

Aufgabe 2:

Ersten 100 Basenpaare des Human T-cell leukemia virus type I:

```
ggctgcgcatc tctccttcac gcgcccgcgc ccttacctga ggccgccatc cagcccggtt gagtcgctt ctgccgcctc  
ccgcctgtgg tgccctccta
```

Aufgabe 3:

Ersten 30 Aminosäuren im 1. 5'3' Frame:

GSHLSFTRPPPYLRPPSTPVESRSAASRLW

- (a) Die Suche in Aminosäuren Sequenzen ist der Suche in Genom Sequenzen vorzuziehen, da sich die Suche um 2/3 verkürzt da eine Aminosäure für ein Basentriplett steht.
- (b) Es ist sinnvoller alle 6. Frames zu durchsuchen, da alle gleich wahrscheinlich sind und so alle möglichen Überlappungen und Leserichtungen berücksichtigt werden.

Aufgabe 4:

Im weiteren Verlauf wurde der open reading frame aus dem 3. 5'3' Frame verwendet, da sonst kein Ergebnis erzielt werden konnte.

Verwendete Sequenz:

```
MGQIFSRASPIPRPPRGLAAHHWLNFLQAAYRLEPGPSSYDFHQLKKFLKIALETPWWICPI  
NYSLLASLLPKGYPRVNEILHILITQAQIPSRPAPPPSSSTHDPDSDPQIPPPYVEPTAP  
QVLPVMHPHGAPPNHRPWQMKDLQAIKQEVSAAPGSPQFMQTRLAVQQFDPTAKDLQD  
LLQ
```

HMM-Logo:

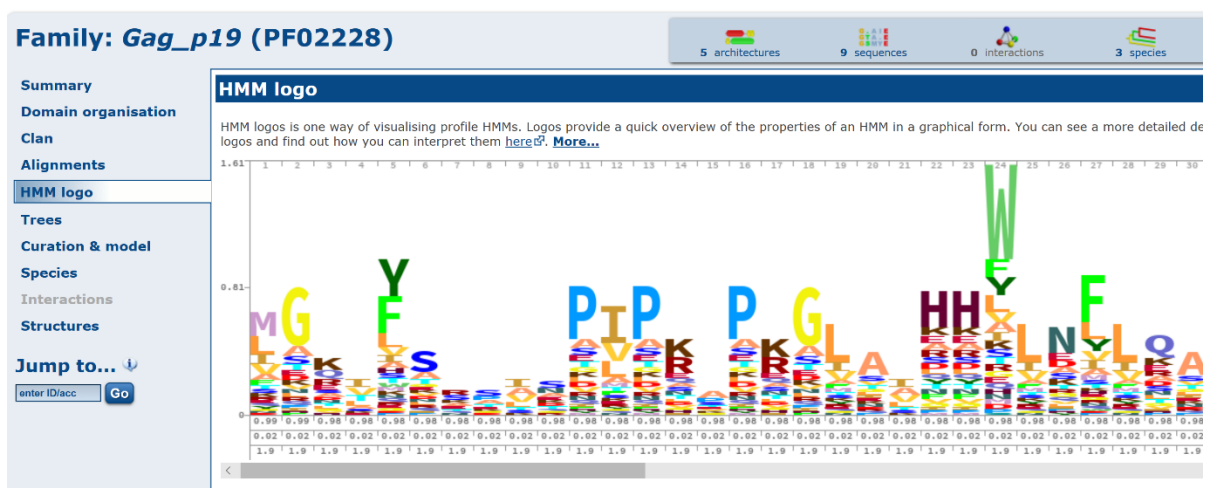


Abbildung 1: HMM-Logo (<http://pfam.xfam.org/family/PF02228.15#tabview=tab4>, (03.07.18, 11:02 Uhr))

Beim Vergleich des Logos mit der eingegeben Sequenz fällt auf, dass sie gut miteinander übereinstimmen. Dabei gibt es Stellen an denen diese nicht so gut übereinstimmen, wie an anderen. Zum Beispiel an Stelle 7-10 ist durch die geringe Größe der Buchstaben sichtbar,

dass die Wahrscheinlichkeit der in der Suchsequenz gegebenen Aminosäuren nicht so hoch ist wie zum Beispiel an Stelle 11-13 oder 22-24. An diesen Stellen sind die in der Suchsequenz vorgegeben Aminosäuren durch ihre Größe im Logo als sehr wahrscheinlich gegeben und stimmen somit sehr gut überein.

