### **Assignment 5**

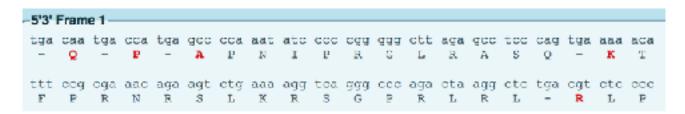
## Aufgabe 2: Genom-Sequenz von "Human T-cell leukemia virus type I"

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

Die ersten 100 Basenpaare:

tgacaatgac catgagccc aaatatcccc cgggggctta gagcctccca gtgaaaaaca tttccgcgaa acagaagtct gaaaaggtca gggcccagac

#### Aufgabe 3:



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 Tripletts (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:

#### Tools:

NCBI Genom-Datenbank:

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

Übersetzungs-Tools "DNA -> Proteinsequenz": <a href="https://web.expasy.org/cgi-bin/translate/dna2aa.cgi">https://web.expasy.org/cgi-bin/translate/dna2aa.cgi</a>

Profil-HMM Such-Tool:

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

## Genomsequenz (1000):

tgacaatgac catgagcccc aaatatcccc cgggggctta gagcctccca gtgaaaaaca tttccgcgaa acagaagtct gaaaaggtca gggcccagac taaggetetg acgteteece eeggagggae ageteageae eggeteagge taggccctga cgtgtccccc tgaagacaaa tcataagctc agacctccgg gaagccaccg gaaccaccca tttcctcccc atgtttgtca agccgccctc aggcgttgac gacaacccct cacctcaaaa aacttttcat ggcacgcata tggctgaata aactaacagg agtctataaa agcgtggaga cagttcagga gggggctcgc atctctctt cacgcgcccg ccgccctacc tgaggccgcc atccacgccg gttgagtcgc gttctgccgc ctcccgcctg tggtgcctcc tgaactgcgt ccgccgtcta ggtaagttta gagctcaggt cgagaccggg cctttgtccg gcgctccctt ggagcctacc tagactcagc cggttctcca cgctttgcct gaccctgctt 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 cctagaaccc ggtccctcca gttacgattt ccaccagttg aaaaaatttc ttaaaatagc tttagaaaca ccggtctgga tctgtcccat taactactcc

## AS-Sequenz:

QPAPNIPRGLRASQKTFPRNRSLKRSGPRLRLRLPPEGQLSTGSGALTCP
PEDKSAQTSGKPPEPPISSPCLSSRPQALTTTPHLKKLFMARIWLNKLTG
VYKSVETVQEGARISPSRARRPTGRHPRRLSRVLPPPACGASTASAVVSL
ELRSRPGLCPALPWSLPRLSRFSTLCLTLLAQLCVFVSFSVLRRYRSKVP
PLSLSFTTDCRLGPRPSTGDSVGSEPATAHSIALSRRETYTVGGSSGIRA
PLYSLGNGPNLFPRPYSAAAPGAGRSSLALPPSGISPRTRSLQLRFPPVE
KISNSFRNTGLDLSHL