

# Computational Biology

Lecturers:

Tanja Stadler, Carsten Magnus & Tim Vaughan

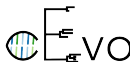
Teaching Assistants:

Jūlija Pečerska, Jérémie Sciré,  
Sarah Nadeau & Marc Manceau

Computational Evolution

Department of Biosystems Science and Engineering

HS 2019



Tree reconstruction

Tree reconstruction

Distance matrices

Hamming distance

JC69 distance

K80 distance

UPGMA algorithm

Exercise

Solution

In reality we observe the sequences at present and the tree is unknown.

[Tree reconstruction](#)[Tree reconstruction](#)[Distance matrices](#)[Hamming distance](#)[JC69 distance](#)[K80 distance](#)[UPGMA algorithm](#)[Exercise](#)[Solution](#)

In reality we observe the sequences at present and the tree is unknown.

How do we build a tree from an alignment?

[Tree reconstruction](#)[Tree reconstruction](#)[Distance matrices](#)[Hamming distance](#)[JC69 distance](#)[K80 distance](#)[UPGMA algorithm](#)[Exercise](#)[Solution](#)

In reality we observe the sequences at present and the tree is unknown.

How do we build a tree from an alignment?

This time we will use UPGMA

[Tree reconstruction](#)[Tree reconstruction](#)[Distance matrices](#)[Hamming distance](#)[JC69 distance](#)[K80 distance](#)[UPGMA algorithm](#)[Exercise](#)[Solution](#)

In reality we observe the sequences at present and the tree is unknown.

How do we build a tree from an alignment?

This time we will use UPGMA:

1. Build the initial distance matrix;

[Tree reconstruction](#)[Tree reconstruction](#)[Distance matrices](#)[Hamming distance](#)[JC69 distance](#)[K80 distance](#)[UPGMA algorithm](#)[Exercise](#)[Solution](#)

In reality we observe the sequences at present and the tree is unknown.

How do we build a tree from an alignment?

This time we will use UPGMA:

1. Build the initial distance matrix;
2. Repeat until all sequences are clustered:

[Tree reconstruction](#)[Tree reconstruction](#)[Distance matrices](#)[Hamming distance](#)[JC69 distance](#)[K80 distance](#)[UPGMA algorithm](#)[Exercise](#)[Solution](#)

In reality we observe the sequences at present and the tree is unknown.

How do we build a tree from an alignment?

This time we will use UPGMA:

1. Build the initial distance matrix;
2. Repeat until all sequences are clustered:
  - 2.1 Cluster two closest nodes;

[Tree reconstruction](#)[Tree reconstruction](#)[Distance matrices](#)[Hamming distance](#)[JC69 distance](#)[K80 distance](#)[UPGMA algorithm](#)[Exercise](#)[Solution](#)

In reality we observe the sequences at present and the tree is unknown.

How do we build a tree from an alignment?

This time we will use UPGMA:

1. Build the initial distance matrix;
2. Repeat until all sequences are clustered:
  - 2.1 Cluster two closest nodes;
  - 2.2 Recalculate the distances to the new merged node.

[Tree reconstruction](#)[Tree reconstruction](#)[Distance matrices](#)[Hamming distance](#)[JC69 distance](#)[K80 distance](#)[UPGMA algorithm](#)[Exercise](#)[Solution](#)



Hamming distance: number of segregating sites.

[Tree reconstruction](#)[Tree reconstruction](#)[Distance matrices](#)[Hamming distance](#)[JC69 distance](#)[K80 distance](#)[UPGMA algorithm](#)[Exercise](#)[Solution](#)

[Tree reconstruction](#)[Tree reconstruction](#)[Distance matrices](#)[Hamming distance](#)[JC69 distance](#)[K80 distance](#)[UPGMA algorithm](#)[Exercise](#)[Solution](#)

Hamming distance: number of segregating sites.

Orangutan	CTGCTACTGAAACCAGACTA
Gorilla	CTGCTGTTTAGAAAAAACTA

[Tree reconstruction](#)[Tree reconstruction](#)[Distance matrices](#)[Hamming distance](#)[JC69 distance](#)[K80 distance](#)[UPGMA algorithm](#)[Exercise](#)[Solution](#)

Hamming distance: number of segregating sites.

