

# Bioinformatik I - Uebung

## Phylogenetische Analyse

*Wenn nicht anders angegeben verwende die Standard-Einstellungen der Programme*

Hintergrund: Die Schwämme (Phylum Porifera) gehören zu den den ältesten lebenden multizellulären Lebewesen auf unserem Planeten. Ökonomisch und ökologisch gesehen sind sie die mit am weitesten entwickelten Lebensformen auf diesem Planeten.



(Quelle: <http://de.wikipedia.org/wiki/Bild:Vaasspons2.jpg>)

Aufgabe: Da du begeisterter Schwammforscher bist, mietest du dir ein Uboot, um nach Schwämmen zu tauchen und diese anhand phylogenetischer Bäume zu klassifizieren.

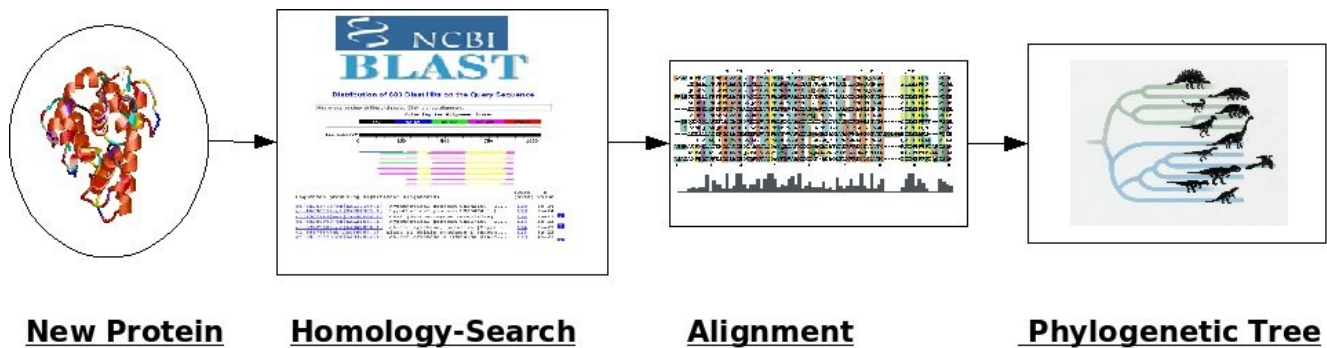
Auf dem Meeresgrund hast du nun Gewebeproben von einem Schwamm genommen, konservierst sie in Ethanol und sequenzierst ein Gen nach der Rückkehr im heimischen Labor.

Folgende Gensequenz konntest du dabei ermitteln:

>unknown\_sponge

```
GACATTTGCGGTATTGTAAAGCCAGGATCTTTGAGCGCGTTGGCAAGCGTACACCGCTGGCCGTGCGCTTC
TCCACTGTTGGCGGAGAATCCGGTAGCGCCGATACCGTCCGTGATCCCCGTGGCTTTGCTGTCAAGTTCTAC
ACTGAGGAAGGCATCTGGGATTTGGTAGGAAACAACACTCCGATTTTCTTCATCCGTGATCCCATACTGTTT
CCCAGCTTCATTCACACGCAGAAGCGCAATCCTCAAACACTCATTTAAGGACGCTGATATGTTTTGGGACTTC
ATTTCCCTTCGCCCCGAGACCACCCATCAGGTGTCGTTTCTCTTCAGTGACCGTGGCATCCCTGACGGTTAT
CGTCACATGAACGGCTACGGCAGCCACACGTTTAAGCTAGTCAACAAGCAGGACAGGCCAGTGTACTGCAAG
TTCCACTACAAGACCGACCAGGGCGTTGAGAACATGCCAGTGGACCGCGCTGCTCAGCTGGCCAGCGCCGAC
CCGGACTACAGCATCCGTGACCTCTTTGAGAGCATCGCCAGAATAACAGCCCATCGTGGACTATGTACTTG
CAAGTTATGACATTTGAGCAAGCTGAGCAGCTCAAGTGGAACCCCTTTGACTTGACCAAGATCTGGCCGCAC
AAGGAGTTTCCGCTGATCAGGTTGGCAAGCTGGTGTGAACCGCAACCCGAGGAACTATTTGCCGAAGTG
GAGCAGCTGGCCTTCTCTCCCGCGCACATGGTGCTGATTGAGCCGTGCGCCAGACAAGATGTTGCAGGGC
```

Ein typischer allgemeine Ablauf einer phylogenetischen Analyse ist wie folgt: Nimm die DNA Sequenz, ueber die du mehr herausfinden moechtest, suche mittels BLAST in Genbank nach homologen Genen (in unserem Fall bei den Schwämmen), berechne daraus ein multiples Sequenzalignment und rekonstruiere anschliessend einen phylogenetischen Baum.



