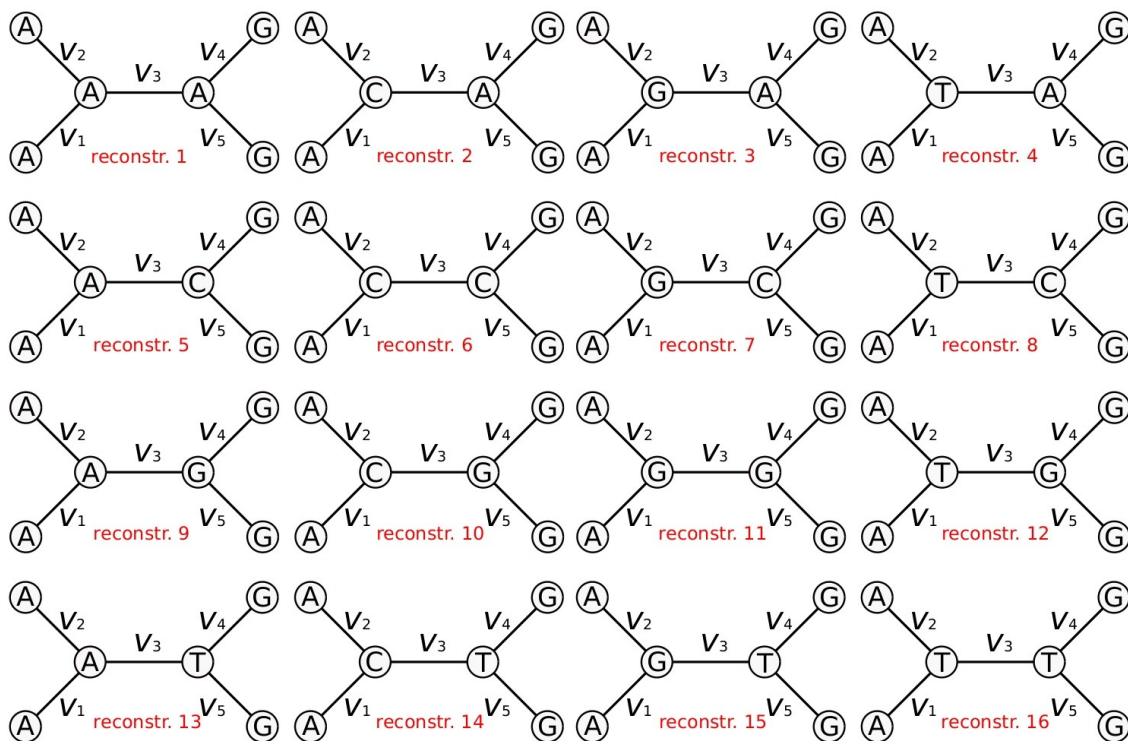
Modelo 3: $v_1 = v_2 = 0,05$, $v_3 = 0,1$ e $v_4 = v_5 = 0,1$.

Atribua os valores correspondentes a cada modelo na planilha em anexo e para cada um deles obtenha as informações que julgar necessárias para responder às seguintes perguntas:

1. Qual topologia é selecionada em cada modelo?
2. Qual(ais) reconstrução(ões) é(são) selecionada(s) em cada modelo e topologia?
3. A seleção de topologias e reconstruções difere em relação aos modelos? Explique.
4. Há congruência de resultados entre o critério de parcimônia e o de máxima verossimilhança?
5. Qual é a relação entre o comprimento relativo dos ramos e a topologia selecionada pelo critério de máxima verossimilhança?

MAPA DE RECONSTRUÇÕES

Topologia 1



Topologia 2

