

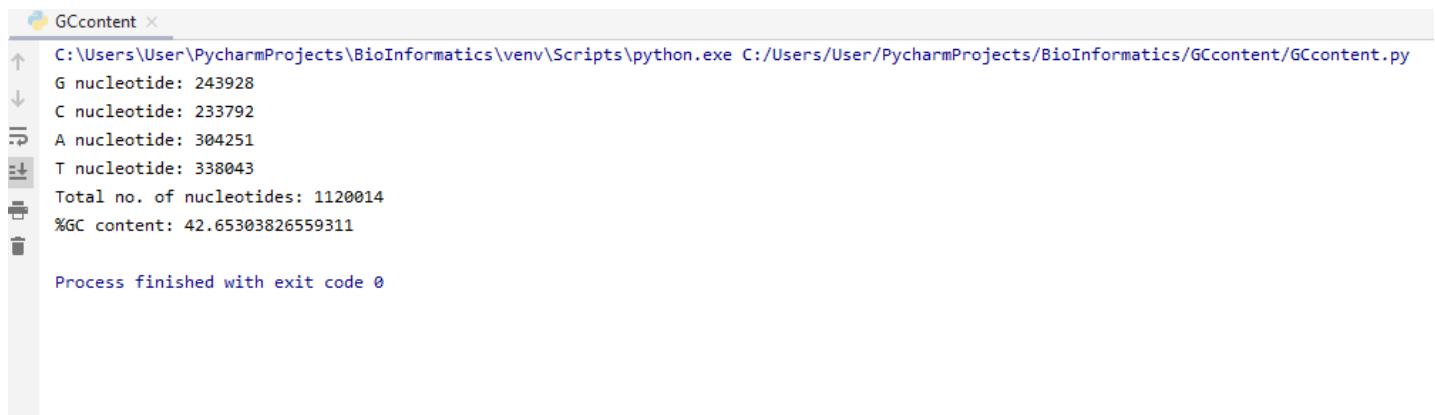
Bioinformatics Lab Manual (Lab2 - Lab9)

Ashish Kumar 2017B1A70854P

--> Lab-1 was list operations in python which was given as buffer.

Lab-2 (GC content)

1) Screenshot - 1 (Python script output)



```
GCcontent x
C:\Users\User\PycharmProjects\BioInformatics\venv\Scripts\python.exe C:/Users/User/PycharmProjects/BioInformatics/GCContent/GCcontent.py
G nucleotide: 243928
C nucleotide: 233792
A nucleotide: 304251
T nucleotide: 338043
Total no. of nucleotides: 1120014
%GC content: 42.65303826559311

Process finished with exit code 0
```

--> The GC content of the DNA is <50%. Hence, one can infer that the gene is not thermostable.

--> GC rich isochores typically include many protein coding regions within them. Existence of GC rich regions can be mapped to gene rich regions. In WWOX raw DNA sequence, about 42% of the sequence contains genes.

--> The average GC-content in human genomes ranges from 35% to 60% across 100-Kb fragments, with a mean of 41%. The GC-content of Yeast (*Saccharomyces cerevisiae*) is 38%, and that of another common model organism, thale cress (*Arabidopsis thaliana*), is 36%. Because of the nature of the genetic code, it is

virtually impossible for an organism to have a genome with a GC-content approaching either 0% or 100%. However, a species with an extremely low GC-content is *Plasmodium falciparum* (GC% = ~20%), and it is usually common to refer to such examples as being AT-rich instead of GC-poor.

Our sequence is close to the mean of the Human beings which is in line to the data that the sequence is indeed from human beings.

Lab - 3

1) Screenshot 1 (Raw Transcript)

```
1 CAGGCCCAUUGUAGGUUAAUCUCCUCCAUGUGUCCGUGGUUUUCAUCAGUCAGCUACCAGAUAGAGAA
2 CAUGUCUAGUCUGUUUCAUGCUGCUGAUAAAGACCUACCCGAGACUGGGAGAAAAAGAGGUUUAAUUG
3 GACUUACAAUAAAAGAGAUUUAAUAGACUUACUCCACAUGGCUGAGGAGGCCUCAGAAUCAUGGCG
4 GGAGGUGAAAGGCACUUCUCCAUUGGUGGCAAGAGAAAUGAGGACGAAGCAAAGUGGAAACCCCU
5 GAUAAACCCAUCAUAUCUCAUGAGACUUACUACUUCACGAGAAUAGCAUGGGAAAGACCAGCUCUCAUG
6 AUUCAGUUACCUCCUGCUGGGUCCCUCACAAACAUUGGGAAUUUUGGGAGAUAAAUAUCAAGUUGAGA
7 UUUGAGUGGAGACACAGCCTAACACAUAAUCAUGUGGUUUGGUUUCUGUCAUGUGUJUAGUUGCAGA
8 GGACAGUGGCUUCCACCUCCAUACGUGUGCCUGCUAGGACAUAAUCUUGUCCUUUCAUGGCUACAU
9 GUAUUCCAUUGGGUAUAUGUACCACAUUUUCCAUUCCAGCUAUCAUCGAUGGGCAUUGGUUUAUUCU
10 AUGUCUUUGUUAAUUGUGAAUAGUGCUGCAAUGAACAUUUGGUCCAUUGUACUUUAUAAUAGAAUGAUU
11 CUGUUACUUUGGUUAUACCAAGUAUUGGAAUUGCUGGUCAAUGGUUUCUGCCUCUAGGUUUG
12 AGGAUUUGCCACACUGUCUCCACAACGGUAAAACUAAUUAACACUCUCACAAAGGGGUAAAAGUGUUU
13 CUUUUCUCCACAACAUUCCAGCAUCUGCUGUUUUUUGACUUUUUAAAUGGCCUUCUGAUUGGCAU
14 GAGAUGGUACUCAAUGUGGUUUGAUUUGCAUAUCUCAAUGAACACUGAUGUUGAGCUCUUCUCAUA
15 UGUUUUCUUGGCUGCAUGAAUGUCUUGAGAAGUGUCUGUUCAUGUCCUUUGCCCACUUUUUAAAUGGGAUU
16 UUUAAAACAAACAUUUCUAAGGCUUUCUAGGGUAACAUUUGGAGAACAUUGCUCAUUCCCAAACAU
17 AGCCUUCUCAAACCGAUAAAACUUAUCCUACCAAGUUAAGUAGGGCUUGCCUGCUGAUAGAUCCUCA
18 UCAGAGCAAAGCUAAAGCUAAAGUUUAGGGGACAGAAACUGAGUAAACAGAGGAAGCCAGUAAGGAGAG
19 AGCAGAAAAGUAGGCCAUGGGAGGGAGACAACCUCAAGAGAAAAGUCCCAACACUUGCAUG
20 CUUACUUCUUUAACCUAUAGAACCCAAACAAAUAAGUUAUAGAAAGGGAGAAGAGGCCGGAGUGGUU
21 CUCAUGUUUGUAAAUCAGCACUUUGGGAGGCCAGGGCGGGUGGAUACAAGGUUCAGGAGUCAAGACCA
22 GCCUGGCCAAGAUGAUGGGAAACCCGUCUCAUAAAUAACAAAAAUAGCUGAGUGGGUGACCU
23 CACCUAUAGUCCAGCUACUUGGGAGGCUGAGGCAGGAGAAUCACUUGAACCCGGAGGUGGAGAUUGCA
24 GUGAGCCAAGAUUGUGCCACUGCUCUCCAGCCUGGAGACUCUGUCAUAGGUUACAGAAC
25 AAAAUGGAGAAGAAAUGUGAUCAUGAUAAACCAUUCGCCAUUACUGGAGCCUGGAGCAGCCGGCAGA
26 UUUCACAGGAUGUCUGACCUUGCAGAGGGCUGGCAAGGUGGGAGGCAUCCAGACCAUGGUUACAGAAC
27 UCAGUAAUCAGAGCCUGUGCUGGAGCCACCAAGCAUCCAGGACCAUGUGAACCUCCUGACAGGAGCA
28 GGCUCCUAAUGACUGAAAAGUAUCAGCAGAACUCCUGUGAUCAUAAGUGACUCAUUUAUCAACUUA
29 UUGUCAGUUAAGAUUCUCCACUGGGAGAAGACGGCAACUUCUGGAGACAUUUUGGUUGGUUACACAUAG
30 GAGAGGGCAUCUACUACUGCAUCUAGUGGUAGAGGCCAGAGAUGCUGCUAAACAUCCUACUGUACAG
31 CCCCACCACAAAGAAUGAUUCUGGGCUCCAGUUAUCGGUAUAGUGCAGGUUGAGAAACCCUGGUGCAGCAG
32 UGGGAGAAGCCUUAGAGGGAAUUCCAAAGUGGUAGGAUACGACAGGAUAGACCCUGAAGUUUAACUU
33 CUUCCAGGGGACCGUUACCCAGUUUGGGAACACCAAGGUUAUACUAGCUGUGGGCUGAUGCCAAGU
34 UACAGGUAAACAGAUGCCAGCAGAACGUCUCCGGAGAGGAAACCAAGAUAAACCUAAGCAGACAGUCCUUG
35 UGAAACUGAACAGAGUAAAAGUAUAGAAUUGGAGAACUUGGCUUACAGUJUUGACCUCAUUCUCCU
```

```
Run: raw_transcript ×
C:\Users\User\PycharmProjects\BioInformatics\venv\Scripts\python.exe C:/Users/User/PycharmPro...
Raw transcript stored in file raw_transcript.txt

Process finished with exit code 0
```

2) Screenshot - 2 (Reverse complement)

```
Run: reverse_compliment ×
C:\Users\User\PycharmProjects\BioInformatics\venv\Scripts\python.exe C:/Users/User/Py...
Open file reverse compliment.txt to get the reverse compliment

Process finished with exit code 0
```

```
1 |
2 |
3 TTTGGGAGGCTGAG
4 GAGGGCAGATCACAAGGTCAAGGAGTTCAAGACCAGCCTGCCAACATAGTGAACCCCGTCTACTTAA
5 AATATGAAAAATTAGCCAGGTGTGGTGTGTACCTGTAATCCCAGCTATTGGTAGGCTGAGGCAGGA
6 GAATCGCGTGAACCTGGGAGGTGGAGGTTGCAGTGAGCTGAGATCATGCCATTGCACTGCAGCCGGCG
7 ACAGTGGTAATGTAGCCTGCTTAGCACAGTGTGGCTCGGGATCATTGTCACAATGATGAAATTATTT
8 TCCTGAACATAAGCCTAGTCACCCAGAATCATTCTTCAGTGGTTATTATAAGTCAGTAAGTTGCT
9 GTGTGACAGGGAGCAGTTCAAGGTACTGTGCCAAGTTGTGAAGAAAACAGACAAATTTCATGGAGTTCT
10 TATAGTGGGTAGACAGAAAAAAATAGACACATAATGTGTACAATAATTAAAGTGTGGAGAAAAAAATA
11 GAGGGGAAGGGGAATAAGAAATGGGGTGACACCATGAAGACAGGGATAACGTGAGCAGAGACTAAAAATCA
12 GTGAGGAAGAAAGGCATGCTGGTTCTGGAAGGAAAGCATTTCAGCAGAGGAAACAGCAAATATGCACCC
13 CTTGAGTTGGGAGTGTGCTTGCATGTCAGGAAAGAGACTAGAGTGGCTCTAGTGGATAAG
14 TGCTGGAAGTTGAGATTAGAGAGGTAGCTGGAGACCAATTATAAGAGAAGTAGGGTAGAAAAAAATTAA
15 CCTTAGTGTGAGCTGAGAACGCTGTATCAAGTTGCTGAGTAACATGTAATCCTGAAACTTAGTGGCT
16 TAAAATAAACTTTACTATCTTTGCAATTGCTGTGAGTGAATGTGAGCTTCTTCCTCTGAATTGGC
17 ACAGCTGGGCTGGATAGTCTGGATGGCTCACTTACATGTCATAACTAACTAAGTGTGGATGCCGGTT
18 CTGGGAGTGAACACGCTCTCATCATCCAGCAGGCTAGCCAGGAGCCGTTCCCTATGATGGAAGTTG
19 CAGGGTTCAAAGATCAACAGGAAAGGCAAGCCCCAGTCAAAGCAGACTTCAAATTGGCCAAGGCCAG
20 TCACATGCCAAGTCAGAGGTCAATGGGAGGGGTGATCCCTCCAGGTGAGATTCTGAGAGTGGAGT
21 TCTGACGGCTGTTTGCAGAACATAATGTCAGGAAAGTGAATCTGAAACTAAGTGTGAGCTTCTTCCTGTA
22 TCCTGTGATAATAAGGATTACGTAAGGAAGACTATAGCCAGATGGGAGGATGTGGCAAGGCTATGGCCA
23 CTGGAGGGCAATGGTGTGTTGACTGCTCTGCACCCACAGATGCAAGGGCAGAGAAGTCAAGAC
24 ATTGTCTGGCCTGAAGCTGCCATGATCAGGAGGGGAACAATGACACCCAGCCTCCACACCTTCTGGTA
25 TTTCAGCAGAACGATTTCTCCGTTGCCAGAGTGGACTTCTCATGTCAGCTGCTGGAAACGGCTGGAA
26 ATCATGGTGTCCATGAGATGGGATGCCAGATAAGAGCTTCTGCACAGCTCCATTCTCAAACATAGAGA
27 TACCCGCTAATCACTAAGTACTTGCTAATTGCAAGGGAGCAGACCATTGAGGACTAAAATAACTCACTGT
28 AAGTAAGTAATAACAAAACATCTTAAGGTGAATGAAACAAATAGTCATTTCCTTTCTTTCTTAA
29 AGAAAATATGTCGGCATGCTAATCGCGTGAAGACAACATTGGCCCACGGACAGCCACAAAGCACAACAA
30 AACACAAACTCACTCAGAGAAGAAACAAGAATCGTGTGTTCTGGCATCTACGAGAAGCGCTTATTAGA
31 AGCGCTTATTTCCTCTAGATTCTTTCTAAAGGAAACAAACTGAAACAAACATT
```

