**M4: Stammbaumrekonstruktion mit MEGA – Wie aus einem Alignment ein Stammbaum wird**

Das Programm MEGA (Molecular Evolutionary Genetics Analysis) bietet neben der automatischen Erstellung und Editierung von Alignments (M3) zudem die Möglichkeit aus den fertigen Alignments Stammbäume zu rekonstruieren und die Knotenpunkte in diesen Stammbäumen auf ihre statistische Verlässlichkeit über die Methode des Bootstrappings hin zu überprüfen (vgl. Knoop & Müller 2006, S. 119f.). Es existieren verschiedene Verfahren, nach denen die Erstellung eines Stammbaums erfolgen kann (vgl. Knoop & Müller 2006, S. 70f.). Der Einfachheit halber erfolgt hier die Berechnung mit dem Distanzverfahren Neighbor Joining. Grundprinzip der Distanzmethode ist die Bestimmung von Distanzen jeweils zweier Sequenzen eines Alignments. Dabei handelt es sich um die Ermittlung eines Distanzmaßes, welches das Verhältnis der Anzahl der unterschiedlichen Nukleotide zweier Sequenzen zur Gesamtzahl der verglichenen Nukleotide beschreibt (vgl. Hall 2001, S. 74).

Beispielsweise werden für das multiple Alignment mit den drei Sequenzen A, B, und C die Anzahl der evolutionären Änderungen in den Nukleotidpositionen paarweise bestimmt, d.h. die Sequenz A und B unterscheiden sich in einer Position, der Position 6, voneinander, während sich die Sequenz C in vier Positionen, den Positionen 3, 6, 7 und 8, von der Sequenz A sowie der Sequenz B unterscheidet.

Das Distanzmaß gibt Auskünfte über den Verwandtschaftsgrad. So zeigt beispielsweise ein Sequenzpaar, welches sich in nur 10% der Nukleotidpositionen unterscheidet einen näheren Verwandtschaftsgrad an als ein Sequenzpaar, welches sich in 40% der Nukleotidpositionen unterscheidet. Die ermittelten Distanzen aller paarweisen Sequenzvergleiche werden in einer Distanzmatrix zusammengefasst (vgl. Hall 2001, S. 74).

In dem Beispiel unterscheiden sich die Sequenzen A und B in einer von 10 verglichenen Nukleotidpositionen, voraus ein sich ein Distanzmaß von 0.1 ergibt. Sequenz C unterscheidet sich von der Sequenz A und B in jeweils 4 von 10 Positionen und besitzt damit ein Distanzmaß von 0.4

Die Sequenzen mit der geringsten Distanz werden gebündelt und dieser Vorgang nach zunehmender Distanz mit den übrigen Sequenzen wiederholt (vgl. Bromham 2008, S. 249ff.).

In dem Beispiel sind die Sequenzen A und B diejenigen mit der größten Ähnlichkeit und werden zu einem Schwestertaxon gebündelt. Die Sequenz C weist mehr Unterschiede zu den Sequenzen A und B auf, als diese beiden untereinander. Folglich wird das Taxon C als Schwestertaxon zu dem gebündelten Schwestertaxon A&B gesetzt.

Aus den gebündelten Sequenzen lässt sich der Stammbaum rekonstruieren, der über die Ähnlichkeiten in der Nukleotidabfolge der betrachteten Taxa deren verwandtschaftliche Beziehung widerspiegelt. Die Länge der Äste des Stammbaumes zeigt in der Regel die Anzahl der evolutionären Unterschiede (Distanzen) zwischen den Sequenzen an (vgl. Bromham 2008, S. 249ff.).

**Anleitung: Wie wird aus einem Alignment ein Stammbaum erstellt?**

**Schritt 1:** Für die nachfolgende Analyse und Stammbaumrekonstruktion ist es erforderlich das editierte und als mas-Dateiformat gesicherte Alignment im Alignment-Explorer über das Data-Menü und „Export Alignment“ ins MEGA-Format zu exportieren (Abbildung 1). Dabei besteht die Möglichkeit dem Alignment einen Titel zu geben. Zudem öffnet sich ein Fenster mit der Frage „Protein-coding nucleotide sequence data?“, die mit „Yes“ bestätigt wird, da eine proteincodierende Nukleotidsequenz analysiert wird.

Abbildung 1: Export des Alignments in ein MEGA-Format.

