

Exercício 2

1. Abaixo desenhe um diagrama binário não direcionado e nomeie seus componentes.

2. Quais são as três etapas principais envolvidas no processo de inferência filogenética?

3. O que é critério de otimização e qual é a sua função em inferência filogenética?

4. O que é Fenética?

5. Qual é a relação existente entre a resolução de diagramas binários e teste de hipóteses?

6. Os únicos atributos passíveis de serem considerados em inferência filogenética são aqueles que consideramos hereditários. Por que essa condição é necessária e suficiente?

7. Considere as seguintes sequências nucleotídicas para 5 espécies de moscas:

Z TAGAGCAATCCCTAACTG-AA
A TAGAGCA-TCGCTA-CTA-AA
B TAGAGCAATCACTAACTG-GA
C TAGAGCTGTCTCTAACAGA--
D TAGAGCTATCACAAACAGAAA

a. Calcule a matriz de distância entre as **OTUs**:

Z TAGAGCAATCCCTAACTG-AA
 A TAGAGCA-TCGCTA-CTA-AA $D_{(Z,A)} = \underline{\hspace{2cm}}$

Z TAGAGCAATCCCTAACTG-AA
 B TAGAGCAATCACTAACTG-GA $D_{(Z,B)} = \underline{\hspace{2cm}}$

Z TAGAGCAATCCCTAACTG-AA
 C TAGAGCTGTCTCTAACAGA-- $D_{(Z,C)} = \underline{\hspace{2cm}}$

Z TAGAGCAATCCCTAACTG-AA
 D TAGAGCTATCACAAACAGAAA $D_{(Z,D)} = \underline{\hspace{2cm}}$

A TAGAGCA-TCGCTA-CTA-AA
 B TAGAGCAATCACTAACTG-GA $D_{(A,B)} = \underline{\hspace{2cm}}$

A TAGAGCA-TCGCTA-CTA-AA
 C TAGAGCTGTCTCTAACAGA-- $D_{(A,C)} = \underline{\hspace{2cm}}$

A TAGAGCA-TCGCTA-CTA-AA
 D TAGAGCTATCACAAACAGAAA $D_{(A,D)} = \underline{\hspace{2cm}}$

B TAGAGCAATCACTAACTG-GA
 C TAGAGCTGTCTCTAACAGA-- $D_{(B,C)} = \underline{\hspace{2cm}}$

B TAGAGCAATCACTAACTG-GA
 D TAGAGCTATCACAAACAGAAA $D_{(B,D)} = \underline{\hspace{2cm}}$

C TAGAGCTGTCTCTAACAGA--
 D TAGAGCTATCACAAACAGAAA $D_{(C,D)} = \underline{\hspace{2cm}}$

	Z	A	B	C	D
Z	-				
A		-			
B			-		
C				-	
D					-

b. Com base nestes dados, utilize o método de ligação média (UPGMA - *Unweighted Pair Group Method with Arithmetic Mean* – método descrito no final deste exercício) para a elaboração de um fenograma.

8. Considere a morfologia das OTUs para as quais você acabou de construir um fenograma baseado em dados moleculares (imagens coloridas disponíveis na sala de aula):

a. Faça uma lista de características morfológicas para as espécies **Z-D** e, em uma tabela, atribua como essa característica se apresenta para cada OTU (e.g., presença/ausência, forma A/ forma B, cor X/cor Y ... etc).

b. A partir destes dados, construa uma matriz de distâncias e elabore um fenograma utilizando o método do UPGMA.

c. Compare os fenogramas obtidos a partir de dados moleculares e morfológicos e responda:

i. Por que as matrizes de distância diferem entre as duas fontes de dados?

ii. Por que o fenograma difere da história evolutiva do grupo?

