

Question 1.

Analysis – “nr” database, (100 results),

- E value = 0.0, for all the results
- 23 results have 100% Query coverage
- There is 1 result with 100% identity
- 66.96% is the Lowest percentage identity

Analysis - “Swiss_PROT” database, (100 results)

- One E-value is 0.0 , E-value increases till it reaches 8.4 for one result.
- Query coverage maximum value = 98% (only 1 result) and goes to min of 6%.
- No result has 100% identity.
- 22.07% is the Lowest percentage identity observed.

Question 2.

Algorithm parameters

General Parameters

Max target sequences: 5000
Select the maximum number of aligned sequences to display

Short queries: ☒ Automatically adjust parameters for short input sequences

Expect threshold: 0.05

Word size: 6

Max matches in a query range: 0

Scoring Parameters

Matrix: BLOSUM62

Gap Costs: Existence: 11 Extension: 1

Compositional adjustments: Conditional compositional score matrix adjustment

Filters and Masking

Filter: ☒ Low complexity regions

Mask: ☐ Mask for lookup table only
☐ Mask lower case letters

General parameters displayed:

1. Max target sequences
2. Expected threshold
3. Word size
4. Maximum matches in a query range

Scoring Parameters

1. Matrix
2. Gap costs
3. Compositional alignments

Filter and Masking

1. Filter
2. Mask

Question 3:

RscC [Pseudomonas fluorescens]

Sequence ID: [AAK81929.1](#) Length: 713 Number of Matches: 1

Range 1: 22 to 690 [GenPept](#) [Graphics](#)

[Next Match](#) [Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
530 bits(1366)	0.0	Compositional matrix adjust.	292/676(43%)	416/676(61%)	28/676(4%)
Query 8		RCRLLGALLMLCATLPAG--AQTPADWKEQSYAYSADRTPLSTVLQDFADGHSVDLHLG			64
		R + +L+ C PA A PA+WK +YAY AD PL VL+DFA L +			
Sbjct 22		RAKWQWLVLGICIMAPAHNLLAAIPAENKNTAYAYEADHKPLREVLEDAQTFTGTLQIE			81
Query 65		NVEDTEVTAKIRAENASAFDLRLALEHHFQWFVYNNNTLYVSPQDEQSSERLEISPDAAPD			124
		+ + +V KIRA + LDRL +EH FQW++YNNNTL+VS D+Q S RLE+S + D			
Sbjct 82		GLLEGDVNGKIRANTPQSMILDRLGVEHRFQWLYNNNTLFVSTLDQESARLEVSSETISD			141
Query 125		IKQALSGIGLLDPRFGWELPDDGVVLVTGPPQYLELVKRFSEQREKKEDRRKVMTFPLR			184
		+KQAL+ IGLLD RFGWELP+DGVVLV+GP Y++ +K+FS +R +++++ V++FPL+			
Sbjct 142		LKQALTDIGLLDSRFGWELPEDGVVLVSGPKTYIDQIKQFSSKRRSADEKQSVLSFPLK			201
Query 185		YASVADRTIHYRDQTVVIPGVATMLNELMNGKRAAPASA-SGIDSTPGGPDNMMQNTQ			243
		+A+ ADR + YR + +V+PGVA +L L+ + A+ + S DS+ P T ++ +			
Sbjct 202		FANAADRKVDYRGEKLVVPGVANILRGLLEPRASLTGMSQPDSSQPSPLTPNVPRLG			261
Query 244		TLLSRLSSRNKTSNRAGGRDN-----EIEDVSGRISADVRNALLIRDDDKRHDEYS			295
		LL ++ N AG D + + R+ ADVRNNA+LI D +R Y			
Sbjct 262		PLLQMLGAN---GNAGQLDTGPTVTPRAPVSKSRIRVEADVRNNAVLIYDLPERQAMYR			318
Query 296		QLIAKIDVPQNLVEIDAVILDIDRTALNRLANWQATLGGVTGGSSLSMSGSTLFVSDFK			355
		LI ++DV + L+EIDA+ILDI+RT L NW GG ++ G+ + D +			
Sbjct 319		DLITQLDVARKLIEIDAIILDIERTQLREFGVNWGFQNSRFRGGVNMAGTSSQVSIDHR			378
Query 356		-RFFADIQALEGEGTASIVANPSVLTLENQPAVIDFSQTAYITATGERVADIQPVTAGTS			414
		RF+AD+ + G+G A++V+NPSVLTLENQPAVIDF++T YI+ G A I PVT GTS			
Sbjct 379		DRFYADMPSTGGQGPATMVSNNPSVLTLENQPAVIDFNRTQYISP-GRDYATILPVTVGTS			437
Query 415		LQVTPRAVGNEGHSIQLMIDIEDGHV-QTNG--DGQATGVKRGTVSTQALISENRLVL			471
		LQV PR G I L++DIEDG++ +TN D V+RG VSTQA++ E R+LV+			
Sbjct 438		LQVPRVTTGRGVHQIHLVVDIEDGNLDETNPDPNHLVRRGKVSTQAVMQEKSLVV			497
Query 472		GGFHVEESADRRRIPLLGDIPLWLGQ-LFSSKRHEISQRQLFILTPRLIGDQTDPTRYV			530
		GGFHV +S+D+ ++IPLLGDIPLG+ L SS ++R+RLFILTPR+IGDQ DP+RY+			
Sbjct 498		GGFHVTDSSDQKKIPLLGDIPLGKTLVSSTERHNNRRERLFILTPRVIGDQDDPSRYL			557
Query 531		TADNRQQLSDAMGRVRRHSS---VNQHDVVENALRDLAEGQSPAGFPQPTSGTRLSEV			586
		D++ +L A+ + RR+S + + D++ R L G+ P F L+ +			
Sbjct 558		PQDDQAEALQAALTPLARRYSPHQPVIKRSDIITTLAR-LVSGEVPKAFNAARMPLGLNTL			616
Query 587		CRSTPALLFESTRGQWYSSSTNGVQLSVGVVRNTSSKPLRFDEANCASKRTLAVAVWPHS			646
		C + L + R QWY+ V +V V+RN + +R DE C++ +TLAV VWP +			
Sbjct 617		CSTRDLLALNTERSQWYAGPDYNV--AVVVLNRNQFKRNVRIDEKECSNSQTLAVTVWPRA			674
Query 647		ALAPGESAEVYLAMP 662			
		L PGE AEV++AM P			
Sbjct 675		WLKPGEEAEVFAMRP 690			

