

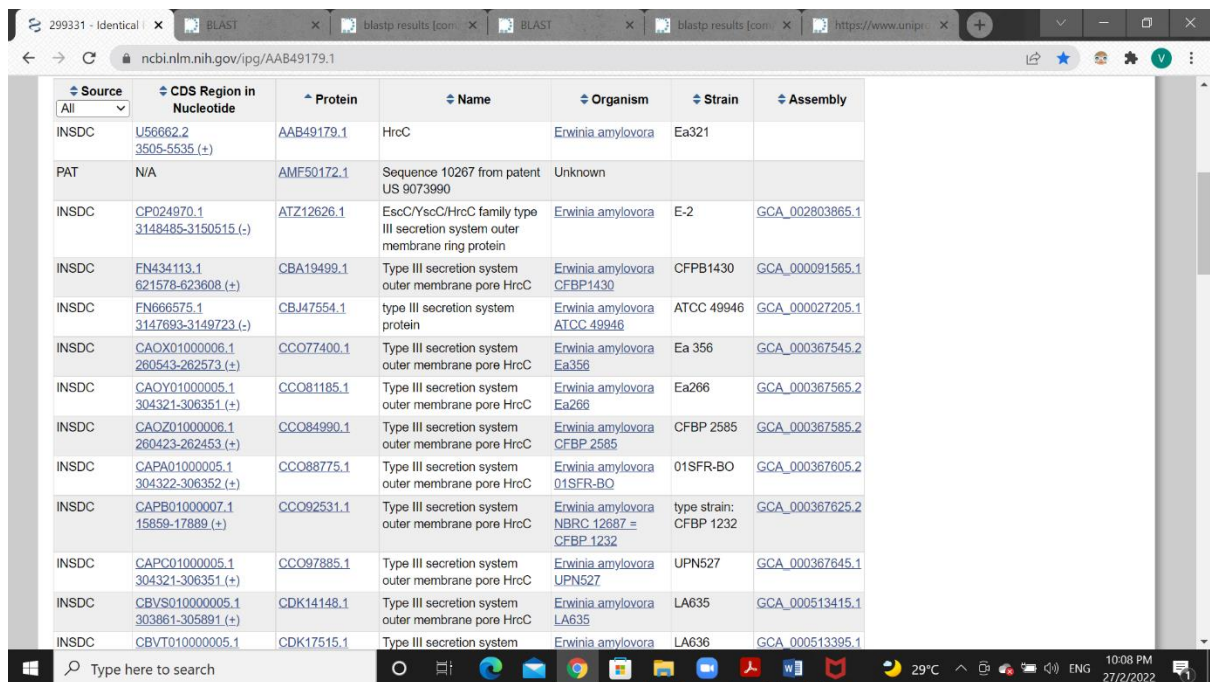
# Practicals – 5

-BS19B032

-R. Vasantha Kumar

1) For the given sequence, the similar proteins were searched:

nr:



Source	CDS Region in Nucleotide	Protein	Name	Organism	Strain	Assembly
INSDC	U56662.2 3505-5535 (+)	AAB49179.1	HrcC	Erwinia amylovora	Ea321	
PAT	N/A	AME50172.1	Sequence 10267 from patent US 9073990	Unknown		
INSDC	CP024970.1 3148485-3150515 (-)	ATZ12626.1	EscC/YscC/HrcC family type III secretion system outer membrane ring protein	Erwinia amylovora	E-2	GCA_002803865.1
INSDC	FN434113.1 621578-623608 (+)	CBA19499.1	Type III secretion system outer membrane pore HrcC	Erwinia amylovora CFBP1430	CFBP1430	GCA_000091565.1
INSDC	FN666575.1 3147693-3149723 (-)	CBJ47554.1	type III secretion system protein	Erwinia amylovora ATCC 49946	ATCC 49946	GCA_000027205.1
INSDC	CAOX01000006.1 260543-262573 (+)	CCO77400.1	Type III secretion system outer membrane pore HrcC	Erwinia amylovora Ea356	Ea 356	GCA_000367545.2
INSDC	CAOY01000005.1 304321-306351 (+)	CCO81185.1	Type III secretion system outer membrane pore HrcC	Erwinia amylovora Ea266	Ea266	GCA_000367565.2
INSDC	CAOZ01000006.1 260423-262453 (+)	CCO84990.1	Type III secretion system outer membrane pore HrcC	Erwinia amylovora CFBP 2585	CFBP 2585	GCA_000367585.2
INSDC	CAPA01000005.1 304322-306352 (+)	CCO88775.1	Type III secretion system outer membrane pore HrcC	Erwinia amylovora 01SFR-BO	01SFR-BO	GCA_000367605.2
INSDC	CAPB01000007.1 15859-17889 (+)	CCO92531.1	Type III secretion system outer membrane pore HrcC	Erwinia amylovora NBRC 12687 = CFBP 1232	type strain: CFBP 1232	GCA_000367625.2
INSDC	CAPC01000005.1 304321-306351 (+)	CCO97885.1	Type III secretion system outer membrane pore HrcC	Erwinia amylovora UPN527	UPN527	GCA_000367645.1
INSDC	CBVSO10000005.1 303861-305891 (+)	CDK14148.1	Type III secretion system outer membrane pore HrcC	Erwinia amylovora LA635	LA635	GCA_000513415.1
INSDC	CBVT010000005.1	CDK17515.1	Type III secretion system	Erwinia amylovora LA636	LA636	GCA_000513395.1

SWISS-PROT:

Entry	Alignment overview	Info	Status
Query: B20220227A084FC58F6BBA219896F365D15F2EB4402FFD7W			
B2VGC0	B2VGC0_ERWT9 - Type 3 secretion system secretin - <i>Erwinia tasmanie...</i> - View alignment	E-value: 0.0 Score: 3,199 Ident.: 93.0%	
A0A014N8S7	A0A014N8S7_9GAMM - Type 3 secretion system secretin <i>Erwinia mallotivora</i> - View alignment	E-value: 0.0 Score: 2,931 Ident.: 84.2%	
A0A0M2KFF6	A0A0M2KFF6_9GAMM - Type 3 secretion system secretin <i>Erwinia tracheiphila</i> - View alignment	E-value: 0.0 Score: 2,933 Ident.: 83.7%	
A0A5R9LED5	A0A5R9LED5_9ENTR - Type 3 secretion system secretin <i>Klebsiella indica</i> - View alignment	E-value: 0.0 Score: 2,876 Ident.: 83.0%	
H3RK02	H3RK02_PANSE - Type III secretion system outermemb... - <i>Pantoea stewarti...</i> - View alignment	E-value: 0.0 Score: 2,648 Ident.: 82.8%	
I1SBB8	I1SBB8_PECAS - Type 3 secretion system secretin - <i>Pectobacterium a...</i> - View alignment	E-value: 0.0 Score: 2,292 Ident.: 68.6%	

In nr, a total of 200 similar proteins were found. Whereas, in SWISS-PROT blast 250 sequences were found, it is because the maximum target sequences is 250. But, most of the sequences are with low sequence identity.

