

Übungszettel 5

Aufgabe 2: Erste 100 Basen des Genoms „Human T-cell leukemia virus type I“

ggctcgcatc tctccttcac gcgcccgcgc ccttacctga ggccgccatc cagcccggtt
gagtcgcggt ctgccgcctc ccgcctgtgg tgctcctga

Aufgabe 3: Die ersten 30 AS der AS-Sequenz des 1. 5'-3'-Reading Frames

GSHLSFTRPPPYLRPPSTPVESRSAASRLW

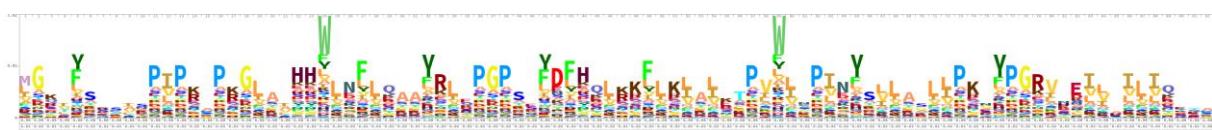
- Die Suche nach Mustern in AS-Sequenzen sind denen in Genomsequenzen vorzuziehen, da der genetische Code degradiert ist. Dies bedeutet, dass mehrere Codons für die gleiche Aminosäure codieren. So kann eine Mutation in der Genomsequenz auch eine stille Mutation sein und die gleiche Aminosäure hervorbringen, wodurch sich das Protein (der Phänotyp) nicht verändert. Eine Suche in den AS-Sequenzen zeigt also offensichtlicher die für Bioinformatiker wichtigen möglichen Veränderungen im Protein auf.
- Es ist sinnvoll, alle 6 Reading-Frames durchzusehen, da man anhand der Sequenzen nicht weiß, wie das Leseraster in der Natur wirklich sein wird (welche 3 Basen ein Codon bilden; wo das „Lesen“ also beginnt).

Aufgabe 4: HMM-Profile innerhalb der Aminosäuresequenz (in Ermangelung von Motiven innerhalb der ersten zwei Frames habe ich den Open Reading Frame des 3. 5'-3'-Frames verwendet)

Suchsequenz:

MGQIFSRASPIPRPPRGLAAHHWLNFLQAAYRLEPGPSSYDFHQLKKFLKIALETP
VWICPINYSLLASLLPKGYPGRVNEILHILIQTQAQIPSRPAPPPSSSTHDPDSDPQ
IPPPYVEPTAPQVLPVMHPHGAPPNHRPWQMKDLQAIKQEVSAAPGSPQFMQTI
RLAVQQFDPTAK

HMM-Logo des **Major core protein p19**



Das Modell ist recht ähnlich zu der Suchsequenz. Die Aminosäuren, die laut HMM-Logo am wahrscheinlichsten an einer Position sind, treten in Realität wirklich an dieser Stelle auf. Wenn die Wahrscheinlichkeiten nicht so klar verteilt sind (ein Buchstabe also nicht sonderlich größer ist als die anderen, sondern alle in etwa gleich groß sind), dann stimmt das HMM-Logo natürlich nicht klar mit der Suchsequenz überein.

Aufgabe 5:

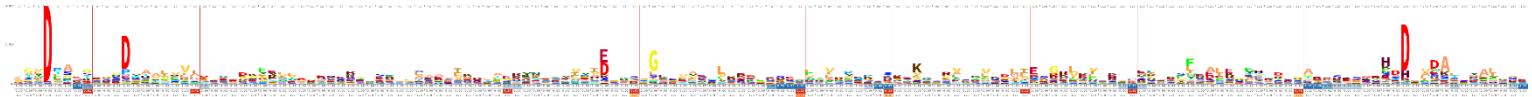
- a) Genomsequenz von Streptococcus phage 2972

2000 tcaagctaga cgataacact ttcttatcaa aacgctatat agactctatc aaagcagcta cgcccaaggg
caaattctac gatagagata tttgggtaa

- b) Zweiter 5'-3'-Frame

MKWTVAEGAIYADYDSKIHVVDELPEMKRCFGGIDWGYTHYGSIVVVGEGV
DGNFYLLDGVA AQFKEIDWWVEQARKLTGIYRNIPFYADSARPEHVARFESE
GFDISNANKSVIAGIELIAKLFKEEKLYVKRGFVPRFFDEIYQYRWKENSTKDE
PLKEFDDVLDVSVRYAIYSDYVIGSTEQASYDDLMSMFR

- c) HMM-Logo der Terminase RNaseH-like domain



Das Modell ist recht ähnlich zu der Suchsequenz. Die Aminosäuren, die laut HMM-Logo am wahrscheinlichsten an einer Position sind, treten in Realität wirklich an dieser Stelle auf. Wenn die Wahrscheinlichkeiten nicht so klar verteilt sind (ein Buchstabe also nicht sonderlich größer ist als die anderen, sondern alle in etwa gleich groß sind), dann stimmt das HMM-Logo natürlich nicht klar mit der Suchsequenz überein.