

# Ausarbeitung zur Implementierungsaufgabe 1 – AMBI 2021

Magdalena Weber und Martin Brand

11. Juni 2021

## Zu Aufgabe 2.1

Laufzeiten zu Berechnung der Distanz-Matrizen mittels der Hamming-Distanz und der Levenshtein-Distanz für die tRNA-Sequenzen von *Aquifer aeolicus*:

- Hamming: 16.9548ms
- Levenshtein: 29.4084s

Beachte: Da die Hamming-Distanz für Sequenzen unterschiedlicher Länge nicht definiert ist, wird sie über die minimale Länge beider Sequenzen berechnet (wenn die Längen 10 und 11 sind, wird die Hamming-Distanz über die ersten 10 Symbole berechnet).

## Zu Aufgabe 2.2

Laufzeiten der hierarchischen agglomerativen Cluster-Algorithmen mit den zwei verschiedenen Distanz-Metriken für die tRNA-Sequenzen von *Aquifer aeolicus*:

	UPGMA	Neighbor-Joining
Hamming	14.9428ms	41.9304ms
Levenshtein	14.0020ms	46.8754ms

Beachte: Die angegebenen Laufzeiten beziehen die Berechnung der Distanz-Matrizen nicht ein, weil sonst die Unterschiede kaum noch deutlich würden. Für die gesamte Laufzeit addiert man einfach die Laufzeit hier mit dem Wert für die jeweilige Distanz (Hamming oder Levenshtein) aus Aufgabe 2.1.

## Zu Aufgabe 2.3

Man sieht natürlich direkt, dass der Neighbor-Joining-Algorithmus mehr als doppelt so viel Laufzeit benötigt. Dieser Punkt spricht für den UPGMA-Algorithmus.

Der Nachteil ist aber, dass er nach dem Prinzip der *molecular clock hypothesis* funktioniert. Das bedeutet, man geht davon aus, dass evolutionäre Verzweigungen in immer gleichen zeitlichen Abständen kommen. In der Realität trifft das jedoch nicht zu. Evolution geschieht ungleichmäßig und nicht mit einer konstanten Rate. Also berechnet der Neighbor-Joining-Algorithmus den wahrscheinlicheren Baum.