Exercise sheet 3: T-Coffee

Exercise 1

You are given the sequences a, b and c

$$a = CACCGG \quad b = ACCAAG \quad c = AACACC$$

The pairwise optimal alignments were calculated as:

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

Question 1A   Calculate the primary library (L)

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 $L_{i,j}^{a,b} = 0\backslash$

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

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

Question 1B   Calculate the extended library (EL)

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

$$EL_{i,j}^{s1,s2} = L_{i,j}^{s1,s2} \qquad \forall L_{i,j}^{s1,s2} \neq 0$$

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 ali- gnments.

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
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