3) Screenshot - 3(Number of start codons)

The screenshot shows the PyCharm interface with a run configuration titled 'no_of_start_codons (1)'. The terminal window displays the following output:

```
C:\Users\User\PycharmProjects\BioInformatics\venv\Scripts\python.exe
Count of AUG in raw transcript: 20520
Process finished with exit code 0
```

Below the terminal, a black box highlights the following additional output:

```
Count of AUG in raw transcript: 20520
Count of UAA in raw transcript: 19943
Count of UGA in raw transcript: 22358
Count of UAG in raw transcript: 14163
```

Interpretations

--> Uracil (U) replaces Thymine (T) in the raw transcript of the DNA. The Watson Crick base pair rule is -

$$A=U \quad G=C$$

--> Reverse compliment is obtained by the reversing the compliment strand of the DNA. It can be used to get the other 3 ORFS of the sequence.

--> The reverse compliment when translated gives the same sequence as the original strand.

--> Also, we use reverse-complement to convert nucleotide sequence to protein sequence instead of reversed sequence.

--> There are about 20520 start codons and 56464 (UAA+UGA+UAG) stop codons.

--> Hence, one can have a total of 20520*56464 translated both in forward as well as in reverse direction.

LABORATORY ASSIGNMENT – 4

ORF Finder analysis:

1. Find number of ORFs
2. Analyze and find the largest ORF
3. Get Six frame transition

1) Screenshot - 1 ,Number of ORFs

When MinLen(ORF) set to 75

Enter Query Sequence

Enter accession number, gi, or nucleotide sequence in FASTA format:

```
>NC_000016.10:78099654-79212667 Homo sapiens chromosome 16, GRCh38.p13 Primary
GCAGTGCAGCGCTGAGCGGTGGGCCCGACGCCGGGGCTCTCGTTGGAGCGGGAGTGAGTCCTG
AGCAGTGGACCCCGCACGGCGATAAGGGGGCAAGGTGCTTCCACAGTCAGCCATGGCAGCGCTGGC
TACGGGGGCTGAGCACACGGAGCATGAAGGAGAACGCTGGGGCTGGAGGAGAGAACCAAGG
ACGGCTGGTTACTACGCAAGTAAGGGGGCGAGTGGGCCGGAGCGCACCTGGGACCTTGACAG
CCCAGACGCCACCTCGCGGGAGGACGCCACTCCAGCGCAGCGCTGGGTGAAAGTAAAGTAA
CTGTTAAGGAGCTTCAGGGAAAGGGTCAAGGGTTCCAGTAGGGGCCGCCCTTGTTGGGCTGGG
TCCAGCGGGGTACCTGGTGGCTCCGGCCGCCCTGCTGTTCAAGATGAGCACTGCGCGGGCG
GCGAGGCAAAGCGGCCATCCCCGCCAAAAATAAGATGTTAAAAGCGCACATGCTCAAGCTCCC
```

From: 1 To: 1110000

Choose Search Parameters

Minimal ORF length (nt): 75

Genetic code: 1. Standard

ORF start codon to use:

"ATG" only

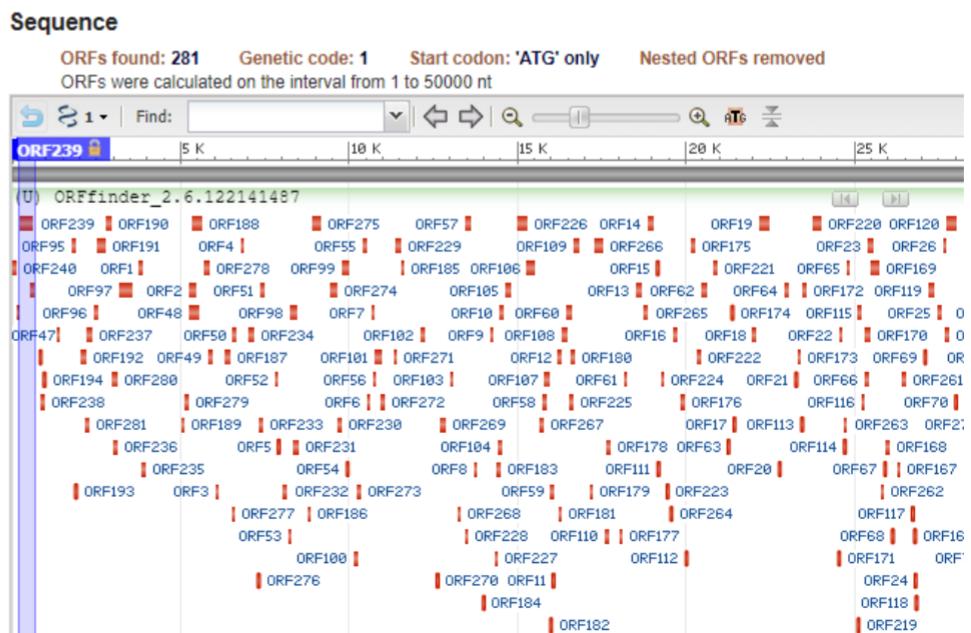
"ATG" and alternative initiation codons

Any sense codon

Ignore nested ORFs:

Start Search / Clear

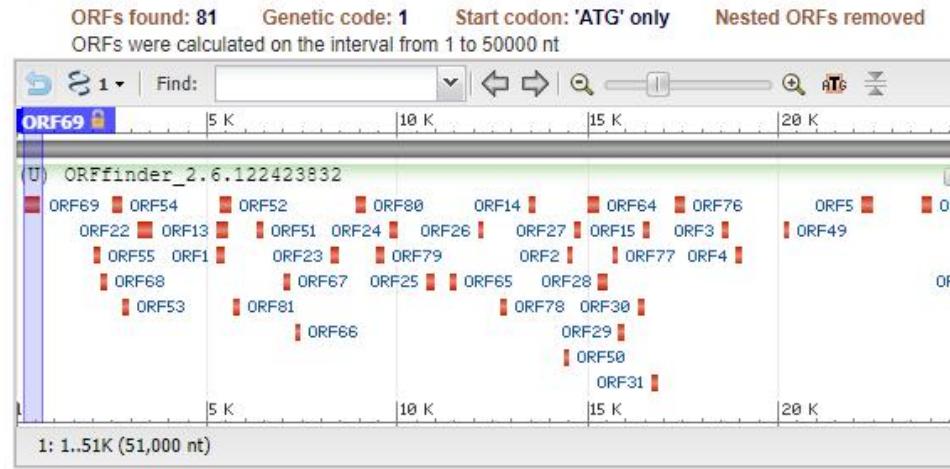
No. of
ORFs = 281



MinLen(ORF) set to 150

No. of
ORFs = 81

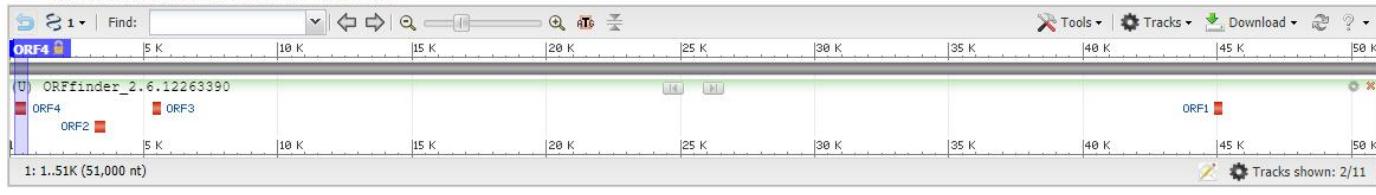
Sequence



MinLen(ORF) set to 300
No. of ORFs = 4

Sequence

ORFs found: 4 Genetic code: 1 Start codon: 'ATG' only Nested ORFs removed
ORFs were calculated on the interval from 1 to 50000 nt



Six-frame translation...

```
>1c1|ORF4 CDS
ATGACTGAGCAATAACCGBGTCTGGACAAAAACAAAAACAAAACACAAACAAAAA
TCCCTATCCCTGCAACCCAGAGCCTGCAGGGGGAGCTGAGCATGATGCCGCT
TTTTAAACATCTTTTTTGCGGGGGATGAGGGCCCTTGTGCCCTCGC
CGCCCGCGCAGTGTGTCATCTCTGAACAGCAAGGGCGCGCCGGAAAGCC
ACCAAGGTGACCCCCGCTGGACCCGAAGGGCCCACCAAGGGGGCGCGCCCCCTA
CTGGGAACCCCTGGACCCCTTTCCCTGAAGCTCTTAACAGTTACTTCA
TTTGCAACGCGACCGCCTGCGCTGGAGGTGCCGCTCTCCCGCGCAGGTGG
CGTCCGCTGGGCTGTGCAAGGGTCCAGGGTGCCTGCCGCCCCACTGCGGC
CCCCCTTACTTGGCGTAG
```

Mark subset...		Marked: 0		Download marked set		as	Protein FASTA ▾
Label	Strand	Frame	Start	Stop	Length (nt aa)		
ORF4	-	2	640	224	417 138		
ORF2	+	3	3192	3560	369 122		
ORF1	+	1	44890	45210	321 106		
ORF3	-	1	5639	5337	303 100		

Maximum ORF length-

--> ORF4; Strand (-) downstream; Length = 417 nt or 138 aa.

Interpretations

--> There are many ORFs in both the DNA strand and they are of varying lengths.

--> A protein sequence may consist of many ORF frames by the mechanism of alternative splicing.

--> Each orf starts from the start codon and ends at the stop codon.

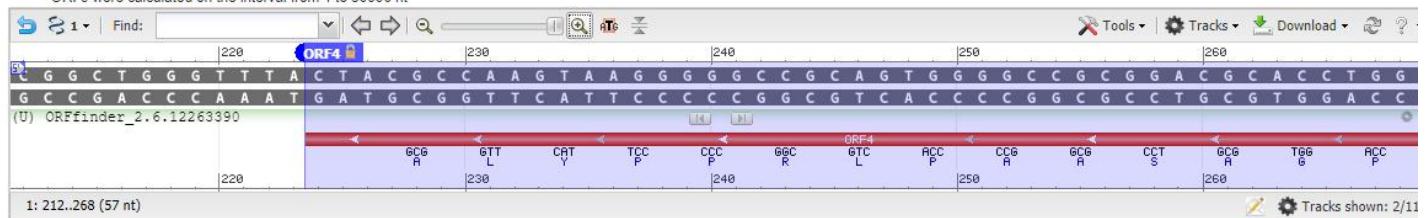
--> ORF is the sequence that is translated by the ribosomes.