**Schritt 2:** Starten Sie das MEGA-Programm und laden Sie das als MEGA-Format gespeicherte Alignment über das Menü „File“ unter „Open A File/Session…“ in das Programm. Im Hauptfenster von MEGA wird das geladene Alignment über das in Abbildung 2 markierte Symbol angezeigt. Durch Anklicken dieses Symbols öffnet sich das „Data Explorer“-Fenster, welches für das Alignment verschiedene Darstellungsoptionen offen hält, die für unsere Analyse jedoch nicht relevant sind.

Abbildung 2: Anzeige einer geöffneten MEGA-Datei im Hauptfenster von MEGA.

**Schritt 3:** Zum Starten der Stammbaumberechnung wird über den Menüpunkt „Phylogeny“ die Option „Construct/Test Neighbor-Joining Tree…“ ausgewählt (Abbildung 3). Es öffnet sich ein Fenster mit der Nachfrage „Would you like to use the currently active data?“, die mit „Yes” bestätigt wird. Des Weiteren öffnet sich eine Rasterkarte „Analysis Preferences“, in der die folgenden Parameter eingestellt werden. Zur Ermittlung der statistischen Wahrscheinlichkeit der Knotenpunkte im Stammbaum, wird unter „Test of Phylogeny“ die Option „Bootstrap method“ ausgewählt und die Wiederholungsanzahl der Testmethode „No. of Bootstrap Replication“ auf 2000 hochgesetzt (Abbildung 4). Es ist darauf zu achten, dass alle drei Codonpositionen sowie die nichtcodierenden Abschnitte berücksichtigt werden (Abbildung 4). Alle weiteren Voreinstellungen bleiben erhalten. Über die Anwahl „Compute“ wird der Stammbaum berechnet.

Abbildung 3: Anzeige einer geöffneten MEGA-Datei im Hauptfenster von MEGA.

Abbildung 4: Einstellung der Parameter für die Stammbaumberechnung.

**Schritt 4:** Nach einer kurzen Berechnungszeit erfolgt die Darstellung des Stammbaums im „Tree Explorer“-Fenster, welches verschiedene Möglichkeiten der graphischen Darstellung bietet, wie z.B. das Drehen und Spiegeln von internen Ästen an Knotenpunkten über die Formatierungssymbole in der senkrechten Symbolleiste. Zudem wird in der Legende des Stammbaums das Distanzmaß, der prozentuale Anteil an Nukleotidaustausche einer Sequenz (Substitutionsrate), und an den Knotenpunkten die Wahrscheinlichkeit der Aufspaltung, der so genannte Bootstrap-Wert, in Prozent angegeben (Abbildung 5). Je höher der Bootstrap-Wert, desto zuverlässiger ist die Aussage über die Aufspaltung einer Art in zwei Schwestertaxa.

Abbildung 5: Darstellung des Stammbaums im „Tree Explorer“-Fenster. Markiert sind der Bootstrap-Wert und die Substitutionsrate. Anmerkung: Diese Abbildung werde ich durch einen fiktiven Stammbaum ersetzen, da sonst das Ergebnis vorweggenommen wird.

**Schritt 5:** Speichern Sie den erstellten Stammbaum im „Tree Explorer“ über das Menü „File“, dann „Save Current Session“ als mts-Datei ab und über den Menüpunkt „Image“, dann „Save as PNG file“, als Bilddatei ab.

**Aufgaben:**

1. Beschreiben Sie die Verwandtschaftsverhältnisse in dem berechneten Chamäleonarten-Stammbaum unter Berücksichtigung der Sequenzunterschiede (Distanzmaß). Welche der Arten bilden Schwestertaxa?
2. Analysieren Sie den berechneten Chamäleonarten-Stammbaum in Hinblick auf die Fragestellung nach dem Ursprungsort (Afrika oder Madagaskar) der Familie Chamaeleonidae. Treffen Sie eine begründete Entscheidung hinsichtlich des hypothetischen Ursprungsortes der Chamaeleonidae und vergleichen Sie Ihr Ergebnis mit Ihrer eingangs aufgestellten Hypothese.
3. Vergleichen Sie die beiden berechneten Stammbäume (auf Basis mitochondrialer und nuklearer DNA) unter Berücksichtigung der Bootstrap-Werte. Lässt sich aus beiden Stammbäumen dieselbe Aussage bezüglich des hypothetischen Ursprungsortes der Chamaeleonidae ableiten?
4. Skizzen Sie einen hypothetischen Stammbaum der Chamaeleonidae, der den gegensätzlichen Ursprungsort prognostiziert als den von Ihnen aus Ihrem berechneten Stammbaum abgeleiteten Ursprungsort.