

Assignment 5

Aufgabe 2)

Human T-cell leukemia virus type I (Accession: NC_001436):

(Quelle: <https://www.ncbi.nlm.nih.gov/nucore/LC183873.1>)

1st 100 Basenpaare

tgacaatgac catgagcccc aaatatcccc cgggggctta gagcctcca

gtgaaaaaca ttccgcgaa acagaagtct gaaaaggta gggcccagac

Aufgabe 3)

5'3' Frame 1

5'3' Frame 1																			
tga	caa	tga	cca	tga	gcc	cca	aat	atc	ccc	cgg	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	tca	ggg	ccc	aga							
F	P	R	N	R	S	L	K	R	S	G	P	R							

5'-Q-P-APNIPRGLRASQ-KTFPRNRSLKRS3'

bzw: 5'QPAPNIPRGLRASQKTFPRNRSLKRSGPRL3'

rot unterlegt: open reading frame

3a) Suche in Aminosäuresequenzen > Suche in Genomsequenzen

- Verkürzung des Strings auf 33% bei Nutzung des Einbuchstabencode für die AS-Sequenz, da nun nicht mehr mit Triplets (Codons) gearbeitet wird
- die Betrachtung des genetischen 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 Splecen)

3b) Sinn alle 6 möglichen Übersetzungs-Frames zu durchsuchen

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

Aufgabe 4

AS-Sequenz: 1.Frame

QPAPNIPRGLRASQKTFPRNRSLKRSGPRLRLRLPPEGQLSTGSGALTCPPEDKSAQTSGK
PPEPPISSPCLSSRPQALTTTPHLKKLFMARIWLNKLTGVYKSVETVQEGARISPSRARRPT
GRHPRRLSRVLP PPACGASTASAVVSLELRSRPG LCPALPWSLPRLSRFSTLCLTLLAQLC
VFVSFSVLRRYRSKVPPLSLSFTTDCRLGPRPSTGDSVGSEPATAHSIALSRRETYTVGGS
SGIRAPLYSLGNPNLFPRPYSAAPGAGRSSLALPPSGISPRTRSLQLRFPPVEKISNSFR
NTGLDLSHLL

➔ Kein Ergebnis bei der Suche

Folgende AS-Sequenz wurde genutzt:

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

Jakobida sp. TP-2015d CDP-DAG synthase (CDS) mRNA, complete cds
(Cytidylyltransferase family)

1 atgagaagaa aggttacgca gcaaatcaaa gatgatactg caagcgacga tgcagacaca
61 tcgatggatg aagaccaaga ggtgtcgtg tctttgcaac ctccaggaaga accgaagaaa
121 tccccaccgg aaagcaagtg gaagaatttt gtctccgca caacgttcac ttgtctgatg
181 attggtttgt ttgcattgat cattgcagct gggcacgtgg taactctgtt ttatcatctt
241 ctggttcagt tcttcatgtt cagagaagtc gtgacgctta gtaccaagcg tggattggac
301 agacgactcg caggttttcg acgcttcacg ggattgaaag ctctcaaatg gatgaattgg
361 tggatatctat ttgccacact gttcttcttc catgggcttt tcttcatgcc gcagtgtcgc
421 aacatcccct acgttggcca cttcatgaag taccattcat tctattctct catgctttat
481 gtgttggggt ttgtaatgtt tgtgctgtcg cttcaaagag ggaagtgtcg tgggcaattc
541 agtcagtttg cattcacaca catggctctt ttgttgattg ttgttcttct ttctgctctc
601 gtgagaaacg tattccacgg gatgatttgg tcttctgtac catgttcatt gattgtgatc
661 aatgacattt gtgcatactt ttttggattc ttcatgggaa gaactccttt gatcaagttg
721 tctccaaga aaacatggga agggttcatc ggcggcctct tctccagat ggtgggagca
781 ttcttctgtg ctgattcct cagcatgttt gatcatatga tctgtcccaa gaataatttg
841 gtgatgttcg aaaggtgtgc atgccacgt cggctgtgt ttattcttca gcttaccat
901 ttctcctctg tgatttcacg tcttgtacca tctttgggac cagaagtcac gatcatgcgc

961 ttccaaatcc atgcaatggc attgggactg ttgctgcca tcattgcgcc atttggtggc
 1021 ttctttgcca gtggattcaa gagagcattc gatttgaaag acttggttc cagtattcct
 1081 gggcatggtg gtatcactga ccgtatggat tgtcaatctc ttatgggatt tttctcgtgg
 1141 gtttatttca tcaactttgt gcaagagtcg attgtgccgc tcgatgaaat cttggaaatg
 1201 gtgttctctc tttcacagaa cgacaagatg gaattgtaca atgcctaca aaagcatttg
 1261 gaattgtga

1st 100 Basenpaare:

atgagaagaa aggttacgca gcaaatcaaa gatgatactg caagcgacga
 tgcagacaca tcgatggatg aagaccaaga ggtgtcgtg tctttgcaac

Die AS-Sequenz des 1. 5'3' Frames das genutzt wurde:

