

[illegible]

```
#####
# Program: water
# Rundate: Mon 18 Oct 2021 07:16:59
# Commandline: water
#   -auto
#   -stdout
#   -asequence emboss_water-I20211018-071656-0609-21294862-p2m.asequence
#   -bsequence emboss_water-I20211018-071656-0609-21294862-p2m.bsequence
#   -datafile EBLOSUM62
#   -gapopen 10.0
#   -gapextend 0.5
#   -aformat3 pair
#   -sprotein1
#   -sprotein2
# Align_format: pair
# Report_file: stdout
#####

#=====
#
# Aligned_sequences: 2
# 1: M57671.1
# 2: AC166935.3
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 900
# Identity:   318/900 (35.3%)
# Similarity: 318/900 (35.3%)
# Gaps:       474/900 (52.7%)
# Score: 1338.0
#
#
#=====

M57671.1          1 GC-----ATTCTGAGGCATTCTCTAAC-----AGGTTCTCGA      32
                  || .|||||.||..||..||..||. || .|||.|||..
AC166935.3       7906 GCCCCCACTTCTG6GGTTATGCCCTGACTCAGACCCACCCACGGTCTCTC      7955
```

Input form	Web services	Help & Documentation	Bioinformatics Tools FAQ
M57671.1	176	G-----AGTGGCT---TCTAT-----AGAC-----	192
AC166935.3	8306	GACTCAGCCAGCCTCCAAGTGACTTGACTCTTCTGGGAGACAAAGGAC	8355
M57671.1	193	-----CCCAC-----GACC--GCCGAGAGCTGGAGGACCTCCAGG	225
AC166935.3	8356	AAGGCTGGGCCCTCCCATCTGCCCTGTCTACAGCTGAAAG-CAGCCAGG	8404
M57671.1	226	TGGAG-----CAGGCA-GAACTGGGTCTGGA-----GG	252
AC166935.3	8405	ATGAGAAGATGAGGTTCCAGCCAAGCACGAACCCGGTCAGGAATCTGGGG	8454
M57671.1	253	C-----AGGC-----GGCCT-----GC--	264
AC166935.3	8455	CATATTCTAGGCAATGTCCCAAAGATCTGTCCAGGGCCTCAAGAAGCTA	8504
M57671.1	265	-----AGCCTTC-----	271
AC166935.3	8505	CCCCGACAAGGATCCGTGGGTGCCCTCTGAGGAGTCAGAGAGAACCTGA	8554
M57671.1	272	-----GGCCTGGAGATGATTCTGCA-GAAGCGCGG--	301
AC166935.3	8555	TGGCGCCAGGGAGAGAGACTGGTAATGACTCCCAGGAAGTGCTGAC	8604
M57671.1	302	-----CATTGTGGATCAGTGCTGTAATAACATTT--GCACA-----	335
AC166935.3	8605	AGGGGGCACATAGAGGAT--GTGCTGAAAGTGCACTTGGGCACAAAGAAT	8652
M57671.1	336	-----TTTAACCAGCTGCAGAAC-----TACTG---	358
AC166935.3	8653	GGGCAGATCCACAGGGGCTGATCCCGCTGAGAGTGCGCCTGTCTGGCT	8702
M57671.1	359	CAATGTC-----CCTTAGACACCTGCCTTGGG-CCTG--	389
AC166935.3	8703	CAGTGTCCACGGTGCCTTTTCACCTCTGACACCCACGTGGGGACCAGGA	8752
M57671.1	390	-GCCTGC-TGCTCT---GCCCTGGCAA-C-CAATAAACCCCTTGAATGAG	432
AC166935.3	8753	AGCCAGCATCCTCCAGGGCTCAGGAAATCTCTATGAACCCC---ACCGAG	8799
#-----			
#-----			

⇒ SIMILARITY PERCENTAGE is found to be - 35.3%

- 2) Given a protein sequence, find out which protein it is and from which organism it belongs using BLAST Program.