2) The algorithm parameters used for searching similar proteins in question 1 are:

- Max target sequences = 250
- Expect Threshold = 0.05
  - Word size = 28
  - Match score = 1
- Mismatch score = -2
- Gap costs = Linear

3) the sequence identity of the query sequence (given in Q1) with AAK81929.1 is:

```
>AAK81929.1 RscC [Pseudomonas fluorescens]
MHNKISKHTCLHIDPPDTSSRRAKWQWLVLGCMAPAHNLLAAIPA EWKNTAYAYEADHKPLREVLEDF
AQTFGTQLQIEGLLEGDVNGKIRANTPQSMLDRLGVEHRFQWYLYNNTLFVSTLDQQESARLEVSSETIS
DLKQALTDIGLLDSRFGWGELPEDGVVLVSGPKTYIDQIKQFSSKRRSADEKQSVLSFPLKFANAADRKV
DYRGEKLVVPGVANILRGLLEPRASASTLTGMSQPDSSQPSPLTPNVPRLG NPLLQMLGANGNAGQLDTG
```

PTVTPRAPVSKSRIRVEADVRRNNAVLIYDLPERQAMYRDLITQLDVARKLIEIDAIILDIERTQLREFGV  
NWGFQNSRFRGGVNMAPGTSSQVSIHHRDRFYADMPSTGGQGPATMVSNSPSVLTLENQPAVIDFNRTQY  
SPGRDYATILPVTVTGSLQVVPRVTTGRGVHQIHLVVDIEDGNLDETNPDPNHLDVRRGKVSTQAVMQ  
EKRSLVVGGFHVTDSSDQKKIPLLGDIPLLGKTLVSSTERHNNRRERLFI LTPRVIGDQDDPSRYLPQD  
DQAEALQALTPARRYSHPQPVIKRSDIITTLARLVSGEVPKAFNAARMPLGLNTLCSTRDLLALNTERS  
QWYAGPDYNVAVVYLRNQFKRNVRIDEKECSNSQTLAVTVWPRAWLKPGEAEVFIAMRPVVKDEHLSVP  
RPSLITPTQKATP

The sequence identity is 41%.

```
Python 3.6.2 Shell
File Edit Shell Debug Options Window Help

Python 3.6.2 (v3.6.2:15fd33b5, Jul 8 2017, 04:57:36) [MSC v.1900 64 bit (AMD64)] on win32
Type "copyright", "credits" or "license()" for more information.
>>>
RESTART: C:\Users\Yvaneth\OneDrive\Desktop\Bioinformatics\Lab practicals\Practical 5\sequence identity.py
Enter your sequence 1:MHNKISKHTCLHIDPPTSSRAKQWQVLVLCGIMAPAHNLLAIPAENWNTAYAYEADHKFLREVLEDAQFTGTLQIEGLLEDVNGKIRANTPQSMDLRLGVHFRQWLYNNLTFLVSTLDQQESARLEV
SEETISLDKALTDIGLDSRFGEWGLPEDGVVLVSGPKTYIDTQIKQFSKKRRADEKQSVLSFLFKLAAADKRVYDGRKGLVPGVANLRLGLEPRASATLTGMSQPDSSQPSPLTNVFRNLGNLQMLGANGNAGOLDTGTVTTPRAPVS
KSRHIVADVRNNALVILDFPGRAGVNRDLTLTQVDAKRLIEIDAILDIERTLOLFEFGVNGQFMSRFFGGVNGQSVSDVSDHRRDRIADMFSTGGGGATPMVNSRNVLTENQPAVIDFNRTQYIYSP-GRDVTATVYTGSLGVPERVT
TGRGVQHILVLDVIEDGNLETFNFERNDHNDLGRGVQVQVHQRSLVQFVGGVHFDSSQGGKIFLLGDITFLPKGLTVLSSTHRRRRIERFILTPFRVGGDDDSRFLYFDQQAELQAALTLPAKRYSPHOPVFKRSIDITTLAR-LVYSEV
SPADAAHGLGLNTICSTLDLALLNENQSWAGTQFHW-AYVYLVNFRKVRNVRIDEEKCSNSQTALTVFWRAKWLKFGEEAEVIAHMFVYDEHLSVFRSLTIFTKQATP
your sequence 2:-NVTGRELKRLGLLALCATLFAQ---ACTPAWKEQSVASADRLTLQVDFADGSGVPLNGVNTVETAKIRAEHSAFADRLDLAHEHFWQVWVNTLYYSPFQGGESERLE
FKAADPAIDKQALISGLLDLRFPGWGLPEDGVVLVPGDYVELVFRFSGEKKEDRKVKTYFLRVASVADRTIHYRDTVVIQVATMLNLMNGKRAAPASA-SGIDSTPGGPTNMGMCNTLTLLSRSSRNKTSNR-AGGRNEDIEDVS
G---RISADVRNNALLIRDDKRHHDEYSQIAKIDVQNLVEVLDITDITRLNRLANQWATLPGVGGGSLMGGSTLVSDFK-FFADIAIQLFEGEATISAVNPSVLTLENQPAVIDFSGTAYITATGERVAADQVPATGASTVQTPFRAV
GNHSSSQILMDIEDHGVTQW---DGAQTGVRGVTSQALISENRLVLDLFGHVEESADRRRIEGLTGDITFWLQO-LFSSKRIEISQRRLQALTEPRYDITADNRQOLQSLDAMGRVRRHSS---VNQRHDEVNALRDLAEGS
PAGFQPTSGTLESEVCRSTPALLFESTRGQWYSSSTNGVGLSVGVNRTSKPLRFDEANCAASKRTLVA/VWHSRALPGESAEVYVAMPD---SRVLHASRESLLNR-----
The sequence identity of given sequences is 41.00418410041841
>>> len(seq1)
717
>>> len(seq2)
717
>>> total_identity
100
```

4) To find how far are hemoglobin sequences in human and chicken similar, I used muscle multiple alignment of sequences:

CLUSTAL multiple sequence alignment by MUSCLE (3.8)

[illegible]

From above results, we could find that these two sequences are very similar. Out of sequence length of 147, only 23 of them are not similar.