2) Screenshot - 2 (ORF nucleotide level zoomed)

--> ORF nucleotide level zoom

Sequence

ORFs found: 4 Genetic code: 1 Start codon: 'ATG' only Nested ORFs removed
ORFs were calculated on the interval from 1 to 50000 nt



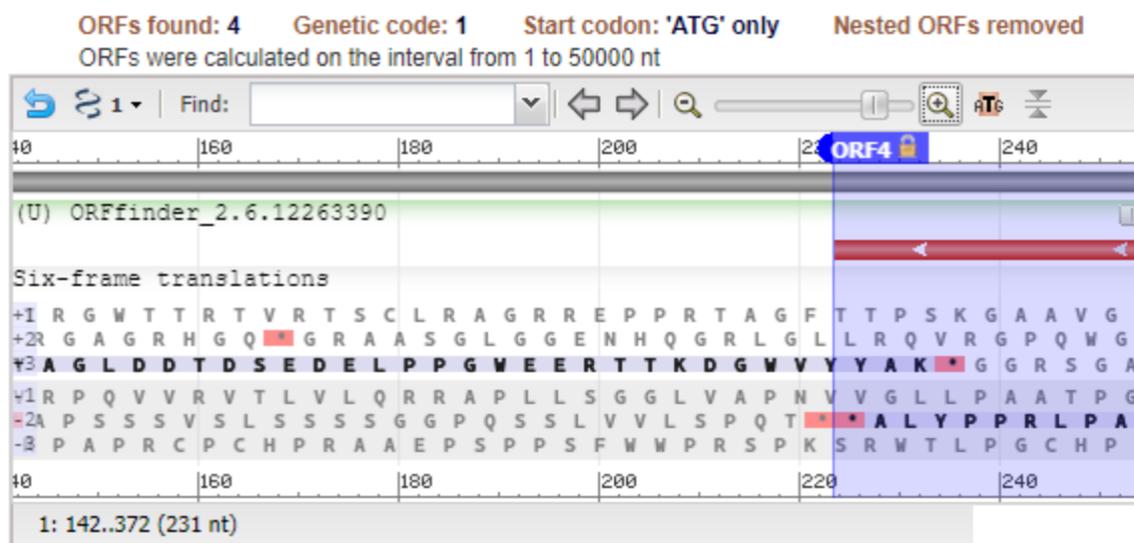
Six-frame translation

--> ORF nucleotide level zoom gives us the accurate depiction of the stop codon out of the 3 stop codons present at the end of each Orf. In the figure above UAG is the stop codon.

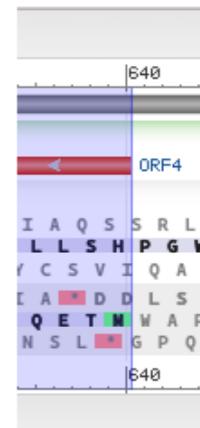
3) Screenshot - 3 (ORF 6 frame transition)

--> ORF Six Frame Transition (Stop codon in Red):

Sequence



--> ORF Six Frame Transition (Start codon in Green):



Interpretations

--> There are in total 6 possible ORFs for each ORF. 3 in the forward strand and 3 in the reverse strand.

The 3 ORFs in 1 strand are obtained by considering the start from all the 3 frames in the nucleic acid code.

LAB ASSIGNMENT - 5

BIO F242
BIOINFORMATICS

NCBI WWOX gene screenshot (Genome data viewer)

Interpretations

--> The gene WWOX in humans is present in the Chromosome 16.

--> The gene location on chromosome 16 :
78,099,654 - 79,212,667

--> The genomic sequence of WWOX is 1,113,014 base pairs in length.

--> But inspite of this, according to NCBI 10 exonic regions only exist for a total base pair length of about 20,000.

--> Most of the genomic sequence is composed of introns which are of major advantage for our genome.

--> There are 3 isoforms of this gene. Gene isoforms are mRNAs that are produced from the same locus but are different in their transcription start sites, protein coding DNA sequences and/or untranslated regions, potentially altering gene function.

--> There are 4 transcript variants of the gene. Transcript variants are mRNA produced after alternative splicing and assembling.

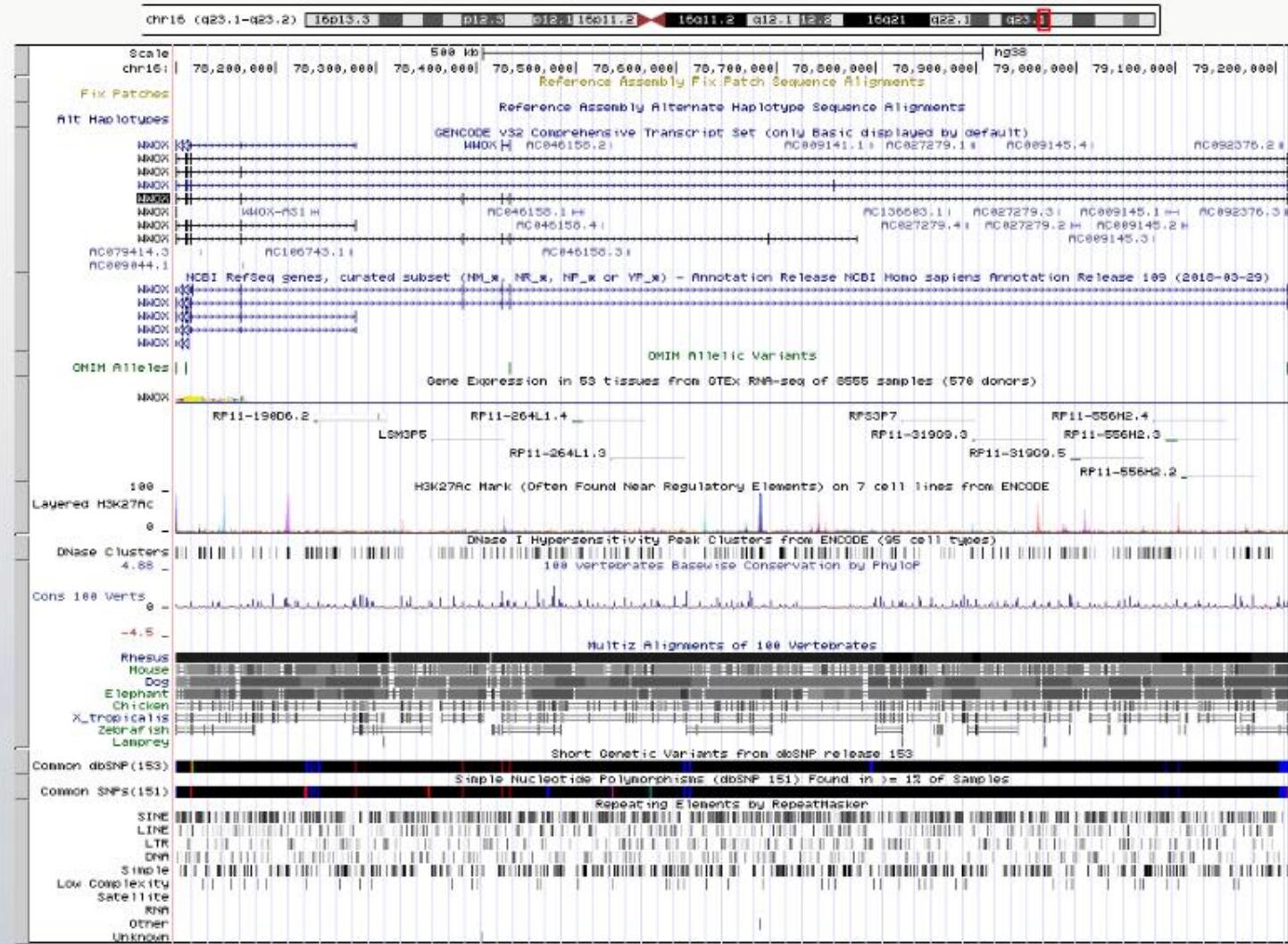
UCSC Interactive Genome Browser

UCSC Genome Browser on Human Dec. 2013 (GRCh38/hg38) Assembly

move <<< << < > >> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x 100x

chr16:78,099,654-79,212,667 1,113,014 bp. enter position, gene symbol, HGVS or search terms

go



HGNC Symbol Report

Symbol report for WWOX

[Report](#)[HGCP homology predictions](#)[Stable symbol](#)

HGNC data for WWOX

Approved symbol	WWOX
Approved name	WW domain containing oxidoreductase
Locus type	gene with protein product
HGNC ID	HGNC:12799
Symbol status	Approved
Previous names	" WW domain-containing oxidoreductase "
Alias symbols	FOR; WOX1; SDR41C1
Alias names	" short chain dehydrogenase/reductase family 41C, member 1 "
Chromosomal location	16q23.1-q23.2
Gene groups	Short chain dehydrogenase/reductase superfamily

Gene resources

Ensembl	ENSG00000186153	Curated
	Ensembl region in detail	
	Ensembl gene sequence	
UCSC	uc002ffk.4	

Nucleotide resources

INSDC	AF187015	Curated
	ENA, GenBank, DDBJ	
CCDS	CCDS42196	Curated

Protein resources

UniProt/Swiss-Prot	Q9NZC7
	InterPro, PDB, Reactome
Enzyme Commission	1.1.1-
	ExPasy, BRENDA, KEGG, IntEnz, IUBMB Enzyme Nomenclature

Orthologs from selected species

Bos taurus	WWOX (VGNC:36977)	
Macaca mulatta	WWOX (VGNC:78805)	
Pan troglodytes	WWOX (VGNC:13810)	
Sus scrofa	WWOX (VGNC:94988)	

Canis familiaris	WWOX (VGNC:48442)	
Mus musculus	Wwox (MGI:1931237)	Curated
Rattus norvegicus	Wwox (RGD:1309927)	

Interpretation

--> The gene has orthologs in various species such as Bos taurus, Macaca mulatta, Pan troglodytes, Sus scrofa. Orthologs are genes that have diverged in the common ancestor.

--> HGNC is a compilation of many different resources for a given gene such as

- OMIM
- Ensembl
- NCBI

GenScan

- As my gene length was greater than 1 million bps, I had to split the sequence file into two and run GenScan individually on them.
- Number of Transcription factor binding sites – (Source – GeneCards) = $357+10+37+65+44 = 513$ TFs

GeneHancer (GH) Regulatory Elements for WWOX Gene
Promoters and enhancers for WWOX Gene 

Filter: (269 results) See all 269 » 

Structural Variation Interpretation using GeneHancer
TGex - GeneCards-powered NGS Analysis & Reporting [LEARN MORE](#)

GeneHancer (GH) Identifier	GH Type	GH Score	GH Sources	Gene Association Score	Total Score	TSS distance (kb)	Number of Genes Away	Size (kb)	Transcription Factor Binding Sites	Gene Targets
GH16J078098	Promoter/Enhancer	2.1★	EPDnew, Ensembl, ENCODE, CraniofacialAtlas	505.4★	1054.99	+0.5	0	2.6	357 TFs  ZBTB40 ZNF217 ...	5 genes  WWOX ...
GH16J078097	Enhancer	0.5	ENCODE	500.7★	271.42	-1.6	0	0.0	10 TFs  NFYB BCL11A ...	3 genes  NONHSAG020051.2 ...
GH16J078122	Enhancer	0.9★	Ensembl, ENCODE	19.5★	17.58	+23.7	4	1.4	37 TFs  ZNF217 CTCF ...	3 genes  ENSG00000276007 ...
GH16J078133	Enhancer	0.8	ENCODE	17.6★	14.46	+34.7	6	1.2	65 TFs  ZNF217 NCOR1 ...	3 genes  WWOX ...
GH16J077961	Enhancer	0.9★	Ensembl, ENCODE	15.1★	14.05	-134.3	9	8.1	44 TFs  SP1 TEAD4 ...	7 genes  CLEC3A ...

* - Elite GeneHancer and/or Elite GeneHancer-gene association [Download GeneHancer data from 2017 publication](#) | [Request up-to-date GeneHancer data \(full dataset\)](#)

- Number of promoters predicted by Genscan = 44
- Number of poly-A tails predicted = 46
- Total number of internal genes predicted = 47
- Number of exons predicted = 230 (Terminal Exons + Internal Exons + Single Exon Genes)

- o The Exons were in Internal exons (1 to n)-....-
- Terminal exon (1)**
- > **Single Exon Gene**
- Number of enhancers = 5

