

# Question 1

NR

NIH U.S. National Library of Medicine  
National Center for Biotechnology Information

COVID-19 Information  
Public health information (CDC) | Research information (NIH) | SARS-CoV-2 data (NCBI) | Prevention and treatment information (HHS) | Español

BLAST® » blastp suite » results for RID-16H9SURF016

Home Recent Results Saved Strategies Help

[Edit Search](#) [Save Search](#) [Search Summary](#) [How to read this report?](#) [BLAST Help Videos](#) [Back to Traditional Results Page](#)

Job Title 1336093(Genbank)Outer membrane integral  
RID 16H9SURF016 Search expires on 02-22 14:54 pm [Download All](#)  
Program BLASTP [Citation](#)  
Database nr [See details](#)  
Query ID ICIQuery\_12710  
Description 1336093(Genbank)Outer membrane integral membrane p...  
Molecule type amino acid  
Query Length 676  
Other reports [Distance tree of results](#) [Multiple alignment](#) [MSA viewer](#)

**Filter Results**

Organism only top 20 will appear ☐ exclude  
Type common name, binomial, taxid or group name  
[Add organism](#)

Percent Identity E value Query Coverage  
 to   to   to   
[Filter](#) [Reset](#)

**Descriptions** Graphic Summary Alignments Taxonomy

Sequences producing significant alignments

Download Select columns Show 100

☒ select all 100 sequences selected

GenPlot Graphics Distance tree of results Multiple alignment MSA Viewer

| Description  | Scientific Name      | Max Score | Total Score | Query Cover | E value | Per. Ident | Acc. Len | Accession      |
|--|----------------------|-----------|-------------|-------------|---------|------------|----------|----------------|
| <input checked="" type="checkbox"/> type III secretion system outer membrane ring subunit ScdC [Erwinia amylovora]       | Erwinia amylovora    | 1305      | 1305        | 100%        | 0.0     | 100.00%    | 676      | WP_004155366.1 |
| <input checked="" type="checkbox"/> type III secretion system outer membrane ring subunit ScdC [Erwinia amylovora]       | Erwinia amylovora    | 1304      | 1304        | 100%        | 0.0     | 99.85%     | 676      | WP_168421624.1 |
| <input checked="" type="checkbox"/> type III secretion system outer membrane ring subunit ScdC [Erwinia amylovora]       | Erwinia amylovora    | 1303      | 1303        | 100%        | 0.0     | 99.85%     | 676      | WP_168385176.1 |
| <input checked="" type="checkbox"/> type III secretion system outer membrane ring subunit ScdC [Erwinia amylovora]       | Erwinia amylovora    | 1289      | 1289        | 100%        | 0.0     | 98.52%     | 677      | WP_004168436.1 |
| <input checked="" type="checkbox"/> Type III secretion system outer membrane pore HrcC [Erwinia amylovora ATCC BAA-2158] | Erwinia amylovora... | 1287      | 1287        | 100%        | 0.0     | 98.38%     | 677      | CBX79367.1     |
| <input checked="" type="checkbox"/> type III secretion system outer membrane ring subunit ScdC [Erwinia sp. Ejp617]      | Erwinia sp. Ejp617   | 1264      | 1264        | 100%        | 0.0     | 96.89%     | 676      | WP_014543268.1 |

[Feedback](#)

Swiss Prot

NIH U.S. National Library of Medicine  
National Center for Biotechnology Information

COVID-19 Information  
Public health information (CDC) | Research information (NIH) | SARS-CoV-2 data (NCBI) | Prevention and treatment information (HHS) | Español

BLAST® » blastp suite » results for RID-16HDTTUF013

Home Recent Results Saved Strategies Help

[Edit Search](#) [Save Search](#) [Search Summary](#) [How to read this report?](#) [BLAST Help Videos](#) [Back to Traditional Results Page](#)

Job Title 1336093(Genbank)Outer membrane integral  
RID 16HDTTUF013 Search expires on 02-22 14:56 pm [Download All](#)  
Program BLASTP [Citation](#)  
Database swissprot [See details](#)  
Query ID ICIQuery\_53522  
Description 1336093(Genbank)Outer membrane integral membrane p...  
Molecule type amino acid  
Query Length 676  
Other reports [Distance tree of results](#) [Multiple alignment](#) [MSA viewer](#)

**Filter Results**

Organism only top 20 will appear ☐ exclude  
Type common name, binomial, taxid or group name  
[Add organism](#)

Percent Identity E value Query Coverage  
 to   to   to   
[Filter](#) [Reset](#)

**Descriptions** Graphic Summary Alignments Taxonomy

Sequences producing significant alignments

Download Select columns Show 100

☒ select all 33 sequences selected

GenPlot Graphics Distance tree of results Multiple alignment MSA Viewer

| Description  | Scientific Name       | Max Score | Total Score | Query Cover | E value | Per. Ident | Acc. Len | Accession |
|--|-----------------------|-----------|-------------|-------------|---------|------------|----------|-----------|
| <input checked="" type="checkbox"/> RecName: Full=Type 3 secretion system secretin; Short=T3SS secretin; AltName: Full=Hypersensitivity response secr... Pseudomonas syc...    | Pseudomonas syc...    | 544       | 544         | 96%         | 0.0     | 44.40%     | 701      | Q01723.2  |
| <input checked="" type="checkbox"/> RecName: Full=Type 3 secretion system secretin; Short=T3SS secretin; AltName: Full=YscC secretin; Flags: Precurs... Yersinia enterocol...  | Yersinia enterocol... | 242       | 242         | 72%         | 4e-70   | 30.86%     | 607      | Q01244.1  |
| <input checked="" type="checkbox"/> RecName: Full=Type 3 secretion system secretin; Short=T3SS secretin; AltName: Full=Type III secretion protein YscC... Yersinia pestis      | Yersinia pestis       | 236       | 236         | 72%         | 6e-68   | 30.80%     | 607      | Q56974.1  |
| <input checked="" type="checkbox"/> RecName: Full=Type 3 secretion system secretin; Short=T3SS secretin; AltName: Full=Hypersensitivity response secr... Ralstonia solanac...  | Ralstonia solanac...  | 201       | 201         | 72%         | 3e-55   | 28.37%     | 568      | Q52498.1  |
| <input checked="" type="checkbox"/> RecName: Full=Type 3 secretion system secretin; Short=T3SS secretin; AltName: Full=Hypersensitivity response secr... Xanthomonas eu...     | Xanthomonas eu...     | 166       | 166         | 73%         | 2e-42   | 26.08%     | 607      | P80151.1  |
| <input checked="" type="checkbox"/> RecName: Full=SPL-1 type 3 secretion system secretin; Short=T3SS-1 secretin; AltName: Full=Protein InvG; Flags: Pr... Salmonella enteri... | Salmonella enteri...  | 163       | 163         | 69%         | 7e-42   | 25.77%     | 562      | P35672.3  |

[Feedback](#)

|   |                          |
|---|--------------------------|
| <b>NR</b>                                   |                          |
| Number of sequences with 100% coverage      | 12                       |
| Number of sequences with E value close to 0 | 100                      |
| Number of sequences with identity 100%      | 1                        |
| Lowest % identity                           | 67.67%                   |
| Max_score range                             | 1305-505                 |
| <b>Swiss Prot</b>                           |                          |
| Number of sequences with 100% coverage      | 0 (max coverage = 96%)   |
| Number of sequences with E value close to 0 | 33                       |
| Number of sequences with identity 100%      | 0 (max identity = 44.4%) |
| Lowest % identity                           | 22.22%                   |
| Max_score range                             | 544-50.8                 |