5) The program to list all the matching pentapeptides (which occur in both the sequences) and their frequency of occurrence in given sequences is:

```
seq1 = input("Enter your sequence1:")
```

```

seq2 = input("Enter your sequence2:")

print("The Matching Pentapeptides are:")

count = 0

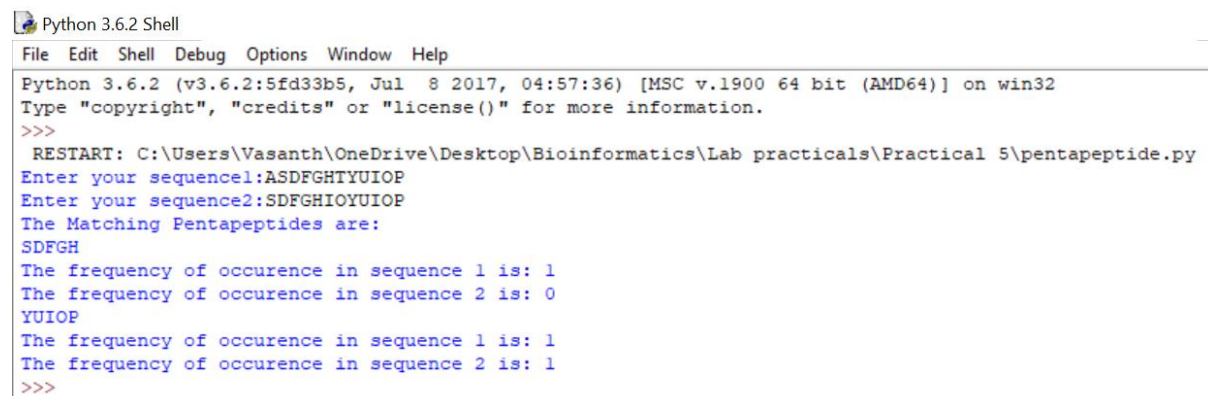
while(count<=(len(seq1)-5)):
    pen_seq1 = seq1[count:count+5]
    if pen_seq1 in seq2:
        print(pen_seq1)
        count2 = 0
        freq1 = 1
        freq2 = 0
        print("The frequency of occurrence in sequence 1 is:",freq1)
        while(count2<=(len(seq2)-5)):
            check_seq = seq2[count2:count2+5]
            if(check_seq == pen_seq1):
                freq2 = freq2 + 1
                count2 = count2 + 1
            print("The frequency of occurrence in sequence 2 is:",freq2)
        count = count + 1

```

For example, the result for these sequences is:

ASDFGHTYUIOP

SDFGHIOYUIOP



```

Python 3.6.2 Shell
File Edit Shell Debug Options Window Help
Python 3.6.2 (v3.6.2:5fd33b5, Jul 8 2017, 04:57:36) [MSC v.1900 64 bit (AMD64)] on win32
Type "copyright", "credits" or "license()" for more information.
>>>
RESTART: C:\Users\Vasanth\OneDrive\Desktop\Bioinformatics\Lab practicals\Practical 5\pentapeptide.py
Enter your sequence1:ASDFGHTYUIOP
Enter your sequence2:SDFGHIOYUIOP
The Matching Pentapeptides are:
SDFGH
The frequency of occurrence in sequence 1 is: 1
The frequency of occurrence in sequence 2 is: 0
YUIOP
The frequency of occurrence in sequence 1 is: 1
The frequency of occurrence in sequence 2 is: 1
>>>

```

6) The program to compute sequence identity, similarity, query coverage and gap percentage from the alignment of human and chicken hemoglobin sequences is:

```
human_seq='MVHLTPEEKSAVTALWGKVVNDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDVAMG  
NPKVKAHGKKVLGAFSDGLAHLNLTGKTFATLSSELHCDKLHVDPENFRLLGNVLVCVLAHHFGKEFT  
PPVQAAAYQKVVAGVANALAHKYH'
```

```
chick_seq='MVHWTAEKQLITGLWGKVNVAECGAEALARLLIVYPWTQRFFASFGNLSSPTAILGNP  
MVRAGHKKVLTSFGDAVKNLNLTGKTFATLSSELHCDKLHVDPENFRLLGDILIVLAHFSKDFTEPCQ  
AAWQKLVRVVAHALARKYH'
```

```
total_identical = 0
```

```
total_similarity = 0
```

```
gap1 = 0
```

```
gap2 = 0
```

```
for i in range(len(human_seq)):
```

```
    if(human_seq[i]=='-'):
```

```
        gap1 = gap1 + 1
```

```
    if(chick_seq[i]=='-'):
```

```
        gap2 = gap2 + 1
```

```
    if(human_seq[i]==chick_seq[i]):
```

```
        total_identical = total_identical + 1
```

```
    elif((human_seq[i]=='A') or (human_seq[i]=='I') or (human_seq[i]=='L') or (human_seq[i]=='M') or  
(human_seq[i]=='F') or (human_seq[i]=='W') or (human_seq[i]=='V')):
```

```
        if((chick_seq[i]=='A') or (chick_seq[i]=='I') or (chick_seq[i]=='L') or (chick_seq[i]=='M') or  
(chick_seq[i]=='F') or (chick_seq[i]=='W') or (chick_seq[i]=='V')):
```

```
            total_similarity = total_similarity + 1
```

```
    elif((human_seq[i]=='K') or (human_seq[i]=='R')):
```

```
        if((human_seq[i]=='K') or (human_seq[i]=='R')):
```

```
            total_similarity = total_similarity + 1
```

```
    elif((human_seq[i]=='S') or (human_seq[i]=='T')):
```

```
        if((human_seq[i]=='S') or (human_seq[i]=='T')):
```

```
            total_similarity = total_similarity + 1
```

```
    elif((human_seq[i]=='H') or (human_seq[i]=='Y')):
```

```
        if((human_seq[i]=='H') or (human_seq[i]=='Y')):
```

```
            total_similarity = total_similarity + 1
```

```

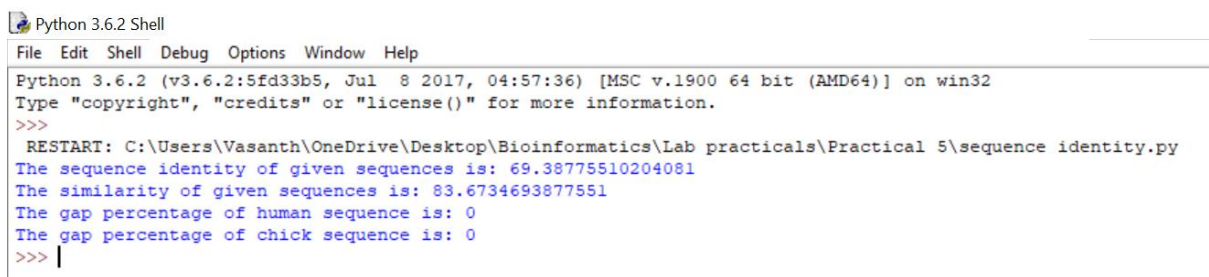
sequence_identity = (total_identical/len(human_seq))*100
print("The sequence identity of given sequences is:",sequence_identity)
similarity = ((total_identical+total_similarity)/len(human_seq))*100
print("The similarity of given sequences is:",similarity)
gap_percentage1 = (gap1/len(human_seq))*100
gap_percentage2 = (gap2/len(chick_seq))*100
print("The gap percentage of human sequence is:",gap1)
print("The gap percentage of chick sequence is:",gap2)

```

The sequence identity is 69.38%.

