

Notes from Melting on 26/11/2021

N

N

N

atgattcatcgtttgtgagcgcgatgctggcggtttgcgctgctgaccattagcctgggtggatggccatagcctgaactaagtgtatccgaaccatccgcgcagcgcgcgcaacaagatcgctgcatggccggatcatcgctgtttgcgcggcgcggttaacgcacotatatattttggcgcgctgggtgocgtttggcaaaatttggcgcaocggcgcgagcgatgcgacc

>CCH57671.1  
Length = 288

Score = 72 bits (174), Expect = 1e-12  
Identities = 46/81 (57%), Positives = 52/81 (64%), Gaps = 3/81 (4%)  
Frame = +1

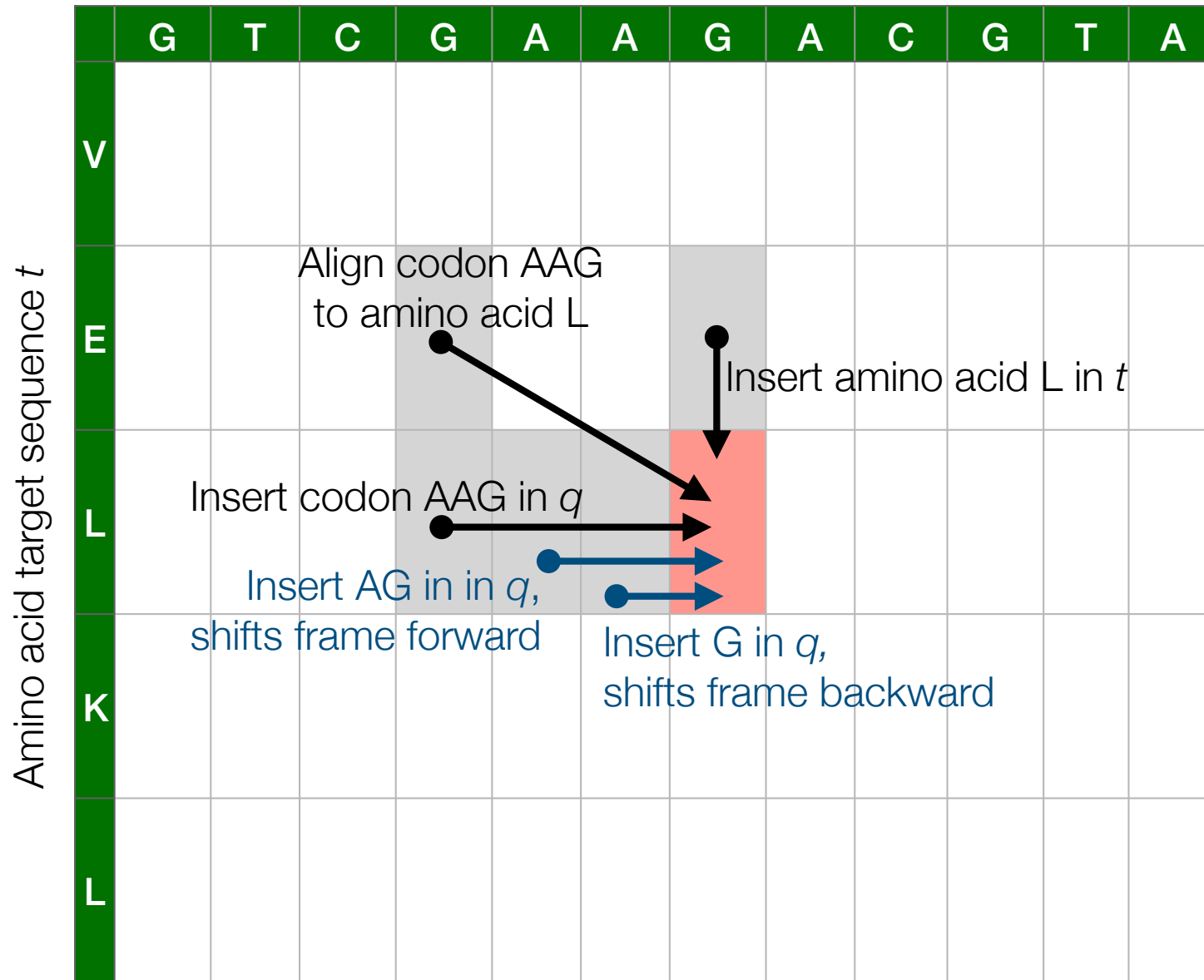
Query:	1	MIHPFVSRMLAFALLTISLVDRHSLN*VYPNHPRQRRSNKDRLHGPDHRLFA--PGG*RTYIFGGLVPFGKIWRTGASDAT	236
		MIHPFVSRMLAFALLTISLV R +L + P + K ++ D + P IFGGLVPFGKIWRTGASDAT	
Sbjct:	1	MIHPFVSRMLAFALLTISLV-RPALAQLSLPEPSPPTTIKQKIGFTDLTIAYSRPAVKGRTIIFGGLVPFGKIWRTGASDAT	80