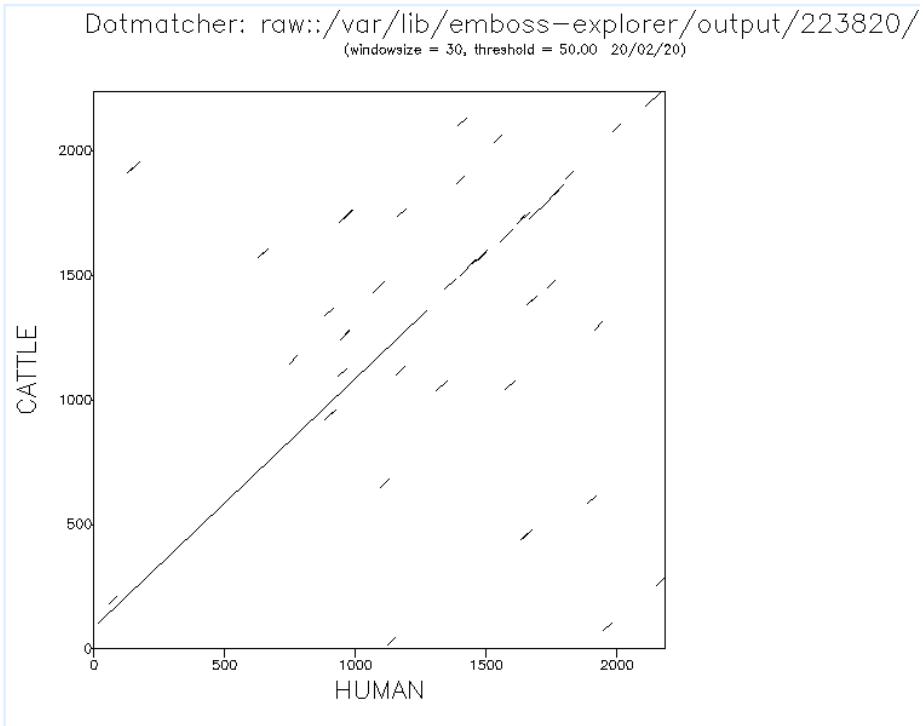
Lab - 6 + Lab -7

Needleman Wunsch Local alignment

Smith Waterman Global alignment

Needleman Wunsch for gene orthologs

Human Cattle :Score 9212.5



```

#####
# Program: needle
# Rundate: Thu 20 Feb 2020 07:10:05
# Commandline: needle
#   -auto
#   -stdout
#   -asequence emboss_needle-I20200220-071002-0655-35472743-p1m.asequence
#   -bsequence emboss_needle-I20200220-071002-0655-35472743-plm.bsequence
#   -datafile EBL0SUM62
#   -gapopen 10.0
#   -gapextend 0.5
#   -endopen 10.0
#   -endextend 0.5
#   -aformat3 pair
#   -sprotein1
#   -sprotein2
# Align_format: pair
# Report_file: stdout
#####

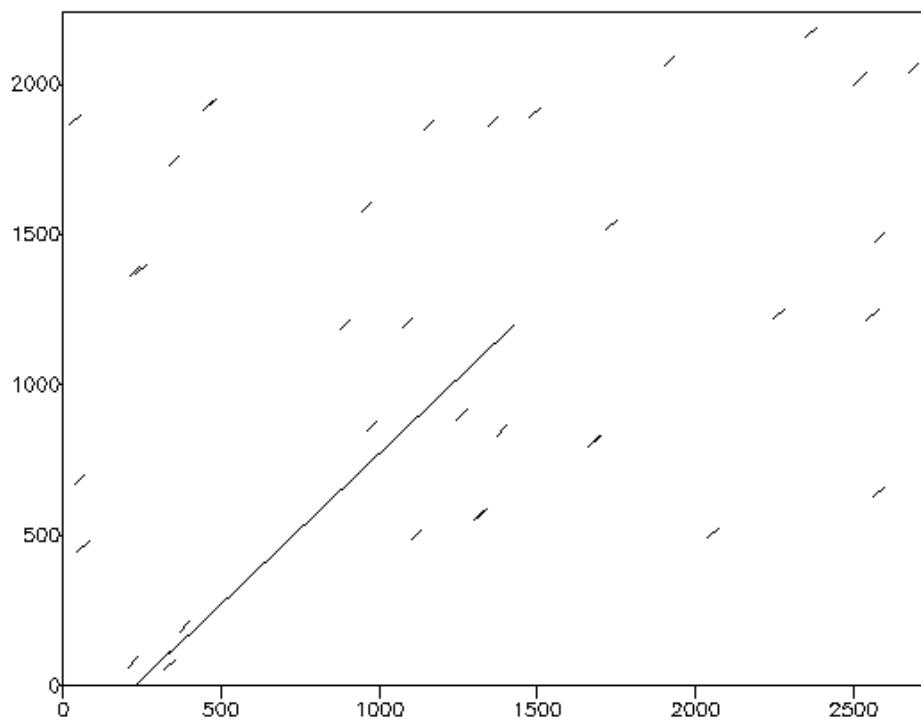
#####
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EBL0SUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 2351
# Identity: 1620/2351 (68.9%)
# Similarity: 1620/2351 (68.9%)
# Gaps: 275/2351 (11.7%)
# Score: 9212.5
#
#
#####

EMBOSS_001      1 GCAGTGCGCAGGCGTGAGCGGTGGGCCCGACGCCGCAGGGTCTCGTT 50
EMBOSS_001      1 ----- 0
EMBOSS_001      51 GGAGCGGGAGTGAGTCCCTGAGCGAGTGGACCCGGCAGCGGGCGATAGGG 100
                           ||.||..||| ||| ||||.|||||
EMBOSS_001      1 -----TGGGGGGCTGG--CGG--GCGAGCGA---- 22
EMBOSS_001      101 GGGCCAGGTGCCCTCACAGTCAGCCATGGCAGCGCTGCGCTACGCGGGC 150
                           || ||.|||.|| | ||||| |||||.||| .||| .
EMBOSS_001      23 ---CC---GCTTACCCCG-CAGCCATGGCAGGCCCTGCGCTACGCTGGCC 64

```

Human Chicken : score 9607

Dotmatcher: raw:::/var/lib/emboss-explorer/output/016
(windowsize = 30, threshold = 50.00 20/02/20)



```

#####
# Program: needle
# Rundate: Thu 20 Feb 2020 07:08:12
# Commandline: needle
#   -auto
#   -stdout
#   -asequence emboss_needle-I20200220-070810-0159-58720721-p2m.asequence
#   -bsequence emboss_needle-I20200220-070810-0159-58720721-p2m.bsequence
#   -datafile EBLOSUM62
#   -gapopen 10.0
#   -gapextend 0.5
#   -endopen 10.0
#   -endextend 0.5
#   -aformat3 pair
#   -sprotein1
#   -sprotein2
# Align_format: pair
# Report_file: stdout
#####

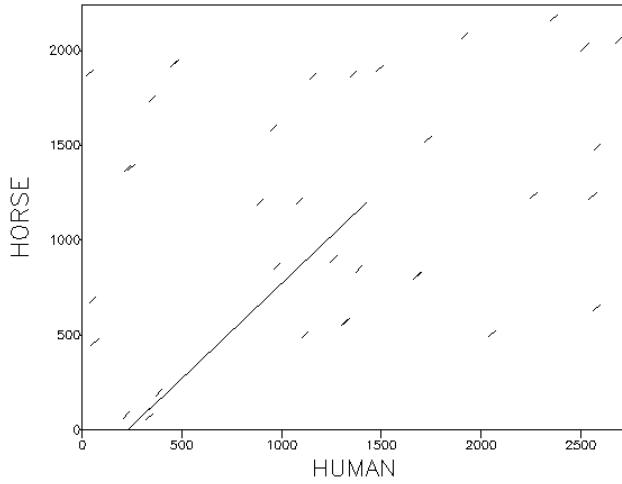
=====
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 2430
# Identity: 1714/2430 (70.5%)
# Similarity: 1714/2430 (70.5%)
# Gaps: 281/2430 (11.6%)
# Score: 9607.0
#
#
=====

EMBOSS_001      1 -----          0
EMBOSS_001      1 GGTCGATTCTTCCCGCGTCCCCACCGCTTTCTAGGCTCCGCCCC 50
EMBOSS_001      1 -----          0

```

Human Horse: Score 6565

Dotmatcher: raw:::/var/lib/emboss-explorer/output/3915:
(windowsize = 30, threshold = 50.00 20/02/20)



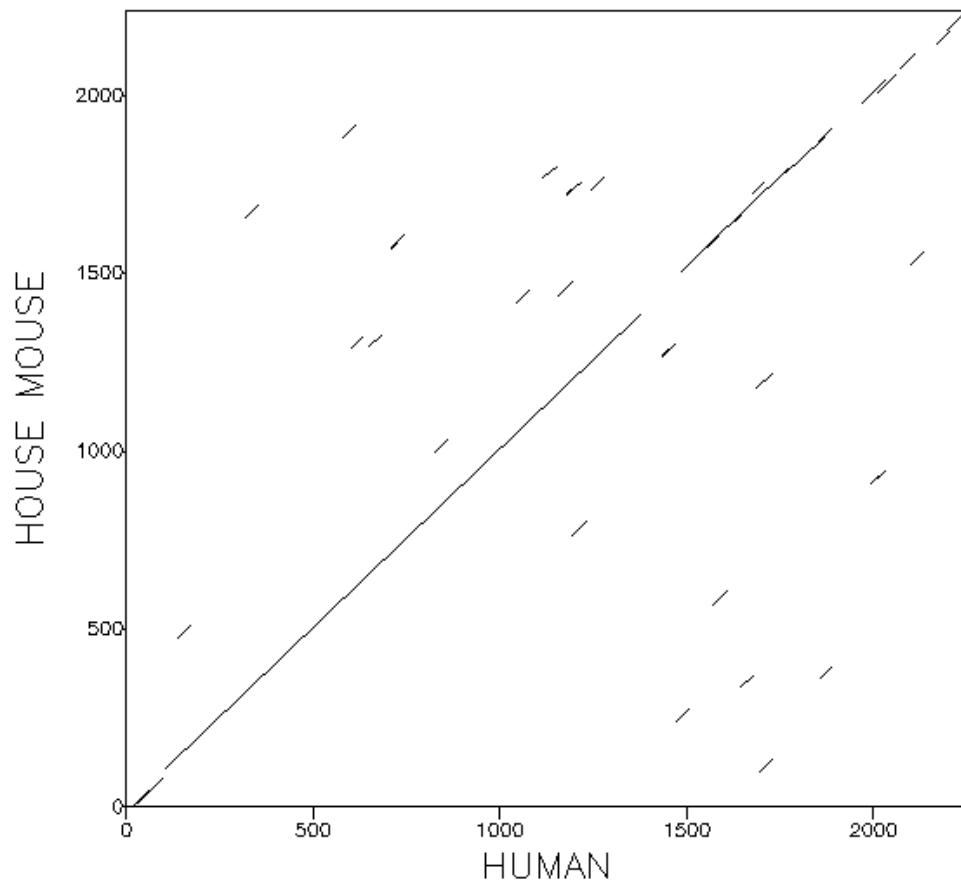
```
#####
# Program: needle
# Rundate: Thu 20 Feb 2020 07:07:29
# Commandline: needle
#   -auto
#   -stdout
#   -asequence emboss_needle-I20200220-070728-0031-866168-p2m.asequence
#   -bsequence emboss_needle-I20200220-070728-0031-866168-p2m.bsequence
#   -datafile EBLOSUM62
#   -gapopen 10.0
#   -gapextend 0.5
#   -endopen 10.0
#   -endextend 0.5
#   -aformat3 pair
#   -sprotein1
#   -sprotein2
# Align_format: pair
# Report_file: stdout
#####

=====
#
# Aligned_sequences: 2
# 1: EMBOS_001
# 2: EMBOS_001
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 2264
# Identity: 1293/2264 (57.1%)
# Similarity: 1293/2264 (57.1%)
# Gaps: 373/2264 (16.5%)
# Score: 6565.0
#
#
=====

EMBOSS_001      1 GCAGTG---CGCAGGCGTGAGCGGTGGGCCCGACGCGCGCGGGCTCG      47
EMBOSS_001      .||| .|||..|||.||| .||| .||| ..|||...||| .||| .|||
                           ATTCATAACGGCTGTGGCGGCTG CCGAATTAACAAACATC      20
```

