

Assignment 5

Aufgabe 2: Genom-Sequenz von „Human T-cell leukemia virus I“

<https://www.ncbi.nlm.nih.gov/nuccore/LC183873.1>

Die ersten 100 Basenpaare:

```
tgacaatgac catgagcccc aaatatcccc cgggggctta gagcctccca  
gtgaaaaaca tttccgcgaa acagaagtct gaaaagggtca gggcccagac
```

Aufgabe 3:

5'3' Frame 1																			
tga	caa	tga	cca	tga	gcc	cca	aat	atc	ccc	ggg	ggg	ctt	aga	gcc	tcg	cag	tga	aaa	aca
-	Q	-	P	-	A	P	N	I	P	R	G	L	R	A	S	Q	-	K	T
ttt	cgg	cga	aac	aga	agt	ctg	aaa	agg	toa	ggg	ccc	aga	ota	agg	ctc	tga	cgt	ctc	ccc
F	P	R	N	R	S	L	K	R	S	G	P	R	L	R	L	-	R	L	P

5'-Q-P-APNIPRGLRASQ-KTFPRNRSLKRS3'

bzw: 5'QPAPNIPRGLRASQKTFPRNRSLKRSGPRL3'

a) AS-Sequenz vs Genomsequenz

- benutzt man den Einbuchstabencode für die AS-Sequenz verkürzt sich der string auf 33%, da nun nicht mehr mit Triplets (Codons) gearbeitet wird
- die Betrachtung des genetisches Codes entfällt bei der AS-Sequenz
- Bei dem Virus hier egal, bei Eukaryoten aber muss Segmentierung der Gene in Exons und Introns beachtet werden (alternatives Splicen)

b) 6 mögliche Übersetzungsframes

- Überprüfung ob die Zuordnung des codogenen Stranges korrekt ist
- Identifikation verschiedener Leseraster, um den ORF zu identifizieren

Aufgabe 4:

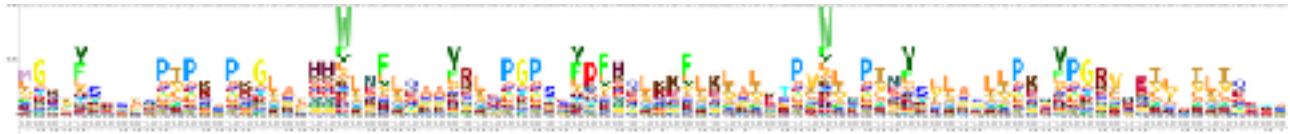
AS-Sequenz: 1.Frame

```
QPAPNIPRGLRASQKTFPRNRSLKRSGPRLRLRLPPEGQLSTGSGALTCPPEDKSAQTSGKPPEPPIS  
SPCLSSRPQALTTTPHLKKLFMARIWLNKLTGVYKSVETVQEGARISPSRARRPTGRHPPRRLSRVLPP  
PACGASTASAVVSLELRSRPGLCPALPWSLPRLSRFSTLCLTLAQLCVFVSFVSLRRYRSKVPPLSLS  
FTTDCRLGPRPSTGDSVGSEPATAHSIALSRRETYTVGGSSGIRAPLYSLGNGPNLFPRPYSAAAPGA  
GRSSLALPPSGISPRTRSLQLRFPPVEKISNSFRNTGLDLSHLL
```

Bei dieser Sequenz erhalte ich mit dem Tool kein Ergebnis. (Vlt ist das ja gewollt als Anspielung auf Frage 3.b)

AS-Sequenz: 2.Frame

```
DNDHEPQISPGGLEPPSEKHFRETEVKGQGPDGSDVSPRRDSSAPAQARPRVPLKTNHKLRLPPGSH  
RNHPFPPHVCQAALRRRQPLTSKNFSWHAYGINQESIKAWRQFRRGLASLLHAPAALPEAAIHAGVA  
FCRLPPVVPPELRPPSRVSSGRDRAFVRRSLGAYLDSAGSPRFAPCLLNSASLFRFLFCAATDRKFRP  
FPFHSRLTAGLAHGQVPATPLARSQRQPILHSPGEKLSTQLGARPGYERPFIPAMGQIFSRASAPIPRP  
PRGLAAHHWLNFLQAAYRLEPGPSSYDFHQLKKFLKIALETVPWWICPINYS
```



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

Der größte Teil beider Sequenzen stimmt über ein, die Ähnlichkeit ist somit gegeben.

Aufgabe 5: Jakobida sp. TP-2015d CDP-DAG synthase (CDS) mRNA

<https://www.ncbi.nlm.nih.gov/nuccore/KT984625.1>

Die ersten 100 Basen:

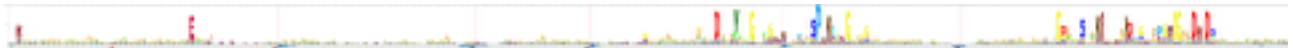
```
atgagaagaa aggttacgca gcaaatcaaa gatgatactg caagcgacga  
tcgagacaca tcgatggatg aagaccaaga gggtgtcgtg tctttgcaac
```

Die ersten 30 AS:

5'MRRKVTQQIKDDTASDDADTSMDEDQEVVV3'

