Bioinformatics I

WS 15/16

Tutor: Alexander Seitz

Jonas Ditz

& Benjamin Schroeder

1	2	3	$\sum$

Assignment 4

(Abgabe am 9. November 2015)

Theoretical Assignment - *Optimal multiple alignment*

To calculate our MSA we use the recursion written down on page 50 in the script.

$$F(i_1, i_2, i_3) = max \begin{cases} F(i_1 - 1, i_2 - 1, i_3 - 1) + s_{SP}(a_{1i_1}, a_{2i_2}, a_{3i_3}) \\ \\ F(i_1 - 1, i_2 - 1, i_3) + s_{SP}(a_{1i_1}, a_{2i_2}, -) \\ F(i_1 - 1, i_2, i_3 - 1) + s_{SP}(a_{1i_1}, -, a_{3i_3}) \\ F(i_1, i_2 - 1, i_3 - 1) + s_{SP}(-, a_{2i_2}, a_{3i_3}) \\ \\ F(i_1 - 1, i_2, i_3) + s_{SP}(a_{1i_1}, -, -) \\ F(i_1, i_2 - 1, i_3) + s_{SP}(-, a_{2i_2}, -) \\ F(i_1, i_2, i_3 - 1) + s_{SP}(-, -, a_{3i_3}) \end{cases}$$

Such a recursion results in a three-dimensional matrix. Since such a matrix is really difficult to sketch, we split it into three two-dimensional matrices.

0	0	C	C	T		T	0	C	C	T		C	0	C	C	T
0	**0**	-2	-4	-6		0	-2	-4	-6	-8		0	-4	-6	-8	-10
C	-2	**-2**	-4	-10		C	-4	-4	-6	-10		C	-6	2	2	-8
T	-4	-8	-8	-6		T	-6	-8	**-4**	2		T	-8	-6	-2	0
T	-6	-10	-12	-10		T	-8	-10	-8	2		T	-10	-8	-6	**-6**

Figure 1: DP matrix of MSA, traceback is shown in red

If we fill the DP matrix using this recursion, we get several optimal alignments. One of them is the following:

$$\begin{pmatrix} C & T & T \\ - & T & C \\ C & C & T \end{pmatrix} \tag{1}$$

with score $\alpha_{SP}(A^*) = S(A, B) + S(A, C) + S(B, C) = -2 + 2 + (-6) = -6$.

Theoretical Assignment - *Progressive alignment*

Progressive alignment methodes are one way to create multiple sequence alignments (MSA), in this task we had to accomplish manualy a progressive alignment. The first step for the progressive alignment was to generate the pairwise global distance matrices. Afterwards the global optima were entered into a new table, which was used to generate the guide tree in the second step.(figure 2).
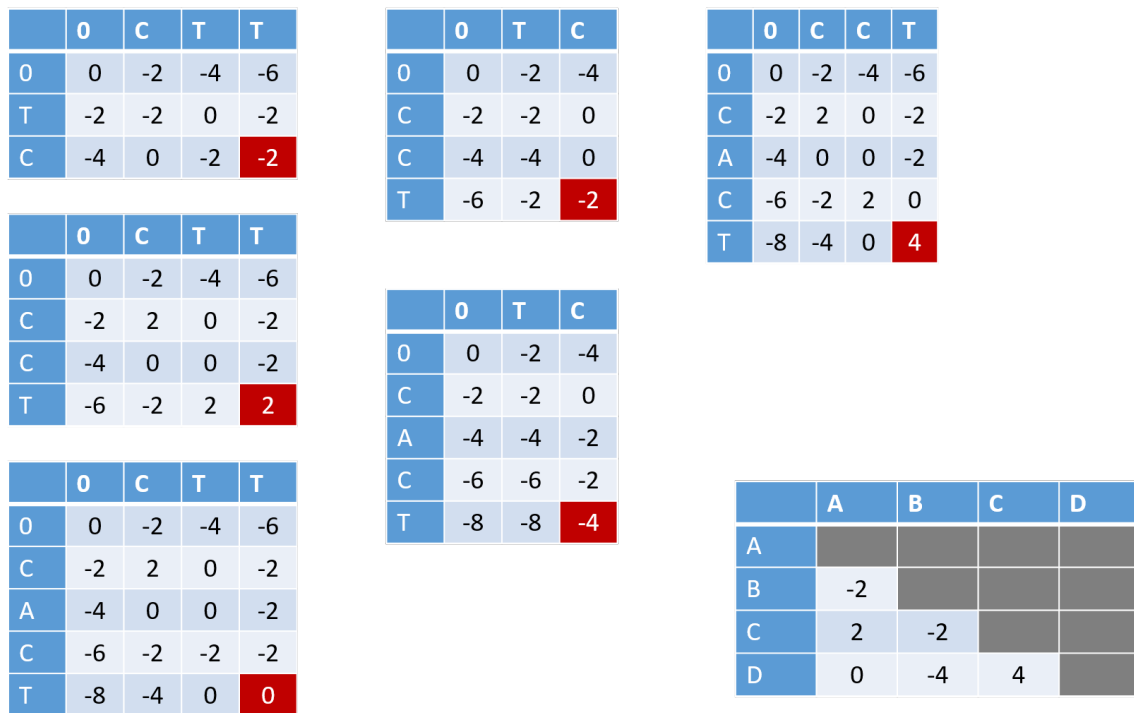
	O	C	T	T
0	0	-2	-4	-6
T	-2	-2	0	-2
C	-4	0	-2	-2