Orangutan    CTGCTACTGAAACCAGACTA

Gorilla      CTGCTGTTTAGAAAAACTA

$$\hat{d}_{\text{hamming}} = 7$$

# Hamming distance matrix

CB

Orangutan (O)	CTGCTACTGAAACCAGACTA
Gorilla (G)	CTGCTGTTTAGAAAAAACTA
Chimpanzee (C)	CTGCTGTTTAGTAAGAACCA
Human (H)	CTGCTGTTTAGTAAGAACTA

Tree reconstruction

Tree reconstruction

Distance matrices

Hamming distance

JC69 distance

K80 distance

UPGMA algorithm

Exercise

Solution

# Hamming distance matrix

CB

Orangutan (O) CTGCTACTGAAACCAGACTA  
Gorilla (G) CTGCTGTTTAGAAAAAACTA  
Chimpanzee (C) CTGCTGTTTAGTAAGAACCA  
Human (H) CTGCTGTTTAGTAAGAACTA

	O	G	C	H
O				
G				
C				
H				

Tree reconstruction

Tree reconstruction

Distance matrices

Hamming distance

JC69 distance

K80 distance

UPGMA algorithm

Exercise

Solution

# Hamming distance matrix

CB

Orangutan (O) CTGCTACTGAAACCAGACTA  
Gorilla (G) CTGCTGTTTAGAAAAAATA  
Chimpanzee (C) CTGCTGTTTAGTAAGAACCA  
Human (H) CTGCTGTTTAGTAAGAACTA

	O	G	C	H
O		7	10	9
G			3	2
C				1
H				

Tree reconstruction

Tree reconstruction

Distance matrices

Hamming distance

JC69 distance

K80 distance

UPGMA algorithm

Exercise

Solution

# Hamming distance matrix

CB

Orangutan (O) CTGCTACTGAAACCAGACTA  
Gorilla (G) CTGCTGTTTAGAAAAAACTA  
Chimpanzee (C) CTGCTGTTTAGTAAGAACCA  
Human (H) CTGCTGTTTAGTAAGAACTA

	O	G	C	H
O	-	7	10	9
G	7	-	3	2
C	10	3	-	1
H	9	2	1	-

Tree reconstruction

Tree reconstruction

Distance matrices

Hamming distance

JC69 distance

K80 distance

UPGMA algorithm

Exercise

Solution

JC69 distance:  $\hat{d} = -\frac{3}{4} \log(1 - \frac{4}{3}\hat{p})$

[Tree reconstruction](#)[Tree reconstruction](#)[Distance matrices](#)[Hamming distance](#)[JC69 distance](#)[K80 distance](#)[UPGMA algorithm](#)[Exercise](#)[Solution](#)



[Tree reconstruction](#)[Tree reconstruction](#)[Distance matrices](#)[Hamming distance](#)[JC69 distance](#)[K80 distance](#)[UPGMA algorithm](#)[Exercise](#)[Solution](#)

JC69 distance:  $\hat{d} = -\frac{3}{4} \log(1 - \frac{4}{3}\hat{p})$

Orangutan	CTGCTACTGAAACCAGACTA
Gorilla	CTGCTGTTTAGAAAAAACTA

[Tree reconstruction](#)[Tree reconstruction](#)[Distance matrices](#)[Hamming distance](#)[JC69 distance](#)[K80 distance](#)[UPGMA algorithm](#)[Exercise](#)[Solution](#)

JC69 distance:  $\hat{d} = -\frac{3}{4} \log(1 - \frac{4}{3}\hat{p})$

Orangutan    CTGCTACTGAAACCAGACTA

Gorilla      CTGCTGTTTAGAAAAACTA

$$\hat{p} = \frac{7}{20} = 0.35$$

[Tree reconstruction](#)[Tree reconstruction](#)[Distance matrices](#)[Hamming distance](#)[JC69 distance](#)[K80 distance](#)[UPGMA algorithm](#)[Exercise](#)[Solution](#)

JC69 distance:  $\hat{d} = -\frac{3}{4} \log(1 - \frac{4}{3}\hat{p})$

Orangutan    CTGCTACTGAAACCAGACTA

Gorilla      CTGCTGTTTAGAAAAACTA

$$\hat{p} = \frac{7}{20} = 0.35$$

$$\hat{d} = -\frac{3}{4} \log(1 - \frac{4}{3}0.35)$$

[Tree reconstruction](#)[Tree reconstruction](#)[Distance matrices](#)[Hamming distance](#)[JC69 distance](#)[K80 distance](#)[UPGMA algorithm](#)[Exercise](#)[Solution](#)