## Question 2

Algorithm parameters

Restore default search parameters

**General Parameters**

Max target sequences: 100  
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

**BLAST** Search database swissprot using Blastp (protein-protein BLAST)  
☒ Show results in a new window

Parameters used:

General parameters displayed in comparison

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

Scoring parameters

- 1. Matrix
- 2. Gap costs
- 3. Compositional alignments

Filter and Masking

- 1. Filter options
- 2. Mask options

# Question 3

tr|Q93PY6|Q93PY6\_PSEFL RscC OS=Pseudomonas fluorescens OX=294 GN=rscC PE=3 SV=1  
Sequence ID: **Query\_63209** Length: **713** Number of Matches: **1**

Range 1: 22 to 690 [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      | RCRLGALLMLCATLPAG---AQTPADWKEQSYAYSADRTPLSTVLQDFADGHSVDLHLG   |              |              | 64         |
| Sbjct          | 22     | R + +L+ C PA A PA+WK +YAY AD PL VL+DFA L +                    |              |              | 81         |
| Query          | 65     | NVEDTEVTAKIRAENASAFDLRLALEHHFQWFVYNNNTLYVSPQDEQSSERLEISPDAAPD |              |              | 124        |
| Sbjct          | 82     | + + +V KIRA + LDRL +EH FQW++YNNTL+VS D+Q S RLE+S + D          |              |              | 141        |
| Query          | 125    | IKQALSIGIGLLDPRFGWGELPDDGVVLVTGPPQYLELVKRFSEQREKKEDRRKVMFTFLR |              |              | 184        |
| Sbjct          | 142    | +KQAL+ IGLLD RFGWGELP+DGVVLV+GP Y++ +K+FS +R +++++ V++FPL+    |              |              | 201        |
| Query          | 185    | YASVADRTIHYRDQTVVIPGVATMLNELMNGKRAAPASA-SGIDSTPGGPDTSMMQNTQ   |              |              | 243        |
| Sbjct          | 202    | +A+ ADR + YR + +V+PGVA +L L+ + A+ + S DS+ P T ++ +            |              |              | 261        |
| Query          | 244    | TLLSRLSSRNKTSNRAGGRDN-----EIEDVSGRISADVRNNALLIRDDDKRHDEYS     |              |              | 295        |
| Sbjct          | 262    | LL ++ N AG D + R+ ADVRNNA+LI D +R Y                           |              |              | 318        |
| Query          | 296    | QLIAKIDVPQNLVEIDAVILDIORTALNRLEANWQATLGGVTGGSSLSMSGSLFVSDFK   |              |              | 355        |
| Sbjct          | 319    | LI ++DV + L+EIDA+ILDI+RT L NW GG ++ G+ + D +                  |              |              | 378        |
| Query          | 356    | -RFFADIQALEGEGTASIVANPSVLTLENQPAVIDFSQYITATGERVADIQPVTAGTS    |              |              | 414        |
| Sbjct          | 379    | RF+AD+ + G+G A++V+NPSVLTLENQPAVIDF++T YI+ G A I PVT GTS       |              |              | 437        |
| Query          | 415    | LQVTPRAVGNEGHSSIQLMIDIEDGHV-QTNG--DGQATGVKRGTVSTQALISENRALVL  |              |              | 471        |
| Sbjct          | 438    | LQV PR G I L++DIEDG++ +TN D V+RG VSTQA++ E R+LV+              |              |              | 497        |
| Query          | 472    | GGFHVEESADRRRIPLLGDIPWLGO-LFSSKRHEISQRQRLFILTPRLIGDQTDPTRYV   |              |              | 530        |
| Sbjct          | 498    | GGFHV +S+D+ ++IPLLGDIP LG+ L SS ++R+RLEFILTPR+IGDQ DP+RY+     |              |              | 557        |
| Query          | 531    | TADNRQQLSDAMGRVERRHSS----VNQHDVVENALRDLAEGQSPAGFPQTSCTRLSEV   |              |              | 586        |
| Sbjct          | 558    | D++ +L A+ + RR+S + + D++ R L G+ P F L+ +                      |              |              | 616        |
| Query          | 587    | CRSTPALLFESTRGQWYSSSTNGVQLSVGVVRNTSSKPLRFDEANCASKRTLAVAVWPHS  |              |              | 646        |
| Sbjct          | 617    | C + L + R QWY+ V +V V+RN + +R DE C++ +TLAV VWP +              |              |              | 674        |
| Query          | 647    | ALAPGESAEVYLAMP 662   |              |              |            |
| Sbjct          | 675    | L PGE AEV++AM P 690   |              |              |            |

Sequence identity = 43%

## Question 4

Alignment view Pairwise Restore defaults Download

1 sequences selected

[Download](#) [Graphics](#) [Next](#) [Previous](#) [Descriptions](#)

**sp|P68871|HBB\_HUMAN Hemoglobin subunit beta OS=Homo sapiens OX=9606 GN=HBB PE=1 SV=2**  
 Sequence ID: **Query\_45853** Length: **147** Number of Matches: **1**

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       | MVHWTAEKQLITGLWGKVNVAECGAEALARLLIVYPWTQRFASFGLSSPTAILGNPM  | 60                           |              |              |           |
| Sbjct 1       | MVH T EEK +T LWGKVVN E G EAL RLL+VYPWTQRF SFG+LS+P A++GNP  | 60                           |              |              |           |
| Query 61      | VRAHGKKVLTSGDAVKNLNDIKNTFSQSELHCDKLHVDPENFRLLGDILIIVLAHFS  | 120                          |              |              |           |
| Sbjct 61      | VKAHGKKVLGAFSDGLAHLNDLKGTFATLSELHCDKLHVDPENFRLLGNVLCVLAHFG | 120                          |              |              |           |
| Query 121     | KDFTPECQAQWQKLVVRVVAHALARKYH                               | 147                          |              |              |           |
| Sbjct 121     | K+FTP QAA+QK+V VA+ALA KYH                                  | 147                          |              |              |           |

The beta subunit of hemoglobin in chicken and human are 69% identical and 82% similar (positives)

## Question 5

In [1]:

```
import re
from DnaUtils import readFasta
import pandas as pd
human = readFasta("human.fasta")[1][0]
chick = readFasta("chicken.fasta")[1][0]
matches = set([human[i:i+5] for i in range(len(human)-5+1)]).intersection(
print("Number of matches =", len(matches))
print("\nMatches:")
freq_h = [len(re.findall(f'(?={match})', human)) for match in matches]
freq_c = [len(re.findall(f'(?={match})', chick)) for match in matches]

df = pd.DataFrame([matches, freq_h, freq_c]).T
df.columns = ["Pentapeptide", "freq_human", "freq_chicken"]
df
```