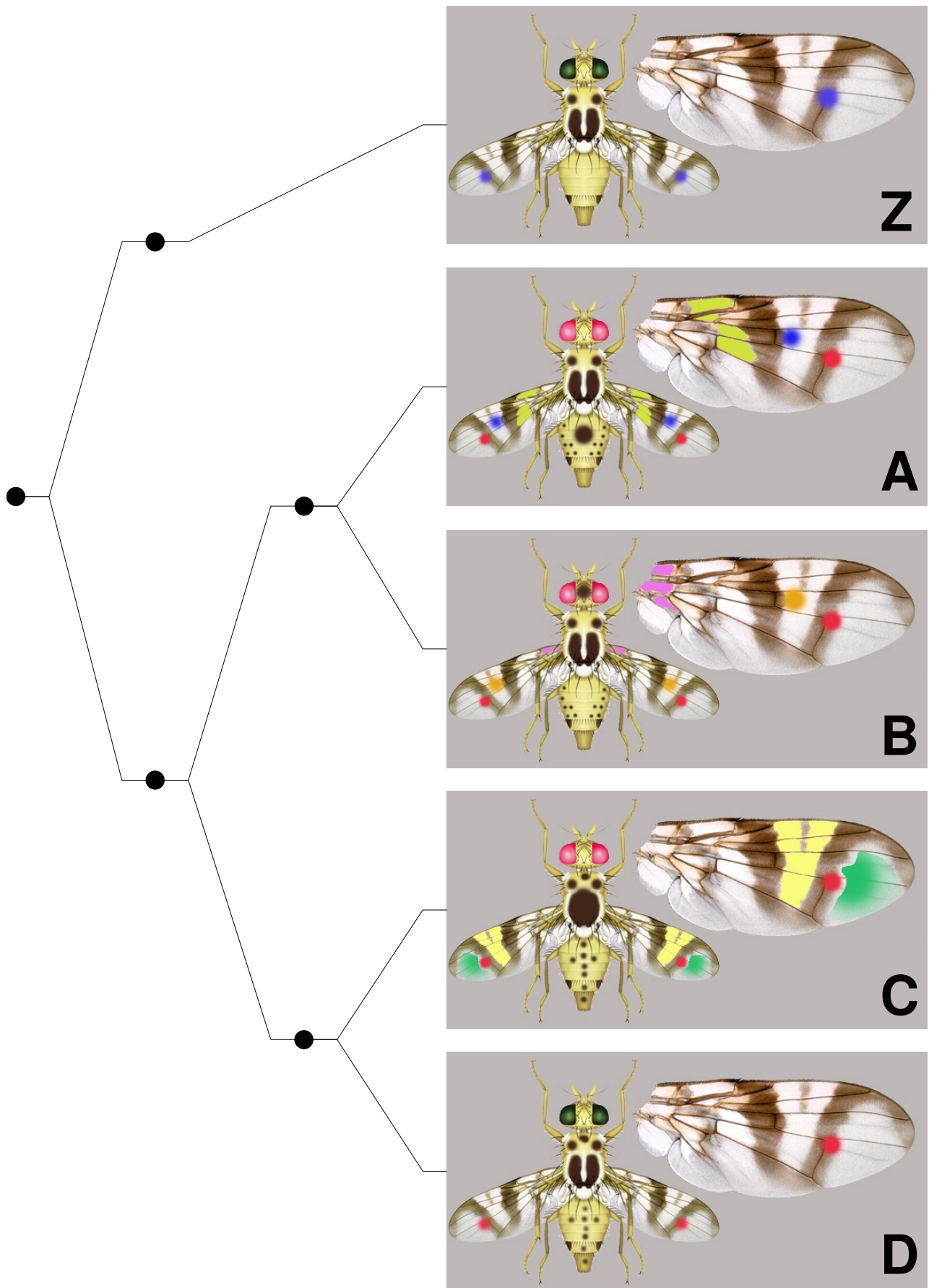


Figura 1. Relações filogenéticas entre Z, A, B, C e D.