JC69 distance:  $\hat{d} = -\frac{3}{4} \log(1 - \frac{4}{3}\hat{p})$

Orangutan    CTGCTACTGAAACCAGACTA

Gorilla      CTGCTGTTTAGAAAAACTA

$$\hat{p} = \frac{7}{20} = 0.35$$

$$\hat{d} = -\frac{3}{4} \log(1 - \frac{4}{3}0.35) \approx 0.4715$$

[Tree reconstruction](#)[Tree reconstruction](#)[Distance matrices](#)[Hamming distance](#)[JC69 distance](#)[K80 distance](#)[UPGMA algorithm](#)[Exercise](#)[Solution](#)

K80 distance:  $\hat{d} = -\frac{1}{2} \log(1 - 2S - V) - \frac{1}{4} \log(1 - 2V)$

[Tree reconstribution](#)[Tree reconstruction](#)[Distance matrices](#)[Hamming distance](#)[JC69 distance](#)[K80 distance](#)[UPGMA algorithm](#)[Exercise](#)[Solution](#)

K80 distance:  $\hat{d} = -\frac{1}{2} \log(1 - 2S - V) - \frac{1}{4} \log(1 - 2V)$

Orangutan    CTGCTACTGAAACCAGACTA

Gorilla      CTGCTGTTTAGAAAAAACTA

[Tree reconstruction](#)[Tree reconstruction](#)[Distance matrices](#)[Hamming distance](#)[JC69 distance](#)[K80 distance](#)[UPGMA algorithm](#)[Exercise](#)[Solution](#)

K80 distance:  $\hat{d} = -\frac{1}{2} \log(1 - 2S - V) - \frac{1}{4} \log(1 - 2V)$

Orangutan CTGCTACTGAAACCAGACTA

Gorilla CTGCTGTTTAGAAAAACTA

Transition proportion ( $T \leftrightarrow C, A \leftrightarrow G$ )  $S = \frac{4}{20} = 0.2$

[Tree reconstruction](#)[Tree reconstruction](#)[Distance matrices](#)[Hamming distance](#)[JC69 distance](#)[K80 distance](#)[UPGMA algorithm](#)[Exercise](#)[Solution](#)

K80 distance:  $\hat{d} = -\frac{1}{2} \log(1 - 2S - V) - \frac{1}{4} \log(1 - 2V)$

Orangutan CTGCTACTGAAACCAGACTA

Gorilla CTGCTGTTTAGAAAAACTA

Transition proportion ( $T \leftrightarrow C, A \leftrightarrow G$ )  $S = \frac{4}{20} = 0.2$

Transversion proportion ( $T \leftrightarrow G, A \leftrightarrow C, A \leftrightarrow T, C \leftrightarrow G$ )

$V = \frac{3}{20} = 0.15$



Tree reconstruction

Tree reconstruction

Distance matrices

Hamming distance

JC69 distance

K80 distance

UPGMA algorithm

Exercise

Solution

K80 distance:  $\hat{d} = -\frac{1}{2} \log(1 - 2S - V) - \frac{1}{4} \log(1 - 2V)$

Orangutan CTGCTACTGAAACCAGACTA

Gorilla CTGCTGTTTAGAAAAACTA

Transition proportion ( $T \leftrightarrow C, A \leftrightarrow G$ )  $S = \frac{4}{20} = 0.2$

Transversion proportion ( $T \leftrightarrow G, A \leftrightarrow C, A \leftrightarrow T, C \leftrightarrow G$ )

$V = \frac{3}{20} = 0.15$

$\hat{d} = -\frac{1}{2} \log(1 - 2 \cdot 0.2 - 0.15) - \frac{1}{4} \cdot \log(1 - 2 \cdot 0.15)$

Tree reconstruction

Tree reconstruction

Distance matrices

Hamming distance

JC69 distance

K80 distance

UPGMA algorithm

Exercise

Solution

K80 distance:  $\hat{d} = -\frac{1}{2} \log(1 - 2S - V) - \frac{1}{4} \log(1 - 2V)$

