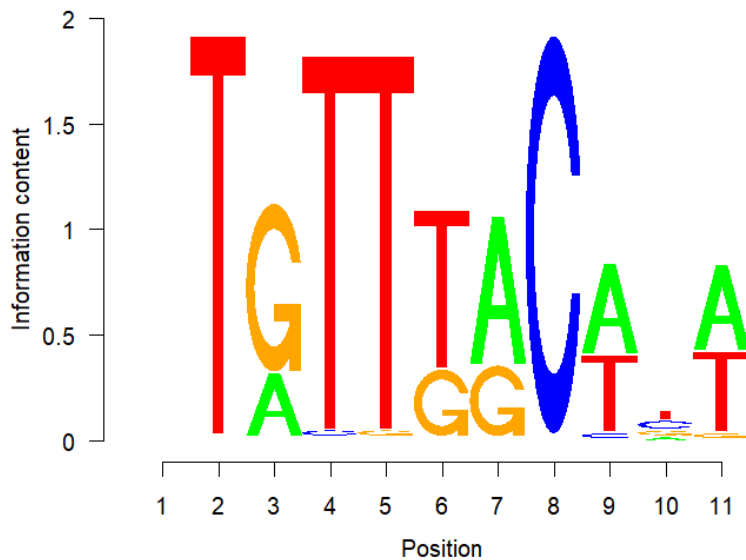


4.2.

Dieses Programm erzeugt über ein in der eingelesenen Library integriertes Muster (GADEM) am Ende ein Motiv. Dieses zeigt die jeweiligen Wahrscheinlichkeiten des Vorkommens der Basen bildlich. Grundsätzlich sucht es in vielen Sequenzen spezifische Sequenzen heraus, die öfter vorkommen. Dabei berechnet es Basenabweichungen ein, was im Motiv zu erkennen ist. (Größe etc. zeigt die Wahrscheinlichkeit der Basen an den jeweiligen Stellen)



Motiv

4.3.

Auf der Website ist auch ein Motiv zu sehen. Unterscheiden tun sich die beiden Motive in den auftretenden Basen (also grundsätzlich wahrscheinlich auch in der zugrunde liegenden Sequenz). Auf der Website sind die ersten Paar Basen scheinbar sehr variabel in den einzelnen gefundenen Sequenzen, wohingegen in unserem Motiv die ersten Paar Basen (vor allem die T's) ziemlich dominant sind, dafür aber am Ende wahrscheinlich die gefundenen Sequenzen keine eindeutige dominante Base zeigen.

In unserer Datei scheinen viele 11-mere vorhanden zu sein.

4.4.

Um unsere Sequenzen in diesem Programm zu nutzen habe ich mir das Package „seqinr“ gedownloadet um mit dem Befehl „write.fasta“ ein neues Fasta file zu erstellen und dieses dann in den Programmcode einzufügen. Leider gibt mir R die ganze Zeit nur Fehlermeldungen z.B. dass die Funktion „write.fasta“ nicht vorhanden ist.