UPGMA (*Unweighted Pair-Group Method using arithmetic Averages*)

Para a construção do fenograma, vários algoritmos são disponíveis para medir a semelhança ou diferença entre as OTUs. Alguns medem a similaridade (coeficientes de similaridade), e os valores são expressos em porcentagem. Outros medem a distância (coeficientes de distância), que pode ser calculada de diversas maneiras. Por exemplo, a distância Manhattan considera valores absolutos; já a distância euclidiana calcula a média entre as OTUs situadas nos vértices de um triângulo. Esses valores podem ser agrupados de diversas formas. Por exemplo: por ligação média sem peso ou UPGMA (*Unweighted Pair-Group Method using arithmetic Averages*); por ligação média com peso ou WPGMA (*Weighted Pair-Group Method using arithmetic Averages*); por ligação simples (*simple linkage*).

No exemplo que segue, será utilizada ligação por 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:

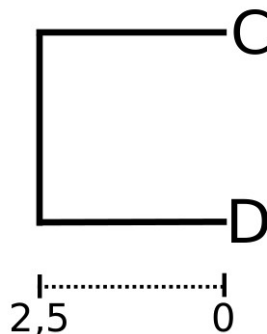
```

A   ACGTACGTACGTACGTACGT
    * * * * * * * * * *
B   CCCTCCCTCCCTCCCTCCCT
  
```

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 abaixo:

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

Com base no menor valor encontrado na matriz (**5**) (menor distância = menor diferença), constrói-se o primeiro núcleo, **C/D**. Para a construção do fenograma, divide-se o valor por 2; $5/2 = 2,5$. Esse valor é colocado no eixo de distância, que acompanha o fenograma em elaboração.



A seguir, constrói-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 as outras OTUs. Devem ser considerados os valores existentes na matriz inicial.

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

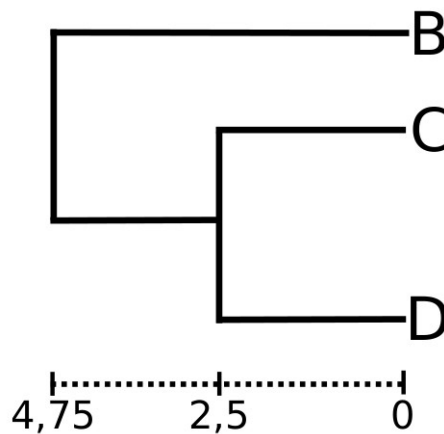
$$D_{(C/D), B} = D_{(C,B)} + D_{(D,B)} / 2 = 9 + 10 / 2 = 9,5.$$

$$D_{(A,B)} = 11 \text{ (observado diretamente na matriz acima).}$$

Nova Matriz de distâncias:

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

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 $9,5/2 = 4,75$

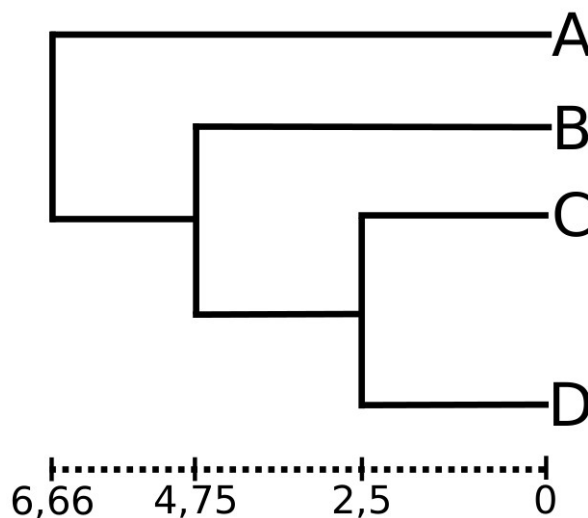


O núcleo **C/D/B** está formado. Como só resta uma OTU, **A**, basta calcular a distância entre **A** e o núcleo **C/D/B**.

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

e $13,33/2 = \mathbf{6,66}$

Abaixo está fenograma final, com seu eixo de distâncias.



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.