	O	T	C
0	0	-2	-4
C	-2	-2	0
C	-4	-4	0
T	-6	-2	-2

	O	C	C	T
0	0	-2	-4	-6
C	-2	2	0	-2
A	-4	0	0	-2
C	-6	-2	2	0
T	-8	-4	0	4

	O	C	T	T
0	0	-2	-4	-6
C	-2	2	0	-2
C	-4	0	0	-2
T	-6	-2	2	2

	O	T	C
0	0	-2	-4
C	-2	-2	0
A	-4	-4	-2
C	-6	-6	-2
T	-8	-8	-4

	O	C	T	T
0	0	-2	-4	-6
C	-2	2	0	-2
A	-4	0	0	-2
C	-6	-2	-2	-2
T	-8	-4	0	0

	A	B	C	D
A				
B	-2			
C	2	-2		
D	0	-4	4	

Figure 2: Global alignments lead to the basis table used to generate the guide tree

In step 2 the guide tree was created by the UPGMA methode in 3 substeps. After every step, which added a new cluster, the distance table was updated. The result from this step was the guide tree, which was used to apply the 'complete alignment' methode in the last step, the progressive alignment step.(figure 3)

The last step was creating the actual MSA. The 'complete alignment' methode was used. This means after every step, the distance of the new sequences was determined to each other sequence in the cluster. The optimal resulting score was used to get the best alignment for the new Sequence. after 3 steps the MSA was completed. (figure 4)

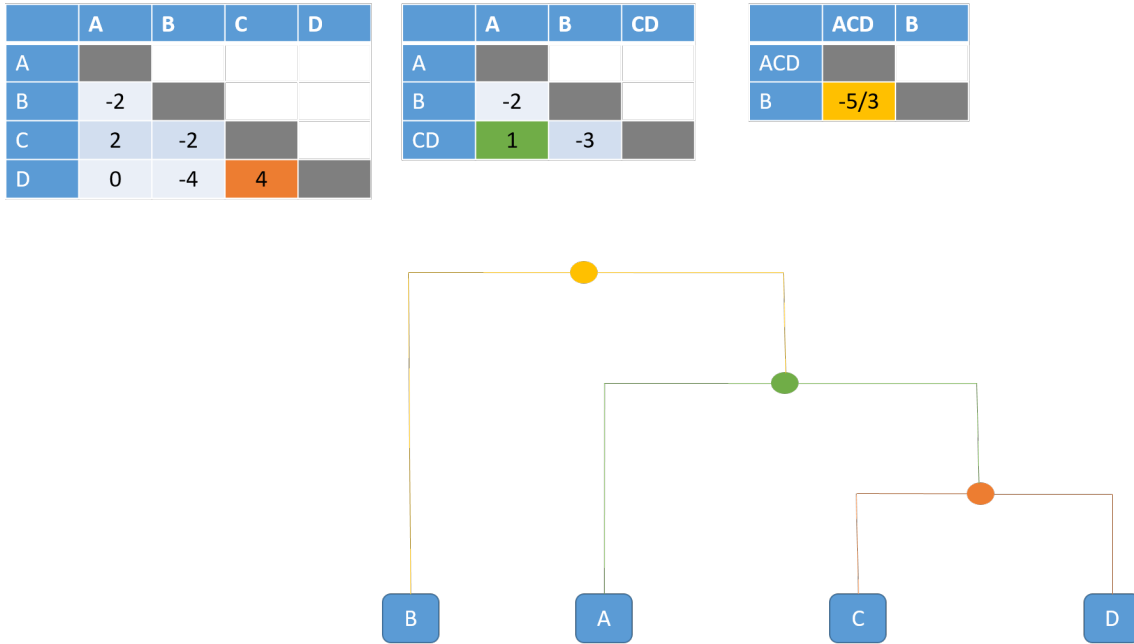# Practical Assignment - *Comparing multiple alignment*