## Aufgabe 1: Homologensuche

Bei dieser Aufgabe werden wir zu einer DNA Sequenz Homologe suchen

1. Gehe zu <http://www.ncbi.nlm.nih.gov/BLAST/> und fuehre eine BLAST-Suche mit dieser Sequenz in der **non-redundant** Datenbank durch. Wähle **blastn** unter **Program selection**
2. Schau dir sowohl die Ergebnisseite der Suche als auch den Taxonomy Bericht dazu an (Link **Taxonomy**) und versuche dir einen Überblick über das Ergebnis der BLAST-Suche zu verschaffen.
3. Auf der Ergebnisseite siehst du Treffer vieler verschiedener Arten (Der Artnamen steht neben jedem Treffer). Wähle alle Schwämme und als Außengruppe "Asterina miniata" (Starfish – Seestern).
4. Oeffne jeden der Treffer in einem neuen Fenster/Tab und lass dir jeweils das FASTA anzeigen.
5. Oeffne nun einen Texteditor (z.B. Gedit) und kopiere alle Sequenzen untereinander.
6. Passe nun den Header jeder Sequenz wie in der letzten Uebung an.

Beispiel:

```
>gi|71370875|gb|DQ087467.1| Leucosolenia sp. KP-2005 catalase mRNA, partial cds
```

zu

```
>Leucosolenia_sp_gi_71370875
```

Warum sollte man das tun?

### Fragen:

Was war der beste BLAST-Treffer?

In wievielen Schwämmen wurde ein potentielles homologes Gen gefunden?

Welches Gen scheint du sequenziert zu haben?

## **Aufgabe 2: Multiples Sequenz Alignment**

Die bei Aufgabe 1 erhaltenen DNA Sequenzen sollen nun mit ClustalW aligniert werden (vgl. Aufgaben zu Alignment).

1. Klicke im Ergebnis auf **Alignment file**, um dir das Alignment anzeigen zu lassen.
2. Kopiere die komplette Seite in einen Texteditor.

### Fragen:

Mit welcher anderen Sequenz erzielte die unbekannte Sequenz den höchsten Score?

## **Aufgabe 3: Einen phylogenetischen Baum konstruieren**

Um aus einem Alignment einen Baum zu berechnen, gibt es verschiedene Methoden.

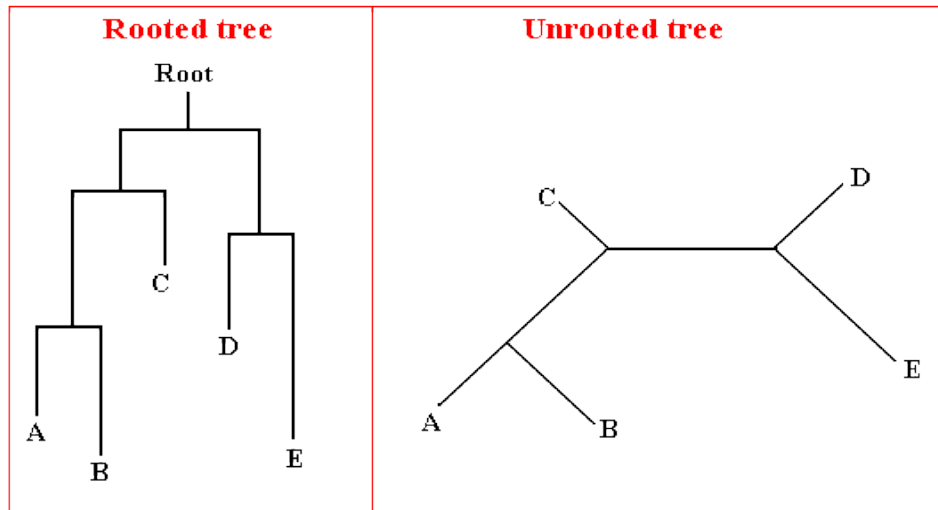
Wir werden im folgenden einen Maximum Likelihood Baum rekonstruieren.

