

# Exercise sheet 3: T-Coffee

## Exercise 1

You are given the sequences  $a$ ,  $b$  and  $c$

$$a = CACCGGb = ACCAAGc = AACACC$$

The pairwise optimal alignments  $A(x, y)$  of the set of sequences  $S$  were calculated as:

a: CACCG_G	a: __CACCGG	b: ACCAAG
:		: : :
b: _ACCAAG	c: AACACC__	c: AACACC

**Question 1A** Calculate the primary library ( $L$ )

**Formulae** init:  $L_{i,j}^{x,y} = 0$

$\forall$  alignments  $A$  of sequences  $x$  and  $y$  of the set  $S$ .

$$weight(A) = \frac{\text{number of matches}}{\min(len(x), len(y))} * 100$$

$\forall$  aligned positions  $i, j$  with  $1 \leq i \leq len(x)$  and  $1 \leq j \leq len(y)$

$$L_{i,j}^{x,y} = L_{i,j}^{x,y} + weight(A)$$

**Solution**  $L_{2,1}^{a,b} = L_{3,2}^{a,b} = L_{4,3}^{a,b} = L_{6,6}^{a,b} = 100 * \frac{4}{6} = 67$  and all other  $L_{i,j}^{a,b} = 0$

$L_{1,3}^{a,c} = L_{2,4}^{a,c} = L_{3,5}^{a,c} = L_{4,6}^{a,c} = 100 * \frac{4}{6} = 67$  and all other  $L_{i,j}^{a,c} = 0$

$L_{1,1}^{b,c} = L_{3,3}^{b,c} = L_{4,4}^{b,c} = 100 * \frac{3}{6} = 50$  and all other  $L_{i,j}^{b,c} = 0$

**Question 1B** Calculate the extended library ( $EL$ )

$$\text{Formulae } EL_{i,j}^{x,y} = L_{i,j}^{x,y} + \sum_{z \in S \setminus \{x,y\}} \sum_{1 \leq k \leq len(z)} \min(L_{i,k}^{x,z}, L_{k,j}^{z,y})$$

**Solution** The original Library doesn't change as there are no edges enforcing certain connections. Hence

$$EL_{i,j}^{x,y} = L_{i,j}^{x,y} \quad \forall L_{i,j}^{x,y} \neq 0$$

and the following weights are added:

a: CACCG\_G  
 |||: |  
 b: \_ACCAAG  
 |:|:|:  
 c: AACACC  
 \* \*

$$EL_{1,3}^{a,b} = EL_{2,4}^{a,b} = 50$$

a: \_\_CACCGG  
 ||||  
 c: AACACC\_\_  
 |:|:|:  
 b: ACCAAG  
 \*\*

$$EL_{2,1}^{a,c} = EL_{4,3}^{a,c} = 50$$

b: ACCAAG  
 |||: |  
 a: CACCG\_G  
 ||||  
 c: AACACC  
 \*\*\*

$$EL_{1,4}^{b,c} = EL_{2,5}^{b,c} = EL_{3,6}^{b,c} = 67$$

**Question 1C** Realign the sequences  $b$  and  $c$  using  $EL$  for scoring and gap costs and mismatch costs of 0

**Solution**

-	-	A	C	C	A	A	G
-	0	0	0	0	0	0	0
A	0	50	50	50	50	50	50
A	0	50	50	50	50	50	50
C	0	50	50	100	100	100	100
A	0	67	67	100	150	150	150
C	0	67	133	133	150	150	150
C	0	67	133	200	200	200	200

**Question 1C** Do the other alignments  $a-b$  and  $a-c$  change? Provide arguments, without calculating new alignments.

**Solution** No. The newly added alignment scores in  $EL$  represent edges that are incompatible with the current best alignments and can not score higher.

**Question 1E** Sketch the Guide Tree

**Solution**  $((a, c), b)$  or  $((a, b), c)$

**Question 1F** Perform a progressive alignment by aligning sequence  $b$  to the already existing alignment  $A(a, c)$ . To score a match between  $b$  and  $A(a, c)$  use the sum  $EL^{a,b} + EL^{b,c}$  with the correct indices. Show the resulting multiple sequence alignment.

**Solution**

–	–	-A	-A	CC	AA	CC	CC	G-	G-
-	0	0	0	0	0	0	0	0	0
<b>A</b>	0	50	50	50	133	133	133	133	133
<b>C</b>	0	50	50	50	133	267	267	267	267
<b>C</b>	0	50	50	150	150	267	400	400	400
<b>A</b>	0	50	50	150	250	267	400	400	400
<b>A</b>	0	50	50	150	250	267	400	400	400
<b>G</b>	0	50	50	150	250	267	400	400	467