Protein Sequence

```
MERGVERRGAALVAAWRSLWERGGLALFRPQCRTGCGACRVQGTRPFSLSAAASAVLG
LGSWGGDSGKQKLTLDVAELIRKKECRRVVMAGAGIS
TPSGIPDFRSPGSGLYSNLEQYNIPYPEAIFELAYFFINPKPFFTLAKELYPGNYRPNYAHY
FLRLLHDKGLLRLTYTQNIDGLERVAGIPDRLVEAHGT
FATATCTVCRKFPGEDFRGDVMADKVPCHRVCTGIVKPDIVFFGEELPQRFFLHMTDF
PMADLLFVIGTSLEVEPFASLAGAVRNSVPRVLINRDLV
GPFAWQQRYNDIAQLGDVVTGVEKMEVLLDWNEEMQTLIQKEKEKLDKDK
```

- ⇒ ORGANISM : [Homo sapiens](#)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;  
 Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates;  
 Haplorrhini;  
 Catarrhini; Hominidae; Homo.

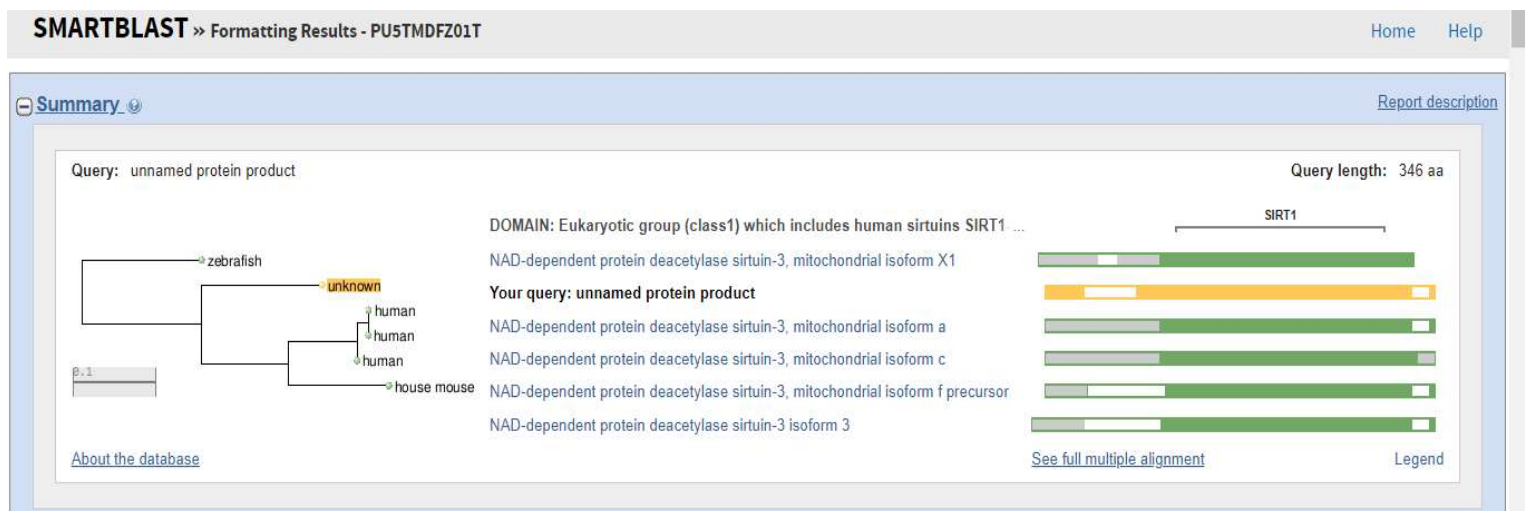
From the given details, find

- What is the length of the query sequence ?

⇒ Length of Query = 346

- What is the number of sequences in the database searched ?

⇒



- What is the E-value ?