The similarity is 83.67%.

The gap percentage of both the sequences is 0.



The screenshot shows a Python 3.6.2 Shell window with a menu bar (File, Edit, Shell, Debug, Options, Window, Help). The command prompt displays the Python version and system information. The user has executed a script, and the output is shown in blue text. The output matches the printed statements in the code block above.

```

Python 3.6.2 Shell
File Edit Shell Debug Options Window Help
Python 3.6.2 (v3.6.2:5fd33b5, Jul 8 2017, 04:57:36) [MSC v.1900 64 bit (AMD64)] on win32
Type "copyright", "credits" or "license()" for more information.
>>>
RESTART: C:\Users\Vasanth\OneDrive\Desktop\Bioinformatics\Lab practicals\Practical 5\sequence identity.py
The sequence identity of given sequences is: 69.38775510204081
The similarity of given sequences is: 83.6734693877551
The gap percentage of human sequence is: 0
The gap percentage of chick sequence is: 0
>>> |

```

7) 20 TIM barrel proteins from different organisms are taken.

>sp|Q4V7R1|TIM9\_XENLA Mitochondrial import inner membrane translocase subunit Tim9 OS=Xenopus laevis OX=8355  
GN=tim9 PE=3 SV=1

MAAQMSQDIKQFKEFLGTYNKLTENCFLDCVKDFTTREVKAEMTCSEHCLQKYLKMT  
QRISMRFQEYHIQQNEALAAKAGLLGQPR

>sp|Q9P7K0|TIM9\_SCHPO Mitochondrial import inner membrane translocase subunit tim9 OS=Schizosaccharomyces pombe  
(strain 972 / ATCC 24843) OX=284812 GN=tim9 PE=3 SV=1

MDRLNVKEQEHLTQVLEAKQLKEYLNMYSTLTQNCFSQDFTSSKLSNKESECIKCA  
DKFLKHSERVGQRFAEFNAKYMGG

>sp|Q9WV97|TIM9\_RAT Mitochondrial import inner membrane translocase subunit Tim9 OS=Rattus norvegicus OX=10116  
GN=Timm9 PE=1 SV=3

MAAQIPESDIKQFKEFLGTYNKLTETCFLDCVKDFTTREVKPEEVTSEHCLQKYLKMT  
QRISMRFQEYHIQQNEALAAKAGLLGQPR

>sp|Q9XGX7|TIM9\_ORYSJ Mitochondrial import inner membrane translocase subunit Tim9 OS=Oryza sativa subsp. japonica  
OX=39947 GN=TIM9 PE=3 SV=2

MDKSMGLDGLDGLPEEDKMRMAAMVDQLQIRDSLRYNSLVERCFTDCVDTFRRTLDKQE  
ESCVRRCAEKFLKHSRVMRFAELNQGVATPD

>sp|Q8J1Z1|TIM9\_NEUCR Mitochondrial import inner membrane translocase subunit tim9 OS=Neurospora crassa (strain  
ATCC 24698 / 74-OR23-1A / CBS 708.71 / DSM 1257 / FGSC 987) OX=367110 GN=tim9 PE=1 SV=1

MDGLTAAESRELDQRLQKRQVKEFMSVFGNLVDNCFACVDDFTSKALSGRESGCISRCV  
LKSMTQTRLGERFGELNAAMTAEMQRR

>sp|Q9WV98|TIM9\_MOUSE Mitochondrial import inner membrane translocase subunit Tim9 OS=Mus musculus OX=10090  
GN=Timm9 PE=1 SV=1

MAAQIPESDIKQFKEFLGTYNKLTETCFLDCVKDFTTREVKPEEVTSEHCLQKYLKMT  
QRISVRFQEYHIQQNEALAAKAGLLGQPR

>sp|Q9XGX8|TIM9\_MESCR Mitochondrial import inner membrane translocase subunit Tim9 OS=Mesembryanthemum  
crystallinum OX=3544 GN=TIM9 PE=3 SV=1

MDKNMLGDLNGLPEEDKLMASIEQLQIRDSLRYNNLVERCFTDCVDSFRRTLDKQE  
ETCVKRCAEKFLKHSRVLRFELNQGAATTD

>sp|Q6CM57|TIM9\_KLULA Mitochondrial import inner membrane translocase subunit TIM9 OS=Kluyveromyces lactis (strain  
ATCC 8585 / CBS 2359 / DSM 70799 / NBRC 1267 / NRRL Y-1140 / WM37) OX=284590 GN=TIM9 PE=3 SV=1

MDQLNGKEQEFQKIVEQKQMKDFMRLYSNLVERCFSDCVNDFTSAKLSKEQNCIMRCS  
EKFLKHSERVGQRFEQNAAMNQTMGR

>sp|Q9Y5J7|TIM9\_HUMAN Mitochondrial import inner membrane translocase subunit Tim9 OS=Homo sapiens OX=9606  
GN=TIMM9 PE=1 SV=1

MAAQIPESDIKQFKEFLGTYNKLTETCFLDCVKDFTTREVKPEETTCSEHCLQKYLKMT  
QRISMRFQEYHIQQNEALAAKAGLLGQPR

>sp|Q4IB65|TIM9\_GIBZE Mitochondrial import inner membrane translocase subunit TIM9 OS=Gibberella zeae (strain ATCC  
MYA-4620 / CBS 123657 / FGSC 9075 / NRRL 31084 / PH-1) OX=229533 GN=TIM9 PE=3 SV=1

MDMLSAAEQRTLEQRMQKRQVKEFMGAFGLVEHCFMSCVDDFTSKAISNRESGCINRCV  
QKWMASQQRISDRFQEHNAQLTAQMNK

>sp|Q9Y8A7|TIM9\_EMENI Mitochondrial import inner membrane translocase subunit tim9 OS=Emicella nidulans (strain  
FGSC A4 / ATCC 38163 / CBS 112.46 / NRRL 194 / M139) OX=227321 GN=tim9 PE=3 SV=3

