

## Algorithmen und Datenstrukturen in der Bioinformatik

### Neuntes Übungsblatt WS 14/15

Abgabe Donnerstag 12:00

#### T-Coffee

Installieren Sie das multiple-Alignment-Programm T-Coffee (es gibt leider keine Windows Version, Sie müssen also Mac oder Linux benutzen). Nach der Installation können Sie in die Konsole `t_coffee -version` eingeben um Ihre erfolgreiche Installation zu überprüfen (evtl. müssen Sie die Konsole neu starten). Gehen Sie auf die Balibase-3 Webseite und laden Sie den Testdatensatz BB20030 1ubi\_ref2 von Referenz 2 herunter. Wahrscheinlich müssen sie die gesamte Balibase herunterladen, Sie werden aber nur die zwei Dateien BB20030.msf und BB20030.tfa brauchen (aus BALiBASE3.0.tar.gz).

- Finden Sie auf der Balibase Webseite heraus, um welche Proteine es sich aus dem Datensatz handelt (falls es geht).
- Verwenden Sie den Standard-Modus von T-Coffee um das multiple Alignment von BB20030.tfa zu bestimmen.
- Verwenden Sie einen anderen Algorithmus, diesmal Muscle. Speichern Sie das Ergebnis. Wie unterscheidet sich Muscle, ein iterativer Algorithmus, von T-Coffee (etwa 100 Worte)?
- Wie unterscheiden sich die Ergebnisse? Benutzen Sie die Vergleichsfunktion im T-Coffee um beide Ergebnisse zu vergleichen: `t_coffee -other_pg aln_compare`. Der `sp` Vergleichsmodus zählt wie viele Matches aus dem ersten Alignment auch in dem zweiten Alignment vorkommen.
- In der Praxis können wir nicht wissen, welches Alignment das beste ist. In unserem Beispiel kennen wir aber das wahre Alignment, welches mit Strukturinformation und Expertenkenntnissen generiert wurde. Vergleichen Sie beide Alignments mit der Referenz BB20030.msf und bestimmen Sie, welches Alignment besser abschneidet.