## Best hits

Select: [All](#) [None](#) Selected: 0

Alignments [GenPept](#)



	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	<a href="#">NAD-dependent protein deacetylase sirtuin-3, mitochondrial isoform a [Homo sapiens]</a>	468	468	80%	1e-164	78.93%	<a href="#">NP_036371.1</a>
<input type="checkbox"/>	<a href="#">NAD-dependent protein deacetylase sirtuin-3 isoform 3 [Mus musculus]</a>	444	444	80%	3e-156	73.93%	<a href="#">NP_001171275.1</a>
<input type="checkbox"/>	<a href="#">NAD-dependent protein deacetylase sirtuin-3, mitochondrial isoform X1 [Danio rerio]</a>	419	419	79%	6e-146	71.27%	<a href="#">XP_009301762.1</a>
<input type="checkbox"/>	<a href="#">NAD-dependent protein deacetylase sirtuin-3, mitochondrial isoform f precursor [Homo sapiens]</a>	457	457	79%	1e-161	78.83%	<a href="#">NP_001357244.1</a>
<input type="checkbox"/>	<a href="#">NAD-dependent protein deacetylase sirtuin-3, mitochondrial isoform c [Homo sapiens]</a>	460	460	80%	3e-161	77.50%	<a href="#">NP_001357239.1</a>

## Additional BLAST Hits

3) Given a protein sequence find out which protein it is and from which organism it belongs from FASTA program based on the identity. Given a protein sequence find out which protein it is and from which organism it belongs from FASTA program based on the identity. Also find the E-value for the given sequence.

EMBL-EBI

Services

Research

Training

Industry

About us

EMBL-EBI

Hinxton

FASTA

Protein

Nucleotide

Genomes

Proteomes

Whole Genome Shotgun

Web services

Also in this section

Feedback

Share

Tools > Sequence Similarity Searching > FASTA

Protein Similarity Search

This tool provides sequence similarity searching against protein databases using the FASTA suite of programs. FASTA provides a heuristic search with a protein query. FASTX and FASTY translate a DNA query. Optimal searches are available with SSEARCH (local), GGSEARCH (global) and GLSEARCH (global query, local database).

STEP 1 - Select your databases

PROTEIN DATABASES

1 Database Selected

X Clear Selection

☐ UniProt Knowledgebase (The UniProt Knowledgebase includes UniProtKB/Swiss-Prot and UniProtKB/TrEMBL)

☒ UniProtKB/Swiss-Prot (The manually annotated section of UniProtKB)

☐ UniProtKB/Swiss-Prot isoforms (The manually annotated isoforms of UniProtKB/Swiss-Prot)

☐ UniProtKB/TrEMBL (The automatically annotated section of UniProtKB)

☐ UniProtKB Reference Proteomes plus Swiss-Prot

☐ UniProtKB COVID-19

UniProtKB Taxonomic Subsets

This website requires cookies, and the limited processing of your personal data in order to function. By using the site you are agreeing to this as outlined in our [Privacy Notice](#) and [Terms of Use](#).

I agree, dismiss this banner



EMBL-EBI Hinxton

FASTA

Protein Nucleotide Genomes Proteomes Whole Genome Shotgun Web services Also in this section

Feedback Share

Tools &gt; Sequence Similarity Searching &gt; FASTA

Results for job fasta-l20211015-143509-0977-52395220-p2m

Summary Table

Tool Output

Visual Output

Functional Predictions

Submission Details

Download

Download in XML format

FASTA searches a protein or DNA sequence data bank  
 version 36.3.08 Aug. 2019  
 Please cite:  
 W.R. Pearson & D.J. Lipman PHAS (1988) 85:2444-2448

