Übung 4

*Von Lucas Niño*

Aufgabe 2

Nach der Aufbereitung der Daten aus der Datei Test\_100.fasta analysiert GADEM in zwei Schritten den Datensatz. In der ersten Zyklus werden die 3 häufigsten 4-,5- Mere gesucht. Dazu werden Consensusmotive gesucht und deren statistische Übereinstimmung angegeben. “0.xxx fitness“ und die Hammingentfernung angegeben (-Zahl)

Außerdem werden vorher noch die Standartvariablen, nach denen gadem operiert angegeben. (diese könnten auch als weitere Argumente in gadem() geändert werden)

Anschließende können noch die Consensussequenzen gefiltert werden und die pwm als Sequenzlogo dargestellt werden.

Aufgabe 3

Es müssen Sequenzen aus DNS-Nucleotidcodes enthalten sein, da sonst der Befehl readDNAString nicht funktioniert. Außerdem müssen sie in Sequenzen geteilt sein, da sonst ein paralleler Vergleich von Consensusmotiven auf verschiedenen Sequenzen nicht möglich ist.

Hier wird wie in dem R-Code eine pwm als Sequenzlogo ausgegeben, jedoch wird hier noch die Auftretenshäufigkeit mit signifikanter Übereinstimmung in den 500 wichtigsten Sequenzen dargestellt. (499/500)

Aufgabe 4

Es wurde wieder wie in Aufgabe 2 das SeqLogo geplottet, jedoch konnte diesmal nur im ersten Zyklus eine Consensussequenz gefunden werden.