Human Mouse: Score 9745

Dotmatcher: raw:::/var/lib/emboss-explorer/output/36297
(windowsize = 30, threshold = 50.00 20/02/20)



```

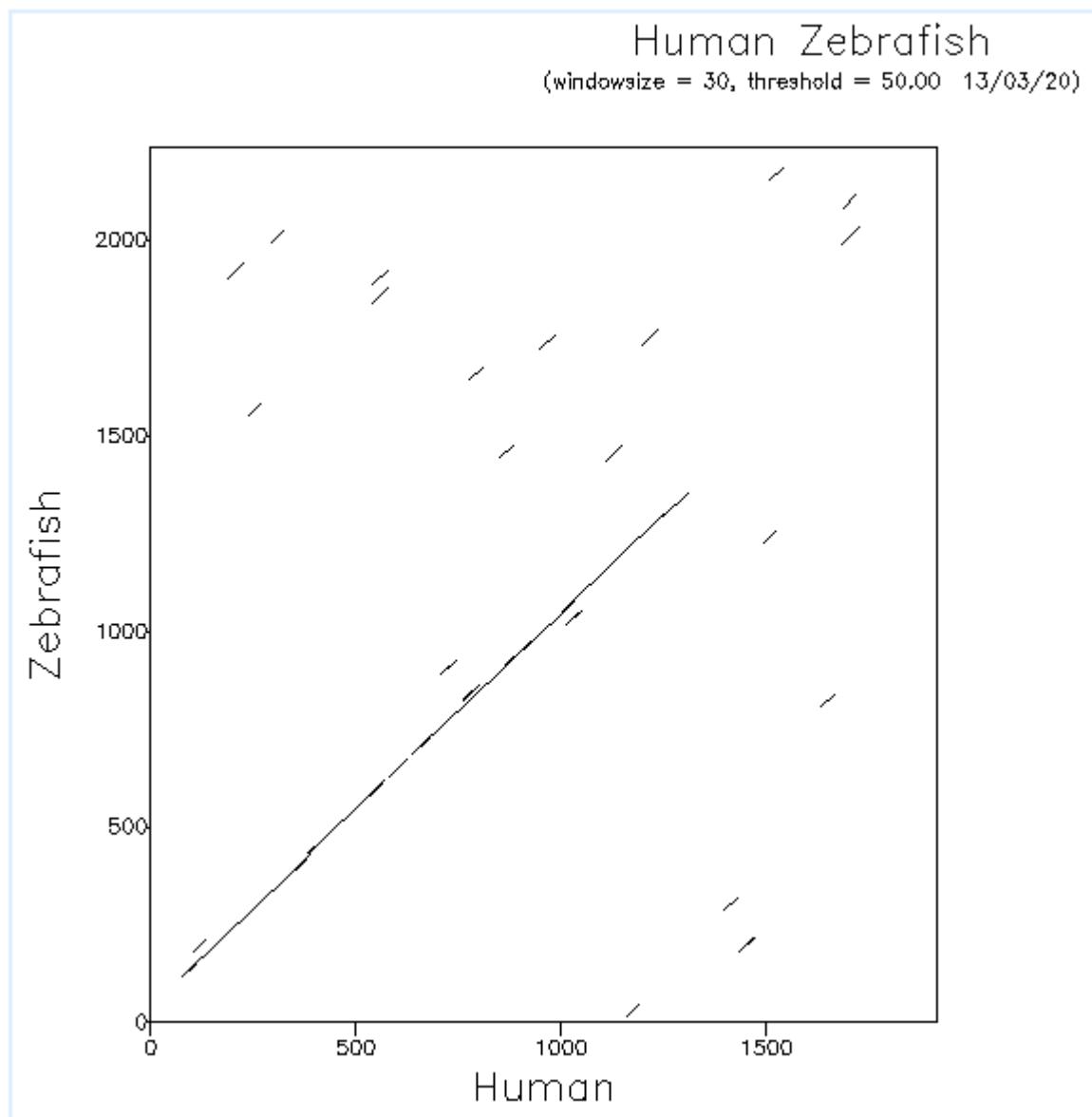
#####
# Program: needle
# Rundate: Thu 20 Feb 2020 07:20:34
# Commandline: needle
#   -auto
#   -stdout
#   -asequence emboss_needle-I20200220-072030-0201-49943196-p2m.asequence
#   -bsequence emboss_needle-I20200220-072030-0201-49943196-p2m.bsequence
#   -datafile EBLOSUM62
#   -gapopen 10.0
#   -gapextend 0.5
#   -endopen 10.0
#   -endextend 0.5
#   -aformat3 pair
#   -sprotein1
#   -sprotein2
# Align_format: pair
# Report_file: stdout
#####

=====
#
# Aligned_sequences: 2
# 1: EMBOS_001
# 2: EMBOS_001
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 2342
# Identity: 1732/2342 (74.0%)
# Similarity: 1732/2342 (74.0%)
# Gaps: 192/2342 ( 8.2%)
# Score: 9745.5
#
#
=====

EMBOSS_001      1 -----GCAGTGCAGGGCGTGAGCGGTC-GGGCCCC---- 30
EMBOSS_001      1 GGACCGGCTGGGCGAGCAGTGCAGGGCGAGCGGTCTGGTCCCCCTCCC 50
EMBOSS_001      31 -GACGCGCGCGGGTCTCGTTGGAGCGGGAGTGAGTTCTGAGCGAGTGG 90

```

Human Zebrafish **(Score = 3531.0)**



```
#####
# Program: needle
# Rundate: Fri 13 Mar 2020 17:14:37
# Commandline: needle
#   -auto
#   -stdout
#   -asequence emboss_needle-I20200313-171434-0893-21401837-p2m.asequence
#   -bsequence emboss_needle-I20200313-171434-0893-21401837-p2m.bsequence
#   -datafile EDNAFULL
#   -gapopen 10.0
#   -gapextend 0.5
#   -endopen 10.0
#   -endextend 0.5
#   -aformat3 pair
#   -snucleotide1
#   -snucleotide2
# Align_format: pair
# Report_file: stdout
#####

=====
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EDNAFULL
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 2398
# Identity: 1332/2398 (55.5%)
# Similarity: 1332/2398 (55.5%)
# Gaps: 641/2398 (26.7%)
# Score: 3531.0
#
#
=====

EMBOSS_001      1 GCAGTG---CGCAGGCGTGAGCGGTGGGCC---CCGACGCGCGCGGGTC      44
                  |.||| ||||..|.||||.||||.||| |..|||.||| |
EMBOSS_001      1 --ATTGTAACGCCTGTGTGCGCGCT---GCCAAACAAACAC-----C      37
```

Summary and Interpretations

Organism	Needleman - Wunsch Score
Human Zebrafish	3531.0
Human Mouse	9745
Human Horse	6565
Human Chicken	9607
Human Cattle	9212.5

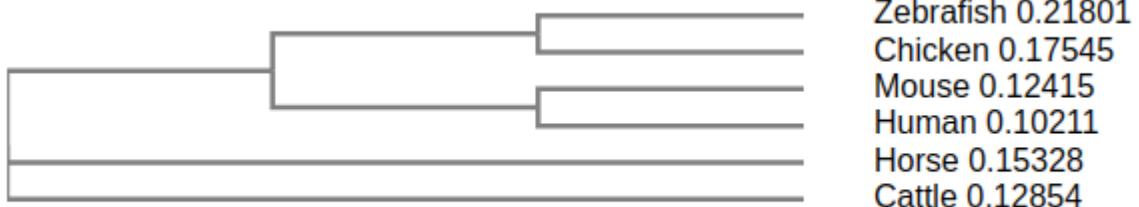
→ EMBOSSTool Needle reads two input sequences and writes their optimal global sequence alignment to file. It uses the Needleman-Wunsch alignment algorithm to find the optimum alignment (including gaps) of two sequences along their entire length.

→ The scoring matrix used is EBLOSUM62

-Default Gap opening score = -10.0

-Gap extension score = -0.5

→ Needleman Wunsch algorithm - The algorithm assigns a score to every possible alignment, and the purpose of the algorithm is to find all possible alignments having the highest score.



- Human and Mouse have the highest alignment score. This is in line with the phylogenetic tree where human and mouse are most closely related.
- Human and Zebrafish are least closely related.
- A diagonal in the dot-plot tells us that the 2 sequences are identical throughout the length of the diagonal.
- A dot-plot has a dot at each position where the 2 sequences have identical nt/aa.

Smith Waterman alignment for gene orthologs

Human - Mouse

(Score = 6548.0)

Human - Zebrafish

(Score = 6565.0)

Human - Cattle

(Score = 9212.5)

```
#####
# Program: water
# Rundate: Fri 13 Mar 2020 17:18:14
# Commandline: water
#   -auto
#   -stdout
#   -asequence emboss_water-I20200313-171812-0726-70337462-p2m.asequence
#   -bsequence emboss_water-I20200313-171812-0726-70337462-p2m.bsequence
#   -datafile EBLOSUM62
#   -gapopen 10.0
#   -gapextend 0.5
#   -aformat3 pair
#   -sprotein1
#   -sprotein2
# Align_format: pair
# Report_file: stdout
#####

=====
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 2268
# Identity: 1620/2268 (71.4%)
# Similarity: 1620/2268 (71.4%)
# Gaps: 192/2268 ( 8.5%)
# Score: 9212.5
#
#
=====

EMBOSS_001      69 TGAGCGAGTGGACCCGGCAGCGGGCGATAGGGGGGCCAGGTGCCTCCACA    118
                  ||.||.|||  |||  |||.|||  ||  ||.||.|.
EMBOSS_001      1 TGGGGGGCTGG---CGG--GCGAGCGA-----CC---GCTTACCCC    33
```

(Human - Horse)

(Score = 8680.5)

```
#####
# Program: needle
# Rundate: Fri 13 Mar 2020 17:22:45
# Commandline: needle
#   -auto
#   -stdout
#   -asequence emboss_needle-I20200313-172243-0110-90738544-p2m.asequence
#   -bsequence emboss_needle-I20200313-172243-0110-90738544-p2m.bsequence
#   -gapopen 10.0
#   -gapextend 0.5
#   -endopen 10.0
#   -endextend 0.5
#   -aformat3 pair
#   -sprotein1
#   -sprotein2
# Align_format: pair
# Report_file: stdout
#####

=====
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 2818
# Identity: 1669/2818 (59.2%)
# Similarity: 1669/2818 (59.2%)
# Gaps: 667/2818 (23.7%)
# Score: 8680.5
#
#
=====

EMBOSS_001      1 ----- 0
EMBOSS_001      1 CTGGACGTCCACGCCCCTTCCCTGCGGCTGGGAGGTTGACGAAGCTCAGC 50
```

(Human - Chicken)

(Score= 8023.0)

```
#####
# Program: needle
# Rundate: Fri 13 Mar 2020 17:25:35
# Commandline: needle
#   -auto
#   -stdout
#   -asequence emboss_needle-I20200313-172532-0045-71378118-p2m.asequence
#   -bsequence emboss_needle-I20200313-172532-0045-71378118-p2m.bsequence
#   -datafile EBLOSUM62
#   -gapopen 10.0
#   -gapextend 0.5
#   -endopen 10.0
#   -endextend 0.5
#   -aformat3 pair
#   -sprotein1
#   -sprotein2
# Align_format: pair
# Report_file: stdout
#####

=====
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 2364
# Identity: 1536/2364 (65.0%)
# Similarity: 1536/2364 (65.0%)
# Gaps: 301/2364 (12.7%)
# Score: 8023.0
#
#
=====

EMBOSS_001      1 GCAGTGCGCAGGCGTGAGCGGTGGGGCCCCGACGCAC-GCGGGTCTCGTT    49
                  |.|.|||.|..|..|||.||.||| |
EMBOSS_001      1 -----GGGGCGGCCGCGGTCCGGCAGGCAGG--CT---- 31
```

Smith Waterman Protein

Human - Cattle **(Score = 46.0)**

```
#####
# Program: water
# Rundate: Fri 13 Mar 2020 17:33:06
# Commandline: water
#   -auto
#   -stdout
#   -asequence emboss_water-I20200313-173305-0747-55859959-p2m.asequence
#   -bsequence emboss_water-I20200313-173305-0747-55859959-p2m.bsequence
#   -datafile EBLOSUM62
#   -gapopen 10.0
#   -gapextend 0.5
#   -aformat3 pair
#   -sprotein1
#   -sprotein2
# Align_format: pair
# Report_file: stdout
#####

=====
#
# Aligned_sequences: 2
# 1: EMBOS_001
# 2: EMBOS_001
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 142
# Identity:      37/142 (26.1%)
# Similarity:    52/142 (36.6%)
# Gaps:          33/142 (23.2%)
# Score:        46.0
#
#
=====

EMBOS_001      43 QWEHPKTGKRKRVA-----GDL---PYGWEQETDENGQVFFVDHINKRTT      84
                  :|.:| :||...||     .|: |.....|....:||.|....|
EMBOS_001      144 EWDAP-SGKTVSVAVKCLKPDVLSQPEAMDDFIREVNAMHSLDHRNLIRL      192
EMBOS_001      85 Y---LDPRLAFTVDDNPTKPTTRQRYDGS-TTAMEILQGRDFTGKV----      126
                  | |.||       |.|...|....|| ...:||....|.:|
```

(Human - Mouse)

(Score = 2071.0)

```

#####
# Program: water
# Rundate: Fri 13 Mar 2020 17:34:59
# Commandline: water
# -auto
# -stdout
# -asequence emboss_water-I20200313-173457-0142-45562810-p2m.asequence
# -bsequence emboss_water-I20200313-173457-0142-45562810-p2m.bsequence
# -datafile EBLOSUM62
# -gapopen 10.0
# -gapextend 0.5
# -aformat3 pair
# -sprotein1
# -sprotein2
# Align_format: pair
# Report_file: stdout
#####

=====
#
# Aligned_sequences: 2
# 1: EMBOS_001
# 2: EMBOS_001
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 413
# Identity: 388/413 (93.9%)
# Similarity: 398/413 (96.4%)
# Gaps: 0/413 (0.0%)
# Score: 2071.0
#
#
=====

EMBOSS_001      1 MAALRYAGLDDTDSEDELPPGWEERTTKDGWVYYANHTEEKTQWEHPKTG
                  |||||||DDTTDSEDELPPGWEERTTKDGWVYYANHTEEKTQWEHPKTG
EMBOSS_001      1 MAALRYAGLDDTDSEDELPPGWEERTTKDGWVYYANHTEEKTQWEHPKTG
                  |||||||DDTTDSEDELPPGWEERTTKDGWVYYANHTEEKTQWEHPKTG