Orangutan CTGCTACTGAAACCAGACTA

Gorilla CTGCTGTTTAGAAAAACTA

Transition proportion ( $T \leftrightarrow C, A \leftrightarrow G$ )  $S = \frac{4}{20} = 0.2$

Transversion proportion ( $T \leftrightarrow G, A \leftrightarrow C, A \leftrightarrow T, C \leftrightarrow G$ )

$V = \frac{3}{20} = 0.15$

$\hat{d} = -\frac{1}{2} \log(1 - 2 \cdot 0.2 - 0.15) - \frac{1}{4} \cdot \log(1 - 2 \cdot 0.15) \approx 0.4884$

# Distance matrix comparison

CB

Hamming:

	O	G	C	H
O		7	10	9
G			3	2
C				1
H				

JC69:

	O	G	C	H
O		0.4715	0.8240	0.6872
G			0.1674	0.1073
C				0.0517
H				

K80:

	O	G	C	H
O		0.4884	0.9324	0.7297
G			0.1702	0.1076
C				0.0527
H				

[Tree reconstruction](#)[Tree reconstruction](#)[Distance matrices](#)[Hamming distance](#)[JC69 distance](#)[K80 distance](#)[UPGMA algorithm](#)[Exercise](#)[Solution](#)

# UPGMA algorithm

CB

**Data:** Distance matrix  $D$

Tree reconstruction

Tree reconstruction

Distance matrices

Hamming distance

JC69 distance

K80 distance

UPGMA algorithm

Exercise

Solution

# UPGMA algorithm

CB

**Data:** Distance matrix  $D$

**Result:** Ultrametric phylogenetic tree

Tree reconstruction

Tree reconstruction

Distance matrices

Hamming distance

JC69 distance

K80 distance

UPGMA algorithm

Exercise

Solution

# UPGMA algorithm

CB

**Data:** Distance matrix  $D$

**Result:** Ultrametric phylogenetic tree

**for**  $i \leftarrow 1$  **to**  $N$  **do**

$n_i \leftarrow 1$ ;  
     $s_i \leftarrow \text{node}(i)$

**end**

Tree reconstruction

Tree reconstruction

Distance matrices

Hamming distance

JC69 distance

K80 distance

UPGMA algorithm

Exercise

Solution

**Data:** Distance matrix  $D$

**Result:** Ultrametric phylogenetic tree

**for**  $i \leftarrow 1$  **to**  $N$  **do**

$n_i \leftarrow 1$ ;

$s_i \leftarrow \text{node}(i)$

**end**

**while**  $\text{size}(D) > (1, 1)$  **do**

    Choose  $s_i, s_j$  such that  $\min(D) = d[s_i, s_j]$ ;

$n_{i,j} \leftarrow n_i + n_j$ ;

$s_{i,j} \leftarrow \{s_i, s_j\}$ ;

$\text{branch}(s_{i,j}, s_i) \leftarrow d[s_i, s_j]/2 - \text{distance\_to\_tip}(s_i)$ ;

$\text{branch}(s_{i,j}, s_j) \leftarrow d[s_i, s_j]/2 - \text{distance\_to\_tip}(s_j)$ ;

**for** all  $m \neq i$  and  $m \neq j$  **do**

$d[s_m, s_{i,j}] \leftarrow \frac{n_i d[s_i, s_m] + n_j d[s_j, s_m]}{n_i + n_j}$ ;

**end**

    Delete node  $s_i$  from  $D$ ;

    Delete node  $s_j$  from  $D$ ;

**end**

Tree reconstruction

Tree reconstruction

Distance matrices

Hamming distance

JC69 distance

K80 distance

UPGMA algorithm

Exercise

Solution

```
while size(D) > (1, 1) do  
  | Choose  $s_i, s_j$  such that  $\min(D) = d[s_i, s_j]$ ;  
  | ...  
end
```

	O	G	C	H
O	-	7	10	9
G	7	-	3	2
C	10	3	-	1
H	9	2	1	-

[Tree reconstruction](#)[Tree reconstruction](#)[Distance matrices](#)[Hamming distance](#)[JC69 distance](#)[K80 distance](#)[UPGMA algorithm](#)[Exercise](#)[Solution](#)



