

Aufgaben zur Bioinformatik

Aufgabe 1: paarweises Sequenzalignment

In einer RNA-Expressionsstudie an menschlichen Zellen wurde folgende Sequenz gefunden.

> Sequenz 1

```
AGCTGCAATTTTTCACATTTTCTAATTCAGAAAGGTTTTCTTATATTAGGGGAAAA
AGTATTTATTTTAATATATAAAATCACTCTGAAAATCACTCTCATAAAAAATGGAGC
GCATGTAAATTTTATCAAAGAAAAATAAACAGGTGAATGGGGGATAGTGATTTTCG
TTTTTCAGCACAGTCTACCTCAGTGTATTGTTAAGATGTGATTCAATCATGGACATC
TTTGAGATTTTCAGAATTCTACCTGGAACCGGTCTGAATCAGGGAACGTGTGTATCAG
CTGATTCGAATGCCAGGGACCAGTAAGAATTTTGAGGGAGGGAGTTGGGATGGAGAA
GGTATGGCCTTTATGCGAGCATAGATCCTTTTCTTCCTGGCTGGTAATATTCTTCTC
TGAGTTTAATCTTCCTTTAAAAAAAATCCTCCATCTATTGTCACTATGTTCCCCAA
ACATAAACTAAGTTCCAGGCTGTCATGATGTATCTGATATATGGGGTAACCCAGCAA
GGTGACCTTCCTTTGGTGAGAGATGGC
```

Finden Sie heraus um welches Gen es sich handelt.

- Welches Programm benutzen sie zu diesem Zweck?
- Welche Suchparameter verwenden Sie und warum?
- Was ist der Name des Gens?
- Entspricht die hier angegebene Sequenz exakt der Sequenz in der Datenbank?
- Können sie ausschliessen dass die gefundene Sequenz ein „Zufallsfund“ war?

Aufgabe 2: paarweises Sequenzalignment

Sie haben die folgenden miteinander verwandten Sequenzen. Versuchen Sie diese Sequenzen auf drei unterschiedliche Arten und Weisen miteinander zu alignen.

- von Hand (verwenden Sie nicht mehr als 5 Minuten)
- mit dem align 2 Sequences Version von BLAST (siehe Tutorial video)
- mit dem auf dem Needleman-Wunsch Algorithmus basierenden Tool auf der Webseite: http://www.ebi.ac.uk/Tools/psa/emboss_needle/nucleotide.html

> Sequenz 1

```
AGTCGTAGACCTGACAGTAGACCTAGATTTACGACAGACGTAGCACATTTGACGAT
```

> Sequenz 2

```
GCTACTAGTCATAGACCCGTACGGACAGTAGTCCTATATTTACGCCAGACGGAGCT
```

- Warum findet der BLAST Algorithmus kein sinnvolles Alignment während der auf dem *dynamic programming* Ansatz beruhende Needleman-Wunsch Algorithmus erfolgreich ist?

- b) Was kann ich an den Suchparametern des BLAST verändern um dennoch ein Alignment zu erhalten?
- c) Welcher Algorithmus gibt das „bessere“ Alignment?
- d) Warum verwendet man den Algorithmus der hier das bessere Alignment produziert nicht für alle Alignment Aufgaben (z.B. suche in grossen Datenbanken)

Aufgabe 3: MSA und Phylogenetischer Baum

Ein menschliches Hämoglobin Molekül setzt sich aus vier einzelnen Polypeptidketten zusammen. Die „Normalform“ besteht dabei aus zwei alpha und zwei beta Ketten die von den entsprechenden Genen kodiert werden.

In unterschiedlichen Tieren finden wir auch mehrere Varianten der Hämoglobin Polypeptidketten aus denen die funktionale Form des Proteins zusammengesetzt ist. Evolutionär bestehen nun zwei Möglichkeiten. Entweder die unterschiedlichen Spezies haben alle ein Ur-Hämoglobin geerbt und die unterschiedlichen Varianten sind dann durch Genduplikation entstanden. Oder die verschiedenen Varianten von Hämoglobinen sind bereits in einem gemeinsamen Vorgänger Organismus entstanden und wurden dann parallel vererbt.

Nutzen Sie öffentlich zugängliche Datenbanken und Softwaretools um heraus zu finden welcher der beiden Prozesse zu den in der heutigen Zeit existierenden Hämoglobinen geführt hat.

Als „Outgroup“-Sequenz verwenden wir hier menschliches Myoglobin.

Spezies	Hämoglobin Alpha	Hämoglobin Beta
Mensch (homo sapiens)	P69905	P68871
Maus (mus musculus)	P01942	P02088
Rind (bos Taurus)	P01966	P02070
Huhn (gallus gallus)	P02001	P02112
Schimpanse(Pan troglodytes)	P69907	P68873
	Myoglobin	
Mensch (homo sapiens)	P02144	

Verfolgen Sie die folgenden Lösungsschritte:

- Gehen Sie zur NCBI Webseite und suchen Sie mit der Suchkategorie „Protein“ nach den oben in der Tabelle angegebenen Proteinsequenzen. Um Missverständnisse zu vermeiden sind die Accession Nummern angegeben.
Tip: Wenn Sie auf der Results Page auf den „FASTA“ Link klicken kommen Sie

zu einer Webpage die nur die Sequenz enthält. Alternativ können Sie die entsprechenden Sequenzen auch aus der KEGG Datenbank herunter laden.

- Mit Hilfe eines Text Editors erstellen Sie eine File im FASTA Format die alle Proteinsequenzen enthält. Dabei ist es vielleicht sinnvoll den recht komplizierten Namen der Sequenz in der Kommentarzeile durch einen leicht verständlichen Namen zu ersetzen (z.B. durch >Mensch_alpha). Diese Namen benutzt das Phylogenie Programm dazu um den Baum zu beschriften.
Tip: Achten Sie darauf dass sie die einzelnen Wörter in dem Namen den Sie den Sequenzen geben mit einem „_“ verknüpfen.
- Gehen Sie zur Webseite <http://www.ebi.ac.uk/Tools/msa/clustalo/> wo Sie eine online Version von CLUSTAL Omega finden. (CLUSTAL Omega ist eine zeitgemässe Version von CLUSTALW). Kopieren Sie die Sequenzen in das Input Feld unter Step 1. Lassen Sie alle Settings auf dem Defaultwert. Und klicken Sie auf „Submit“.
- Nach einigen Sekunden erscheint das Resultat Fenster mit dem MSA. Inspizieren Sie das Alignment. Es sollte alles gut aussehen...
- Klicken Sie nun auf den Knopf „Send to Simple_Phylogeny“
- Wählen Sie in Step 2 die Clustering-Method „UPGMA“. Alle anderen Settings können im Default bleiben.
- Klicken Sie „Submit“ .
- Wie sieht der Phylogenetische Baum aus? Wie Clustern die Sequenzen, nach Spezies oder nach Hämoglobin Typ? Was bedeutet dies für unsere Frage nach der Evolution der Hämoglobin Gene?

Falls Sie das Thema der evolutionären Prozesse von Hämoglobinen interessiert finden Sie auf Moodle einen weiterführenden Artikel „Evolution of Hemoglobin and its Genes“ (Harrison, Cold Spring Harb Perspect Med 2012) der dieses Thema weiterführend behandelt.