AS-Sequenz (1. 5'3' Frame)

```
MRRKVTQQIKDDTASDDADTSMDEDQEVVVSLQPQEEPCKSPPEKWKNFVVRTTFTCLMIGLFALII  
AAGHVVTLFFIFLVQFFMFREVTLSTKRGLDRRLAGFRRFTGLKALKWMNWWYLFATLFFFHGLFLM  
PQLRNIPYVGHFMKYHSFYSLMLYVLGFVMFVLSLQRGKLRGQFSQFAFTHMALLIVVLSSALVRNV  
FHGMIWFLLPCSLIVINDICAYLFGFFMGRTPLIKLSPKKTWEGFIGGLFSTMVGAFFLARFLSMFDHMI  
CPKNNLVMFERLSCPRPAVFIPQAYHVPSVISSLVPFLGPEVTIMPFQIHAMALGLFASIIAPFGGFFASG  
FKRAFDLKDFGSSIPGHGGITDRMDCQSLMGFFSWVYFINFVQESIVPLDEILEMVFSLSQNDKMELY  
NALQKHLEL
```



<http://pfam.xfam.org/family/PF01148.19#tabview=tab4>

Die Übereinstimmung der beiden Sequenzen nimmt von links nach rechts ab.

Tools:

NCBI Genom-Datenbank:

<https://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?taxid=10239>

Übersetzungs-Tools „DNA -> Proteinsequenz“:

<https://web.expasy.org/cgi-bin/translate/dna2aa.cgi>

Profil-HMM Such-Tool:

<https://www.ebi.ac.uk/Tools/hmmer/search/hmmscan>

Verwendete Genomsequenz Aufgabe 2

tgacaatgac	catgagcccc	aaatatcccc	cgggggctta	gagcctccca
gtgaaaaaca	tttccgcgaa	acagaagtct	gaaaagggtca	gggcccagac
taaggctctg	acgtctcccc	ccggagggac	agctcagcac	cggctcaggc
taggccctga	cgtgtccccc	tgaagacaaa	tcataagctc	agacctccgg
gaagccaccg	gaaccaccca	tttcctcccc	atgtttgtca	agccgccctc
aggcgttgac	gacaaccctt	cacctcaaaa	aacttttcat	ggcacgcata
tggctgaata	aactaacagg	agtctataaa	agcgtggaga	cagttcagga
gggggctcgc	atctctcctt	cacgcgcccg	ccgccctacc	tgaggccgcc
atccacgccg	gttgagtcgc	gttctgcccg	ctcccgcctg	tgggtgcctcc
tgaactgcgt	ccgccgtcta	ggtaagttta	gagctcaggt	cagagaccggg
cctttgtccg	gcgctccctt	ggagcctacc	tagactcagc	cggttctcca
cgttttgctt	gacctgctt	gctcaactct	gcgtctttgt	ttcgttttct
gttctgcgcc	gctacagatc	gaaagttccg	cccctttccc	tttcattcac
gactgactgc	cggcttggcc	cacggccaag	taccggcgac	tccgttggct
cggagccagc	gacagcccat	tctatagcac	tctccaggag	agaaacttag
tacacagttg	ggggctcgtc	cgggatacga	gcgccccttt	attccctagg
caatgggcca	aatcttttcc	cgtagcgcta	gccctattcc	gcggccgccc
cgggggctgg	ccgctcatca	ctggcttaac	ttcctccaag	cggcatatcg
cctagaacct	ggccctcca	gttacgattt	ccaccagttg	aaaaaatttc
ttaaaatagc	tttagaaaca	ccggtctgga	tctgtcccat	taactactcc

Verwendete Genomsequenz Aufgabe 5

atgagaagaa	aggttacgca	gcaaatcaaa	gatgatactg	caagcgacga
tgcagacaca	tcgatggatg	aagaccaaga	ggttgtcgtg	tctttgcaac
ctcaggaaga	accgaagaaa	tccccaccgg	aaagcaagtg	gaagaatttt
gtcgtccgca	caacgttcac	ttgtctgatg	attggtttgt	ttgcattgat
cattgcagct	gggcacgtgg	taactctgtt	tttcatcttt	ctggttcagt
tcttcatgtt	cagagaagtc	gtgacgctta	gtaccaagcg	tggattggac
agacgactcg	caggtttttcg	acgcttcacg	ggattgaaag	ctctcaaattg
gatgaattgg	tggatatctat	ttgccacact	gttcttcttc	catgggcttt
tcctcatgcc	gcagttgcgc	aacatcccct	acgttggtca	cttcatgaag
taccattcat	tctattctct	catgctttat	gtgttggggg	ttgtaatgtt
tgtgctgtcg	cttcaaagag	ggaagttgcg	tgggcaattc	agtcagtttg
cattcacaca	catggctctt	ttgttgattg	ttgttctttc	ttctgctctc
gtgagaaacg	tattccacgg	gatgatttgg	ttcttgctac	catgttcatt
gattgtgatc	aatgacattt	gtgcatatct	ttttggattc	ttcatgggaa
gaactccttt	gatcaagttg	tctcccaaga	aaacatggga	agggttcatt
ggcggcctct	tctccacgat	ggtgggagca	ttcttccttg	ctcgattcct
cagcatgttt	gatcatatga	tctgtcccaa	gaataatttg	gtgatgttcg
aaaggttgtc	atgcccacgt	ccggctgtgt	ttattcctca	ggcttaccat
gttccctctg	tgatttcac	tcttgtacca	ttcttgggac	cagaagtcac
gatcatgccg	ttccaaatcc	atgcaatggc	attgggactg	tttgcgtcca