Enter the accession number in the first box select "Align 2 or multiple sequences" from the drop-down menu.

Type the accession number of the second one, select BLAST.

Result - 43.20%

Question: 4

Range 1: 1 to 147 [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
221 bits(564)	1e-80	Compositional matrix adjust.	102/147(69%)	121/147(82%)	0/147(0%)
Query 1	MVHWTAEKQLITGLWGKVNVAECGAEALARLLIVYPWTQRFFASFGNLSSTAILGNPM	60			
Sbjct 1	MVHLTPEEKSAVTALWGKVNVDDEVGGEALGRLLVYPWTQRFFESFGDLSTPDAMGNPK	60			
Query 61	VRAHGKKVLTSGDAVKNLNDNIKNTFSQLSELHCDKLHVDPENFRLLGDILIIIVLAAHFS	120			
Sbjct 61	VKAHGKKVLGAFSDGLAHLNLIKGTATLSELHCDKLHVDPENFRLLGNVLCVLAHHFG	120			
Query 121	KDFTPECQAAWQKLVRVVAHALARKYH	147			
Sbjct 121	KEFTPPVQAAAYQKVAGVANALAHKYH	147			

- Get both sequences from uniprot.
- Blast both the sequences and their uniprot IDs(prior query)4
- Identity is not the same as similarity.
- Nature / properties of two Amino acids.

Question 5:

h(sequence1) =

'MVHLTPEEKSAVTALWGKVNVDDEVGGEALGRLLVYPWTQRFFESFGDLSTPDAMGNPKVK
AHGKKVLGAFS D

GLAHLNLIKGTATLSELHCDKLHVDPENFRLLGNVLCVLAHHFGKEFTPPVQAAAYQKVAGV
ANALAHKYH' c(sequence2) =

'MVHWTAEKQLITGLWGKVNVAECGAEALARLLIVYPWTQRFFASFGNLSSTAILGNPMVRA
HGKKVLTSGD
AVKNLNDNIKNTFSQLSELHCDKLHVDPENFRLLGDILIIIVLAAHFSKDFTPECQAAWQKLVRVVAH
ALARKYH'

```
human= "MVHLTPEEKSAVTALWGKVNVDDEVGGEALGRLLVYPWTQRFFESFGDLSTPDAMGNPKVKAHGKKVLGAFSDGLAHLNLIKGTATLSELHCDKLHVDPENFRLLGNVLCVLAHHFGKEFTPPVQAAAYQKVAGVANALAHKYH"
chicken= "MVHWTAEKQLITGLWGKVNVAECGAEALARLLIVYPWTQRFFASFGNLSSTAILGNPMVRAHGKKVLTSGDAVKNLNDNIKNTFSQLSELHCDKLHVDPENFRLLGDILIIIVLAAHFSKDFTPECQAAWQKLVRVVAHALARKYH"
def match(human,chicken):
    human_seq_len= len(human)
    chicken_seq_len= len(chicken)
    occurrence_h= [0 for i in range(human_seq_len-4)]
    occurrence_c=[0 for i in range(chicken_seq_len-4)]
    for i in range(human_seq_len-4):
        penta_p= human[i:i+5]
        for j in range(chicken_seq_len-4):
            if human[j:j+5]==penta_p:
                occurrence_h[i]+=1
            if chicken[j:j+5]==penta_p:
                occurrence_c[j]+=1
    if occurrence_h[i]>0 and occurrence_c[i]>0:
        print("The occurrence of sequence",penta_p,"in human sequence:", occurrence_h[i],"in chicken sequence:", occurrence_c[i])
match(human,chicken)
```