# Step 1

CB

```
while size(D) > (1, 1) do  
  | Choose  $s_i, s_j$  such that  $\min(D) = d[s_i, s_j]$ ;  
  | ...  
end
```

	O	G	C	H
O	-	7	10	9
G	7	-	3	2
C	10	3	-	<b>1</b>
H	9	2	<b>1</b>	-

Tree reconstruction

Tree reconstruction

Distance matrices

Hamming distance

JC69 distance

K80 distance

UPGMA algorithm

Exercise

Solution

```
while size(D) > (1, 1) do  
  | Choose  $s_i, s_j$  such that  $\min(D) = d[s_i, s_j]$ ;  
  | ...  
end
```

	O	G	C	H
O	-	7	10	9
G	7	-	3	2
C	10	3	-	<b>1</b>
H	9	2	<b>1</b>	-

$\Rightarrow$  Merge Human and Chimp.

[Tree reconstruction](#)[Tree reconstruction](#)[Distance matrices](#)[Hamming distance](#)[JC69 distance](#)[K80 distance](#)[UPGMA algorithm](#)[Exercise](#)[Solution](#)

```
while size(D) > (1, 1) do
```

```
  ...
```

```
   $n_{i,j} \leftarrow n_i + n_j;$ 
```

```
   $s_{i,j} \leftarrow \{s_i, s_j\};$ 
```

```
   $\text{branch}(s_{i,j}, s_i) \leftarrow d[s_i, s_j]/2 - \text{distance\_to\_tip}(s_i);$ 
```

```
   $\text{branch}(s_{i,j}, s_j) \leftarrow d[s_i, s_j]/2 - \text{distance\_to\_tip}(s_j);$ 
```

```
  ...
```

```
end
```

Tree reconstruction

Tree reconstruction

Distance matrices

Hamming distance

JC69 distance

K80 distance

UPGMA algorithm

Exercise

Solution

**while** size(D) > (1, 1) **do**

...

$n_{i,j} \leftarrow n_i + n_j;$

$s_{i,j} \leftarrow \{s_i, s_j\};$

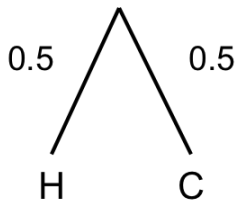
$\text{branch}(s_{i,j}, s_i) \leftarrow d[s_i, s_j]/2 - \text{distance\_to\_tip}(s_i);$

$\text{branch}(s_{i,j}, s_j) \leftarrow d[s_i, s_j]/2 - \text{distance\_to\_tip}(s_j);$

...

**end**

	O	G	C	H
O	-	7	10	9
G	7	-	3	2
C	10	3	-	<b>1</b>
H	9	2	<b>1</b>	-



Tree reconstruction

Tree reconstruction

Distance matrices

Hamming distance

JC69 distance

K80 distance

UPGMA algorithm

Exercise

Solution

```
while size(D) > (1, 1) do  
  |  
  ... for all  $m \neq i$  and  $m \neq j$  do  
    |  $d[s_m, s_{i,j}] \leftarrow \frac{n_i d[s_i, s_m] + n_j d[s_j, s_m]}{n_i + n_j};$   
  end  
  Delete node  $s_i$  from D;  
  Delete node  $s_j$  from D;  
end
```

Tree reconstruction

Tree reconstruction

Distance matrices

Hamming distance

JC69 distance

K80 distance

UPGMA algorithm

Exercise

Solution

**while** size(D) > (1, 1) **do**

    ... **for** all  $m \neq i$  and  $m \neq j$  **do**

$d[s_m, s_{i,j}] \leftarrow \frac{n_i d[s_i, s_m] + n_j d[s_j, s_m]}{n_i + n_j};$

**end**

    Delete node  $s_i$  from D;

    Delete node  $s_j$  from D;

**end**

$$d[\text{O}, \text{H.C}] = \frac{1 \cdot 10 + 1 \cdot 9}{1 + 1} = 9.5$$

Tree reconstruction

Tree reconstruction

Distance matrices

Hamming distance

JC69 distance

K80 distance

UPGMA algorithm

Exercise

Solution

**while** size(D) > (1, 1) **do**

... **for** all  $m \neq i$  and  $m \neq j$  **do**

$d[s_m, s_{i,j}] \leftarrow \frac{n_i d[s_i, s_m] + n_j d[s_j, s_m]}{n_i + n_j};$