MDGLNAAEQRELNRMERKQMKFMTMYSKLVQRCFDDCVNDFTTKSLISREEGCVMRCV  
DKFMKGSQRLNERFQEQNAAMMQSGQLPGR



## Clustal Omega:

CLUSTAL O(1.2.4) multiple sequence alignment

```
sp|Q9VYD7|TIM9_DROME      -----MAKTPENIAIDQLDKDQIKTFSDFLMSYNKLSSETCFDTCIRDFTTRDVKDSE 52
sp|Q9W762|TIM9_DANRE      -----MAAQVTESDQIKQFKFELGTYNKLTENCMDVCVKDFTTREVKPEE 45
sp|Q5ZIR8|TIM9_CHICK      -----MAGQISSETDQIKQFKFELGTYNKITENCMDTCIRDFTTREVKPEE 45
sp|Q4V7R1|TIM9_XENLA      -----MAAQMSSEDQIKQFKFELGTYNKLTENCFLDCVKDFTTREVKAEE 45
sp|Q9Y5J7|TIM9_HUMAN      -----MAAQIPESDQIKQFKFELGTYNKLTETCFLDCVKDFTTREVKPEE 45
sp|Q2KIV2|TIM9_BOVIN      -----MAAQIPESDQIKQFKFELGTYNKLTETCFLDCVKDFTTREVKPEE 45
sp|Q9WV97|TIM9_RAT        -----MAAQIPESDQIKQFKFELGTYNKLTETCFLDCVKDFTTREVKPEE 45
sp|Q9WV98|TIM9_MOUSE      -----MAAQIPESDQIKQFKFELGTYNKLTETCFLDCVKDFTTREVKPEE 45
sp|Q9XGX9|TIM9_ARATH      MDASMMAGLDGLPEEDKAKMASMIDQLQLRDSLRMYNSLVERCFVDCVDSFTRKSLQKQE 60
sp|Q9XGX7|TIM9_ORYSJ      MDKSMLGDLGLPEEDKMRMAAMVDQLQIRDSLRMYNSLVERCFDTCVDTFRRKTLDKQE 60
sp|Q9XGX8|TIM9_MESCR      MDKNMLGDLNLPEEDKLKMASMIEQLQIRDSLRMYNNLVERCFDTCVDSFRRKTLDKQE 60
sp|Q8J1Z1|TIM9_NEUCR      -----MDGLTAAESRELDQRLQKRQVKFMSVFGNLVDNCFACVDDFTSKALSGRE 52
sp|Q4IB65|TIM9_GIBZE      -----MDMLSAAEQRTLQRMQKRQVKFEMGAFGGLVEHCFMSCVDDFTSKAISNRE 52
sp|P0CR96|TIM9_CRYNJ      -----MDFSQFNGAEQAHSKVIKKQMDFMRLYSGLVEKCFNACAQDFTSKALTINE 54
sp|Q9Y8A7|TIM9_EMENI      -----MDGLNAAEQRELANRMERKQMKFMTMYSKLVQRCFDDCVNDFTTKSLISRE 52
sp|Q9P7K0|TIM9_SCHPO      -----MDRLNVKEQEHLTQVLEAKQLKEYLNMYSTLTQNCFSDCVQDFTSSKLSNKE 52
sp|Q6CM57|TIM9_KLULA      -----MDQLNGKEQQEFQKIVEQKQMKDFMRLYNSLVERCFSDCVNDFTSKALTSKE 52
sp|Q6FRT3|TIM9_CANGA      -----MDQLNAKEQQEFQKLVEQKQMKDFMRLYSGLVERCFDTCVNDFTSKALTSKE 52
sp|Q6BU42|TIM9_DEBHA      -----MDQLNVKEQQDFQQIVEQKQMKDFMRLYNSLVSKCFDDCVNDFTSNNLTKE 52
sp|Q59R24|TIM9_CANAL      -----MDQLNVKEQQEFQKIVEQKQMKDFMRLYNSLVSRCFDDCVNDFTSNLSLTSKE 52
                               : . : : : . : * * * * : *

sp|Q9VYD7|TIM9_DROME      EKCSLNCMEKYLKMNQRVSRFQFEFQVIAHENALAMAQKTGKL ---- 95
sp|Q9W762|TIM9_DANRE      TTCSESLQKYLKMTQRISMRFQEYHIQQNERWP ---QKPDY----- 84
sp|Q5ZIR8|TIM9_CHICK      ITCSEHCLQKYLKMTQRISMRFQEYHIQQNEALA ---AKAGLLSQPR 89
sp|Q4V7R1|TIM9_XENLA      MTCSEHCLQKYLKMTQRISMRFQEYHIQQNEALA ---AKAGLLGQPR 89
sp|Q9Y5J7|TIM9_HUMAN      TTCSEHCLQKYLKMTQRISMRFQEYHIQQNEALA ---AKAGLLGQPR 89
sp|Q2KIV2|TIM9_BOVIN      TTCSEHCLQKYLKMTQRISMRFQEYHIQQNEALA ---AKAGLLGQPR 89
sp|Q9WV97|TIM9_RAT        VTCSEHCLQKYLKMTQRISMRFQEYHIQQNEALA ---AKAGLLGQPR 89
sp|Q9WV98|TIM9_MOUSE      VTCSEHCLQKYLKMTQRISVRFQEYHIQQNEALA ---AKAGLLGQPR 89
sp|Q9XGX9|TIM9_ARATH      ETCVMRCAEKFLKHTMRVGMRFaelNQNAPTQ ---D----- 93
sp|Q9XGX7|TIM9_ORYSJ      ESCVRRCAEKFLKHSMRVGMRFaelNQGVATP ---D----- 93
sp|Q9XGX8|TIM9_MESCR      ETCVKRCAEKFLKHSMRVGLRFAELNQGAATT ---D----- 93
sp|Q8J1Z1|TIM9_NEUCR      SGCISRCVLKSMSTQTRLGERFGElnAAMTAE ---MQRR----- 88
sp|Q4IB65|TIM9_GIBZE      SGCINRCVQKWMASQQRISDRFQEhNAQLTAQ ---MNK----- 87
sp|P0CR96|TIM9_CRYNJ      TTCVQNCTDKFLKHSErVGARFAEHNAEQMQG ---AGQ----- 89
sp|Q9Y8A7|TIM9_EMENI      EGCVMRCVDKFMKGSQRLNERFQEQAAMMQS ---GQLPGR----- 90
sp|Q9P7K0|TIM9_SCHPO      SECIAKCADKFLKHSErVGQRFaEFNAKYMgQ ----- 84
sp|Q6CM57|TIM9_KLULA      QNCIMRCSEKFLKHSErVGQRFQEQAAMNQT ---MGR----- 87
sp|Q6FRT3|TIM9_CANGA      ESCILKCEKFLKHSErVGQRFQEQAALGQG ---LGR----- 87
sp|Q6BU42|TIM9_DEBHA      TGCITKCEKFLKHSErVGQRFQEQAALLMQN ---MQKR----- 88
sp|Q59R24|TIM9_CANAL      TSCIAKCEKFLKHSErVGQRFQEQAALLMQQ ---GPK----- 87
                               * * * : * : * * * :
```