Number of matches = 27

Matches:

| Out[1]: | Pentapeptide | freq_human | freq_chicken |
|---------|--------------|------------|--------------|
| 0       | WTQRF        | 1          | 1            |
| 1       | NFRLL        | 1          | 1            |
| 2       | GKVVN        | 1          | 1            |
| 3       | HGKKV        | 1          | 1            |
| 4       | VDPEN        | 1          | 1            |
| 5       | LWGKV        | 1          | 1            |
| 6       | HCDKL        | 1          | 1            |
| 7       | PWTQR        | 1          | 1            |
| 8       | VYPWT        | 1          | 1            |
| 9       | ENFRL        | 1          | 1            |
| 10      | CDKLH        | 1          | 1            |
| 11      | PENFR        | 1          | 1            |
| 12      | FRLLG        | 1          | 1            |
| 13      | YPWTQ        | 1          | 1            |
| 14      | TQRFF        | 1          | 1            |
| 15      | AHGKK        | 1          | 1            |
| 16      | SELHC        | 1          | 1            |
| 17      | LHVDP        | 1          | 1            |
| 18      | LHCDK        | 1          | 1            |
| 19      | KLHVD        | 1          | 1            |
| 20      | GKKVL        | 1          | 1            |
| 21      | DPENF        | 1          | 1            |
| 22      | LSELH        | 1          | 1            |
| 23      | DKLHV        | 1          | 1            |
| 24      | ELHCD        | 1          | 1            |
| 25      | HVDPE        | 1          | 1            |
| 26      | WGKVN        | 1          | 1            |

## Question 6

```
blast_seq = "MVH T EEK +T LWGKVN E G EAL RLL+VYPWTRQRF SFG+LS+P A++GNP V-  
LSELHCDKLHVDPENFRLLG++L+ VLA HF K+FTP QAA+QK+V VA+ALA KYH"  
n = len(blast_seq)  
identity = (n - blast_seq.count(' ') - blast_seq.count('+')) / n * 100  
similarity = (n - blast_seq.count(' ')) / n * 100  
query_coverage = n / len(human) * 100  
gaps = blast_seq.count('-') / n * 100  
df = pd.DataFrame([identity, similarity, query_coverage, gaps])  
df.columns = ['Value (%)']  
df.index = ["Identity", "Similarity", "Query Coverage", "Gap Percentage"]  
df
```

|                       | Value (%)  |
|-----------------------|------------|
| <b>Identity</b>       | 69.387755  |
| <b>Similarity</b>     | 82.312925  |
| <b>Query Coverage</b> | 100.000000 |
| <b>Gap Percentage</b> | 0.000000   |

## Question 7

```
ids, seqs = readFasta("uniprot-id_Q9Y5J7+OR+id_074700+OR+id_Q9WV98+OR+id_Q9Y5J7")
orgs = [" ".join(i.split('OS=')[1].split(' ')[0:2]) for i in ids]
prots = [i.split('|')[2].split(' ')[0] for i in ids]
df = pd.DataFrame([prots, orgs]).T
df.columns = ["Protein", "Organism"]
df
```

Out [3]:

|    | Protein    | Organism                  |
|----|------------|---------------------------|
| 0  | TIM9_HUMAN | Homo sapiens              |
| 1  | TIM9_YEAST | Saccharomyces cerevisiae  |
| 2  | TIM9_CAEEL | Caenorhabditis elegans    |
| 3  | TIM9_MOUSE | Mus musculus              |
| 4  | TIM9_RAT   | Rattus norvegicus         |
| 5  | TIM9_ARATH | Arabidopsis thaliana      |
| 6  | T10B_BOVIN | Bos taurus                |
| 7  | TIM9_NEUCR | Neurospora crassa         |
| 8  | TIM9_CANGA | Candida glabrata          |
| 9  | TIM9_DROME | Drosophila melanogaster   |
| 10 | TIM9_BOVIN | Bos taurus                |
| 11 | TIM9_CAEBR | Caenorhabditis briggsae   |
| 12 | TIM9_DEBHA | Debaryomyces hansenii     |
| 13 | TIM9_SCHPO | Schizosaccharomyces pombe |
| 14 | TIM9_DANRE | Danio rerio               |
| 15 | TIM9_DICDI | Dictyostelium discoideum  |
| 16 | TIM9_EMENI | Emericella nidulans       |
| 17 | TIM9_CHICK | Gallus gallus             |
| 18 | TIM9_XENLA | Xenopus laevis            |
| 19 | TIM9_ORYSJ | Oryza sativa              |

