

## Exercise sheet 3: T-Coffee



### Exercise 1

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

$$a = CACCGG = ACCAAG = AACACC$$

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

**1a)**

Calculate the primary library ( $L$ )

**Hide**

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

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

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

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

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

**Solution**  $L_{2,1}^{a,b} = L_{3,2}^{a,b} = L_{4,3}^{a,b} = L_{6,6}^{a,b} = 100 * \frac{4}{6} = 66.6\bar{6} \approx 67$  and all other  $L_{i,j}^{a,b} = 0$

$L_{1,3}^{a,c} = L_{2,4}^{a,c} = L_{3,5}^{a,c} = L_{4,6}^{a,c} = 100 * \frac{4}{6} = 66.6\bar{6} \approx 67$  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  $L_{i,j}^{b,c} = 0$

**1b)**

Calculate the extended library ( $EL$ )

**Hide**

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

**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: __CACCGG
    ||||
c: AACACC__
    |:||:|
b: ACCAAG
    **

```

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

```
a: CACCG_G
    |||: |
b: _ACCAAG
    |:||:|
c: AACACC
    * *

```

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

```
b: ACCAAG
    |||: |
a: CACCG_G
    ||||
c: AACACC
    ***

```

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

**1c)**

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

**Hide**

## Formulae

$$\begin{aligned}
i &\in [1, |x|] \\
j &\in [1, |y|] \\
L(0, 0) &= 0 \\
L(i, 0) &= w(x_i, -) * i \quad \text{or} = L(i - 1, 0) + w(x_i, -) \\
L(0, j) &= w(-, y_j) * j \quad \text{or} = L(0, j - 1) + w(-, y_j) \\
L(i, j) &= \max \begin{cases} L(i - 1, j) + w(x_i, -) \\ L(i, j - 1) + w(-, y_j) \\ L(i - 1, j - 1) + w(x_i, y_j) \end{cases} \\
w(x_i, y_j) &= \begin{cases} EL_{i,j}^{x,y} & \text{match-score if } (x_i = y_j) \\ 0 & \text{insert/deletion-score if } (x_i = - \vee y_j = -) \\ 0 & \text{mismatch-score else } (x_i \neq y_j) \end{cases}
\end{aligned}$$

## Solution

-1	-	A1	C2	C3	A4	A5	G6
-	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

Alignment:

b: \_\_\_ACCAAG  
      |||  
 c: AACACC\_\_\_

1d)

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

Hide

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

1e)

Sketch a Guide Tree (either Sketch or Newick format)

**Hide**

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

**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.

**Hide**

-	-	-	-	C	A	C	C	G	G
-	-	A	A	C	A	C	C	-	-
-	0	0	0	0	0	0	0	0	0
A	0	50	50	50	133	133	133	133	133
C	0	50	50	50	133	267	267	267	267
C	0	50	50	150	150	267	400	400	400
A	0	50	50	150	250	267	400	400	400
A	0	50	50	150	250	267	400	400	400
G	0	50	50	150	250	267	400	400	467

**Solution**

One Possible Alignment:

a: \_\_CACC\_GG  
       ||||  
 c: AACACC\_\_  
       |||  
 b: \_\_\_ACCAAG

---