## MAFFT:



```
>sp|Q4V7R1|TIM9_XENLA Mitochondrial import inner membrane translocase subunit Tim9 OS=Xenopus laevis OX=8355 GN=timm9 PE=3 SV=1
MA-----AQMSESDQIKQKFEFLGTYNKLTEGFLDCVKDFTTREVKAEE
MTCSEHCLQKYLKMTQRTISMRFQEYHIQQNE---ALAAKAGLLGQPR
>sp|Q9WV97|TIM9_RAT Mitochondrial import inner membrane translocase subunit Tim9 OS=Rattus norvegicus OX=10116 GN=Timm9 PE=1 SV=3
MA-----AQIPESDQIKQKFEFLGTYNKLTEGFLDCVKDFTTREVKPEE
VTCSEHCLQKYLKMTQRTISMRFQEYHIQQNE---ALAAKAGLLGQPR
>sp|Q9WV98|TIM9_MOUSE Mitochondrial import inner membrane translocase subunit Tim9 OS=Mus musculus OX=10090 GN=Timm9 PE=1 SV=1
MA-----AQIPESDQIKQKFEFLGTYNKLTEGFLDCVKDFTTREVKPEE
VTCSEHCLQKYLKMTQRTISMRFQEYHIQQNE---ALAAKAGLLGQPR
>sp|Q9Y5J7|TIM9_HUMAN Mitochondrial import inner membrane translocase subunit Tim9 OS=Homo sapiens OX=9606 GN=TIMM9 PE=1 SV=1
MA-----AQIPESDQIKQKFEFLGTYNKLTEGFLDCVKDFTTREVKPEE
TTCSEHCLQKYLKMTQRTISMRFQEYHIQQNE---ALAAKAGLLGQPR
>sp|Q2KIV2|TIM9_BOVIN Mitochondrial import inner membrane translocase subunit Tim9 OS=Bos taurus OX=9913 GN=TIMM9 PE=3 SV=1
MA-----AQIPESDQIKQKFEFLGTYNKLTEGFLDCVKDFTTREVKPEE
TTCSEHCLQKYLKMTQRTISMRFQEYHIQQNE---ALAAKAGLLGQPR
>sp|Q5ZIR8|TIM9_CHICK Mitochondrial import inner membrane translocase subunit Tim9 OS=Gallus gallus OX=9031 GN=TIMM9 PE=3 SV=1
MA-----GQTSETDQIKQKFEFLGTYNKLTEGFLDCVKDFTTREVKPEE
ITCSEHCLQKYLKMTQRTISMRFQEYHIQQNE---ALAAKAGLLGQPR
>sp|Q9W762|TIM9_DANRE Mitochondrial import inner membrane translocase subunit Tim9 OS=Danio rerio OX=7955 GN=timm9 PE=3 SV=2
MA-----AQVTESDQIKQKFEFLGTYNKLTEGFLDCVKDFTTREVKPEE
TTCSEHCLQKYLKMTQRTISMRFQEYHIQQNE---RMPQKPDY-----
>sp|Q9YVY7|TIM9_DROME Mitochondrial import inner membrane translocase subunit Tim9 OS=Drosophila melanogaster OX=7227 GN=Tim9a PE=3 SV=1
MA-----KTPENIATDQLDKQIKTFSDFLMSYNKLSETCTDCIRDFTTREVKPEE
EKCSLNCMEKYLKMTQRTISMRFQEYHIQQNE---ALAAKAGLLGQPR
>sp|Q9P7K0|TIM9_SCHPO Mitochondrial import inner membrane translocase subunit tim9 OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) OX=284812 GN=tim9 PE=3 SV=1
MD-----RLIMVFEQHLTVLEAKQLKEVLHMYSTLTQNCFSDCVQDFTSSKSNKE
SECIACADQKFLKHSERVGQRFQEQA-----KYMQ-----
>sp|Q6CM57|TIM9_KLULA Mitochondrial import inner membrane translocase subunit TIM9 OS=Kluyveromyces lactis (strain ATCC 8585 / CBS 2359 / DSM 70799 / NBRC 1267 / NRRL Y-1140 / WM37)
MD-----QLNGFEQEFQKIVEQKQKDFMRLYSNLVERCFSDCVDSFTRKSLQKQE
QNCIMRCSEKFLKHSERVGQRFQEQA-----APNQTMGR-----

>sp|Q6FRT3|TIM9_CANGA Mitochondrial import inner membrane translocase subunit TIM9 OS=Candida glabrata (strain ATCC 2001 / CBS 138 / JCM 3761 / NBRC 0622 / NRRL Y-65)
MD-----QLNAKEQQEFQKVLVEQKQKDFMRLYSGLVERCFDTCVNDFTSSKSLTSKE
ESCILKCSKFLKHSERVGQRFQEQA-----ALGQGLGR-----
>sp|Q6BU42|TIM9_DEBHA Mitochondrial import inner membrane translocase subunit TIM9 OS=Debaryomyces hansenii (strain ATCC 36239 / CBS 767 / BCRC 21394 / JCM 1990 / NBRC 100001)
MD-----QLNVKEQQDFQQIVEQKQKDFMRLYSNLVSKCFDDCVNDFTSNLTTKE
TGCITKCSKFLKHSERVGQRFQEQA-----LLMQNMQR-----
>sp|Q59R24|TIM9_CANAL Mitochondrial import inner membrane translocase subunit TIM9 OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=TIM9 PE=3 SV=2
MD-----QLNVKEQQEFQQIVEQKQKDFMRLYSNLVSKCFDDCVNDFTSNLTSKE
TSCIAKCSKFLKHSERVGQRFQEQA-----LLMQQGPK-----
>sp|P0CR96|TIM9_CRYNJ Mitochondrial import inner membrane translocase subunit TIM9 OS=Cryptococcus neoformans var. neoformans serotype D (strain JEC21 / ATCC MYA-565)
MDFS-----QFNGAEQAHMVKVIEKKQMDFMRLYSGLVEKCFNACAQDFTSKALTTNE
TTCVQNCIDKFLKHSERVGARFAEHNA-----EQMQGAGQ-----
>sp|Q9Y8A7|TIM9_EMENI Mitochondrial import inner membrane translocase subunit tim9 OS=Emmericella nidulans (strain FGSC A4 / ATCC 38163 / CBS 112.46 / NRRL 194 / M139)
MD-----GLNAAEQRLANRMRKQKQKFMFMYSKLVRCFDDCVNDFTTKSLISRE
EGCVMRVDKFMKGSRQLNERFQEQA-----AMMQSGQLPG--R
>sp|Q8J1Z1|TIM9_NEUCR Mitochondrial import inner membrane translocase subunit tim9 OS=Neurospora crassa (strain ATCC 24698 / 74-OR23-1A / CBS 708.71 / DSM 1257 / FGSC 9650)
MD-----GLTAAESRELDQRLQKRVKEFMSVFGNLVDNCFACVDDFTSKALSGRE
SGCISRCVLKSMSTQTRLGERFGELNA-----AMTAEMQRR-----
>sp|Q4IB65|TIM9_GIBZE Mitochondrial import inner membrane translocase subunit TIM9 OS=Gibberella zeae (strain ATCC MYA-4620 / CBS 123657 / FGSC 9075 / NRRL 31084 / PH-2)
MD-----MLSAAEQRTLEQRMQRQVKEFMGAFGGGLVEHCFMSCVDDFTSKAISNRE
SGCINRCVQKMMASQQRISDRFQEHA-----QLTAQMNK-----
>sp|Q9XGX7|TIM9_ORYSJ Mitochondrial import inner membrane translocase subunit Tim9 OS=Oryza sativa subsp. japonica OX=39947 GN=TIM9 PE=3 SV=2
MDKSNLGDGLPEEDKMRMAAMVDQLIRDSLRYMNSLVERCFDTCVDSFRRKTLQKQE
ESCVRRCAEKFLKHSMRVGMFAELNQ-----GVATPD-----
>sp|Q9XGX8|TIM9_MESCR Mitochondrial import inner membrane translocase subunit Tim9 OS=Mesembryanthemum crystallinum OX=3544 GN=TIM9 PE=3 SV=1
MDKNMLGDLDLPEEDKLMASMEQLQIRDSLRYMNSLVERCFDTCVDSFRRKTLQKQE
ETCVKRCAEKFLKHSMRVGLRFAELNQ-----GAATD-----
>sp|Q9XGX9|TIM9_ARATH Mitochondrial import inner membrane translocase subunit TIM9 OS=Arabidopsis thaliana OX=3702 GN=TIM9 PE=1 SV=2
MDASMMAGLDGLPEEDKAKMASMDQLQLRDSLRYMNSLVERCFDTCVDSFTRKSLQKQE
ETCVMRCAEKFLKHTMRVGMFAELNQ-----NAPTQD-----
```