## Clustal Omega

```

sp|Q559H1|TIM9_DICDI      -----MDRRLSKKEEERIVNELNKLQMIEMVDTSVNLTKCFQSCITNFRIRKLDDE
sp|Q3SZW4|T10B_BOVIN     -----MEQQQQQQQQQQQLRNLRDPLLVDYNRMTLCLFQRCVPSLHHRALDAE
sp|Q9XGX9|TIM9_ARATH     MDASMMAGLD-GLPEEDKAKMASMIDQLQLRDSLRMYNSLVERCFVDCVDSFTTRKSLQKQ
sp|Q9XGX7|TIM9_ORYSJ     MDKSMLGDL-GLPEEDKMRMAAMVDQLQIRDSLRMYNSLVERCFVDCVDSFTTRKSLQKQ
sp|Q8J1Z1|TIM9_NEUCR     -----MD-GLTAAESRELDQRLQKRQVKEFMSVFGNLVDNCFACVDDFTSKALSGR
sp|Q9Y8A7|TIM9_EMENI     -----MD-GLNAAEQRELANRMERKQMKFMTMYSKLVQRCFDDCVNDFTTSLISR
sp|Q9P7K0|TIM9_SCHPO     -----MD-RLNVKEQEHLTQVLEAKQLKEYLNMYSTLTQNCFSDCVQDFTSSKLSNK
sp|Q6BU42|TIM9_DEBHA     -----MD-QLNVKEQQDFQQIVEQKQMKDFMRLYSNLVSKCFDDCVNDFTSNNLTTK
sp|Q74700|TIM9_YEAST     -----MD-ALNSKEQQEFQKVVEQKQMKDFMRLYSNLVRCFVDCVNDFTTSLKLTNK
sp|Q6FRT3|TIM9_CANGA     -----MD-QLNAKEQQEFQKLVEQKQMKDFMRLYSGLVERCFVDCVNDFTTSSKLTSK
sp|Q17754|TIM9_CAEEL     -----MTSEQNIQTFRDFTLTQYNLVAEQCFNSCVNEFGSRTVSGK
sp|Q61TH2|TIM9_CAEBR     -----MASEQNIQTFRDFTLTQYNLVAEQCFNSCVNEFGSRTVNAK
sp|Q9VYD7|TIM9_DROME     -----MA-KTPENIAIDQLDKDQIKTFSDFLMSYNKLSSETCFDTCIRDFTTRDVKDS
sp|Q9W762|TIM9_DANRE     -----MAAQVTESDQIKQFKEFLGTYNKLTENCFCMDCKVDFTTREVKPE
sp|Q5ZIR8|TIM9_CHICK     -----MAGQISETDQIKQFKEFLGTYNKLTENCFCMDCKIRDFTTREVKEPE
sp|Q4V7R1|TIM9_XENLA     -----MAAQMESDQIKQFKEFLGTYNKLTENCFLDCVKDFTTREVKAPE
sp|Q9WV98|TIM9_MOUSE     -----MAAQIPESDQIKQFKEFLGTYNKLTETCFDLCVKDFTTREVKEPE
sp|Q9Y5J7|TIM9_HUMAN     -----MAAQIPESDQIKQFKEFLGTYNKLTETCFDLCVKDFTTREVKEPE
sp|Q2KIV2|TIM9_BOVIN     -----MAAQIPESDQIKQFKEFLGTYNKLTETCFDLCVKDFTTREVKEPE
sp|Q9WV97|TIM9_RAT       -----MAAQIPESDQIKQFKEFLGTYNKLTETCFDLCVKDFTTREVKEPE
                                : . : : : . ** * : :

sp|Q559H1|TIM9_DICDI      EQLCVYKCKVEKNMFFTSALNNHFMKLSNEGMP-----
sp|Q3SZW4|T10B_BOVIN     EEACLHSCAGKLIHSNHRLMAYVQLPALVQRRMADYEAAASAVPHATAEQLETSFSPRSL
sp|Q9XGX9|TIM9_ARATH     EETCMVRCAEKFLKHTMRVGMRFPAELNQNAFTQD-----
sp|Q9XGX7|TIM9_ORYSJ     EESCVRRCAEKFLKHSRMRVGMRFPAELNQGVATPD-----
sp|Q8J1Z1|TIM9_NEUCR     ESGCISRCVLKSMSTQTRLGERFGLNAAMTAEM-----Q-----R-----
sp|Q9Y8A7|TIM9_EMENI     EEGCVMRQVDFKPMKGSQRLNERFQEQNAAMMQSG-----Q-----L-----
sp|Q9P7K0|TIM9_SCHPO     ESECIACADKFLKHSERVQGRFAEFNAKYMGO-----
sp|Q6BU42|TIM9_DEBHA     ETGCITKCKSEKFLKHSERVQGRFQEQNALLMQNM-----Q-----K-----
sp|Q74700|TIM9_YEAST     EQTCIMKCKSEKFLKHSERVQGRFQEQNAALGQGL-----G-----R-----
sp|Q6FRT3|TIM9_CANGA     EESCILKCKSEKFLKHSERVQGRFQEQNAALGQGL-----G-----R-----
sp|Q17754|TIM9_CAEEL     EESCANNCLDKFLKMTQRVSRFQEQHQLLNAQANGAAIKVE-----N-----
sp|Q61TH2|TIM9_CAEBR     EESCANNCLDKFLKMTQRVSRFQEQHQLLNAQANGAAM-----
sp|Q9VYD7|TIM9_DROME     EEKCSLNCMEKYLKMNQRVSRFQEQFQVIAHENALAM-----AQ-----K-----
sp|Q9W762|TIM9_DANRE     ETTCSSECLQKYLKMTQRIISMRFQEQYHIQQNERWP-----Q-----K-----
sp|Q5ZIR8|TIM9_CHICK     EITCSEHCLQKYLKMTQRIISMRFQEQYHIQQNEALA-----A-----K-----
sp|Q4V7R1|TIM9_XENLA     EMTCSHCLQKYLKMTQRIISMRFQEQYHIQQNEALA-----A-----K-----
sp|Q9WV98|TIM9_MOUSE     EVTCSHCLQKYLKMTQRIISVRFQEQYHIQQNEALA-----A-----K-----
sp|Q9Y5J7|TIM9_HUMAN     ETTCSHCLQKYLKMTQRIISMRFQEQYHIQQNEALA-----A-----K-----
sp|Q2KIV2|TIM9_BOVIN     ETTCSHCLQKYLKMTQRIISMRFQEQYHIQQNEALA-----A-----K-----
sp|Q9WV97|TIM9_RAT       EVTCSHCLQKYLKMTQRIISMRFQEQYHIQQNEALA-----A-----K-----
                                * * * * : : :

sp|Q559H1|TIM9_DICDI      -----
sp|Q3SZW4|T10B_BOVIN     PSGNLGKGGAG
sp|Q9XGX9|TIM9_ARATH     -----
sp|Q9XGX7|TIM9_ORYSJ     -----
sp|Q8J1Z1|TIM9_NEUCR     -R-----
sp|Q9Y8A7|TIM9_EMENI     -PGR-----
sp|Q9P7K0|TIM9_SCHPO     -----
sp|Q6BU42|TIM9_DEBHA     -R-----
sp|Q74700|TIM9_YEAST     -----
sp|Q6FRT3|TIM9_CANGA     -----
sp|Q17754|TIM9_CAEEL     -GGKINKIQ--
sp|Q61TH2|TIM9_CAEBR     -----
sp|Q9VYD7|TIM9_DROME     -TGKL-----
sp|Q9W762|TIM9_DANRE     -PDY-----
sp|Q5ZIR8|TIM9_CHICK     -AGLLSQPR--
sp|Q4V7R1|TIM9_XENLA     -AGLLGQPR--
sp|Q9WV98|TIM9_MOUSE     -AGLLGQPR--
sp|Q9Y5J7|TIM9_HUMAN     -AGLLGQPR--
sp|Q2KIV2|TIM9_BOVIN     -AGLLGQPR--
sp|Q9WV97|TIM9_RAT       -AGLLGQPR--

```

## MAFFT



```

sp|Q9Y5J7|TIM9_  MAA-----QIPESDQIKQKPEFLGTYNKLTETCFDLCVKDFTTREVKPEE
sp|Q2KIV2|TIM9_  MAA-----QIPESDQIKQKPEFLGTYNKLTETCFDLCVKDFTTREVKPEE
sp|Q9WV98|TIM9_  MAA-----QIPESDQIKQKPEFLGTYNKLTETCFDLCVKDFTTREVKPEE
sp|Q9WV97|TIM9_  MAA-----QIPESDQIKQKPEFLGTYNKLTETCFDLCVKDFTTREVKPEE
sp|Q4V7R1|TIM9_  MAA-----QMSESDQIKQKPEFLGTYNKLTENCFLDCVKDFTTREVKAEE
sp|Q5ZIR8|TIM9_  MAG-----QISETDQIKQKPEFLGTYNKITENCFMDCIRDFTTREVKPEE
sp|Q9W762|TIM9_  MAA-----QVTESDQIKQKPEFLGTYNKLTENCFMDCVKDFTTREVKPEE
sp|Q9VYD7|TIM9_  MAK-----TPENIAIDQLDKDQIKTFSDFLMSYNKLSETCFDTCIRDFTTRDVKDSE
sp|Q17754|TIM9_  MTS-----EQNIQTFRDFLTQYNLVAEQCFNSCVNEFGSRTVSGKE
sp|Q61TH2|TIM9_  MAS-----EQNIQTFRDFLTQYNLVAEQCFNSCVNEFGSRTVNAKE
sp|Q74700|TIM9_  MDA-----LNSKEQQEFQKVVEQKQMKDFMRLYSNLVERCFDTCVNDFTTSKLTNKE
sp|Q6FRT3|TIM9_  MDQ-----LNAKEQQEFQKLVEQKQMKDFMRLYSGLVERCFDTCVNDFTSSKLTNKE
sp|Q6BU42|TIM9_  MDQ-----LNVKEQQDFQQIVEQKQMKDFMRLYSNLVSKCFDDCVNDFTSNNLTKE
sp|Q9P7K0|TIM9_  MDR-----LNVKEQEHLTQVLEAKQLKEYLNMYSLTQNCFSDCVQDFTSSKLSNKE
sp|Q9Y8A7|TIM9_  MDG-----LNAAEQRELANRMERKQMKEFMTMYSKLVQRCFDDCVNDFTTKSLISRE
sp|Q8J1Z1|TIM9_  MDG-----LTAAESRELDQRLQKRQVKEFMSVFGNLVDNCFACVDDFTSKALSGRE
sp|Q9XGX9|TIM9_  MDASMMAGLDGLPEEDKAKMASMIDQLQLRDSLRMYNSLVERCFVDCVDSFTRKSLQKQE
sp|Q9XGX7|TIM9_  MDKSMGLDGLDGLPEEDKMRMAAMVDQLQIRDSLRMYNSLVERCFDTCVDTFRRKTLQKQE
sp|Q3SZW4|T10B_  MEQ-----QQQQQQQQQQLRNLRDPLLVDYRMTELCPQRCVPSLHHRALDAEE
sp|Q559H1|TIM9_  MDRR-----LSKKEERIVNELNKLQMIEMVDTSVNLTKCFQSCITNFRIRKLDDEE