**end**

Delete node  $s_i$  from D;

Delete node  $s_j$  from D;

**end**

$$d[\text{O}, \text{H.C}] = \frac{1 \cdot 10 + 1 \cdot 9}{1 + 1} = 9.5$$

$$d[\text{G}, \text{H.C}] = \frac{1 \cdot 3 + 1 \cdot 2}{1 + 1} = 2.5$$

Tree reconstruction

Tree reconstruction

Distance matrices

Hamming distance

JC69 distance

K80 distance

UPGMA algorithm

Exercise

Solution

**while** size(D) > (1, 1) **do**

    ... **for** all  $m \neq i$  and  $m \neq j$  **do**

$d[s_m, s_{i,j}] \leftarrow \frac{n_i d[s_i, s_m] + n_j d[s_j, s_m]}{n_i + n_j};$

**end**

    Delete node  $s_i$  from D;

    Delete node  $s_j$  from D;

**end**

$$d[\text{O}, \text{H.C}] = \frac{1 \cdot 10 + 1 \cdot 9}{1 + 1} = 9.5$$

$$d[\text{G}, \text{H.C}] = \frac{1 \cdot 3 + 1 \cdot 2}{1 + 1} = 2.5$$

	O	G	C	H	H.C
O	-	7	10	9	9.5
G	7	-	3	2	2.5
C	10	3	-	1	0
H	9	2	1	-	0
H.C	9.5	2.5	0	0	-

Tree reconstruction

Tree reconstruction

Distance matrices

Hamming distance

JC69 distance

K80 distance

UPGMA algorithm

Exercise

Solution



**while** size(D) > (1, 1) **do**

... **for** all  $m \neq i$  and  $m \neq j$  **do**

$$d[s_m, s_{i,j}] \leftarrow \frac{n_i d[s_i, s_m] + n_j d[s_j, s_m]}{n_i + n_j};$$

**end**

Delete node  $s_i$  from D;

Delete node  $s_j$  from D;

**end**

$$d[O, H.C] = \frac{1 \cdot 10 + 1 \cdot 9}{1 + 1} = 9.5$$

$$d[G, H.C] = \frac{1 \cdot 3 + 1 \cdot 2}{1 + 1} = 2.5$$

	O	G	C	H	H.C
O	-	7	10	9	9.5
G	7	-	3	2	2.5
C	10	3	-	1	0
H	9	2	1	-	0
H.C	9.5	2.5	0	0	-

	O	G	H.C
O	-	7	9.5
G	7	-	2.5
H.C	9.5	2.5	-

Tree reconstruction

Tree reconstruction

Distance matrices

Hamming distance

JC69 distance

K80 distance

UPGMA algorithm

Exercise

Solution

# Repeat...

CB

	O	G	H.C
O	-	7	9.5
G	7	-	2.5
H.C	9.5	2.5	-

Tree reconstruction

Tree reconstruction

Distance matrices

Hamming distance

JC69 distance

K80 distance

UPGMA algorithm

Exercise

Solution

# Repeat...

CB

	O	G	H.C
O	-	7	9.5
G	7	-	<b>2.5</b>
H.C	9.5	<b>2.5</b>	-

Tree reconstruction

Tree reconstruction

Distance matrices

Hamming distance

JC69 distance

K80 distance

UPGMA algorithm

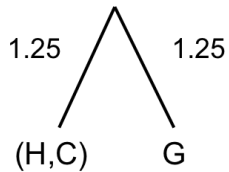
Exercise

Solution

# Repeat...

CB

	O	G	H.C
O	-	7	9.5
G	7	-	<b>2.5</b>
H.C	9.5	<b>2.5</b>	-



Tree reconstruction

Tree reconstruction

Distance matrices

Hamming distance

JC69 distance

K80 distance

UPGMA algorithm

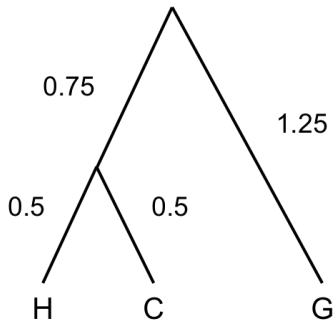
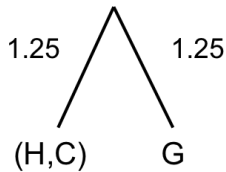
Exercise