## MUSCLE:

CLUSTAL multiple sequence alignment by MUSCLE (3.8)

```

sp|Q9VYD7|TIM9_DROME      -----MAKTPENIAIDQLDKDQIKTFSDFLMSYNKLSSETCTDCIRDFTTRDVKDSE
sp|Q9W762|TIM9_DANRE      -----MAAQVTESDQI-----KQFKEFLGTYNKLTENCFMDCVKDFTTREVVKPEE
sp|Q5ZIR8|TIM9_CHICK      -----MAQGISETDQI-----KQFKEFLGTYNKITENCFMDCIRDFTTREVVKPEE
sp|Q4V7R1|TIM9_XENLA      -----MAAQMSESDQI-----KQFKEFLGTYNKLTENCFLDCVKDFTTREVKAEE
sp|Q9WV98|TIM9_MOUSE      -----MAAQIPESDQI-----KQFKEFLGTYNKLTETCFLDCVKDFTTREVVKPEE
sp|Q9Y5J7|TIM9_HUMAN      -----MAAQIPESDQI-----KQFKEFLGTYNKLTETCFLDCVKDFTTREVVKPEE
sp|Q2KIV2|TIM9_BOVIN      -----MAAQIPESDQI-----KQFKEFLGTYNKLTETCFLDCVKDFTTREVVKPEE
sp|Q9WV97|TIM9_RAT        -----MAAQIPESDQI-----KQFKEFLGTYNKLTETCFLDCVKDFTTREVVKPEE
sp|Q9XGX9|TIM9_ARATH      MDASMMAGLDGLPEEDKAKMASMIDQLQLRDSLRMYNSLVERCFVDCVDSFTRKSLQKQE
sp|Q9XGX7|TIM9_ORYSJ      MDKSMLGDLGLPEEDKMRMAAMVDQLQIRDSLRMYNSLVERCFDTCVDFRRKTLQKQE
sp|Q9XGX8|TIM9_MESCR      MDKNMLGDLGLPEEDKLKMASMIEQLQIRDSLRMYNNLVERCFDTCVDSFRRKTLQKQE
sp|Q8J1Z1|TIM9_NEUCR      -----MDGLTAAESRELDQRLQKRQVKFMSVFGNLVDNCFACVDDFTSKALSGRE
sp|Q4IB65|TIM9_GIBZE      -----MDMLSAAEQRTLEQRMQKRQVKFEMGAFGGLVEHCFMSCVDDFTSKAISNRE
sp|Q9Y8A7|TIM9_EMENI      -----MDGLNAAEQRELANRMERKQMKFMTMYSKLVQRCFDDCVNDFTTKSLISRE
sp|P0CR96|TIM9_CRYNJ      -----MDFSQFNGAEQAHMSKVIEKKQMDFMRLYSGLVEKCFNACAQDFTSKALTTNE
sp|Q9P7K0|TIM9_SCHPO      -----MDRLNVKEQEHLTQVLEAKQLKEYLNMYSTLTQNCFSDCVQDFTSSKLSNKE
sp|Q6BU42|TIM9_DEBHA      -----MDQLNVKEQQDFQQIVEQKQMKDFMRLYSNLVSKCFDDCVNDFTSNNLTTSKE
sp|Q59R24|TIM9_CANAL      -----MDQLNVKEQQEFQQIVEQKQMKDFMRLYSNLVSRCFDDCVNDFTSNLSLSKE
sp|Q6CM57|TIM9_KLULA      -----MDQLNGKEQQEFQKIVEQKQMKDFMRLYSNLVERCFSDCVNDFTSAKLTSKE
sp|Q6FRT3|TIM9_CANGA      -----MDQLNAKEQQEFQKLVEQKQMKDFMRLYSGLVERCFDTCVNDFTSSKLSKE
                                     . : : : . ** * * : *

