

Bioinformatik- Assignment 5- Shelly Harel

2) ersten 100 Basen aus der Genom-Sequenz von „*Human T-cell leukemia virus I*“:

ggctcgcatctctcttcacgcgcccgccttacctgaggccgcatccacgccggttgagtcgcgttctgccgcctccgcct
gtggtgcctcctga

3) ersten 30 Aminosäuren des 1. 5'3' Frames aus dem ersten Gen (450bp-4638bp):










MGQIFSRASPIPRPPRGLAAHHWLNFLQA

- (a) Es ist sinnvoller in Aminosäuresequenzen zu suchen als in Genomsequenzen, da der genetische Code degeneriert ist. Dies bedeutet, dass mehrere Basentriplets für die selbe Aminosäure codieren. In diesem Zusammenhang kann es also sein, dass beim Vergleich zweier Genome unterschiedlicher Spezies zwar Unterschiede auftreten, diese sich jedoch nicht auf die Aminosäuresequenz auswirken. Mutationen in der DNA, die genau aufgrund dieses Mechanismus keine Auswirkung auf die Proteinbiosynthese haben nennt man auch stille Mutationen.
- (b) Es ist sinnvoll alle sechs Reading Frames zu analysieren, da man erstens nicht weiß, ob der einem gegebenen Strang oder dessen komplementäre Strang zu einer mRNA transkribiert wird. Zweitens weiß man auch nicht ob man die Tripletts beim ersten, zweiten oder dritten Nukleotid anfängt zu lesen.

4) HMM-Ergebnisse:

Pfam Matches

Advanced

Family		Clan	Description	Cross-references	Start	End	Domain E-values		
Id	Accession						Ind.	Cond.	
>	Gag_p24	PF00607.19	CL0148	gag gene protein p24 (core nucleocapsid protein)	  	147	344	1.2e-64	2.2e-68
>	Gag_p19	PF02228.15	CL0074	Major core protein p19	  	1	92	7.3e-57	1.3e-60
>	zf-CCHC	PF00098.22	CL0511	Zinc knuckle	  	355	372	8.5e-07	1.5e-10

Your search took: 0.1 secs

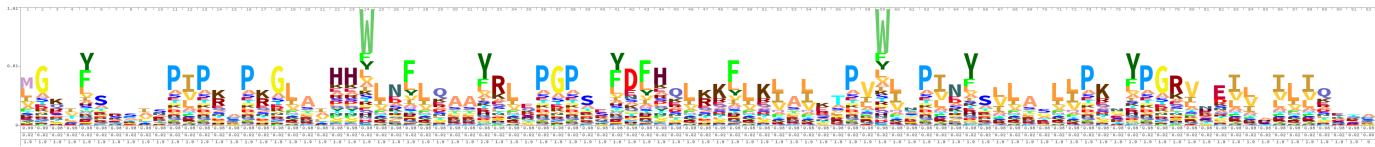
HMM-Logo zu *gag gene protein p24 (core nucleocapsid protein)*:



<http://pfam.xfam.org/family/PF00607.19#tabview=tab4>

Einschätzung: Hier ist der Vergleich des HMM-Logos mit der Aminosäuresequenz sehr viel ungenauer als bei bspw. Major core protein p19, da hier die meisten Aminosäuren nicht eindeutig zu erkennen sind. Die Entropie der einzelnen Positionen ist deutlich höher.