>CCH57671.1 hypothetical protein BN8\_p06879 plasmid [Fibrisoma limi BUZ 3]  
Length = 288

Score = 177 (bits 448), Expect = 3e-41  
Identities = 100/104 (96%), Positives = 100/104 (96%), Gaps = 4/104 (4%)  
Frame = +1

Query:	1	MIHPFVSRMLAFALLTISLV	RPALAQLSLPEPSPTTIKQ	KIGFTDLTIAYSRPAVKGRT	IFGGLVPFGKIWRTGASDAT	236
		MIHPFVSRMLAFALLTISLV	RPALAQLSLPEPSPTTIKQ	KIGFTDLTIAYSRPAVKGRT	IFGGLVPFGKIWRTGASDAT	
Sbjct:	1	MIHPFVSRMLAFALLTISLV	RPALAQLSLPEPSPTTIKQ	KIGFTDLTIAYSRPAVKGRT	IFGGLVPFGKIWRTGASDAT	80

DNA query sequence  $q$



- Subst. matrix  
- gap penalty  
- frame-shift penalty

# Tasks

- Main program agents  
I/O
- Utilities
  - FastA
  - BLAST
  - DNA - codon lookup  
amino
- DP implementation
- Test cases

# Ex hnt of project:

- nice Command-line program

Can set penalties, choose sub. matrix,  
provide DNA & protein sequence pair

- get alignmt using \ / notation -

- does frame-shift DP

- test cases

- some thing to support that program  
works as intended

- speed assessment



Final product:

github repository  
with documentation,  
index examples.

Things to form a view upon:

- Would affine gap scoring be easy to implement?
- Would linear space be easy to implement?



>WP\_006339994.1  
Length = 612

Score = 963 bits (2489), Expect = 0  
Identities = 503/618 (81%), Positives = 563/618 (91%), Gaps = 12/618 (2%)  
Frame = -1