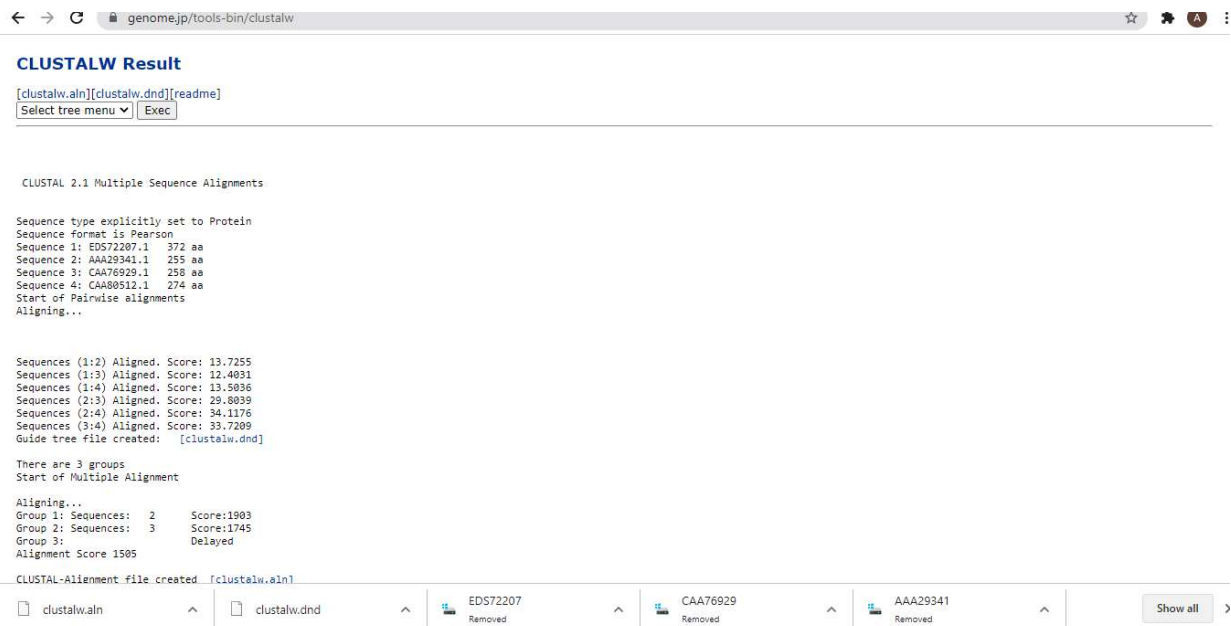
Query: @  
 1>>>E06055.001 - 344 aa  
 Library: UniProtKB  
 203850821 residues in 565254 sequences

Statistics: Expectation\_n\_fit: rho(ln(x))= 7.6607/-0.000143; mu= 6.3409+/- 0.008  
 mean\_var=64.2774/-13.592, 0's: 517 Z-trim(116.5): 704 B-trim: 736 in 1/63  
 Lambda= 0.1559972  
 statistics sampled from 60000 (63357) to 83966 sequences  
 Algorithm: FASTA (3.8 Nov 2011) [optimized]  
 Parameters: BL50 matrix (15/-5), open/pen: -10/-2  
 ktop: 2, E-join: 1 (0.482), E-opt: 0.2 (0.149), width: 16  
 Scan time: 9.840

The best scores are:  
 (opt bits E(565254))

Q291G7	SI33_HUMAN	NAD-dependent protein deacetylase	(399)	1557	367.9	1.4e-100
Q291Q4	SI33_MOUSE	NAD-dependent protein deacetylase	(334)	1475	349.1	5.4e-95
Q291S2	H5T2_SCHPO	NAD-dependent protein deacetylase	(332)	900	216.4	4.8e-55
Q5A695	H5T21_EMENI	NAD-dependent protein deacetylase	(361)	898	215.9	7.3e-55
Q7V7K3	SI32_DANRE	NAD-dependent protein deacetylase	(379)	884	212.6	7.3e-54
Q58FB1	SI32_PONAB	NAD-dependent protein deacetylase	(352)	877	211.0	2e-53

4) Collect the protein sequences CAA80512, AAA29341, CAA76929, and EDS72207 from the enterz database, keep the sequences together in FASTA format file, align the sequences each other and report the pair wise score using CLUSTALW?



```
CLUSTAL 2.1 Multiple Sequence Alignments

Sequence type explicitly set to Protein
Sequence format is Pearson
Sequence 1: EDS72207.1 372 aa
Sequence 2: AAA29341.1 255 aa
Sequence 3: CAA76929.1 258 aa
Sequence 4: CAA80512.1 274 aa
Start of Pairwise alignments
Aligning...

Sequences (1:2) Aligned. Score: 13.7255
Sequences (1:3) Aligned. Score: 12.4031
Sequences (1:4) Aligned. Score: 13.5036
Sequences (2:3) Aligned. Score: 29.8039
Sequences (2:4) Aligned. Score: 34.1176
Sequences (3:4) Aligned. Score: 33.7209
Guide tree file created: [clusterw.dnd]

There are 3 groups
Start of Multiple Alignment

Aligning...
Group 1: Sequences: 2 Score:1903
Group 2: Sequences: 3 Score:1745
Group 3: Delayed
Alignment Score 1505

CLUSTAL-Alignment file created [clusterw.aln]
```