```

\* : . : : : . \*\* \* : : \*

```

sp|Q9Y5J7|TIM9_  TTCSEHCLOKYLKMTQR-----ISMRFQYHIQQ--NEALAACA-----
sp|Q2KIV2|TIM9_  TTCSEHCLOKYLKMTQR-----ISMRFQYHIQQ--NEALAACA-----
sp|Q9WV98|TIM9_  VTCSEHCLOKYLKMTQR-----ISVRFQYHIQQ--NEALAACA-----
sp|Q9WV97|TIM9_  VTCSEHCLOKYLKMTQR-----ISMRFQYHIQQ--NEALAACA-----
sp|Q4V7R1|TIM9_  MTCSEHCLOKYLKMTQR-----ISMRFQYHIQQ--NEALAACA-----
sp|Q5ZIR8|TIM9_  ITCSEHCLOKYLKMTQR-----ISMRFQYHIQQ--NEALAACA-----
sp|Q9W762|TIM9_  TTCSESCLOKYLKMTQR-----ISMRFQYHIQQ--NERWPQKP-----
sp|Q9VYD7|TIM9_  EKCSLNCMEKYLKMNQR-----VSQRFQEFQVIAHENALAMAQKT-----
sp|Q17754|TIM9_  ESCANNCLDKFLKMTQR-----VSQRFQEHQLLNAQ--ANGAAIKVEN-----
sp|Q61TH2|TIM9_  ESCANNCLDKFLKMTQR-----VSQRFQEHQILNAQ--ANGAAM-----
sp|Q74700|TIM9_  QTCIMKCSEKFLKHSER-----VGQRFQEQN-----A-ALGQGL-----
sp|Q6FRT3|TIM9_  ESCILKCSEKFLKHSER-----VGQRFQEQN-----A-ALGQGL-----
sp|Q6BU42|TIM9_  TGCITKCSEKFLKHSER-----VGQRFQEQN-----A-LLMQNM-----
sp|Q9P7K0|TIM9_  SECIAKCADKFLKHSER-----VGQRFQEFN-----AKYMGQ-----
sp|Q9Y8A7|TIM9_  EGCVMRCVDKFMKGSQR-----LNERFQEQN-----A-AMMQSG-----
sp|Q8J1Z1|TIM9_  SGCISRCVLKSMSTQTR-----LGERFQELN-----A-AMTAEM-----
sp|Q9XGX9|TIM9_  ETCVMRCAEKFLKHTMR-----VGMRFAELN-----QNAPTQD-----
sp|Q9XGX7|TIM9_  ESCVRRCAEKFLKHSMR-----VGMRFAELN-----QGVATPD-----
sp|Q3SZW4|T10B_  EACLHSCAGKLIHSNHRMAAYVQLMPALVQRRMADYEAASAV--PHATAEQLETSPSRSL
sp|Q559H1|TIM9_  QLCVYKCVKKNMFF TSA-----LNNHFMKLS-----NEGMF-----

```

\* \* \* : : : .

```

sp|Q9Y5J7|TIM9_  --GLLGQPR--
sp|Q2KIV2|TIM9_  --GLLGQPR--
sp|Q9WV98|TIM9_  --GLLGQPR--
sp|Q9WV97|TIM9_  --GLLGQPR--
sp|Q4V7R1|TIM9_  --GLLGQPR--
sp|Q5ZIR8|TIM9_  --GLLSQPR--
sp|Q9W762|TIM9_  --DY-----
sp|Q9VYD7|TIM9_  --GKL-----
sp|Q17754|TIM9_  --GGKINKIQ--
sp|Q61TH2|TIM9_  -----
sp|Q74700|TIM9_  --GR-----
sp|Q6FRT3|TIM9_  --GR-----
sp|Q6BU42|TIM9_  --QKR-----
sp|Q9P7K0|TIM9_  -----
sp|Q9Y8A7|TIM9_  --QLPGR-----
sp|Q8J1Z1|TIM9_  --QRR-----
sp|Q9XGX9|TIM9_  -----
sp|Q9XGX7|TIM9_  -----
sp|Q3SZW4|T10B_  PSGNLGKGGAG
sp|Q559H1|TIM9_  -----