```

Human - Zebrafish

(Score = 1638.0)

```
#####
# Program: water
# Rundate: Fri 13 Mar 2020 17:36:02
# Commandline: water
#   -auto
#   -stdout
#   -asequence emboss_water-I20200313-173557-0721-87757291-p2m.asequence
#   -bsequence emboss_water-I20200313-173557-0721-87757291-p2m.bsequence
#   -datafile EBLOSUM62
#   -gapopen 10.0
#   -gapextend 0.5
#   -aformat3 pair
#   -sprotein1
#   -sprotein2
# Align_format: pair
# Report_file: stdout
#####

=====
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 408
# Identity: 297/408 (72.8%)
# Similarity: 348/408 (85.3%)
# Gaps: 3/408 ( 0.7%)
# Score: 1638.0
#
#
=====

EMBOSS_001      1 MAALRYAGLDDTDSEDELPPGWEERTTKDGWVYYANHTEEKTQWEHPKTG      50
                  ||||:||||:|||||:||||||:|||||:|||||||.|||:|||||||
EMBOSS_001      1 MAALKYAGMEDTDSEDELPPGWEERSTKDGWVYYANHEEMKTQWEHPKTG      50
```

Human - Chicken

(Score = 1887.0)

```
#####
# Program: water
# Rundate: Fri 13 Mar 2020 17:37:22
# Commandline: water
#   -auto
#   -stdout
#   -asequence emboss_water-I20200313-173720-0447-11806600-p2m.asequence
#   -bsequence emboss_water-I20200313-173720-0447-11806600-p2m.bsequence
#   -datafile EBLOSUM62
#   -gapopen 10.0
#   -gapextend 0.5
#   -aformat3 pair
#   -sprotein1
#   -sprotein2
# Align_format: pair
# Report_file: stdout
#####

=====
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 413
# Identity: 344/413 (83.3%)
# Similarity: 385/413 (93.2%)
# Gaps: 0/413 ( 0.0%)
# Score: 1887.0
#
#
=====

EMBOSS_001      1 MAALRYAGLDDTDSEDELPPGWEERTTKDGWVVYANHTEEKTQWEHPKTG      50
                  ||||:||||:||||:|||||||||||||.||||||||:|
EMBOSS_001      1 MAALKYAGLEDTDSEEELPPGWEERTTKDGWVVYANHLEEKTQWEHPKSG      50
```

Human - Horse

(Score = 2125.0)

```
#####
# Program: water
# Rundate: Fri 13 Mar 2020 17:38:30
# Commandline: water
#   -auto
#   -stdout
#   -asequence emboss_water-I20200313-173827-0010-64491273-p2m.asequence
#   -bsequence emboss_water-I20200313-173827-0010-64491273-p2m.bsequence
#   -datafile EBLOSUM62
#   -gapopen 10.0
#   -gapextend 0.5
#   -aformat3 pair
#   -sprotein1
#   -sprotein2
# Align_format: pair
# Report_file: stdout
#####

=====
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 412
# Identity: 396/412 (96.1%)
# Similarity: 403/412 (97.8%)
# Gaps: 0/412 ( 0.0%)
# Score: 2125.0
#
#
=====

EMBOSS_001      1 MAALRYAGLDDTDSEDELPPGWEERTTKDGWVYYANHTEEKTQWEHPKTG      50
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
EMBOSS_001      1 MAALRYAGLDDTDSEDELPPGWEERTTKDGWVYYANHTEEKTQWEHPKTG      50
```

Organism	Nt local alignment score	Peptide local alignment score
Human Mouse	6548.0	2071.0
Human Cattle	9212.5	46.0
Human Zebrafish	6565.0	1638.0
Human Chicken	8023.0	1887.0
Human Horse	8680.5	2125.0

→ We see that the Peptide local alignment scores are significantly less than nt counterparts-

- Because peptide sequence is shorter due to the fact that 1 codon covers 3 nucleotide.

→ Also, one can see that in spite of a high local nt alignment between Human Cattle, the peptide local alignment score is quite low. This may be attributed to the fact that the difference in the nt sequence must have been in the 1st or 2nd nt of the codon, giving a completely different amino acid.

BLASTn

WWOX transcript sequence -1 (2241 bp)

Taxonomy

Program BLASTN [?](#) Citation [▼](#)

Database nt [See details](#) [▼](#)

Query ID Icl|Query_37337

Description None

Molecule type dna

Query Length 2241

Other reports Distance tree of results MSA viewer [?](#)

Organism only top 20 will appear exclude

Type common name, binomial, taxid or group name

+ Add organism

Percent Identity to E value to Query Coverage to

Filter **Reset**

Descriptions Graphic Summary Alignments **Taxonomy**

Reports Lineage Organism **Taxonomy**

100 sequences selected [?](#)

Description	Score	E value	Accession
Homo sapiens (human) [primates]			▼ Next ▲ Previous ◀ First
Homo sapiens WW domain containing oxidoreductase (WWOX), transcript variant 1, mRNA	4042	0.0	NM_016373
Homo sapiens WW domain-containing protein WWOX mRNA, complete cds	4042	0.0	AF211943
Homo sapiens cDNA FLJ314720 f1, clone NTZRP3001495, highly similar to WW domain-containing oxidoreductase (EC 1.1.1.-)	4036	0.0	AK027626
Homo sapiens FOR II protein mRNA, complete cds	3984	0.0	AF227527
Homo sapiens WOX8 isoform 8 (WWOX) mRNA, complete cds	3856	0.0	AY256821
Homo sapiens WW domain containing oxidoreductase (WWOX), transcript variant 4, mRNA	3507	0.0	NM_001291997
Homo sapiens WWOX isoform 1 (WWOX) mRNA, complete cds, alternatively spliced	3412	0.0	MH400250
Homo sapiens WWOX isoform 2 (WWOX) mRNA, complete cds, alternatively spliced	3407	0.0	MH400251
Homo sapiens cDNA FLJ77452 complete cds, highly similar to Homo sapiens WW domain containing oxidoreductase (WWOX), transcript variant 1, n	3134	0.0	AK290438
Homo sapiens cDNA FLJ56924 complete cds, highly similar to WW domain-containing oxidoreductase (EC 1.1.1.-)	2843	0.0	AK298322
Homo sapiens cDNA FLJ78250 complete cds, highly similar to Homo sapiens WW domain containing oxidoreductase (WWOX), transcript variant 1, n	2764	0.0	AK291300
Human oxidoreductase (HHCMA56) mRNA, complete cds	2553	0.0	U13395
	2047	0.0	AK027626

Interpretations

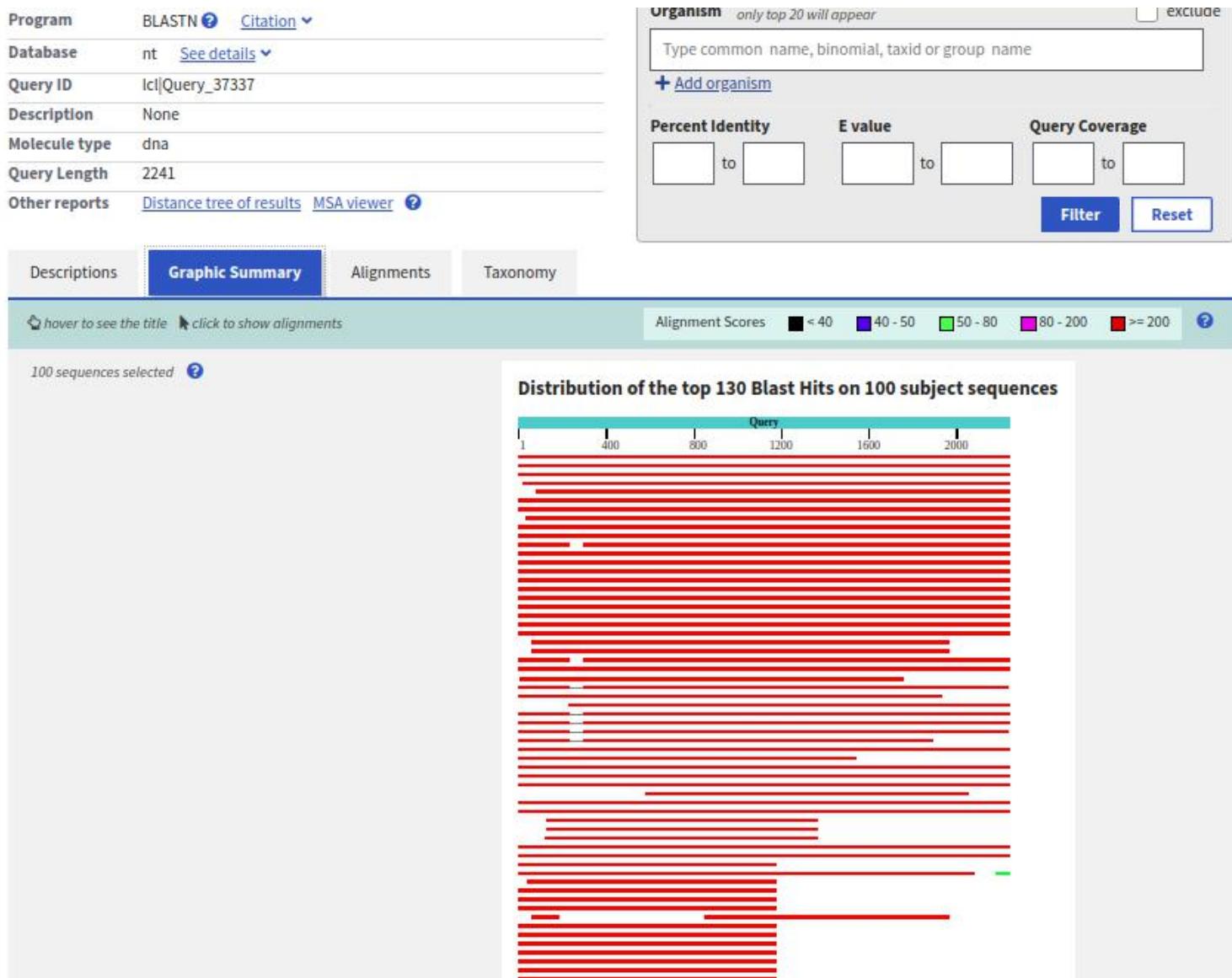
--> Taxonomy is the science of naming, defining and classifying groups of biological organisms on the basis of shared characteristics.

--> The most closely taxonomic sequence according to BLASTm for my transcript is Homo sapiens WWOX transcript variant 1 mRNA.

--> BLASTn is an efficient algorithm for comparing primary biological sequences based on nucleotide data.

--> The transcript variant 1 has the same score as cds which is listed as 2nd. This tells us that the whole transcript might be the coding domain sequence.

BLASTn graphical Summary



--> The top hits all have alignment scores ≥ 200 at all ranges of their suquence with the query sequence.

--> Below, not shown in screenshot, some of the sequences had scores in the range 80-200 and also in the range 50-80.

--> The more red a hit has, the more closely related the sequence is the inference of this screenshot.

BLASTn E-value

Program BLASTN [?](#) [Citation](#) [▼](#)

Database nt [See details](#) [▼](#)

Query ID Icl|Query_37337

Description None

Molecule type dna

Query Length 2241

Other reports Distance tree of results [MSA viewer](#) [?](#)

Organism only top 20 will appear exclude

Type common name, binomial, taxid or group name

+ Add organism

Percent Identity to E value to Query Coverage to

[Filter](#) [Reset](#)

