Bioinformatik Übungsblatt 5

Human T-cell leukemia virus type I:

Die ersten 100 Basenpaare:

ggetegeate teteetteae gegeeegeeg cettacetga ggeegeeate eaegeeggtt gagtegegtt etgeegeete eegeetgtgg tgeeteetga

Die ersten 30 Aminosäuren des 1.5'-3' Leserasters:

GSHLSFTRPPPYLRPPSTPVESRSAASRLW

Vergleich der Aminosäuresequenz mit dem HMM-Profil:

Ich habe als Sequenz den Open Reading Frame aus dem 3. 5'-3' Frame genommen, da es keine Übereinstimmung der AS-Sequenz mit der PFAM-Datenbank gab.

Sequenz:

AMGQIFSRSASPIPRPPRGLAAHHWLNFLQAAYRLEPGPSSYDFHQLKKFLKIALETPVWIC PINYSLLASLLPKGYPGRVNEILHILIQTQAQIPSRPAPPPPSSSTHDPPDSDPQIPPPYVEPTAP QVLPVMHPHGAPPNHRPWQMKDLQAIKQEVSQAAPGSPQFMQTIRLAVQQFDPTAKDLQ DLLQ

HMM-Logo:

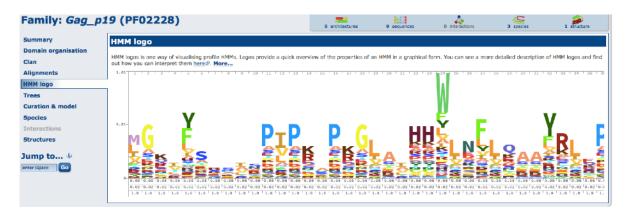


Abbildung 1: HMM Logo des Gag_p19 Proteins (http://pfam.xfam.org/family/PF02228.15#tabview=tab4)

Beim Vergleich des Logos mit der Suchsequenz fällt auf, dass diese teilweise sehr gut übereinstimmen. Beispielsweise an den Stellen 24-35 und 56-59 herrscht eine komplette Übereinstimmung zwischen Suchsequenz und Logo. Zudem ist aber deutlich erkennbar, dass bei kleinerer Buchstabengröße im Logo und folglich einer kleineren Wahrscheinlichkeit für das Auftreten dieser Aminosäure, das Logo oftmals nicht mit der Suchsequenz übereinstimmt.

Mouse mammary tumor virus (complete genome; Accession: NC_001503.1): Die ersten 100 Basenpaare:

geaacagtee taacatteae etetegtgt tttgtgtetg ttegecatee egteteeget egteaettat eetteaettt eeagagggte eeeeegeaga

Die ersten 30 Aminosäuren des 1.5'-3' Leserasters:

ATVLTFTSRVFVSVRHPVSARHLSFTFQRV

Vergleich der Aminosäuresequenz mit dem HMM-Profil:

Ich habe als Sequenz den Open Reading Frame aus dem 1.5'-3' Frame genommen.

Sequenz:

MGVSGSKGQKLFVSVLQRLLSERGLHVKESSAIEFYQFLIKVSPWFPEEGGLNLQDWKRVG REMKRYAAEHGTDSIPKQAYPIWLQLREILTEQSDLVLLSAEAKSVTEEELEEGLTGLLSTSS QEKTYGTRGTAYAEIDTEVDKLSEHIYDEPYEEKEKADKNEEKDHVRKIKKVVQRKENSEG KRKEKDSKAFLATDWNDDDLSPEDWDDLEEQAAHYHDDDELILP

HMM-Logo:

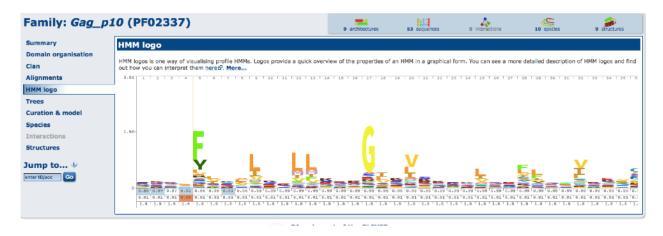


Abbildung 2: HMM Logo des Gag_p10 Proteins (http://pfam.xfam.org/family/PF02337.16#tabview=tab4)

Beim Vergleich des Logos mit der Suchsequenz fällt auf, dass diese nicht gut zusammenpassen. Selbst bei den Buchstaben die mit hoher Wahrscheinlichkeit vorkommen sollten, gibt es in der Suchsequenz starke Abweichungen zum Logo.

3a) Die Suche in Aminosäuresequenzen der Suche in Genomsequenzen vorzuziehen, da eine Aminosäure immer aus einem Basentriplett gebildet wird und sich die Suchsequenz somit um 2/3

verkürzt. Zudem zeigt die Übersetzung in Aminosäuren den Open Reading Frame der Sequenz an.

3b) Es ist sinnvoll alle möglichen Übersetzungsframes zu durchsuchen, da wir ja nicht wissen wo sich der Ablesestart der Genomsequenz im Organismus befindet. Um keine Möglichkeit auszuschießen ist es daher notwendig alle Übersetzungsframes zu durchsuchen.