Query:	2299870	MAKQYDESSV/KTLDALSHIRLRPGMYIGRL/GDGAHPDDGIYVMLKEVIDNSVDEFIMKQKKISIELDEETQTVSVRDYGRGIPLGKVIDCVSQINTGAKYNDDVFQFSVGLNGVGTK	2299519
		MAK+YDESSV KTL AL HIRLR GMYIGRL GDG+HPDDGIY+M+KEVIDNSVDEFIMK G +I I +D E +SVRD+GRGIPLGKVIDCVSQINTGAKYNDDVFQFSVGLNGVGTK	
Sbjct:	1	MAKKYDESSV-KTLGALEHIRLRSGMYIGRL-GDGSHPDGIYIMVKEVIDNSVDEFIMKHGNRIIIGVDPEQGLISVRDFGRGIPLGKVIDCVSQINTGAKYNDDVFQFSVGLNGVGTK	118
Query:	2299518	AVNALSSHFVVRSHRDGKYLEADFNRGILKQ/KKEGTTKEPNGTYIQFIPDGELFK/KFRFQKEYILKRLWHYAYLNTGLVLSFNGEIIESKNGLLDLLNAEVNVDDRLYEPLHYRGKML	2299167
		AVNALSSHF+VRSHR+GK +EA F++G L + +K T EP+GT++FIPD E+FK KFRF+ EYI KRLWHYAYLNTGL+L +NG+ I S+NGLLDLLN EV +DRLYEPL+YRGK+L	
Sbjct:	119	AVNALSSHFVVRSHRREGKSMEAVFSQGLIR-EKSQKTSEPDGTYVEFIPDKEIFK-KFRFEPEYIQKRLWHYAYLNTGLILEYNGQDIYSENGLLDLLNEEV-TEDRLYEPLYRGKLL	235
Query:	2299166	EFAF/LHTQSYGETYFSFVNGQYTSDDGGTHLSAFREGILKGVNEYSKKNFQGVVDVREGIVGTIILVRVKDPLFESQTKNKLGNSELRGPVVQEVKDAVVDLLYKNPTTANQIIERIVFNEK	2298811
		EFAF LHTQSYGE+YFSFVNGQYTSDDGGTHLSAFREGILKGVNEYSKKNFQGVVD+REGIVGT+ VR+KDP+FESQTKNKLGN+ELRGP+VQEVKDAVV++LYKN +TAN IIERIVFNEK	
Sbjct:	236	EFAF-LHTQSYGESYFSFVNGQYTSDDGGTHLSAFREGILKGVNEYSKKNFQGVDLREGIVGTAVRIKDPVFESQTKNKLGNTELRGPLVQEVKDAVVNMLYKNSSTANSIIERIVFNEK	354
Query:	2298810	LRKELANV/KKEAKEKQKKISFKIPKLRDCKFHFQDGTGTEHGDRSMIFLTEGDSASASIVAARDPLTQAVF\FSLRGKPLNVHGMKLDQLYKNEEMFNLMSALNVEDEIEKLRQ/KIILA	2298461
		LRKELA V KKEAKEKQKKISFKIPKLRDCK+H QD + G+++MIFLTEGDSASASIVA+RDPLTQAV FSLRGKPLNVHGMK+DQLYKNEEMFNLMSALN+ED+IE LRYQ K+ILA	
Sbjct:	355	LRKELAAV-KKEAKEKQKKISFKIPKLRDCKYHAQDKADVGEKTMIFLTEGDSASASIVASRDPLTQAV--FSLRGKPLNVHGMKIDQLYKNEEMFNLMSALNIEDDIENLRYQ-KVILA	470
Query:	2298460	TDADVDGMHIRNLLITFFLTTFEGLVLNGHLYILETPLFKVRNKE/KTFYCYNEEERDKAIEKL/KKGIEITRFKGLGEISPNEFKQFISKDIRLIPVTVMNFSDIKATLEFYMGKNTPE	2298109
		TDADVDGMHIRNLLITFFLTTFEGLVLNGHL+ILETPLFKVRN+E KT YCY+EEERDKAI++L KKGIE+TRFKGLGEISPNEFKQFI KDIRLIPVT+++FSDIK TL+FYMGNKTP+	
Sbjct:	471	TDADVDGMHIRNLLITFFLTTFEGLVLNGHLFILETPLFKVRNRE-KTLYCYSEEERDKAIKQL-KKGIEVTRFKGLGEISPNEFKQFIGKDIRLIPVTISLFSDIKTTLDFYMGKNTPK	588
Query:	2298108	RKQFIMRNLVNEDEEALK 2298055	
		RKQFIM NL+NE E L+	
Sbjct:	589	RKQFIMENLINELEPTLE 606	