

Aufgabe 2:

Human T-cell leukemia virus type I

Die ersten 100 Basenpaare:

ggctcgcatc tctcctcac gcgcccgcg ccttaccta ggccgccatc cagccgggt gagtcgcgt ctgccgcctc
ccgctgtgg tgcctctga

Aufgabe 3:

Die ersten 30 AS des 1. 5'3' Frames:

GSHLSFTRPPPYLRPPSTPVESRSAASRLW

Um ein sinnvolles Ergebnis zu erzielen werden die AS aus dem 3. Frame verwendet, da es sonst zu keiner Übereinstimmung kommt (s. Aufgabe 4)

- Die Suche in Aminosäuresequenzen ist der Suche in Genomsequenzen vorzuziehen, da jeweils 3 Basen für eine Aminosäure codieren und sich die zu untersuchende Sequenz somit verkürzt. Zudem codiert nicht jede Base für eine Aminosäure, denn schon allein durch die Wobble Position fällt oft die dritte Base weg, wodurch die Aminosäuresequenz endgültiger und unveränderbarer ist.
- Es ist sinnvoll, alle 6 möglichen Übersetzungs-Frames zu untersuchen, da nicht klar ist, welcher Frame letztendlich verwendet wurde. Um alle Wahrscheinlichkeiten zu betrachten muss beachtet werden, dass es durch die Verschiebung des Leserasters letztendlich 3 Möglichkeiten für den Organismus anzusetzen und nicht klar ist, wo genau die Translation startet.

Aufgabe 4

Folgende Sequenz wurde verwendet:

MGQIFSRASPIPRPPRGLAAHHWLNFLQA

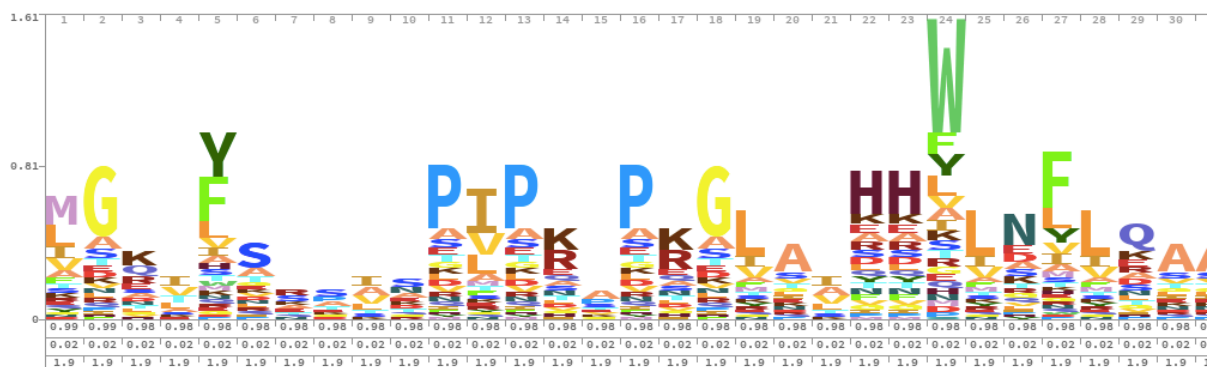


Abbildung 1: HMM Logo[1]

Das Logo stimmt grob mit der Sequenz überein, beispielsweise bei 1-2 ist die Sequenz MG auch im Logo zu finden. Des Weiteren erkennt man bei 10-13 im Logo die Sequenz PIP. Auch die die Sequenz HHW ist durch die Größe ihrer Buchstaben bei 22-24 im Logo vorhanden. Andere Teile der Sequenz wie GLAA sind zwar auch im Logo zu finden, allerdings ist dort das letzte A durch ein T ersetzt (18-21). Auch zu Beginn der Sequenz ist die Wahrscheinlichkeit an der 5. Stelle für Y größer als für F, letztendlich kommt in der Sequenz aber F vor. Somit entspricht das Logo in etwa der Sequenz.

5. Aufgabe

Haloarcula hispanica pleomorphic virus 2

Die ersten 100 Basenpaare:

```
atgagttctg aacctgcccc tggcggggcg gactctggc caccacagtc ttactccgcg ctggacagtt ccgggtctgt  
acctgtcgaa tcggtccctg
```

Die ersten 30 AS aus dem ORF aus 5'3' Frame 1:

MSSEPAPGGADSRPPQSYSALDSSGSVPVE

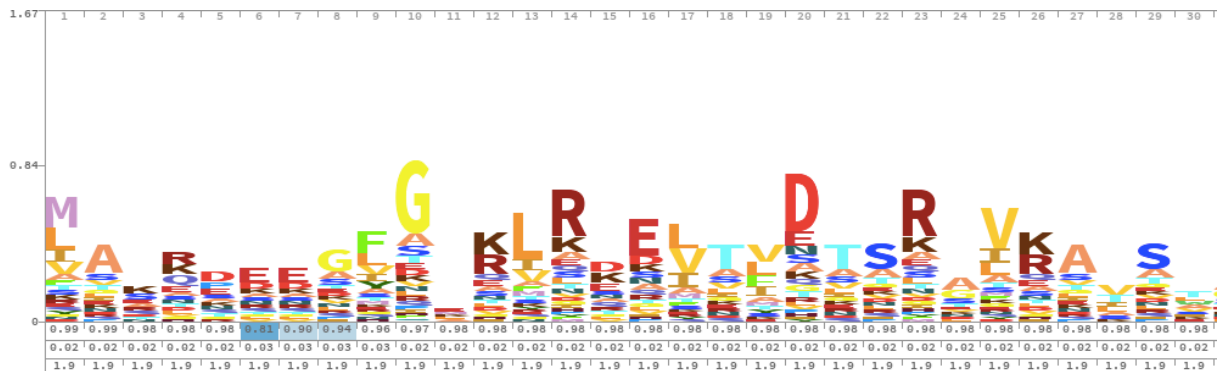


Abbildung 2: HMM-Logo^[2]

Schon bei der zweiten Aminosäure ist in der Abbildung die Wahrscheinlichkeit für A größer als für das in der Sequenz vorkommende S. Erst an 8. Stelle stimmt das G überein, danach ist zwar meistens eine gewisse Wahrscheinlichkeit für die jeweilige Aminosäure im Logo abgebildet, wirklich ähnlich sind sich die Sequenz und das Logo allerdings nicht.

[1] <http://pfam.xfam.org/family/PF02228.15#tabview=tab4> (3.7.2018, 11:03 Uhr)

[2] <http://pfam.xfam.org/family/PF07232.10#tabview=tab4> (3.7.2018, 11:55 Uhr)