clusterw.aln ^ clusterw.dnd ^ EDS72207 Removed CAA76929 Removed AAA29341 Removed Show all >



```
genome.jp/tools-bin/clusterw
CLUSTAL-Alignment file created [clusterw.aln]

clusterw.aln
CLUSTAL 2.1 multiple sequence alignment

CAA76929.1  -----HIRFTLALVIGVTFAS-----
CAAB0512.1  -----HSNKIAIILAVLVAVVCAEAQANQR-----HRLVR
AAA29341.1  -----RLFLALLALGFAAFAVP-----
EDS72207.1  MREKSGLTIFFLGIIGLIFGSAFFVTINRMEKIDKIGTVVQGGNSGSSSATT
          : * . : .

CAA76929.1  -----TPQETHPHILETIGGDANIIDYVPHQISFQRLH-----HFCGGFLISDTWVTA
CAAB0512.1  PPSFSPRPARYAVQRTVGFEDVSDAPYQVSLQYIKR-----HNCGSVLSSKIM/LTA
AAA29341.1  -----ANPQRITVGGSTTTIQVPTIALLFSRINHTFFQACGGTILNINRW/LTA
EDS72207.1  LKSSSTTEKVAHVTPAVVGITETELSKDHFIRREYESSGVGTGFIVDKNGIIVTNQHVVS
          : : * : : . * : : .

CAA76929.1  AHCTYEG-----YSOTENLNIIRVGSSEHSAGKLDHVK--
CAAB0512.1  AHCTAG-----ASTSSLTVRLGTSRHSAGGTVVRVA--
AAA29341.1  AHCPHG-----DAVIRIRVRSGSTFANSGGAVHNLN--
EDS72207.1  NPQKQINVTLDGTSYEAKVFKSDTVDLAVIKINAKILPVNLDGSEKINIGQTATIGIH
          : . * : .

CAA76929.1  -----RYITHPQYNIITNDNDIALLELALPDLNQSVRPALPVAGQEPDQAQLTIT
CAAB0512.1  -----RVVQHPKYDSSSIDFDYSLLELEDELTFSDAVQVGLPKQDETVDGTHHTVS
AAA29341.1  -----SVRIHPNIRNLNDIAIHTASIIAFHIAAQPARIAGAHVNLGDQVVA
EDS72207.1  PLGLTFERSVTQGIISALNRSIAVDNKTIAEDLIQTDAINAGSGGPLNLAGEVIGIH
          : . * . : : : .

CAA76929.1  GNGATVVGSEYETLQVVTIPTVINWVQ-----SAITNDT--ITNI
CAAB0512.1  GNGITQSAEENAVLRAHVPTVINQKECH-----KAYSDFGG-VTD
AAA29341.1  GNGAIRSGGSPSEQLRHVQVITVQATCR-----SRVYSGRTVTDI
EDS72207.1  TYKIQSGEGLGFAIPINIVKPIINQKEEGSFPTVTIGIEGVDKIARYYTEKLSVDK
          : . * . : : .

CAA76929.1  MFCAGLIGVGGDSCSGDSGG--PAVTDGQVGVISVHGSCADPKYPGITYKVSAFRDN
CAAB0512.1  MLCAGYQ--QGGKDACQDQSGG--PLVADGKLVGVVSHGVCAGQYPGVYSRAVVRDN
AAA29341.1  MLCSGHLDVGGDRDQDQSGG--PLYHGVVGVGCSHGEECALARFPGVNIARVTRYSNI
EDS72207.1  MLCSGHLDVGGDRDQDQSGG--PLYHGVVGVGCSHGEECALARFPGVNIARVTRYSNI
```

```
genome.jp/tools-bin/clusterw
clusterw.dnd
(
  (
    EDS72207.1:0.52641,
    AAA29341.1:0.33634
  )
  :0.81266,
  CAA76929.1:0.34493,
  CAAB0512.1:0.31786
);

Select tree menu  Exec

clusterw.aln  clusterw.dnd  EDS72207 Removed  CAA76929 Removed  AAA29341 Removed  Show all
```

\*\*\* END \*\*\*