```

## MUSCLE

```

sp|Q559H1|TIM9_DICDI      -----MDRRLSKKEEERIVNELNKLQMIEMVDTSVNLTKCFQSCITNFRIRKLDDEE
sp|Q3SZW4|T10B_BOVIN     -----MEQQQQQQQQQQQ-----LRNLRDFLLVYNRMTLFCFQRCVPSLHHRALDAEE
sp|Q9XGX9|TIM9_ARATH     MDASMMAGLDGLPEEDKAKMASMIDQLQLRDSLRMYNSLVERCFVDCVDSFTRKSLQKQE
sp|Q9XGX7|TIM9_ORYSJ     MDKSMLGDLGLPEEDKMRMAAMVDQLQIRDSLRMYNSLVERCFVDCVDSFTRKSLQKQE
sp|Q8J1Z1|TIM9_NEUCR     -----MDGLTAAESRELDQRLQKRQVKEFMSVFGNLVNCFTACVDDFTSKALSGRE
sp|Q9Y8A7|TIM9_EMENI     -----MDGLNAAEQRELANRMERKQMKFMTMYSKLVQRCFDDCVNDFTTKSLISRE
sp|Q9P7K0|TIM9_SCHPO     -----MDRLNVKEQEHLTQVLEAKQLKEYLNMYSTLTQNCFSDCVQDFTSSKLSNKE
sp|Q6BU42|TIM9_DEBHA     -----MDQLNVKEQQDFQQIVEQKQMKDFMRLYSNLVSKCFDDCVNDFTSNNLTKE
sp|O74700|TIM9_YEAST     -----MDALNSKEQQEFQKVVEQKQMKDFMRLYSNLVERCFVDCVNDFTTSKLTNKE
sp|Q6FRT3|TIM9_CANGA     -----MDQLNAKEQQEFQKLVEQKQMKDFMRLYSGLVERCFVDCVNDFTSSKLTKE
sp|Q17754|TIM9_CAEEL     -----MTSEQN-----IQTFRDFLTQYNLVAEQCFNSCVNEFGSRTVSGKE
sp|Q61TH2|TIM9_CAEBR     -----MASEQN-----IQTFRDFLTQYNLVAEQCFNSCVNEFGSRTVNAKE
sp|Q9VYD7|TIM9_DROME     -----MAKTPENIAIDQLDKDQIKTFSDFLMSYNKLSSETCFVDCVNDFTTRDVKDSE
sp|Q9W762|TIM9_DANRE     -----MAAQVTESDQ-----IKQFKEFLGTYNKLTENCFCMDCKVDFTTREVKPEE
sp|Q5ZIR8|TIM9_CHICK     -----MAGQISETDQ-----IKQFKEFLGTYNKLTENCFCMDCKVDFTTREVKPEE
sp|Q4V7R1|TIM9_XENLA     -----MAAQMESDQ-----IKQFKEFLGTYNKLTENCFLDCVKDFTTREVKAEE
sp|Q9WV98|TIM9_MOUSE     -----MAAQIPESDQ-----IKQFKEFLGTYNKLTETCFLDCVKDFTTREVKPEE
sp|Q9Y5J7|TIM9_HUMAN     -----MAAQIPESDQ-----IKQFKEFLGTYNKLTETCFLDCVKDFTTREVKPEE
sp|Q2KIV2|TIM9_BOVIN     -----MAAQIPESDQ-----IKQFKEFLGTYNKLTETCFLDCVKDFTTREVKPEE
sp|Q9WV97|TIM9_RAT       -----MAAQIPESDQ-----IKQFKEFLGTYNKLTETCFLDCVKDFTTREVKPEE
                                . : : : . ** * : : *

sp|Q559H1|TIM9_DICDI      QLCVYKCVKEKNMFFTSA-----LNNHFMKLS--NEG--MF-----
sp|Q3SZW4|T10B_BOVIN     EACLHSCAGKLIHSHNRLMAAYVQLMPALVQRRMADYEAAAVPHATAEQLETSFRRSLP
sp|Q9XGX9|TIM9_ARATH     ETCVMRCAEKFLKHTMR-----VGMRFaelnQNAPTQD-----
sp|Q9XGX7|TIM9_ORYSJ     ESCVRRCAEKFLKHSR-----VGMRFaelnQGV-----AT-----P
sp|Q8J1Z1|TIM9_NEUCR     SGCISRCVLKSMSTQTR-----LGERFQELNAAMTAE--MQ-----R
sp|Q9Y8A7|TIM9_EMENI     EGCVMRCVDKFMKGSQR-----LNERFQEQNAAMQSGQLPG-----
sp|Q9P7K0|TIM9_SCHPO     SECIAKCADKFLKHSE-----VGQRFQEFNAKYMGO-----
sp|Q6BU42|TIM9_DEBHA     TGCITKCEKFLKHSE-----VGQRFQEQNALLMQN--MQ-----K
sp|O74700|TIM9_YEAST     QTCIMKCEKFLKHSE-----VGQRFQEQNAALGQG--LG-----
sp|Q6FRT3|TIM9_CANGA     ESCILKCEKFLKHSE-----VGQRFQEQNAALGQG--LG-----
sp|Q17754|TIM9_CAEEL     ESCANNCLDKFLKMTQR-----VSQRFQEHQLLNAANGAAIKVENGKINKI
sp|Q61TH2|TIM9_CAEBR     ESCANNCLDKFLKMTQR-----VSQRFQEHQILNAANGAAM-----
sp|Q9VYD7|TIM9_DROME     EKCSLNCMEKYLKMNQR-----VSQRFQEFQVIAHENALAMA--QKTGKL--
sp|Q9W762|TIM9_DANRE     TTCSESLQKYLKMTQR-----ISMRFQEYHIQQNER-----WPQKP
sp|Q5ZIR8|TIM9_CHICK     ITCSEHCLQKYLKMTQR-----ISMRFQEYHIQQNEA--LA--AKAGLLSQP
sp|Q4V7R1|TIM9_XENLA     MTCSEHCLQKYLKMTQR-----ISMRFQEYHIQQNEA--LA--AKAGLLGQP
sp|Q9WV98|TIM9_MOUSE     VTCSEHCLQKYLKMTQR-----ISVRFQEYHIQQNEA--LA--AKAGLLGQP
sp|Q9Y5J7|TIM9_HUMAN     TTCSEHCLQKYLKMTQR-----ISMRFQEYHIQQNEA--LA--AKAGLLGQP
sp|Q2KIV2|TIM9_BOVIN     TTCSEHCLQKYLKMTQR-----ISMRFQEYHIQQNEA--LA--AKAGLLGQP
sp|Q9WV97|TIM9_RAT       VTCSEHCLQKYLKMTQR-----ISMRFQEYHIQQNEA--LA--AKAGLLGQP
                                * * * : : : .

```

```

sp|Q559H1|TIM9_DICDI      -----
sp|Q3SZW4|T10B_BOVIN     SGNLGKGGAG
sp|Q9XGX9|TIM9_ARATH     -----
sp|Q9XGX7|TIM9_ORYSJ     D-----
sp|Q8J1Z1|TIM9_NEUCR     R-----
sp|Q9Y8A7|TIM9_EMENI     R-----
sp|Q9P7K0|TIM9_SCHPO     -----
sp|Q6BU42|TIM9_DEBHA     R-----
sp|O74700|TIM9_YEAST     R-----
sp|Q6FRT3|TIM9_CANGA     R-----
sp|Q17754|TIM9_CAEEL     Q-----
sp|Q61TH2|TIM9_CAEBR     -----
sp|Q9VYD7|TIM9_DROME     -----
sp|Q9W762|TIM9_DANRE     DY-----
sp|Q5ZIR8|TIM9_CHICK     R-----
sp|Q4V7R1|TIM9_XENLA     R-----
sp|Q9WV98|TIM9_MOUSE     R-----
sp|Q9Y5J7|TIM9_HUMAN     R-----
sp|Q2KIV2|TIM9_BOVIN     R-----
sp|Q9WV97|TIM9_RAT       R-----

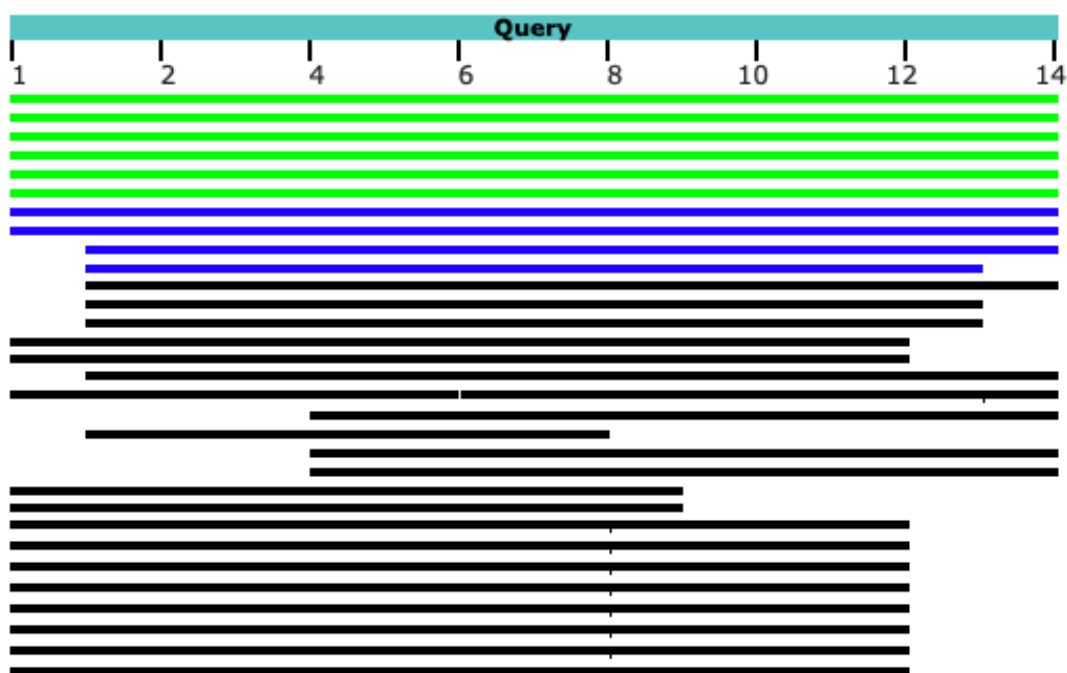
```