Aufgabe 5:

Ersten 100 Basenpaare des Haloarcula hispanica pleomorphic virus 1:

```
acttcggcct agctacacca cagttttggt tacgcttata ttitcaaagc ttaagttggt ttaattttt gccgagagac
cgtgaggttg aaccgggcaa
```

Ersten 30 Aminosäuren im 1. 5'3' Frame:

MSSEPAPSTGGSRPPQSNSALDSSGSVPVV

Verwendeter Open Reading Frame aus dem 1. 5'3' Frame:

```
MSSEPAPSTGGSRPPQSNSALDSSGSVPVVSVDERPDLDELPDPSERVREFVDRLGERA
HLPMSDVHGRMTMRDCCDEEWSYTVEDSSGSTKREERVNVTAPLYQVVSMDLHWHE
DYLRLSTLRLEYGERADPEHKLLDVPLDNSWMIQYQKKERARLKAMERETAGYRTCDECDT
RYCTEQDEHSTEYVFGEFDDPYVVLGTGRTAAGDGAPPVDHARSIAEAWTGEHGNDGAGRS
LRYVMSEKLGLESDWVRWTQGEPHTGKRRQAGGYGNLGYHHAHDVIVLDGAAASVEPT
AATFRTVIEKHVEECAGAGREAHDLDKSAEEWKNGDVDVTEV
```

HMM-Logo:

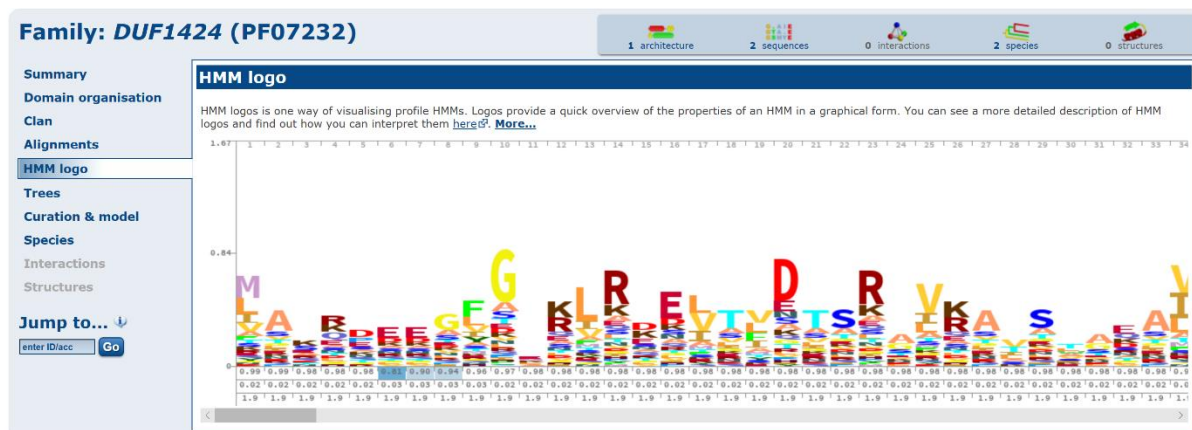


Abbildung 2: HMM-Logo (<http://pfam.xfam.org/family/PF07232.10#tabview=tab4>, 03.07.18, 11:38 Uhr)

Beim Vergleich des HMM-Logos und der Suchsequenz fällt auf, dass diese nicht miteinander übereinstimmen. Die erste Stelle der Sequenz passt einigermaßen, da das M am größten ist, aber es sind auch viele andere größere Buchstaben sichtbar. Die restliche Sequenz stimmt nicht mit dem Logo überein. Auch die Buchstaben, die am größten sind und damit mit sehr hoher Wahrscheinlichkeit vorkommen (Stelle 40) stimmt nicht mit der Suchsequenz überein.