BIOINFORMATIK ÜBUNG 5 VON PAUL WULF

Aufgabe 2: die Genom-Sequenz von "Human T-cell leukemia virus type I" erste $100~\mathrm{BP}$:

ggctcgcatc tctccttcac gcgcccgccg ccttacctga ggccgccatc cacgccggt gagtcgcgtt ctgccgcctc ccgcctgtgg tgcctcctga actacgtccg ccgtctaggt

Aufgabe 3:

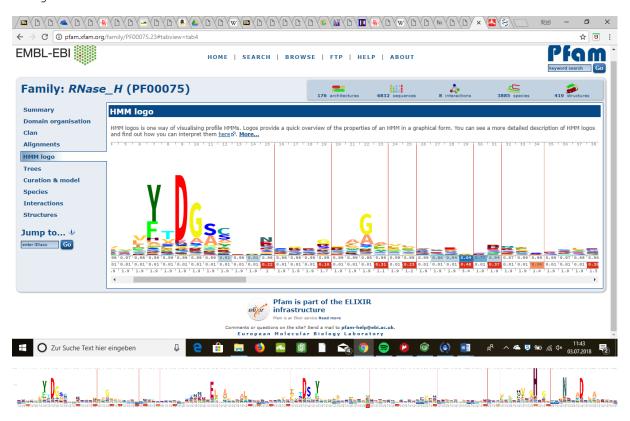
Ersten ca. 30 AS des 1. 5'-3' Frames:

GSHLSFTRPPPYLRPPSTPVESRSAASRLWCLLNYVRRLGKFRAQVETGPLSGAPLEPT

Suche in AS-Sequenz ist Genomsequenz vorzuziehen, da in diese um 1/3 kürzer ist. Dadurch sollte die Suchzeit klar verringert werden.

Es ist sinnvoll alle Übersetzungsframes zu durchsuchen, da dadurch sichergestellt wird, dass keinerlei Profil übersehen werden kann.

Aufgabe 4:



http://pfam.xfam.org/family/PF00075.23#tabview=tab4

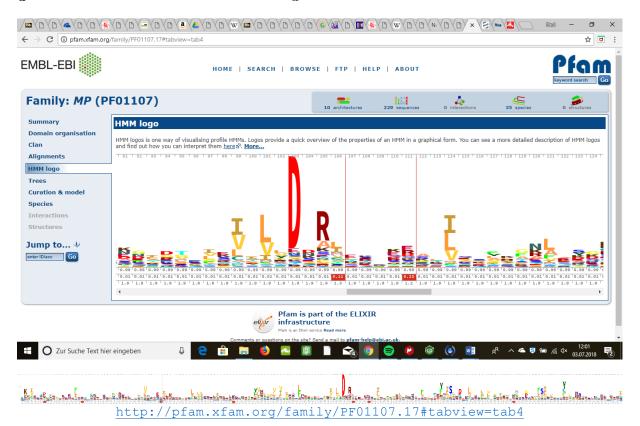
Das HMM-Logo sieht durchaus relativ ähnlich zur Suchsequenz aus. Da es nicht viele Punkte gibt, bei der eine starke Häufigkeit gegeben ist, geht das relativ gut. Nun die Aufgaben erneut, nur für ein anderen Organismus:

Organismus: Rice tungro bacilliform virus (NC_001914.1)

Die ersten ca. 100 BP:

tggtatcaga gcgatgttcg aactttaagg gaaaaataga atacaatcta gtcaaactaa ttacataagt tccaaagagg gatcttattt ctcaaaatat tgagtcacgt tacgaaaaac

Die ersten ca. 30 AS: WYQSDVRTLREK-NTI-SN-LHKFQRGILFLKILSHVTKNLNS-T-QSGEKRKNKNIYYLPIILAFIVTLILVKLVNQKESIPFILIISNSIVL



Das HMM Logo passt erneut halbwegs zur untersuchten Sequenz.