

Aula 7: Filogenética

Instrutor: Berilhes Borges Garcia

Escriba: Rafael Gabeira Cola

DRAFT**1 Introdução**

Características comportamentais e anatômicas foram os critérios dominantes utilizados para derivar o relacionamento entre as espécies desde Darwin. Equipados com análises baseadas nestas observações relativamente subjetivas, os relacionamentos evolutivos foram freqüentemente inconclusivos ou equivocados.

Por exemplo, durante aproximadamente 100 anos os cientistas não conseguiram determinar a qual família o panda gigante pertencia. Em 1985, Steve O'Brien et al resolveu o problema da classificação do panda gigante utilizando algoritmos e seqüências de DNA.

A aproximadamente 20 anos que a construção de árvores filogenéticas a partir de DNA tem sido extensamente pesquisada e debatida. No entanto as raízes desta pesquisa podem ser traçadas ao trabalho seminal de Emile Zuckerkand e Linus Pauling.

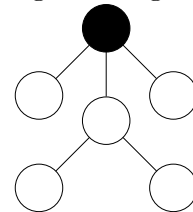
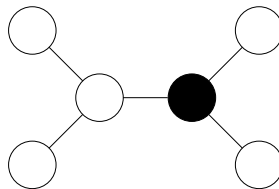
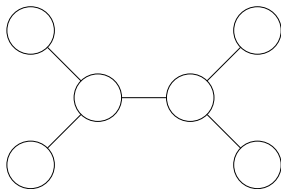
Ao mesmo tempo que o quebra-cabeça do urso panda foi resolvido, um modelo baseado em DNA da árvore evolutiva humana conduziu à chamada hipótese fora da África. Segundo esta hipótese nossos ancestrais viveram na África a aproximadamente 200.000 anos.

Como são estas árvores construídas a partir de seqüências?

Vamos começar com alguns fundamentos.

- Árvores Enraizadas
 - Infere um ancestral evolutivo
 - As folhas representam as espécies existentes
 - Vértices internos representam ancestrais hipotéticos
 - Estas árvores podem ser vistas como árvores dirigidas da raiz para as folhas
- Árvores Não Enraizadas

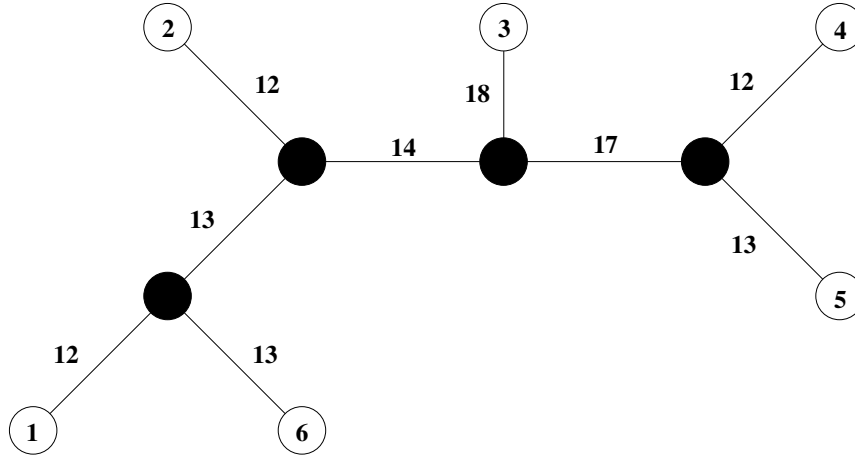
Não inferem um ancestral evolutivo, portanto não podem ser vistas como um grafo dirigido.



As arestas podem ter peso, os quais refletem:

- O número de mutações no caminho evolutivo de uma espécie para outra, ou
- O tempo estimado para a evolução de uma espécie em outra.

Em uma árvore T com n folhas, ns freqüentemente computamos o tamanho de um caminho entre as folhas i e j , $d_{ij}(T)$. d_{ij} refere-se distância entre i e j e é a soma dos pesos das arestas entre i e j .



$$d_{14} = 12 + 13 + 14 + 17 + 12 = 68$$

Dado n espécies, nós podemos computar a matriz de distância D_{ij} .

Existe uma árvore correspondente, mas nós não conhecemos a árvore. Com tal matriz nós necessitamos um método para projetar uma árvore que se ajusta a esta matriz de distância D_{ij} .

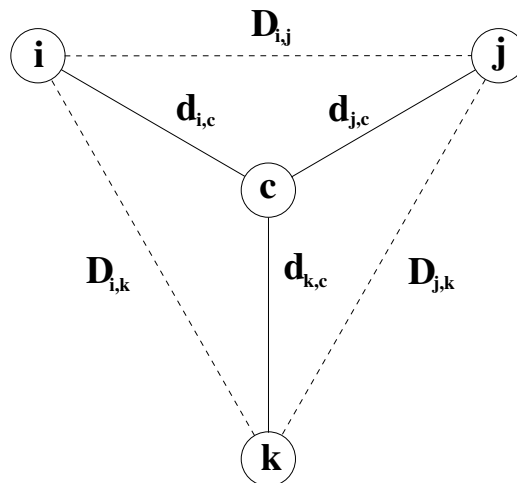
Ajustar significa

$$D_{ij} = d_{ij}(T) \tag{1}$$

onde D_{ij} é a distância entre as espécies e $d_{ij}(T)$ é o tamanho dos caminhos na árvore.

D_{ij} pode ser definido como a distância de edição entre um gene na espécie i e espécie j , onde o gene de interesse é sequenciado para todas as n espécies.

A reconstrução da árvore para qualquer matriz 3×3 é direta. Nós temos folhas i , j e k e um vértice central c .



Observe que

$$d_{ic} + d_{jc} = D_{ij}$$

$$d_{ic} + d_{kc} = D_{ik}$$

$$d_{jc} + d_{kc} = D_{jk}$$

$$\begin{aligned}d_{ic} + d_{jc} &= D_{ij} \\+d_{ic} + d_{kc} &= D_{ik} \\2d_{ic} + \underbrace{d_{jc} + d_{kc}} &= D_{ij} + D_{ik} \\2d_{ic} + D_{jk} &= D_{ij} + D_{ik} \\d_{ic} &= \frac{D_{ij} + D_{ik} - D_{jk}}{2}\end{aligned}$$

Similarmente

$$d_{jc} = \frac{D_{ij} + D_{jk} - D_{ik}}{2}$$

$$d_{kc} = \frac{D_{ki} + D_{kj} - D_{ij}}{2}$$