```

```

sp|Q9VYD7|TIM9_DROME      EKCSLNCMEKYLKMNQRVSRFQEFQVIAHENALAMAQKTGKL----
sp|Q9W762|TIM9_DANRE      TTCSESLQKYLKMTQRISMRFQEYHIQQNE---RWPQKPDY-----
sp|Q5ZIR8|TIM9_CHICK      ITTCEHCLQKYLKMTQRISMRFQEYHIQQNE---ALAAGAGLLSQPR
sp|Q4V7R1|TIM9_XENLA      MTTCEHCLQKYLKMTQRISMRFQEYHIQQNE---ALAAGAGLLGQPR
sp|Q9WV98|TIM9_MOUSE      VTTCEHCLQKYLKMTQRISVRFQEYHIQQNE---ALAAGAGLLGQPR
sp|Q9Y5J7|TIM9_HUMAN      TTCSEHCLQKYLKMTQRISMRFQEYHIQQNE---ALAAGAGLLGQPR
sp|Q2KIV2|TIM9_BOVIN      TTCSEHCLQKYLKMTQRISMRFQEYHIQQNE---ALAAGAGLLGQPR
sp|Q9WV97|TIM9_RAT        VTTCEHCLQKYLKMTQRISMRFQEYHIQQNE---ALAAGAGLLGQPR
sp|Q9XGX9|TIM9_ARATH      ETCVMRCAEKFLKHTMRVGMRFALN-----QNAPTQD-----
sp|Q9XGX7|TIM9_ORYSJ      ESCVRRCAEKFLKHSRMRVGMRFALN-----QGVATPD-----
sp|Q9XGX8|TIM9_MESCR      ETCVKRCAEKFLKHSRMRVGLRFAELN-----QGAATTD-----
sp|Q8J1Z1|TIM9_NEUCR      SGCISRCVLKSMSTQTRLGERFGELN-----AAMTAEM---QRR
sp|Q4IB65|TIM9_GIBZE      SGCINRCVQKWMASQQRISDRFQEHN-----AQLTAQM---NK
sp|Q9Y8A7|TIM9_EMENI      EGCVMRCVDKFMKGSQRLNERFQEQN-----AAMMQSGQL-PGR
sp|P0CR96|TIM9_CRYNJ      TTCVQNCTDKFLKHSERVGARFAEHN-----AEQMQGA---GQ
sp|Q9P7K0|TIM9_SCHPO      SECIACADKFLKHSERVQGRFAEFN-----AKYMGQ-----
sp|Q6BU42|TIM9_DEBHA      TGCITKCEKFLKHSERVQGRFQEQN-----ALLMQNM---QKR
sp|Q59R24|TIM9_CANAL      TSCIACKEKFLKHSERVQGRFQEQN-----ALLMQQG---PK
sp|Q6CM57|TIM9_KLULA      QNCIMRCSEKFLKHSERVQGRFQEQN-----AAMNQTM---GR
sp|Q6FRT3|TIM9_CANGA      ESCILKCEKFLKHSERVQGRFQEQN-----AALGQGL---GR
                                     * * * : * : * * :

```

Five residue positions that are aligned differently in these methods are:





1. The first residue (M) in all sequences
2. 11 th position (D)
3. The last residue (R)

4. 12 th position (Q)

5. 13 th position (L)

8) When blasted the given sequence 'EPDMRTPIAHTMAW', with the PDB database the results are:

Entry	Alignment overview	Info	Status
<input type="checkbox"/> Query: B202202274ABAA9BC7178C81CEBC9459510EDDEA3030429F			
<input type="checkbox"/> P45568	DXR_ECOLI - 1-deoxy-D-xylulose 5-phosphate redu... - <a href="#">Escherichia coli...</a> - <a href="#">View alignment</a>	E-value: 1.5e-10 Score: 118 Ident.: 100.0%	
<input type="checkbox"/> Q9X5F2	DXR_ZYMMO - 1-deoxy-D-xylulose 5-phosphate redu... - <a href="#">Zymomonas mobili...</a> - <a href="#">View alignment</a>	E-value: 1.2e-6 Score: 92 Ident.: 84.6%	
<input type="checkbox"/> B1JQG4	DXR_YERPY - 1-deoxy-D-xylulose 5-phosphate redu... - <a href="#">Yersinia pseudot...</a> - <a href="#">View alignment</a>	E-value: 3.3e-6 Score: 89 Ident.: 91.7%	
<input type="checkbox"/> Q8ZH62	DXR_YERPE - 1-deoxy-D-xylulose 5-phosphate redu... <a href="#">Yersinia pestis</a> - <a href="#">View alignment</a>	E-value: 3.3e-6 Score: 89 Ident.: 91.7%	
<input type="checkbox"/> B7H1U5	DXR_ACIB3 - 1-deoxy-D-xylulose 5-phosphate redu... - <a href="#">Acinetobacter ba...</a> - <a href="#">View alignment</a>	E-value: 3e-4 Score: 76 Ident.: 76.9%	
<input type="checkbox"/> D5VAW2	D5VAW2_MORCB - 1-deoxy-D-xylulose 5-phosphate redu... - <a href="#">Moraxella catarr...</a> - <a href="#">View alignment</a>	E-value: 1.7e-3 Score: 71 Ident.: 75.0%	
<input type="checkbox"/> Q8DBF5	DXR_VIBVU - 1-deoxy-D-xylulose 5-phosphate redu... - <a href="#">Vibrio vulnificu...</a> - <a href="#">View alignment</a>	E-value: 1.1e-1 Score: 59 Ident.: 75.0%	
<input type="checkbox"/> Q8L6Z4	RPS9M_ARATH - 30S ribosomal protein S9, mitochond... - <a href="#">Arabidopsis thal...</a> - <a href="#">View alignment</a>	E-value: 2.6e0 Score: 50 Ident.: 100.0%	
<input type="checkbox"/> P32297	ACHA3_HUMAN - Neuronal acetylcholine receptor sub... <a href="#">Homo sapiens (Human)</a> - <a href="#">View alignment</a>	E-value: 2.6e0 Score: 50 Ident.: 63.6%	
<input type="checkbox"/> Q9WZZ1	DXR_THEMA - 1-deoxy-D-xylulose 5-phosphate redu... - <a href="#">Thermotoga marit...</a> - <a href="#">View alignment</a>	E-value: 3.7e0 Score: 49 Ident.: 85.7%	
<input type="checkbox"/> P04757	ACHA3_RAT - Neuronal acetylcholine receptor sub... - <a href="#">Rattus norvegicu...</a> - <a href="#">View alignment</a>	E-value: 3.7e0 Score: 49 Ident.: 63.6%	
<input type="checkbox"/> P12390	ACHB2_RAT - Neuronal acetylcholine receptor sub... - <a href="#">Rattus norvegicu...</a> - <a href="#">View alignment</a>	E-value: 3.7e0 Score: 49 Ident.: 63.6%	

<input type="checkbox"/>	<a href="#">P17787</a>	ACHB2_HUMAN - Neuronal acetylcholine receptor sub... <a href="#">Homo sapiens (Human)</a> - <a href="#">View alignment</a>	E-value: 3.7e0 Score: 49 Ident.: 63.6%	
				
<input type="checkbox"/>	<a href="#">O93655</a>	TRF3_THEAC - Tricorn protease-interacting factor... - <a href="#">Thermoplasma aci...</a> - <a href="#">View alignment</a>	E-value: 3.7e0 Score: 49 Ident.: 66.7%	
				

A total of 14 hits were found with the PDB database.

The most significance of the results is that it has 100% sequence identity with two entries **P45568**, which is a reductoisomerase in E.coli strain and **Q8L6Z4**, which is a ribosomal protein in *Arabidopsis thaliana*. Also, it has two more sequences with sequence identity of more than 90.