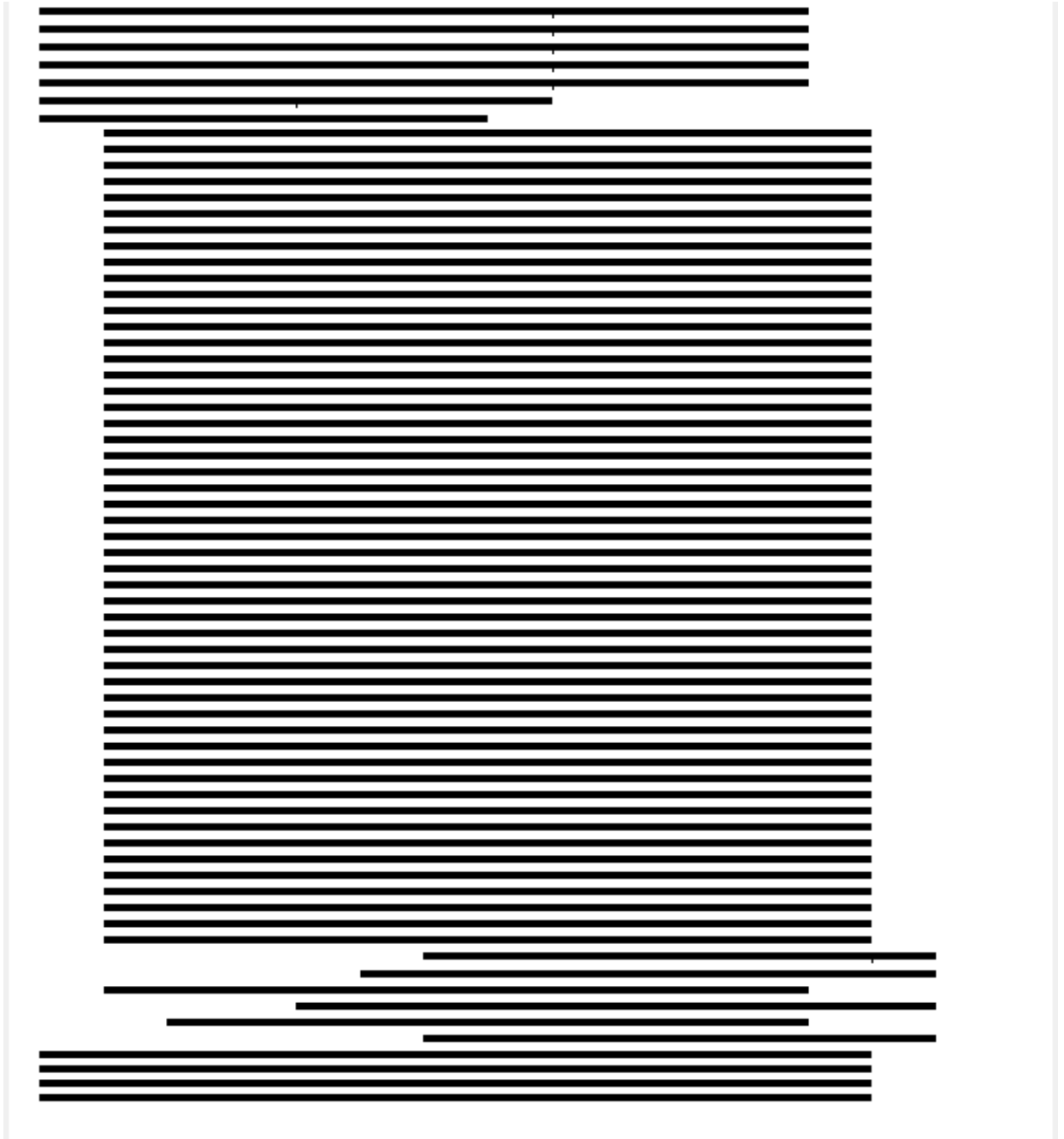
Positions 8, 9, 10, 11, 12 are aligned differently

