

TUTORIAL - UPGMA

(Unweighted Pair-Group Method using arithmetic Averages)

Existe uma série de algoritmos que lhe permite medir a semelhança ou diferença entre as OTUs e derivar um fenograma das matrizes de distâncias. Alguns medem a similaridade entre OTUs – coeficientes de similaridade, e os valores são expressos em porcentagem. Outros medem a distância – coeficientes de distância, que podem ser calculadas de diversas maneiras. Por exemplo, a distância Manhattan considera valores absolutos. A distância euclidiana é a média entre as OTUs situadas nos vértices de um triângulo. Uma vez calculada as semelhanças ou diferenças entre as OTUs, esses valores podem ser utilizados para gerar agrupamentos (i.e., fenogramas) de diversas maneiras. Uma delas seria por ligação média sem ponderação (pesagem) por um algoritmo conhecido como UPGMA (*Unweighted Pair-Group Method using arithmetic Averages*). Outra possibilidade seria calcular estes fenogramas utilizando a ligação média com ponderação – WPGMA (*Weighted Pair-Group Method using arithmetic Averages*). Finalmente, poderíamos ainda fazer o mesmo por ligação simples (*simple linkage*).

No exemplo abaixo iremos utilizar o algoritmo de UPGMA e as distâncias serão consideradas em seus valores absolutos (distância Manhattan).

O cálculo desta distância é feita aos pares entre todas as OTUs. Por exemplo, considere duas OTUs com as seguintes sequências nucleotídicas:

```
A      ACGTACGTACGTACGTACGT
      * * * * * * * * * *
B      CCCTCCCTCCCTCCCTCCCC
```

Teríamos que a distância entre A e B é 11 (*s). Com base nestas comparações, construímos então a matriz inicial de distâncias Manhattan e teríamos a matriz da Tabela 1.

| | A | B | C | D |
|---|----|----|---|---|
| A | – | | | |
| B | 11 | – | | |
| C | 14 | 9 | – | |
| D | 15 | 10 | 5 | – |

Tabela 1: Matriz de distância.

Com base no menor valor encontrado na matriz (5) (menor distância = menor diferença), constroi-se o primeiro núcleo (ou componente), C/D. Para a construção do fenograma, divide-se o valor por 2; ou seja $D_{(C,D)} = \frac{5}{2} = 2,5$. Esse valor é colocado no eixo de distância, que acompanha o fenograma em elaboração (Figurea 1).

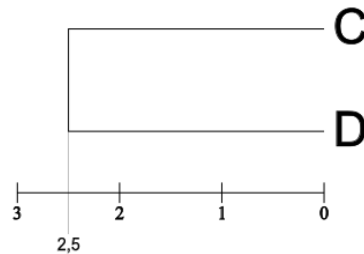


Figura 1: Fenograma parcial para o componete C/D.

A seguir, constroi-se uma nova matriz, considerando-se o núcleo C/D.

Para encontrar os valores da matriz, calcula-se a distância do núcleo C/D para a as outras OTUs. Devem ser considerados os valores existentes na matriz inicial:

$$D_{(C/D,A)} = \frac{D_{(C,A)}+D_{(D,A)}}{2} = \frac{14+15}{2} = 14,5$$

$$D_{(C/D,B)} = \frac{D_{(C,B)}+D_{(D,B)}}{2} = \frac{9+10}{2} = 9,5$$

$$D_{(A,B)} = 11^1$$

De acordo com as distâncias calculadas a matriz na Tabela 2.

| | C/D | A | B |
|-----|------------|----|---|
| C/D | – | | |
| A | 14,5 | – | |
| B | 9,5 | 11 | – |

Tabela 2: Matriz com as distâncias recalculadas.

Como o menor valor obtido é 9,5, B é a nova OTU que vai ser ligada ao núcleo C/D. A distância será igual a $\frac{9,5}{2} = 4,75$ (veja Figura 2).

¹ observado diretamente na matriz da Tabela 1

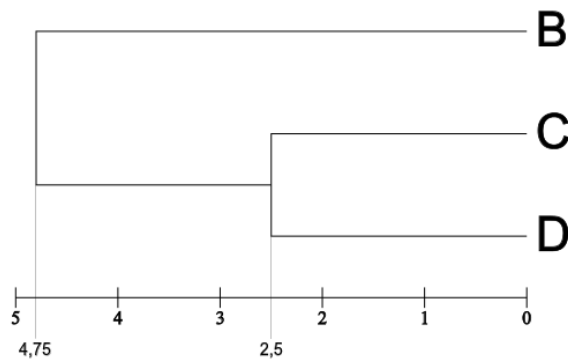


Figura 2: Fenograma parcial para o componetes C/D e C/D/B.

O núcleo C/D/B está formado. Como só resta uma OTU (*i.e.*, A), basta calcular a distância entre A e o núcleo C/D/B:

$$D_{(B/C/D),A} = \frac{D_{(B,A)} + D_{(C,A)} + D_{(D,A)}}{3} = \frac{11 + 14 + 15}{3} = \frac{40}{3} = 13,33$$

A inserção de este último terminal resulta no fenograma ilustrado na Figura 3. A distância reflete a similaridade entre duas OTUs. Se possuírem uma distância zero, isso significa que as OTUs são iguais. Quanto maior for a distância que separa duas OTUs, menos similares elas serão.

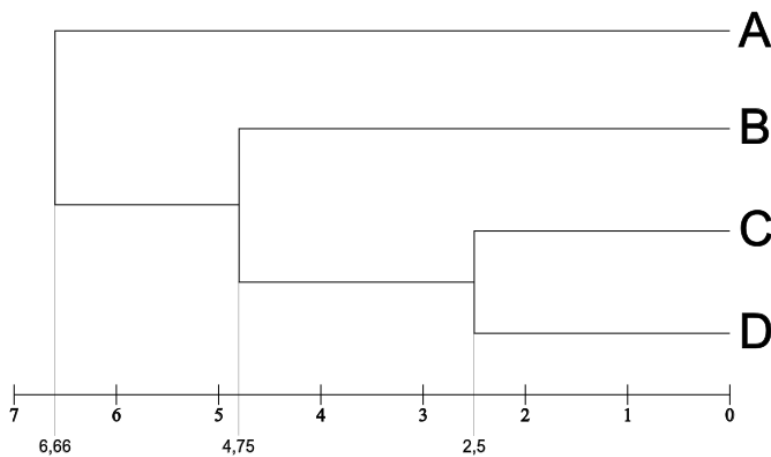


Figura 3: Fenograma final gerado por UPGMA de acordo com a matriz de distância da Tabela 1.

