

Hausaufgabe 4

Aufgabe 4.1 Progressives Alignment (50%)

① Zuerst werden 6 pairwise alignments berechnet:

1 A A T C G
2 A A C G
3 C G
4 A A

$\alpha(1,2)$

```

- A A T C G
- 0 -2 -4 -6 -8 -10
A -2 -1 -3 -5 -7
A -4 -1 -2 -0 -2 -4
C -6 -3 0 1 -1 -3
G -8 -5 -2 -1 0 2
        
```

$p = 0,2$

$\alpha(1,3)$

```

- A A T C G
- 0 -2 -4 -6 -8 -10
C -2 -1 -3 -5 -7
G -4 -3 -2 -4 -6 -4
        
```

$\alpha(2,3)$

```

A A C G
- - C G
        
```

$p = 0,5$

$\alpha(3,4)$

```

- C G
A -2 -1 -3
A -4 -3 -2
        
```

$\alpha(2,4)$

```

A A C G
A A - -
        
```

$p = 0,5$

$\alpha(1,4)$

```

A A T C G
A A - - -
        
```

$p = 0,6$

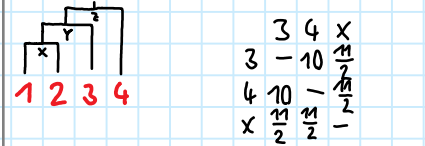
② Es ergibt sich folgende Distanzmatrix:

1	2	3	4
1	-	0,2	0,6
2	0,2	-	0,5
3	0,6	0,5	-
4	0,6	0,5	1

\Rightarrow

1	2	3	4
1	-	2	6
2	2	-	5
3	6	5	-
4	6	5	10

③ Damit wird der Guide Tree erstellt:



④ Anhand des Guide Trees werden die Sequenzen aligned

```

1 A A T C G
2 A A - C G
3 - - - C G
4 - - - A A
    
```

Konsens Alignment:

Sum-of-pairs-Score:

$A A - C G$

-30

auch $AA - CG$ möglich, wenn im Guide Tree 3 u. 4. vertauscht werden

Aufgabe 4.2 Multiple Sequence Alignment mit CLUSTAL OMEGA (30%)

Die zugehörigen Dateien finden sich im GitHub-Repository.

Das Tool CLUSTAL OMEGA besaß die Funktion, sowohl den Guide Tree als auch einen phylogenetischen Baum aus den Sequenzen zu erstellen. Dabei unterscheiden sich die beiden Outputs: Im Guide Tree sind - wie zu erwarten war - Löwe und Amur Leopard sehr eng verwandt, ebenso die Gans und die Amsel. Im Phylogenetischen Baum ist zwischen Amsel und Gans keinerlei Verwandtschaft zu erkennen, die Amsel ist tatsächlich näher mit allen anderen Tieren verwandt (mit Ausnahme der Moschusente). Die Güte des phylogenetischen Baums ist damit anzuzweifeln, weswegen ich mich im Folgenden auf den Guide Tree beziehen werde.

Hier ist zu erkennen, dass alle Vögel in einem Cluster vereint sind und deren nächster Verwandter die Schlange ist, was evolutionär gesehen durchaus Sinn ergibt. Die Säugetiere bilden ebenfalls einen eigenen Cluster, einzig das Siam-Krokodil besitzt in dieser Auswahl keinen näheren Verwandten.

Über MView ließ sich das MSA noch besser darstellen. Hier wurde die Proteinsequenz des Krokodils als Referenz genutzt, da diese Sequenz keine näheren Verwandten in der Auswahl besaß und so gut stark konservierte Regionen erkenntlich werden. Es ist erkenntlich, dass mittels der Hämoglobin beta Untereinheit ein sehr hoher Selektionsdruck wirkt, da viele konservierte Bereiche vorliegen. Viele Aminosäuren sind in keiner einzigen Stelle verändert, es existieren mehrere Bereiche mehrerer Aminosäuren, wo nur maximal in einer Sequenz eine andere Aminosäure eingefügt ist.