

Übungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, WS 2025/26

Prof. Dr. Jens Stoye

Leonard Bohnenkämper

Tutorien: Lennart Finke, Sofie Jans

<https://gi.cebitec.uni-bielefeld.de/teaching/2025winter/sa1>

Übungsblatt 6 vom 04.12.2025

Abgabe bis 11.12.2025 bis 9:30 Uhr per Mail an den Tutor/die Tutorin

Aufgabe 1 (MSA verwandter Sequenzen mit Clustal Omega)

(9 Punkte)

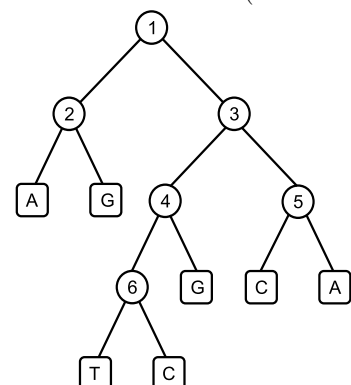
Benutze für diese Aufgabe *Clustal Omega* auf dem EBI Webserver: <https://www.ebi.ac.uk/jdispatcher/msa/clustalo>.

1. Accession.version-Nummern sind eindeutige Kennungen in biologischen Datenbanken (wie beim NCBI: <https://www.ncbi.nlm.nih.gov/>) für DNA-, RNA- oder Proteinsequenzen. Die Accession identifiziert die Sequenz, die Versionsnummer nach dem Punkt zeigt die Version (Updates der Sequenz) an. Lade dir von der NBI-Webseite die Proteinsequenzen des Haemoglobin-Alpha-Gens des Menschen, der Hausmaus, der Schopffente, der Spitzschwanzente und des Atlantischen Lachses heraus. Verwende dazu entweder die *Accession*-Nummern: *CAA00181.1*, *NP_032244.2*, *ACT80500.1*, *ACM67760.1* und *NP_001117134.1* oder die Links im Übungsrepository¹.
2. Berechne mit den Sequenzen ein multiples Alignment mit *Clustal Omega* mit Standardeinstellungen. Beschreibe die Ausgabe kurz.
3. Was kannst du über die Optimalität dieses Alignments sagen?
4. Als letzte Zeile im multiplen Alignment findest du verschiedene Zeichen. Welche sind es und was bedeuten sie?
5. Lasse dir jetzt die Farben im multiplen Alignment anzeigen. Was zeigen sie an? (Eine kurze Erklärung genügt.)
6. Bilde die beiden phylogenetischen Bäume in deiner Abgabe ab. Ist die Haemoglobin-Sequenz des Menschen ähnlicher der von Vögeln oder der von Fischen? Woran kannst du das erkennen?

Aufgabe 2 (Fitch-Algorithmus)

(4 Punkte)

Gegeben sei der nebenstehende phylogenetische Baum. Berechne für den Baum die sparsamste Beschriftung der inneren Knoten mit Hilfe des Fitch-Algorithmus. Gib dabei für die inneren Knoten jeweils die mit ihnen assoziierten Informationen der Bottom-Up- und der Top-Down-Phase an.



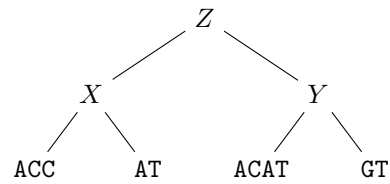
(Bitte wenden!)

¹https://gitlab.ub.uni-bielefeld.de/lbohnkaemper/sqa-ex/-/blob/main/w07/proteins.txt?ref_type=heads

Aufgabe 3 (Progressives Alignment und Baumalignment)

(7 Punkte)

Gegeben sei folgender Baum:



1. Berechne ein progressives Alignment unter Einheitskosten mithilfe des Baums als Guide Tree. Nutze dafür die bereits vorberechneten Alignments $X = \begin{pmatrix} A & C & C \\ A & T & - \end{pmatrix}$ und $Y = \begin{pmatrix} A & C & A & T \\ G & - & - & T \end{pmatrix}$
2. Berechne die Baumalignmentkosten unter dem oben gegebenen Baum des resultierenden Alignments aus Teilaufgabe 1. Solltest du in Teilaufgabe 1 kein Ergebnis haben, nutze folgendes Alignment:

$$Z = \begin{pmatrix} A & C & C & - \\ A & - & T & - \\ A & C & A & T \\ G & - & - & T \end{pmatrix}$$

Tipp: Betrachte dazu jede Spalte einzeln und wende den Fitch-Algorithmus an.

3. Finde (ohne Algorithmus) ein Alignment mit geringeren Baumalignmentkosten. Könnte dieses von einem progressiven Alignment-Algorithmus unter Einheitskosten gefunden werden? Warum/ warum nicht?