Descriptions Graphic Summary Alignments Taxonomy

Sequences producing significant alignments Download Manage Columns Show 100 [?](#)

select all 100 sequences selected

	Description	Max Score	Total Score	Query Cover	E value	Per. ident	Accession
<input checked="" type="checkbox"/>	Homo sapiens WW domain containing oxidoreductase (WWOX), transcript variant 1, mRNA	4042	4042	100%	0.0	100.00%	NM_016373.4
<input checked="" type="checkbox"/>	Homo sapiens WW domain-containing protein WWOX mRNA, complete cds	4042	4042	100%	0.0	100.00%	AF211943.1
<input checked="" type="checkbox"/>	Homo sapiens cDNA FLJ14720 fis, clone NT2RP3001495, highly similar to WW domain-containing oxidoreductase (EC 1.1.1.-)	4036	4036	99%	0.0	99.96%	AK027626.1
<input checked="" type="checkbox"/>	Homo sapiens FOR II protein mRNA, complete cds	3984	3984	99%	0.0	99.77%	AF227527.1
<input checked="" type="checkbox"/>	Homo sapiens WOX8 isoform 8 (WWOX) mRNA, complete cds	3856	3856	96%	0.0	99.77%	AY256821.1
<input checked="" type="checkbox"/>	PREDICTED: Pan troglodytes WW domain containing oxidoreductase (WWOX), transcript variant X2, mRNA	3773	3773	100%	0.0	96.46%	XM_001144696.6
<input checked="" type="checkbox"/>	PREDICTED: Gorilla gorilla gorilla WW domain containing oxidoreductase (WWOX), mRNA	3769	3769	100%	0.0	96.69%	XM_019012536.1
<input checked="" type="checkbox"/>	Pongo abelii WW domain containing oxidoreductase (WWOX), mRNA	3632	3632	98%	0.0	96.35%	NM_001132377.1
<input checked="" type="checkbox"/>	PREDICTED: Hylobates moloch WW domain containing oxidoreductase (WWOX), transcript variant X1, mRNA	3615	3615	99%	0.0	95.80%	XM_032156362.1
<input checked="" type="checkbox"/>	PREDICTED: Nomascus leucogenys WW domain containing oxidoreductase (WWOX), transcript variant X1, mRNA	3577	3577	99%	0.0	95.12%	XM_030819181.1
<input checked="" type="checkbox"/>	Homo sapiens WW domain containing oxidoreductase (WWOX), transcript variant 4, mRNA	3507	3926	97%	0.0	100.00%	NM_001291997.2
<input checked="" type="checkbox"/>	PREDICTED: Macaca mulatta WW domain containing oxidoreductase (WWOX), transcript variant X1, mRNA	3502	3502	99%	0.0	94.78%	XM_001105944.4
<input checked="" type="checkbox"/>	PREDICTED: Macaca fascicularis WW domain containing oxidoreductase (WWOX), transcript variant X1, mRNA	3498	3498	99%	0.0	94.66%	XM_005592616.2
<input checked="" type="checkbox"/>	PREDICTED: Rhinopithecus roxellana WW domain containing oxidoreductase (WWOX), transcript variant X2, mRNA	3475	3475	99%	0.0	94.52%	XM_010381243.1
<input checked="" type="checkbox"/>	PREDICTED: Cercopithecus atys WW domain containing oxidoreductase (WWOX), transcript variant X1, mRNA	3470	3470	99%	0.0	94.47%	XM_012090452.1
<input checked="" type="checkbox"/>	PREDICTED: Macaca nemestrina WW domain containing oxidoreductase (WWOX), transcript variant X1, mRNA	3470	3470	99%	0.0	94.47%	XM_011757425.2
<input checked="" type="checkbox"/>	PREDICTED: Chlorocebus sabaeus WW domain containing oxidoreductase (WWOX), transcript variant X1, mRNA	3461	3461	99%	0.0	94.29%	XM_007994143.1
<input checked="" type="checkbox"/>	PREDICTED: Theropithecus gelada WW domain containing oxidoreductase (WWOX), transcript variant X1, mRNA	3454	3454	99%	0.0	94.29%	XM_025369939.1
<input checked="" type="checkbox"/>	PREDICTED: Colobus angolensis palliatus WW domain containing oxidoreductase (WWOX), transcript variant X1, mRNA	3452	3452	99%	0.0	94.24%	XM_011942139.1
<input checked="" type="checkbox"/>	PREDICTED: Panio anubis WW domain containing oxidoreductase (WWOX), transcript variant X1, mRNA	3450	3450	99%	0.0	94.21%	XM_017953716.2

Interpretations

--> E-value - The BLAST E-value is the number of expected hits of similar quality (score) that could be found just by chance. E-value of 10 means that up to 10 hits can be expected to be found just by chance, given the same size of a random database.

--> Thus the lower the E-value, the lower the chance of random hitting and hence the better the alignment.

--> E-value is inversely proportional to the percent identity.

--> Max score is the score when a sequence is aligned with itself.

--> Percent identity = Alignment score/Maximum score

Lab-7 (27/3)

BLASTp (WWOX refseq-1 from NCBI)

Screenshot – BLASTp taxonomy- organism

Job Title: WWOX blastp

RID: 7WC5Z4ZU01R Search expires on 03-29 00:58 am [Download All](#)

Program: BLASTP [?](#) Citation [▼](#)

Database: nr [See details](#) [▼](#)

Query ID: lcl|Query_36050

Description: None

Molecule type: amino acid

Query Length: 414

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](#)

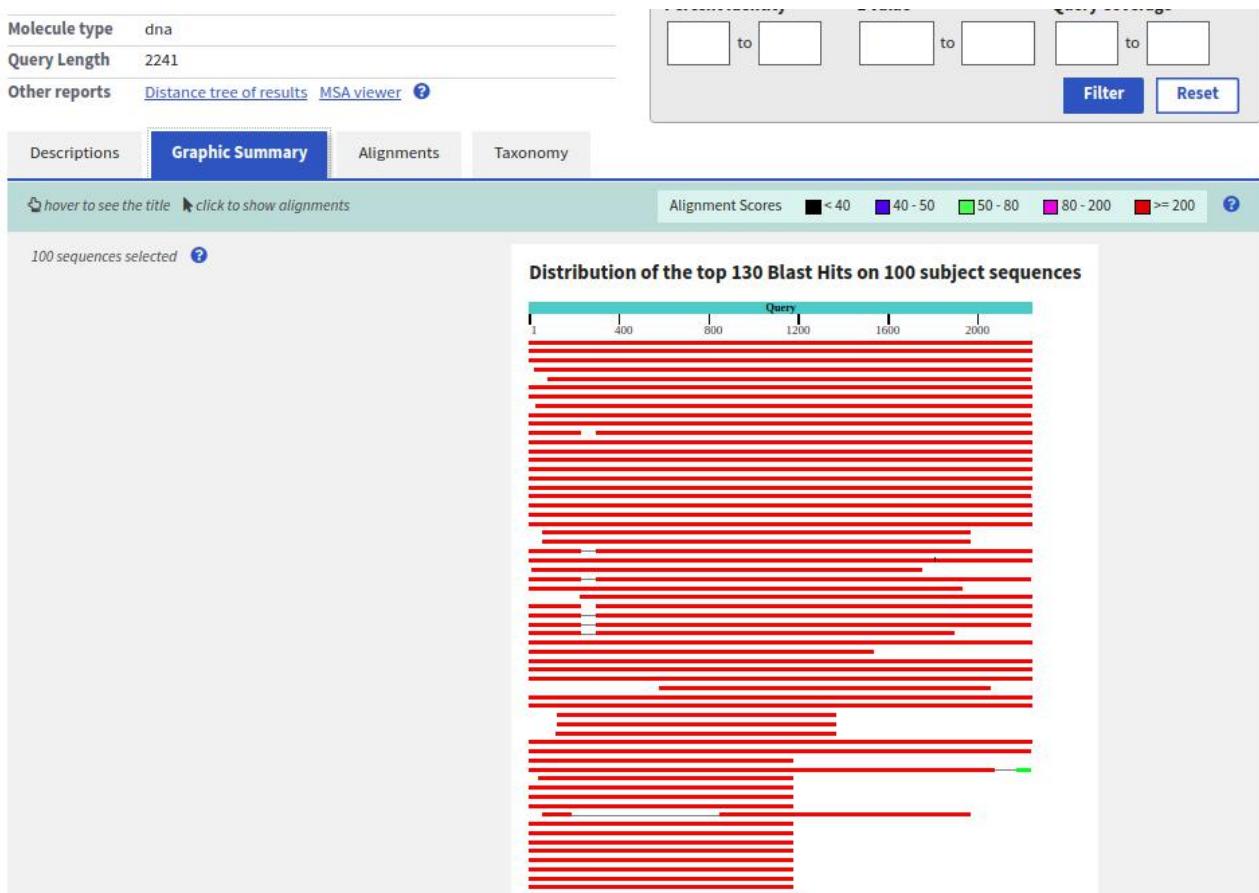
[Descriptions](#) [Graphic Summary](#) [Alignments](#) **Taxonomy**

Reports [Lineage](#) [Organism](#) [Taxonomy](#)

100 sequences selected [?](#)

Description	Score	E value	Accession
Homo sapiens (human) [primates]			
WW domain-containing oxidoreductase isoform 1 [Homo sapiens]	871	0.0	NP_057457
RecName: Full=WW domain-containing oxidoreductase; AltName: Full=Frugle site FRA1QD oxidoreductase; AltName: Full=Short chain dehydrogenase/reductase family 19 member 1	871	0.0	Q9NZC7
WW domain-containing protein WWOX [Homo sapiens]	871	0.0	AAF27049
WW domain-containing oxidoreductase isoform F01RII [Homo sapiens]	871	0.0	AAI05449
WWOX isoform 1 [Homo sapiens]	871	0.0	QBB67478
unnamed protein product [Homo sapiens]	871	0.0	BAF83127
hCG2042882, isoform CRA_a [Homo sapiens]	870	0.0	EAW95578
unnamed protein product [Homo sapiens]	869	0.0	BAF83389
unnamed protein product [Homo sapiens]	867	0.0	BAG51361
FOR II protein [Homo sapiens]	867	0.0	AAF82054
synthetic construct [other sequences]			
WW domain containing oxidoreductase, partial [synthetic construct]	870	0.0	BAJ20228
WW domain containing oxidoreductase, partial [synthetic construct]	868	0.0	AAX36701
WW domain containing oxidoreductase [synthetic construct]	868	0.0	AAX41075
Gorilla gorilla gorilla (western lowland gorilla) [primates]			
WW domain-containing oxidoreductase [Gorilla gorilla gorilla]	867	0.0	XP_018860801
Pongo abelii (Sumatran orangutan) [primates]			
WWOX Isoform 1 [Pongo abelii]	867	0.0	PNJ01942
WW domain-containing oxidoreductase [Pongo abelii]	862	0.0	NP_001125849
RecName: Full=WW domain-containing oxidoreductase [Pongo abelii]	862	0.0	Q9R9W5
hypothetical protein [Pongo abelii]	862	0.0	CAH91445
Pan troglodytes (chimpanzee) [primates]			
WW domain-containing oxidoreductase isoform X2 [Pan troglodytes]	865	0.0	XP_001144896
WWOX Isoform 1 [Pan troglodytes]	865	0.0	PNJ43605
WW domain-containing oxidoreductase isoform X1 [Pan troglodytes]	852	0.0	XP_009425059
Pan paniscus (pygmy chimpanzee) [primates]			
WW domain-containing oxidoreductase isoform X1 [Pan paniscus]	865	0.0	XP_008972621
Macaca fascicularis (crab-eating macaque) [primates]			
PREDICTED: WW domain-containing oxidoreductase isoform X1 [Macaca fascicularis]	859	0.0	XP_003582073
Chlorocebus sabaeus (green monkey) [primates]			

Screenshot-2 (BLASTp graphical summary)



Screenshot – 3 (BLASTp E-value and Identity)

Query ID	Icl Query_1623
Description	None
Molecule type	amino acid
Query Length	414
Other reports	Distance tree of results Multiple alignment MSA viewer ?

[+ Add organism](#)

Percent Identity <input type="text"/> to <input type="text"/>	E value <input type="text"/> to <input type="text"/>	Query Coverage <input type="text"/> to <input type="text"/>
Filter Reset		

Descriptions [Graphic Summary](#) [Alignments](#) [Taxonomy](#)

Sequences producing significant alignments

Download Manage Columns Show 100 ?

select all 100 sequences selected [GenPept](#) [Graphics](#) [Distance tree of results](#) [Multiple alignment](#)

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/>	WW domain-containing oxidoreductase isoform 1 [Homo sapiens]	871	871	100%	0.0	100.00%	NP_057457.1
<input checked="" type="checkbox"/>	hCG2042882, isoform CRA_a [Homo sapiens]	870	870	100%	0.0	99.76%	EAW95578.1
<input checked="" type="checkbox"/>	unnamed protein product [Homo sapiens]	869	869	100%	0.0	99.76%	BAF83989.1
<input checked="" type="checkbox"/>	WW domain containing oxidoreductase [synthetic construct]	868	868	100%	0.0	99.76%	AAX36701.1
<input checked="" type="checkbox"/>	WW domain containing oxidoreductase [synthetic construct]	868	868	100%	0.0	99.76%	AAX41075.1
<input checked="" type="checkbox"/>	WW domain-containing oxidoreductase [Gorilla gorilla gorilla]	867	867	99%	0.0	99.76%	XP_018868081.1
<input checked="" type="checkbox"/>	unnamed protein product [Homo sapiens]	867	867	100%	0.0	99.76%	BAG51361.1
<input checked="" type="checkbox"/>	WWOX isoform 1 [Pongo abelii]	867	867	100%	0.0	99.28%	PNJ61942.1
<input checked="" type="checkbox"/>	FOR II protein [Homo sapiens]	867	867	100%	0.0	99.52%	AAE82054.1
<input checked="" type="checkbox"/>	WW domain-containing oxidoreductase isoform X1 [Hylobates moloch]	865	865	100%	0.0	98.79%	

[Feedback](#)

Interpretations

--> BLASTP performs protein-protein sequence comparison, and its algorithm is the basis of many other types of BLAST searches such as BLASTX and TBLASTN.

--> The hits here also have alignment score ≥ 200 .

--> The maximum Percent identity = Score/Max score of 1st sequence.