The occurrence of sequence LWGKV in human sequence: 1 in chicken sequence: 1

The occurrence of sequence WGKVN in human sequence: 1 in chicken sequence: 1

The occurrence of sequence GKVNV in human sequence: 1 in chicken sequence: 1

The occurrence of sequence VYPWT in human sequence: 1 in chicken sequence: 1

The occurrence of sequence YPWTQ in human sequence: 1 in chicken sequence: 1

The occurrence of sequence PWTQR in human sequence: 1 in chicken sequence: 1

The occurrence of sequence WTQRF in human sequence: 1 in chicken sequence: 1

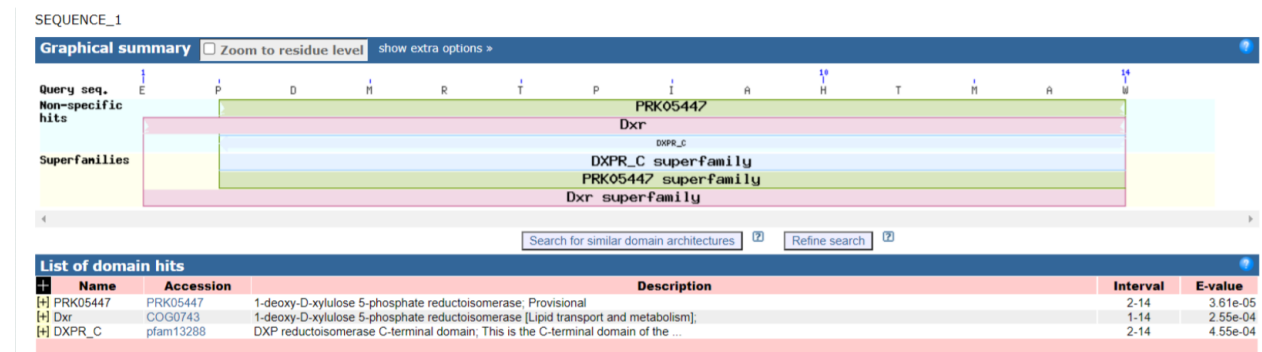
The occurrence of sequence TQRFF in human sequence: 1 in chicken sequence: 1

The occurrence of sequence AHGKK in human sequence: 1 in chicken sequence: 1
The occurrence of sequence HGKKV in human sequence: 1 in chicken sequence: 1
The occurrence of sequence GKKVL in human sequence: 1 in chicken sequence: 1
The occurrence of sequence LSELH in human sequence: 1 in chicken sequence: 1
The occurrence of sequence SELHC in human sequence: 1 in chicken sequence: 1
The occurrence of sequence ELHCD in human sequence: 1 in chicken sequence: 1
The occurrence of sequence LHCDK in human sequence: 1 in chicken sequence: 1
The occurrence of sequence HCDKL in human sequence: 1 in chicken sequence: 1
The occurrence of sequence CDKLH in human sequence: 1 in chicken sequence: 1
The occurrence of sequence DKLHV in human sequence: 1 in chicken sequence: 1
The occurrence of sequence KLVHD in human sequence: 1 in chicken sequence: 1
The occurrence of sequence LHVDP in human sequence: 1 in chicken sequence: 1
The occurrence of sequence HVDPE in human sequence: 1 in chicken sequence: 1
The occurrence of sequence VDPEN in human sequence: 1 in chicken sequence: 1
The occurrence of sequence DPENF in human sequence: 1 in chicken sequence: 1
The occurrence of sequence PENFR in human sequence: 1 in chicken sequence: 1
The occurrence of sequence ENFRL in human sequence: 1 in chicken sequence: 1
The occurrence of sequence NFRLI in human sequence: 1 in chicken sequence: 1
The occurrence of sequence FRLIG in human sequence: 1 in chicken sequence: 1

Question 7:

