

Mouse mammary tumor virus (complete genome; Accession: NC_001503.1):

Die ersten 100 Basenpaare:

gcaacagtcc taacattcac ctctegtgtg ttgtgtctg ttgccatcc cgtctccgct cgtcacttat ccttcacttt ccagagggtc
cccccgaga

Die ersten 30 Aminosäuren des 1. 5'-3' Leserasters:

ATVLTFTSRV FVSVRHPVSARHLSFTFQRV

Vergleich der Aminosäuresequenz mit dem HMM-Profil:

Ich habe als Sequenz den Open Reading Frame aus dem 1. 5'-3' Frame genommen.

Sequenz:

MGVSGSKGQKLFVSVLQRLLSERGLHVKESSAIEFYQFLIKVSPWFPEEGGLNLQDWKRVG
REMKRYAAEHGTD SIPKQAYPIWLQLREILTEQSDLVLLSAEAKSVTEEELEEGLTGLLSTSS
QEKTYGTRGTAYAEIDTEVDKLSEHIYDEPYEEKEKADKNEEKDHVRKIKKVVQRKENSEG
KRKEKDSKAFLATDWNDDDLSPEDWDDLEEQA AHYHDDDELILP

HMM-Logo:

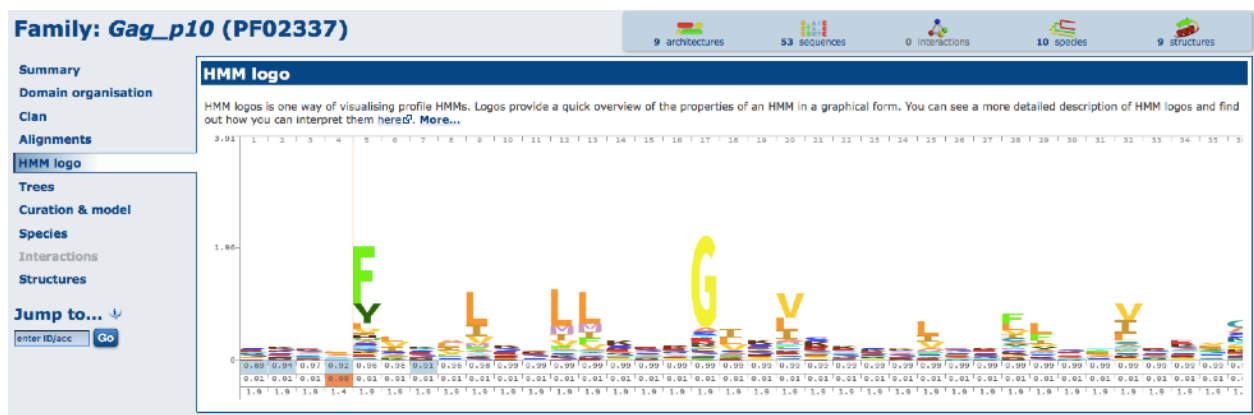


Abbildung 2: HMM Logo des Gag_p10 Proteins (<http://pfam.xfam.org/family/PF02337.16#tabview=tab4>)

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 $\frac{2}{3}$ 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 auszuschließen ist es daher notwendig alle Übersetzungsframes zu durchsuchen.