# Question 8

|   | Description   | Scientific Name               | Max Score | Total Score | Query Cover | E value | Per. Ident | Acc. Len | Accession               |
|---|---|-------------------------------|-----------|-------------|-------------|---------|------------|----------|-------------------------|
| ✓ | Crystal structure of 1-deoxy-D-xylulose 5-phosphate reductoisomerase complexed with a magnesium ion, NADPH and...           | <i>Escherichia coli</i> st... | 53.2      | 53.2        | 100%        | 2e-10   | 100.00%    | 424      | <a href="#">2EGH_A</a>  |
| ✓ | Crystal structure of 1-deoxy-D-xylulose 5-phosphate reductoisomerase (DXR) complexed with pyridin-2-ylmethylphosp...        | <i>Escherichia coli</i> K...  | 53.2      | 53.2        | 100%        | 2e-10   | 100.00%    | 420      | <a href="#">3ANL_A</a>  |
| ✓ | IspC in complex with an N-methyl-substituted hydroxamic acid [Escherichia coli K-12]  | <i>Escherichia coli</i> K...  | 53.2      | 53.2        | 100%        | 2e-10   | 100.00%    | 410      | <a href="#">3R0I_A</a>  |
| ✓ | Crystal structure of DXR in complex with fosmidomycin [Escherichia coli]  | <i>Escherichia coli</i>       | 53.2      | 53.2        | 100%        | 2e-10   | 100.00%    | 406      | <a href="#">1Q0L_A</a>  |
| ✓ | 1-deoxy-D-xylulose 5-phosphate reductoisomerase [Escherichia coli K-12]   | <i>Escherichia coli</i> K...  | 53.2      | 53.2        | 100%        | 2e-10   | 100.00%    | 398      | <a href="#">1K5H_A</a>  |
| ✓ | Crystal Structure of the Reductoisomerase Complexed with a Bisphosphonate [Escherichia coli]                                | <i>Escherichia coli</i>       | 53.2      | 53.2        | 100%        | 2e-10   | 100.00%    | 398      | <a href="#">1T1R_A</a>  |
| ✓ | Crystal structure of selenomethionine-labelled DXR in complex with fosmidomycin [Escherichia coli]                          | <i>Escherichia coli</i>       | 43.1      | 43.1        | 100%        | 7e-07   | 85.71%     | 406      | <a href="#">1Q0H_A</a>  |
| ✓ | Crystal structure of 1-deoxy-D-xylulose 5-phosphate reductoisomerase, a target enzyme for antimalarial drugs [Escheri...    | <i>Escherichia coli</i>       | 43.1      | 43.1        | 100%        | 7e-07   | 85.71%     | 400      | <a href="#">1JVS_A</a>  |
| ✓ | Chain A, 1-deoxy-D-xylulose 5-phosphate reductoisomerase [Zymomonas mobilis]  | <i>Zymomonas mobilis</i>      | 42.2      | 42.2        | 92%         | 1e-06   | 84.62%     | 388      | <a href="#">1R0K_A</a>  |
| ✓ | 1-deoxy-D-xylulose 5-phosphate reductoisomerase from Yersinia pestis [Yersinia pseudotuberculosis YPIII]                    | <i>Yersinia pseudotu...</i>   | 40.9      | 40.9        | 85%         | 4e-06   | 91.67%     | 401      | <a href="#">3IIE_A</a>  |
| ✓ | Chain A, 1-deoxy-D-xylulose 5-phosphate reductoisomerase [Acinetobacter baumannii AB307-0294]                               | <i>Acinetobacter bau...</i>   | 35.4      | 35.4        | 92%         | 3e-04   | 76.92%     | 406      | <a href="#">4ZN6_A</a>  |
| ✓ | Chain A, 1-deoxy-D-xylulose 5-phosphate reductoisomerase [Moraxella catarrhalis]  | <i>Moraxella catarrh...</i>   | 33.3      | 33.3        | 85%         | 0.002   | 75.00%     | 432      | <a href="#">4ZQE_A</a>  |
| ✓ | Chain A, 1-deoxy-D-xylulose 5-phosphate reductoisomerase [Moraxella catarrhalis]  | <i>Moraxella catarrh...</i>   | 33.3      | 33.3        | 85%         | 0.002   | 75.00%     | 415      | <a href="#">4ZQG_A</a>  |
| ✓ | Chain A, 1-deoxy-D-xylulose 5-phosphate reductoisomerase [Vibrio vulnificus CMCP6]  | <i>Vibrio vulnificus C...</i> | 28.2      | 28.2        | 85%         | 0.13    | 75.00%     | 427      | <a href="#">5KQO_A</a>  |
| ✓ | 1-deoxy-D-xylulose 5-phosphate reductoisomerase from Vibrio vulnificus in complex with Mn(2+) [Vibrio vulnificus CM...      | <i>Vibrio vulnificus C...</i> | 28.2      | 28.2        | 85%         | 0.13    | 75.00%     | 405      | <a href="#">5KRR_A</a>  |
| ✓ | Arabidopsis thaliana glutamate dehydrogenase isoform 1 in apo form [Arabidopsis thaliana]                                   | <i>Arabidopsis thaliana</i>   | 24.4      | 37.3        | 92%         | 2.9     | 64.29%     | 414      | <a href="#">6YEH_A</a>  |
| ✓ | Chain Bh, 30S ribosomal protein S9, mitochondrial [Arabidopsis thaliana]  | <i>Arabidopsis thaliana</i>   | 24.4      | 24.4        | 42%         | 2.9     | 100.00%    | 430      | <a href="#">6XYW_Bh</a> |
| ✓ | Human alpha3beta4 nicotinic acetylcholine receptor in complex with nicotine [Homo sapiens]                                  | <i>Homo sapiens</i>           | 24.4      | 24.4        | 71%         | 2.9     | 63.64%     | 525      | <a href="#">6PV7_A</a>  |
| ✓ | Crystal structure of DXR from Thermooga maritima, in complex with fosmidomycin and NADPH [Thermotoga maritima]              | <i>Thermotoga marit...</i>    | 24.0      | 24.0        | 50%         | 4.2     | 85.71%     | 376      | <a href="#">3A06_A</a>  |
| ✓ | CryoEM structure of human alpha4beta2 nicotinic acetylcholine receptor in complex with varenicline [Homo sapiens]           | <i>Homo sapiens</i>           | 24.0      | 24.0        | 71%         | 4.2     | 63.64%     | 397      | <a href="#">6UR8_B</a>  |
| ✓ | X-ray structure of the human Alpha4Beta2 nicotinic receptor [Homo sapiens]  | <i>Homo sapiens</i>           | 24.0      | 24.0        | 71%         | 4.2     | 63.64%     | 403      | <a href="#">5KXI_B</a>  |
| ✓ | Chain A, Tricorn protease-interacting factor F3 [Thermoplasma acidophilum DSM 1728]   | <i>Thermoplasma ac...</i>     | 24.0      | 24.0        | 64%         | 4.2     | 66.67%     | 780      | <a href="#">3QTJ_A</a>  |
| ✓ | Chain A, Tricorn protease interacting factor F3 [Thermoplasma acidophilum]  | <i>Thermoplasma ac...</i>     | 24.0      | 24.0        | 64%         | 4.2     | 66.67%     | 780      | <a href="#">1Z1W_A</a>  |
| ✓ | Chain A, Alpha-ketoglutarate-dependent dioxygenase alkB [Escherichia coli K-12]   | <i>Escherichia coli K...</i>  | 23.5      | 35.2        | 85%         | 5.9     | 75.00%     | 201      | <a href="#">3BI3_A</a>  |
| ✓ | Chain A, Alpha-ketoglutarate-dependent dioxygenase alkB [Escherichia coli K-12]   | <i>Escherichia coli K...</i>  | 23.5      | 35.2        | 85%         | 5.9     | 75.00%     | 201      | <a href="#">3BKZ_A</a>  |
| ✓ | Chain A, Alpha-ketoglutarate-dependent dioxygenase alkB [Escherichia coli K-12]   | <i>Escherichia coli K...</i>  | 23.5      | 35.2        | 85%         | 5.9     | 75.00%     | 202      | <a href="#">3BIE_A</a>  |
| ✓ | Complex structure of AlkB/rhein [Escherichia coli]  | <i>Escherichia coli</i>       | 23.5      | 35.2        | 85%         | 5.9     | 75.00%     | 203      | <a href="#">4RFR_B</a>  |
| ✓ | Crystal structure of AlkB protein with cofactors bound to dsDNA containing m6A [Escherichia coli]                           | <i>Escherichia coli</i>       | 23.5      | 35.2        | 85%         | 5.9     | 75.00%     | 205      | <a href="#">4NID_A</a>  |
| ✓ | Crystal structure of AlkB D135I mutant protein with cofactors bound to dsDNA containing m6A/A [Escherichia coli]            | <i>Escherichia coli</i>       | 23.5      | 35.2        | 85%         | 5.9     | 75.00%     | 205      | <a href="#">4NII_A</a>  |
| ✓ | Iron-Catalyzed Oxidation Intermediates Captured in A DNA Repair Dioxygenase [Escherichia coli K-12]                         | <i>Escherichia coli K...</i>  | 23.5      | 35.2        | 85%         | 5.9     | 75.00%     | 206      | <a href="#">3O1M_A</a>  |
| ✓ | Crystal structure of AlkB in complex with Fe(III) and 2-(3-hydroxypicolinoimido)acetic acid [Escherichia coli K-12]         | <i>Escherichia coli K...</i>  | 23.5      | 35.2        | 85%         | 5.9     | 75.00%     | 206      | <a href="#">3T3Y_A</a>  |
| ✓ | Crystal Structure of AlkB in complex with Fe(II), 2-oxoglutarate, and methylated trinucleotide T-meA-T [Escherichia coli... | <i>Escherichia coli K...</i>  | 23.5      | 35.2        | 85%         | 5.9     | 75.00%     | 211      | <a href="#">2FD8_A</a>  |
| ✓ | Crystal Structure of AlkB T208A mutant protein in complex with Co(II), 2-oxoglutarate, and methylated trinucleotide T-...   | <i>Escherichia coli K...</i>  | 23.5      | 35.2        | 85%         | 5.9     | 75.00%     | 211      | <a href="#">4ZHN_A</a>  |

## Distribution of the top 117 Blast Hits on 100 subject sequences





|  |               |
|--|---------------|
| <b>Max Score</b>   | <b>53.2</b>   |
| <b>Min Score</b>   | <b>21.4</b>   |
| <b>Matches with &gt;90% Query coverage</b>                       | <b>15</b>     |
| <b>Matches with E value in the order of <math>10^{-3}</math></b> | <b>13</b>     |
| <b>Matches with &gt;90% identity</b>                             | <b>8</b>      |
| <b>Min E value</b>   | <b>34</b>     |
| <b>Min Identity</b>  | <b>41.67%</b> |

Since the query sequence is of a short length, there is a high chance that this particular sequence can be found by pure chance, that is why the E values are quite high.

Residues 2-14 are evolutionarily conserved (PRK05447) in E. coli and other organisms.