1. Gehe auf das Webinterface des Programmpakets **Phylip** unter <http://bioweb.pasteur.fr/seqanal/phylogeny/phylip-uk.html>
2. Klicke **fastDNAmI** unter **Programs for molecular sequence data**, um aus dem Alignment eine Distanzmatrix zu berechnen.
3. Gib deine Emailadresse (muss keine gueltige sein) und das Alignment ein und druecke **Run fastDNAmI**  
alternative ULR: <http://cbi.labri.fr/outils/Pise/fastdnaml.html>
4. Nachdem das Ergebnis da ist, kannst du dir unter **treefile** den berechneten ungewurzelten (unrooted) phylogenetischen Baum ansehen.
5. Speichere das **treefile** in einem Editor (oder lass das Browserfenster einfach offen), da du es später noch brauchen wirst

## **Aufgabe 4: Rooted vs. unrooted Tree**

Ein *gewurzelter* (rooted) phylogenetischer Baum ist ein gerichteter Baum mit einem einzelnen Knoten, der dem nächsten gemeinsamen Vorfahren aller Organismen

(Gene) im Baum entspricht, während ein *ungewurzelter* (unrooted) Baum keinen ausgezeichneten nächsten gemeinsamen Vorfahren besitzt, sondern soll lediglich die Verwandtschaftsnähe oder -ferne der einzelnen Arten darstellen.



(Quelle: <http://www.bioinf2.leeds.ac.uk/msc2004/trees.gif>)

Da wir aber konkret an eben dieser Vorgänger-Nachfolger-Beziehung interessiert sind, werden wir den ungewurzelten Baum wurzeln

Um die Unterschiede zwischen gewurzelten und ungewurzelten Bäumen deutlich zu machen fahren wir wie folgt mit dem in Aufgabe 3 berechneten Baum fort:

1. Klicke auf der Resultseite von Aufgabe 3 **Run the selected program on outtree**. Auf der folgenden Seite klickst du auf **Advanced consensus form** und gibts du unter **Other options** die Nummer der Outgroup-Sequenz im **treefile**.
2. Drücke **Run consensus** um den Baum zu wurzeln
3. Speichere auch den gewurzelten Ergebnisbaum (outtree).

## Aufgabe 5: Den Baum betrachten

In dieser Aufgabe wollen wir uns die erzeugten Bäume anschauen und analysieren.

1. Gehe zu <http://cgi-www.daimi.au.dk/cgi-chili/phyfi/go> und kopiere den Baum in das Textfeld.
2. Druecke **Draw**, um den Baum anzeigen zu lassen.  
(Bei Bedarf kannst du den Baum nun als png speichern.)
3. Mache das gleiche nun fuer den anderen gespeicherten Baum aus.

Fragen:

Handelt es sich bei dem sequenzierten Tier um einen Schwamm?

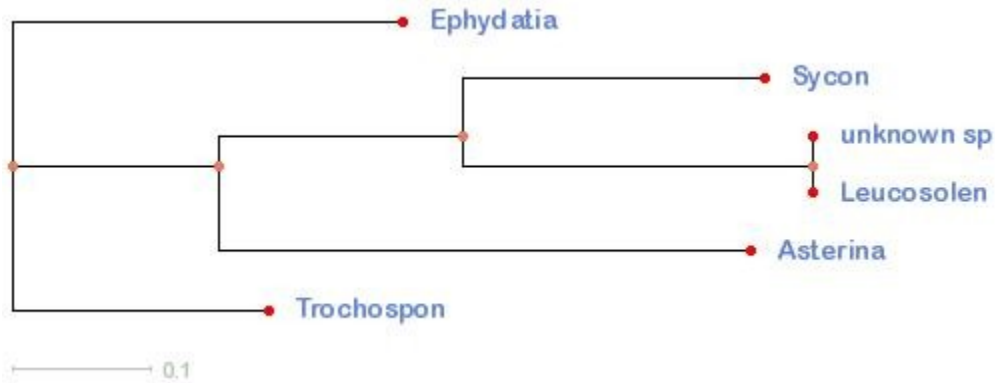
Mit welchem Schwamm ist das unbekannte Tier am nächsten verwandt?

### Anmerkung

Der ungewurzelte Baum stellt nur eine Gruppierung der Daten dar.

Es ist nicht möglich eine Vorgänger-Nachfolger-Beziehung abzulesen.

Die Schwämme bilden in diesem Baum keine monophyletische Gruppe und widersprechen damit bekanntem Wissen. Generell sollte man gewurzelte Bäume bevorzugen, da die Vorgänger-Nachfolger-Beziehung sehr wichtig bei der Analyse von Verwandtschaftsbeziehungen ist.



Im gewurzelten Baum lassen sich Vorgänger-Nachfolger-Beziehungen ablesen

So lässt sich beispielsweise sagen, dass A der gemeinsame Vorfahre von den Schwämmen und der Seesterne war. B zeigt, dass die Schwämme eine monophyletische Gruppe bilden und C zeigt, dass es sich bei dem gesuchten Organismus tatsächlich um einen Schwamm hat und dieser am nächsten mit dem Schwamm Leucosolenia ist.

