

Softwarewerkzeuge der Bioinformatik

Prof. Dr. Volkhard Helms
PD Dr. Michael Hutter, Markus Hollander,
Marie Detzler
Wintersemester 2020/2021

Universität des Saarlandes
Zentrum für Bioinformatik

Übungsblatt 3

Sequenzanalyse: Multiple Sequence Alignment (MSA) und Phylogenie

***Lernziel:** Sie sollen lernen, wie man multiple Sequenzalignments erstellt, wie man diese beispielsweise bzgl. Konservierung interpretiert und für welche Fragestellungen man sie anwenden kann. Weiterhin sollen Sie den Sankoff Algorithmus anwenden und lernen, mit phylogenetischen Bäumen umzugehen.*

Aufgabe 3.1: Homologe Sequenzen, gemeinsame Domänen und phylogenetischer Baum

Tools zur Berechnung von multiplen Alignments: <http://www.ebi.ac.uk/Tools/msa>

- a) Speichern Sie die Sequenz des Proteins **Q38856** zusammen mit 9 homologen Sequenzen im Multi-Fasta-Format.
- b) Finden Sie hochkonservierte Bereiche dieser Sequenzen mit dem Tool Ihrer Wahl.
- c) Ist es immer nötig, dass alle Aminosäuren hoch konserviert sind, um daraus schließen zu können, dass die Proteine homolog sind?
- d) Angenommen Sie möchten das aktive Zentrum eines Proteins lokalisieren, haben aber nur die Proteinsequenz und keine Struktur. Wie kann ein multiples Sequenzalignment dabei helfen, dieses Problem zu lösen?
- e) Erstellen Sie nun ein MSA von 50 homologen Sequenzen mit demselben Tool.
- f) Welche Unterschiede stellen Sie zwischen den beiden Alignments fest?
- g) Schauen Sie sich den phylogenetischen Baum zu den Sequenzen aus 3.1.e) an und finden Sie drei biologische Gruppen (Pflanzen, Pilze, Tiere).

Aufgabe 3.2: Vergleich verschiedener Tools

Die folgenden MSAs wurden mit verschiedenen Tools erstellt:

Tool	Protein	Alignment
ClustalW	FOS_Rat	MMFSGFNADYEASSSRCSSASPAGDSL SY YHSPADSFSSMGSPVNTQDFC
	FOS_MOU	MMFSGFNADYEASSSRCSSASPAGDSL SY YHSPADSFSSMGSPVNTQDFC
	FOS_CHIC	MMYQGFAGEYEAPSSRCSSASPAGDSL TY YYPSPADSFSSMGSPVNSQDFC
	FOSB_MOU	-MFQAFP GDYD-SGSRCSS-SPSAESQ--YLSSVDSFGSPPTAAASQE-C
	FOSB_HU	-MFQAFP GDYD-SGSRCSS-SPSAESQ--YLSSVDSFGSPPTAAASQE-C *:. . . * . : * : . . * * * * * * * : : : * * * . * * * . * : . . . : * : *
MAFFT	FOS_Rat	MMFSGFNADYEASSSRCSSASPAGDSL SY YHSPADSFSSMGSPVNTQDFC
	FOS_MOU	MMFSGFNADYEASSSRCSSASPAGDSL SY YHSPADSFSSMGSPVNTQDFC
	FOS_CHIC	MMYQGFAGEYEAPSSRCSSASPAGDSL TY YYPSPADSFSSMGSPVNSQDFC
	FOSB_MOU	-MFQAFP GDYD-SGSRCSS-SPSAES--QYLSSVDSFGSPPTAAASQE-C
	FOSB_HU	-MFQAFP GDYD-SGSRCSS-SPSAES--QYLSSVDSFGSPPTAAASQE-C *:. . . * . : * : . . * * * * * * * : : : * * * . * * * . * : . . . : * : *
MUSCLE	FOS_Rat	MMFSGFNADYEASSSRCSSASPAGDSL SY YHSPADSFSSMGSPVNTQDFC
	FOS_MOU	MMFSGFNADYEASSSRCSSASPAGDSL SY YHSPADSFSSMGSPVNTQDFC
	FOS_CHIC	MMYQGFAGEYEAPSSRCSSASPAGDSL TY YYPSPADSFSSMGSPVNSQDFC
	FOSB_MOU	-MFQAFP GDYD-SGSRCSS-SPSAESQ--YLSSVDSFGSPPTAAASQE-C
	FOSB_HU	-MFQAFP GDYD-SGSRCSS-SPSAESQ--YLSSVDSFGSPPTAAASQE-C *:. . . * . : * : . . * * * * * * * : : : * * * . * * * . * : . . . : * : *
Clustal Omega	FOS_Rat	MMFSGFNADYEASSSRCSSASPAGDSL SY YHSPADSFSSMGSPVNTQDFC
	FOS_MOU	MMFSGFNADYEASSSRCSSASPAGDSL SY YHSPADSFSSMGSPVNTQDFC
	FOS_CHIC	MMYQGFAGEYEAPSSRCSSASPAGDSL TY YYPSPADSFSSMGSPVNSQDFC
	FOSB_MOU	-MFQAFP GDYD SGS -RCSS SPSA -- ES QYLSSVDSFGSPPTA-AASQEC
	FOSB_HU	-MFQAFP GDYD SGS -RCSS SPSA -- ES QYLSSVDSFGSPPTA-AASQEC *:. . . * . : * : . * * * * * : * : : * * . * * * . * : : : * :

Vergleichen Sie die MSAs miteinander.