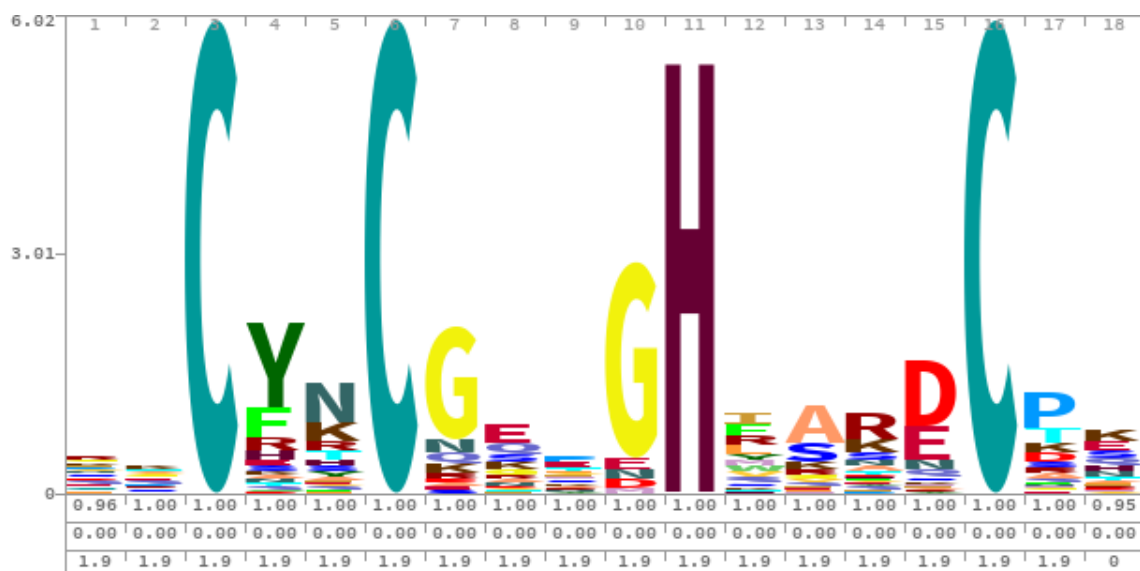
HMM-Logo zu **Major core protein p19**:



<http://pfam.xfam.org/family/PF02228.15#tabview=tab4>

Einschätzung: Vergleicht man die Aminosäuresequenz aus Expsy mit dem HMM-Logo, könnte man sagen, dass diese sich sehr ähneln, wenn man die wahrscheinlichsten Aminosäuren innerhalb des Motif Logo annehmen würde.

HMM-Logo zu **Zinc knuckle**:



<http://pfam.xfam.org/family/PF00098.22>

Einschätzung: Nach einem Vergleich des HMM-Logos mit der Aminosäuresequenz kann man sagen, dass diese beiden identisch sind, wenn man annimmt, dass die wahrscheinlichsten Aminosäuren aus dem Motif Logo vorkommen.

5) ersten 100 Basen aus der Genom-Sequenz von "**Haloarcula hispanica pleomorphic virus 1**":

ATGAGTTCTGAACCTGCGCCGTCCACGGGCGGCTCTCGGCCACCACAGTCTAACTCCGCGCTCGATA
GTAGCGGCTCTGTACCTGTCGTTTCGGTTTCTG


ersten 30 Aminosäuren des 1. 5'3' Frames aus dem ersten Gen (1bp-2352bp):

MSSEPAPSTGGSRPPQNSALDSSGSPVV

HMM-Ergebnisse:

Pfam Matches

Advanced

Family		Clan	Description	Cross-references	Start	End	Domain E-values		
Id	Accession						Ind.	Cond.	
>	DUF1424	PF07232.10	CL0169	Putative rep protein (DUF1424)		178	422	5.5e-10	3.3e-14

Your search took: 0.1 secs

HMM-Logo zu *Putative rep protein (DUF1424)*:



<http://pfam.xfam.org/family/PF07232.10#tabview=tab4>

Einschätzung: Vergleicht man das HMM-Logo mit der Aminosäuresequenz im gegebenen Bereich so ergibt sich keine eindeutige Ähnlichkeit. Da die Höhe der Aminosäuren im Motif Logo jedoch auch nicht wirklich hoch sind, bedeutet es, dass die Entropie groß ist und somit diese Aminosäureposition nicht konserviert ist.