Solution

# Repeat...

CB

	O	G	H.C
O	-	7	9.5
G	7	-	<b>2.5</b>
H.C	9.5	<b>2.5</b>	-



Tree reconstruction

Tree reconstruction

Distance matrices

Hamming distance

JC69 distance

K80 distance

UPGMA algorithm

Exercise

Solution

# Repeat steps

$$d[\text{O}, \text{H.C.G}] = \frac{1 \cdot 7 + 2 \cdot 9.5}{1+2} = 8.67$$

CB

Tree reconstruction

Tree reconstruction

Distance matrices

Hamming distance

JC69 distance

K80 distance

UPGMA algorithm

Exercise

Solution

# Repeat steps

CB

$$d[\text{O}, \text{H.C.G}] = \frac{1 \cdot 7 + 2 \cdot 9.5}{1+2} = 8.67$$

	O	G	H.C	H.C.G
O	-	7	9.5	8.67
G	7	-	2.5	0
H.C	9.5	2.5	-	0
H.C.G	8.67	0	0	-

Tree reconstruction

Tree reconstruction

Distance matrices

Hamming distance

JC69 distance

K80 distance

UPGMA algorithm

Exercise

Solution

# Repeat steps

CB

$$d[\text{O}, \text{H.C.G}] = \frac{1 \cdot 7 + 2 \cdot 9.5}{1+2} = 8.67$$

	O	G	H.C	H.C.G
O	-	7	9.5	8.67
G	7	-	2.5	0
H.C	9.5	2.5	-	0
H.C.G	8.67	0	0	-

	O	H.C.G
O	-	8.67
H.C.G	8.67	-

Tree reconstruction

Tree reconstruction

Distance matrices

Hamming distance

JC69 distance

K80 distance

UPGMA algorithm

Exercise

Solution



# Repeat steps

CB

$$d[\text{O}, \text{H.C.G}] = \frac{1 \cdot 7 + 2 \cdot 9.5}{1+2} = 8.67$$

	O	G	H.C	H.C.G
O	-	7	9.5	8.67
G	7	-	2.5	0
H.C	9.5	2.5	-	0
H.C.G	8.67	0	0	-

	O	H.C.G
O	-	8.67
H.C.G	8.67	-

	O	H.C.G
O	-	8.67
H.C.G	8.67	-

Tree reconstruction

Tree reconstruction

Distance matrices

Hamming distance

JC69 distance

K80 distance

UPGMA algorithm

Exercise

Solution

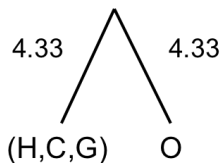
# Repeat steps

$$d[\text{O}, \text{H.C.G}] = \frac{1 \cdot 7 + 2 \cdot 9.5}{1+2} = 8.67$$

	O	G	H.C	H.C.G
O	-	7	9.5	8.67
G	7	-	2.5	0
H.C	9.5	2.5	-	0
H.C.G	8.67	0	0	-

	O	H.C.G
O	-	8.67
H.C.G	8.67	-

	O	H.C.G
O	-	8.67
H.C.G	8.67	-



Tree reconstruction

Tree reconstruction

Distance matrices

Hamming distance

JC69 distance

K80 distance

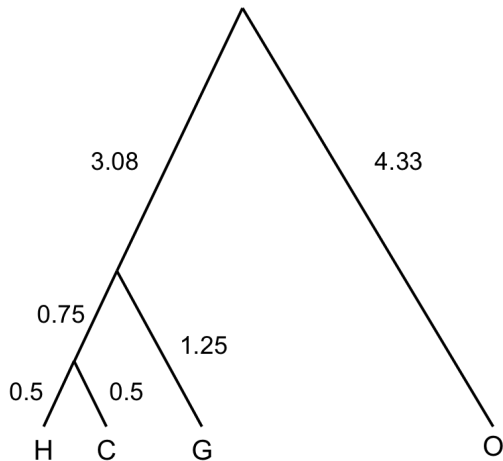
UPGMA algorithm

Exercise

Solution

# Final tree

CB



Tree reconstruction

Tree reconstruction

Distance matrices

Hamming distance

JC69 distance

K80 distance

UPGMA algorithm

Exercise

Solution

# Pen and paper exercise

CB

```
for i ← 1 to N do
  | ni ← 1;
  | si ← node(i)
end
while size(D) > (1,1) do
  | Choose si, sj such that min(D) = d[si, sj];