	A	B	C	D
A				
B	-2			
C	2	-2		
D	0	-4	4	

	A	B	CD
A			
B	-2		
CD	1	-3	

	ACD	B
ACD		
B	-5/3	

Figure 3: The guide tree is made from the final table from step 1

B:TC

A: CTT

C-CT
CACT

C-CT
CACT
-TC-

C-CT
CACT
-TC-
C-TT

	0	C	C	T
0	0	-2	-4	-6
C	-2	2	0	-2
A	-4	0	0	-2
C	-6	-2	2	0
T	-8	-4	0	4

	0	T	C
0	0	-2	-4
C	-2	-2	0
-	-4	-4	-2
C	-6	-6	-2
T	-8	-8	-6

	0	C	T	T
0	0	-2	-4	-6
C	-2	2	0	-2
-	-4	0	0	0
C	-6	-2	-2	-2
T	-8	-4	0	0

	0	C	T	T
0	0	-2	-4	-6
C	-2	2	0	-2
A	-4	0	0	-2
C	-6	-2	-2	-2
T	-8	-4	0	0

	0	T	C
0	0	-2	-4
C	-2	-2	0
A	-4	-4	-2
C	-6	-6	-2
T	-8	-8	-4

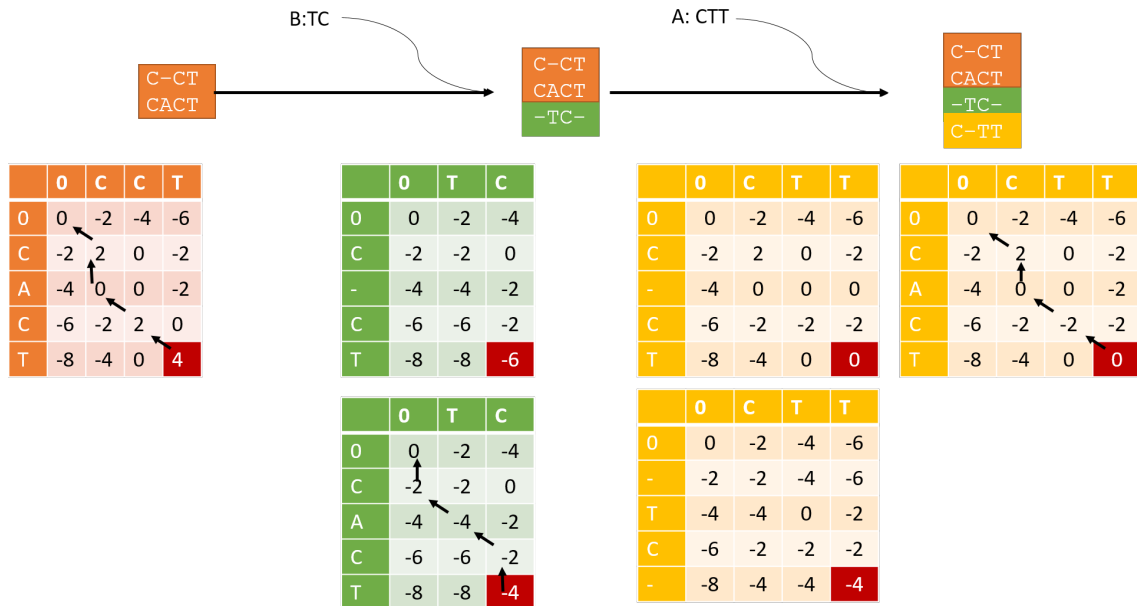|	0	C	T	T
0	0	-2	-4	-6
-	-2	-2	-4	-6
T	-4	-4	0	-2
C	-6	-2	-2	-2
-	-8	-4	-4	-4

Figure 4: Complete alignment methode is used to get an MSA from the guidetree from step 2