MRRKVTQQIKDDTASDDADTSMDEDQEVVSLQPQEEP KKSPPE SKWKNFVVRTT
 FTCLMIGLFALIIAAGHVVTLFFIFLVQFFMFREVVT LSTKRGLDRRLAGFRRFTGLKA
 LKWMNWWYLFATLFFFHGLFLMPQLRNIPYVGHFMKYHSFYSLMLYVLGFVMFVL
 SLQRGKLRGQFSQFAFTHMALLLIVVLSSALVRNVFHGMWFLLP CSLIVINDICAYLF
 GFFMGRTP LIKLSPKKTWEGFIGGLFSTMVGAFFLARFLSMFDH MICPKNNLVMFE
 RLSCPRPAVFIPQAYHVPSVISSLVPFLGPEVTIMPFQIHAMALGLFASIIAPFGGFFA
 SGFKRAFDLKD FGSSIPGHGGITDRMDCQSLMGFFSWVYFINFVQESIVPLDEILEM
 VFSLSQNDKMELYNALQKHLEL



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

Ähnlichkeit: Wo es am Anfang noch Übereinstimmungen zwischen dem Logo und der Suchsequenz gibt, scheint mit wachsender Länge die Ähnlichkeiten abzunehmen.

<https://www.ncbi.nlm.nih.gov/nuccore/KR049084.1>

Artemisia sp. ZYL-2016 chrysanthemyl diphosphate synthase (CDS) mRNA, complete cds

(polyprenyl synthetase)

Origin:

1 atggcatgct ctagtagtct ttcttcaaa tgggcttctt ggaatgcctc ttcttttccg
61 caccatcag ttcaaccttt tgtgactcgg aagaatgtgg tacggtacca taaacaaca
121 tctgagtcta ctatcactac gacattgagc agcaatctaa actcacaatt catgcaagtt
181 tatgagactt tgaaatcaga gctaattcac gaccgctcat ttgagtttga tgacgattct
241 cgtcaatggg tggagcggat gatcgactac actgtacctg gaggaagat ggttcgaggc
301 tattccgttg ttgacagcta ccaattgctt aaaggacaag aattgactga agaagaagcc
361 ttctctgtgt gtgctcttgg ttggtgcact gaatggtttc aagcctttat tcttgacat
421 gatgacataa tggatggctc gcacacacgc agaggtcaac cctgttggtt cagactaccc
481 gaggttgag tagttgctat aaatgatggt gttcttcttc gtaacatgt ccacagaatc
541 ctaagaagaa atttccaagg aaaggcttat tacgtgcatc ttgtggacct ctcaatgag
601 actgaatttc aaacaatctc tggacaaatg attgatacga tcgctagact agctggacag
661 aaagatctct caaagtatac aatgtctctt aaccgtaaga ttgttcagta caaaggttct
721 tactactctt gttaccttcc gattgcgtgt gcactcctta tgttggaga gaatctggaa
781 gatcatgttc aagtgaaga catccttgta gaattgggta tgtattatca aattcagaat
841 gattatctcg acacttttgg tgatcctaata gtttttgga agactggaac agatattgaa
901 gaatgcaagt gctcatggtt gattgcgaaa gcactggaac ttccaacga ggagcaaaag
961 aaaattttaa gcgaaaacta tgggaaaaat gatccagcaa aggtagcaaa agttaaggaa
1021 gtataccatg ctctcgatct aaagagtgcg tatgaagatt atgagacaaa ttgtacgag
1081 aagtcgata aagcaattaa agctcttcca agcattacag tgcaagcggg gttgaaatct
1141 tgtctggaaa agatgtatga gggacataag taa

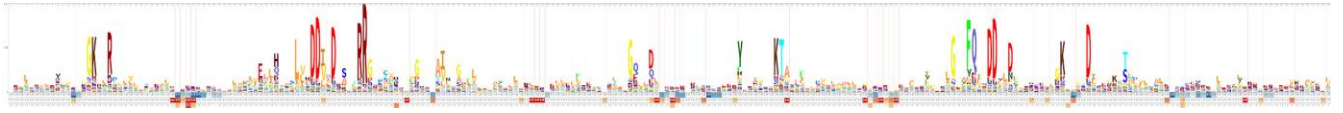
1st 100 Basenpaare:

atgagaagaa aggttacgca gcaaatcaaa gatgatactg caagcgacga
tgcagacaca caccatcag ttcaaccttt tgtgactcgg aagaatgtgg

Die AS-Sequenz des benutzten 1. 5'3' Frames:

MACSSSLSSKWASWNASSFPHPSVQPFVTRKNVVRYHKQTSESTITTTTLSSNLNS
QFMQVYETLKSELIHDPSEFDDDSRQWVERMIDYTVPGGKMVRGYSVVD SYQLL
KGQELTEEEAFLVCALGWCTEWFQAFILAHDDIMDGSHTRRGQPCWFRLPEVGVV
AINDGVLLRNHVRILKKHFQGKAYYVHLVDFNETEFQTISGQMIDTIARLAGQKDL
SKYTMSLNRKIVQYKGSYYSCYLPIACALLMFGENLEDHVQVKDILVELGMYYYQIQN

DYLDTFGDPNVFGKTGTDIEECKCSWLIKALELSNEEQKKILSENYGKNDDPAKVAK
VKEVYHALDLKSAYEDYETNLYEKSMKAIKALPSITVQAVLKSCLEKMYEGHK



Quelle: <http://pfam.xfam.org/family/PF00348.16#tabview=tab4>

Ähnlichkeit: Die Suchsequenz und das Logo haben kaum Übereinstimmungen.