  | ni,j ← ni + nj;
  | si,j ← {si, sj};
  | branch(si,j, si) ← d[si, sj]/2 - distance_to_tip(si);
  | branch(si,j, sj) ← d[si, sj]/2 - distance_to_tip(sj);
  | for all m ≠ i and m ≠ j do
  |   | d[sm, si,j] ←  $\frac{n_i d[s_i, s_m] + n_j d[s_j, s_m]}{n_i + n_j}$ ;
  | end
  | Delete node si from D;
  | Delete node sj from D;
end
```

	O	G	C	H
O		0.4884	0.9324	0.7297
G			0.1702	0.1076
C				0.0527
H				

Tree reconstruction

Tree reconstruction

Distance matrices

Hamming distance

JC69 distance

K80 distance

UPGMA algorithm

Exercise

Solution

# Exercise solution: distances and branches (1)

	O	G	C	H
O		0.4884	0.9324	0.7297
G			0.1702	0.1076
C				<b>0.0527</b>
H				

$$\text{branch}[C.H, C] = \frac{0.0527}{2} = 0.02635$$

$$\text{branch}[C.H, H] = \frac{0.0527}{2} = 0.02635$$

$$d[C.H, O] = \frac{1 \cdot 0.9324 + 1 \cdot 0.7297}{1+1} = 0.83105$$

$$d[C.H, G] = \frac{1 \cdot 0.1702 + 1 \cdot 0.1076}{1+1} = 0.1389$$

	O	G	C.H
O		0.4884	0.83105
G			0.1389
C.H			

Tree reconstruction

Tree reconstruction

Distance matrices

Hamming distance

JC69 distance

K80 distance

UPGMA algorithm

Exercise

Solution

# Exercise solution: distances and branches (2)

CB

Tree reconstruction

Tree reconstruction

Distance matrices

Hamming distance

JC69 distance

K80 distance

UPGMA algorithm

Exercise

Solution

	O	G	C.H
O		0.4884	0.83105
G			<b>0.1389</b>
C.H			

$$\text{branch}[\text{C.H.G}, \text{G}] = \frac{0.1389}{2} = 0.06945$$

$$\text{branch}[\text{C.H.G}, \text{C.H}] = \frac{0.1389}{2} - 0.02635 = 0.0431$$

$$d[\text{C.H.G}, \text{O}] = \frac{1 \cdot 0.4884 + 2 \cdot 0.83105}{1+2} = 0.71683$$

	O	C.H.G
O		0.71683
C.H.G		

# Exercise solution: distances and branches (3)

CB

Tree reconstruction

Tree reconstruction

Distance matrices

Hamming distance

JC69 distance

K80 distance

UPGMA algorithm

Exercise

Solution

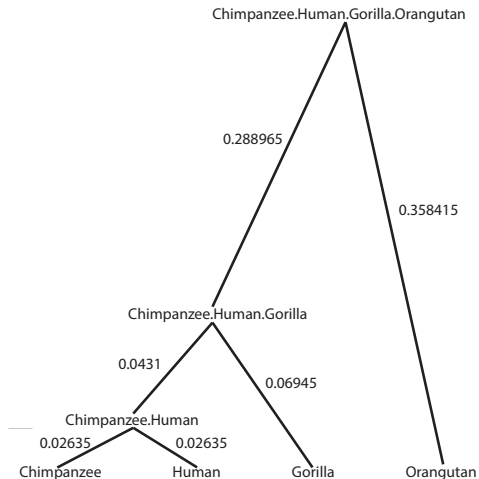
	O	C.H.G
O		<b>0.71683</b>
C.H.G		

$$\text{branch}[\text{C.H.G.O}, \text{C.H.G}] = \frac{0.71683}{2} - 0.06945 = 0.288965$$

$$\text{branch}[\text{C.H.G.O}, \text{O}] = \frac{0.71683}{2} = 0.358415$$

# Exercise solution: tree

CB



Tree reconstruction

Tree reconstruction

Distance matrices

Hamming distance

JC69 distance

K80 distance

UPGMA algorithm

Exercise

Solution