Entry	Entry name	Status	Protein names	Gene names	Organism	Length
Q9Y5I7	TIM9_HUMAN	reviewed	Mitochondrial import inner membrane translocase subunit Tim9	TIMM9 TIM9 TIM9A TIMM9A	Homo sapiens (Human)	89
P87108	TIM10_YEAST	reviewed	Mitochondrial import inner membrane translocase subunit TIM10 (Mitochondrial intermembrane protein MRS11)	TIM10 MRS11 YHR005C-A YHR005BC	Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's yeast)	93
P62072	TIM10_HUMAN	reviewed	Mitochondrial import inner membrane translocase subunit Tim10	TIMM10 TIM10	Homo sapiens (Human)	90
Q60220	TIM8A_HUMAN	reviewed	Mitochondrial import inner membrane translocase subunit Tim8 A (Deafness dystonia protein 1) (X-linked deafness dystonia protein)	TIMM8A DDP DDP1 TIM8A	Homo sapiens (Human)	97
O74700	TIM9_YEAST	reviewed	Mitochondrial import inner membrane translocase subunit Tim9	TIM9 YEL020W-A YEL020BW	Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's yeast)	87
P53299	TIM13_YEAST	reviewed	Mitochondrial import inner membrane translocase subunit Tim13	TIM13 YGR181W G7157	Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's yeast)	105
Q9Y5L4	TIM13_HUMAN	reviewed	Mitochondrial import inner membrane translocase subunit Tim13	TIMM13 TIM13B TIMM13A TIMM13B	Homo sapiens (Human)	95
P57744	TIM8_YEAST	reviewed	Mitochondrial import inner membrane translocase subunit Tim8	TIM8 YJR135W-A YJR135BW	Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's yeast)	87
Q9Y5I9	TIM8B_HUMAN	reviewed	Mitochondrial import inner membrane translocase subunit Tim8 B (DDP-like protein) (Deafness dystonia protein 2)	TIMM8B DDP2 DDP1 TIM8B	Homo sapiens (Human)	83
Q17754	TIM9_CAEL	reviewed	Mitochondrial import inner membrane translocase subunit Tim9	tim-9.1 tim9a tim-9 C06C3.11	Caenorhabditis elegans	90
Q9WV98	TIM9_MOUSE	reviewed	Mitochondrial import inner membrane translocase subunit Tim9	Timm9 Tim9 Tim9a Timm9a	Mus musculus (Mouse)	89
Q9Y0V6	TIM10_CAEL	reviewed	Mitochondrial import inner membrane translocase subunit Tim10	tim-10 tim-10 Y66D12A.22	Caenorhabditis elegans	86
Q9WVA1	TIM8A_RAT	reviewed	Mitochondrial import inner membrane translocase subunit Tim8 A (Deafness dystonia protein 1 homolog)	Timm8a Ddp1 Tim8a	Rattus norvegicus (Rat)	97
Q9WV97	TIM9_RAT	reviewed	Mitochondrial import inner membrane translocase subunit Tim9	Timm9 Tim9 Tim9a Timm9a	Rattus norvegicus (Rat)	89
Q9KH48	TIM13_ARATH	reviewed	Mitochondrial import inner membrane translocase subunit Tim13	TIM13 At1g61570 T25B24.8 T25B24_16	Arabidopsis thaliana (Mouse-ear cress)	87
Q9N408	TIM8_CAEL	reviewed	Mitochondrial import inner membrane translocase subunit Tim8	ddp-1 tim-8 Y39A3C0.4	Caenorhabditis elegans	83
Q9XGK9	TIM9_ARATH	reviewed	Mitochondrial import inner membrane translocase subunit Tim9 (Protein EMBRYO DEFECTIVE 2474)	TIM9 EMB2474 At3g46560 F12A12.80	Arabidopsis thaliana (Mouse-ear cress)	93
Q9WVA2	TIM8A_MOUSE	reviewed	Mitochondrial import inner membrane translocase subunit Tim8 A (Deafness dystonia protein 1 homolog)	Timm8a1 Ddp1 Tim8a Timm8a	Mus musculus (Mouse)	97
P62073	TIM10_MOUSE	reviewed	Mitochondrial import inner membrane translocase subunit Tim10	Timm10 Tim10	Mus musculus (Mouse)	90
Q9XGY4	TIM8_ARATH	reviewed	Mitochondrial import inner membrane translocase subunit Tim8	TIM8 At5g50810 K7B16.3	Arabidopsis thaliana (Mouse-ear cress)	77

Question 8:



Residue 2-14 is a domain named PRK05447 which is conserved across many organisms. BLAST results show this sequence is commonly found in - Escherichia coli, Klebsiella pneumoniae. 100 sequences, only 10% of the sequences have very minimal E value.

Highest E value is 28, this sequence has 57% query coverage and 75% identity with the given sequence.