- Gibt es Unterschiede bei der Anordnung der Gaps?
- Ändert sich dadurch der Konservierungsgrad der farblich markierten Spalten?

Aufgabe 3.3: Konservierte Motive

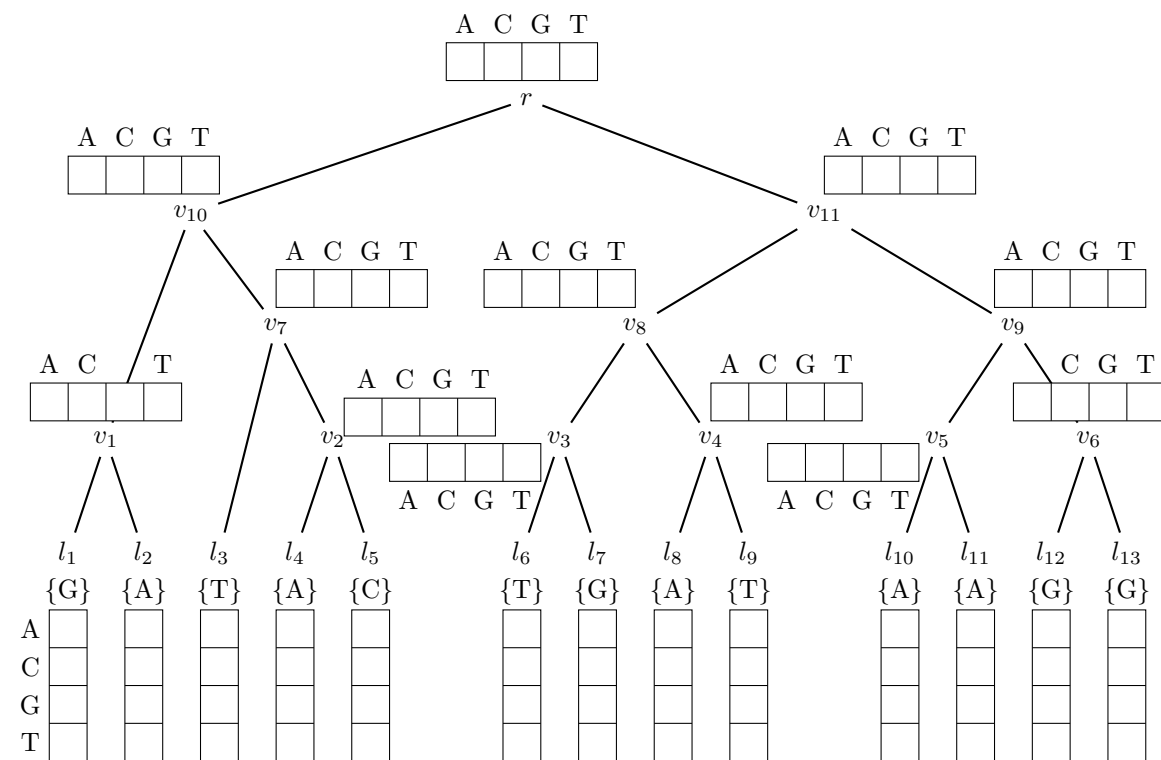
Erstellen Sie mit den Sequenzen der Vorlesungsseite (sequences1.fasta) unter Verwendung von Clustal Omega ein MSA und Sie ununterbrochene, hoch konservierten Bereiche mit mindestens Länge 10. Speichern Sie diese als potentielle Motive für Übungsblatt 4 (ausgehend von FOSB_MOUSE).

Aufgabe 3.4: Outgroup

- Erstellen Sie ein MSA aus den Sequenzen der Vorlesungsseite (sequences2.fasta).
- Ist überall Konservierung zu erkennen?
- Welche Spezies unterscheidet sich von den anderen?
- Erstellen Sie einen phylogenetischen Baum.

Aufgabe 3.5: Sankoff Algorithmus

Welche Base hatte die Urvorgängersequenz wahrscheinlich an der gegebenen Stelle eines Alignments? Verwenden Sie dazu den Sankoff Algorithmus und die folgende Kostenfunktion.



→ Base der Urvorgängersequenz:

Kostenfunktion:

	A	C	G	T
A	0	2	1	2
C	2	0	2	1
G	1	2	0	2
T	2	1	2	0

Viel Spaß!