MSA(ClustalW)

(MSA- 6 organisms 414 aa)

Screenshot-1 (Organisms)

```
>Homo sapiens
MAALRYAGLDDTDSEDELPPGWEERTTKDGWYYANHTEEKTQWEHPKTGKRKRVAGD
DFTGKVVVVTGANSIGFETAKSFALHGAHVILACRNMARASEAVSRILEEWHKAKVE
QLLQDVLCRSAPARVIVVSSESFRFTDINDSLGKLDLSRSLSPKTNDYWAMLAYNRSKL
YCAAVPELEGLGGMFNNCCRCMPSPEAQSEETARTLWALSERLIQERLGSQSG

>Gorilla
MAALRYAGLDDTDSEDELPPGWEERTTKDGWYYANHTEEKTQWEHPKTGKRKRVAGD
AKSFALHGAHVILACRNMARASEAVSRILEEWHKAKVEAMTLDALLRSVQHFAEAFK
SPTKNDYWAMLAYNRSKLCNLFSNELHRRRLSPRGVTSNAVHPGNMMYSNIHRSWWVY

>Pongo abelii
MAALRYAGLDDTDSEDELPPGWEERTTKDGWYYANHTEEKTQWEHPKTGKRKRVAGD
AKSFALHGAHVILACRNMARASEAVSRILEEWHKAKVEAVTLDALLRSVQHFAEAFK
SPTKNDYWAMLAYNRSKLCNLFSNELHRRRLSPRGVTSNAVHPGNMMYSNIHRSWWVY

>Pan troglodytes
MAALRYAGLDDTDSEDELPPGWEERTTKDGWYYANHTEEKTQWEHPKTGKRKRVAGD
AKSFALHGAHVILACRNMARASEAVSRILGEWHKAKVEAMTLDALLRSVQHFAEAFK
SPTKNDYWAMLAYNRSKLCNLFSNELHRRRLSPRGVTSNAVHPGNMMYSNIHRSWWVY

>Macaca fascicularis
MAALRYAGLDDTDSEDELPPGWEERTTKDGWYYANHTEEKTQWEHPKTGKRKRVAGD
AKSFALHGAHVILACRNMARASEAVSRILEEWHKAKVEAMALDLALLRSVQHFAEAFK
SPSKNDYWAMLAYNRSKLCNLFSNELHRRRLSPRGVTSNAVHPGNMMYSNIHRSWWVY

>Nomascus leucogenys
MAALRYAGLDDTDSEDELPPGWEERTTKDGWYYANHTEEKTQWEHPKTGKRKRVAGD
AKSFALHGAHVILACRNMARASEAVSRILEEWHKAKVEAMTLDALLRSVQHFAEAFK
SPSRNDYWAMLAYNRSKLCNLFSNELHRRRLSPRGVTSNAVHPGNMMYSNIHRSWWVY
```

Screenshot-2 (Colored Results)

No. of conserved regions = 86

Input form	Web services	Help & Documentation	Bioinformatics Tools FAQ
Macaca	MAALRYAGLDOTDSEDELPPGWEERTTKDGWVYYANHTEEKTQWEHPKTGKRKRVAGDLP		60
Nomascus	MAALRYAGLDOTDSEDELPPGWEERTTKDGWVYYANHTEEKTQWEHPKTGKRKRVAGDLP		60
Pan	MAALRYAGLDOTDSEDELPPGWEERTTKDGWVYYANHTEEKTQWEHPKTGKRKRVAGDLP		60
Homo	MAALRYAGLDOTDSEDELPPGWEERTTKDGWVYYANHTEEKTQWEHPKTGKRKRVAGDLP		60
Gorilla	MAALRYAGLDOTDSEDELPPGWEERTTKDGWVYYANHTEEKTQWEHPKTGKRKRVAGDLP		60
Pongo	MAALRYAGLDOTDSEDELPPGWEERTTKDGWVYYANHTEEKTQWEHPKTGKRKRVAGDLP		60

Macaca	YGWEQETDENGQVFVFDHINKRTTYLDPRLAFTVDONPTKPCTTQRYDGSTTAMEILQGR		120
Nomascus	YGWEQETDENGQVFVFDHINKRTTYLDPRLAFTVDONPTKPATRQRYDGSTTAMEILQGR		120
Pan	YGWEQETDENGQVFVFDHINKRTTYLDPRLAFTVDONPTKPCTTQRYDGSTTAMEILQGR		120
Homo	YGWEQETDENGQVFVFDHINKRTTYLDPRLAFTVDONPTKPCTTQRYDGSTTAMEILQGR		120
Gorilla	YGWEQETDENGQVFVFDHINKRTTYLDPRLAFTVDONPTKPCTTQRYDGSTTAMEILQGR		120
Pongo	YGWEQETDENGQVFVFDHINKRTTYLDPRLAFTVDONPTKPCTTQRYDGSTTAMEILQGR		120

Macaca	DFTGKVVVTGANSGIGFETAKSFALHGAHVILACRNMARASEAVSRILEEWHKAKVEAM		180
Nomascus	DFTGKVVVTGANSGIGFETAKSFALHGAHVILACRNMARASEAVSRILEEWHKAKVEAM		180
Pan	DFTGKVVVTGANSGIGFETAKSFALHGAHVILACRNMARASEAVSRILEEWHKAKVEAM		180
Homo	DFTGKVVVTGANSGIGFETAKSFALHGAHVILACRNMARASEAVSRILEEWHKAKVEAM		180
Gorilla	DFTGKVVVTGANSGIGFETAKSFALHGAHVILACRNMARASEAVSRILEEWHKAKVEAM		180
Pongo	DFTGKVVVTGANSGIGFETAKSFALHGAHVILACRNMARASEAVSRILEEWHKAKVEAM		180

Macaca	ALDLALLRSVQHFAEAFKAKNVPLHVLCVNAAFALPWSLTKDGLETTFQVNHLGHFYLV		240
Nomascus	TLDLALLRSVQHFAEAFKAKNVPLHVLCVNAATFALPWSLTKDGLETTFQVNHLGHFYLV		240
Pan	TLDLALLRSVQHFAEAFKAKNVPLHVLCVNAATFALPWSLTKDGLETTFQVNHLGHFYLV		240
Homo	TLDLALLRSVQHFAEAFKAKNVPLHVLCVNAATFALPWSLTKDGLETTFQVNHLGHFYLV		240
Gorilla	TLDLALLRSVQHFAEAFKAKNVPLHVLCVNAATFALPWSLTKDGLETTFQVNHLGHFYLV		240
Pongo	TLDLALLRSVQHFAEAFKAKNVPLHVLCVNAATFALPWSLTKDGLETTFQVNHLGHFYLV		240

Macaca	QLLQDVLCRASAPARVIVVSSESHRTDINDSLGKLFDSRLSPSKNDYWAMLAYNRSKLCN		300
Nomascus	QLLQDVLCRASAPARVIVVSSESHRTDINDSLGKLFDSRLSPSRNDYWAMLAYNRSKLCN		300
Pan	QLLQDVLCRASAPARVIVVSSESHRTDINDSLGKLFDSRLSPTKNDYWAMLAYNRSKLCN		300
Homo	QLLQDVLCRASAPARVIVVSSESHRTDINDSLGKLFDSRLSPTKNDYWAMLAYNRSKLCN		300
Gorilla	QLLQDVLCRASAPARVIVVSSESHRTDINDSLGKLFDSRLSPTKNDYWAMLAYNRSKLCN		300
Pongo	QLLQDVLCRASAPARVIVVSSESHRTDINDSLGKLFDSRLSPTKNDYWAMLAYNRSKLCN		300

Macaca	VLFSNELHRRLLSPRGVTNSNAVHPGNMIMYSNIHRSWWVYTLLFTLARPFTKSMQQGAATTV		360
Nomascus	VLFSNELHRRLLSPRGVTNSNAVHPGNMIMYSNIHRSWWVYTLLFTLARPFTKSMQQGAATTV		360
Pan	ILFSNELHRRLLSPRGVTNSNAVHPGNMIMYSNIHRSWWVYTLLFTLARPFTKSMQQGAATTV		360
Homo	ILFSNELHRRLLSPRGVTNSNAVHPGNMIMYSNIHRSWWVYTLLFTLARPFTKSMQQGAATTV		360
Gorilla	ILFSNELHRRLLSPRGVTNSNAVHPGNMIMYSNIHRSWWVYTLLFTLARPFTKSMQQGAATTV		360
Pongo	VLFSNELHRRLLSPRGVTNSNAVHPGNMIMYSNIHRSWWVYTLLFTLARPFTKSMQQGAATTV		360

Macaca	YCAAAPLEGLGGMYFNICCRCMPSPAEQSEETARALWALSERLIQERLGQSOG		414
Nomascus	YCAAAPLEGLGGMYFNICCRCMPSPAEQSEETARTLWALSERLIQERLGQSOG		414
Pan	YCAAAPLEGLGGMYFNICCRCMPSPAEQSEETARTLWALSERLIQERLGQSOG		414
Homo	YCAAAPLEGLGGMYFNICCRCMPSPAEQSEETARTLWALSERLIQERLGQSOG		414
Gorilla	YCAAAPLEGLGGMYFNICCRCMPSPAEQSEETARTLWALSERLIQERLGQSOG		414
Pongo	YCAAAPLEGLGGMYFNICCRCMPSPAEQSEETARTLWALSERLIQERLGQSOG		414

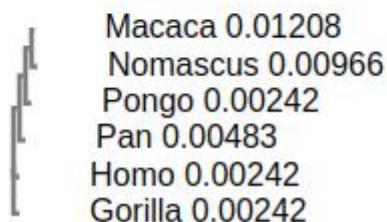
Screenshot -3 (Phylogenetic Tree)



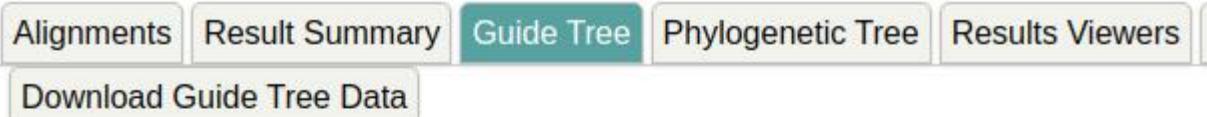
Phylogenetic Tree

This is a Neighbour-joining tree without distance corrections.

Branch length: Cladogram Real

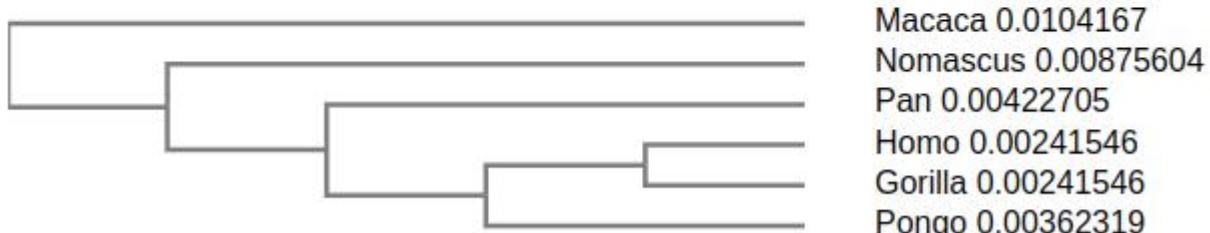


Screenshot-4 (Guide Tree)



Phylogram

Branch length: Cladogram Real



--> Human and Gorilla are most closely related.

--> Human and macaca are farthest in the phylogenetic tree.

Screenshot -5 (MView)

	Input form	Web services	Help & Documentation	Bioinformatics Tools FAQ	
1	Macaca	100.0%	100.0%	MAALRYAGLDDTDSEDELPPGWEERTTKDWWYYANHTEEKTOWEHKTCRKRJAGDLPYGEQETDENQFFDHIN	80
2	Nomascus	100.0%	97.8%	MAALRYAGLDDTDSEDELPPGWEERTTKDWWYYANHTEEKTOWEHKTCRKRJAGDLPYGEQETDENQFFDHIN	
3	Pan	100.0%	97.8%	MAALRYAGLDDTDSEDELPPGWEERTTKDWWYYANHTEEKTOWEHKTCRKRJAGDLPYGEQETDENQFFDHIN	
4	Homo	100.0%	98.1%	MAALRYAGLDDTDSEDELPPGWEERTTKDWWYYANHTEEKTOWEHKTCRKRJAGDLPYGEQETDENQFFDHIN	
5	Gorilla	100.0%	98.1%	MAALRYAGLDDTDSEDELPPGWEERTTKDWWYYANHTEEKTOWEHKTCRKRJAGDLPYGEQETDENQFFDHIN	
6	Pongo	100.0%	98.3%	MAALRYAGLDDTDSEDELPPGWEERTTKDWWYYANHTEEKTOWEHKTCRKRJAGDLPYGEQETDENQFFDHIN	
consensus/100%			MAALRYAGLDDTDSEDELPPGWEERTTKDWWYYANHTEEKTOWEHKTCRKRJAGDLPYGEQETDENQFFDHIN		
consensus/90%			MAALRYAGLDDTDSEDELPPGWEERTTKDWWYYANHTEEKTOWEHKTCRKRJAGDLPYGEQETDENQFFDHIN		
consensus/80%			MAALRYAGLDDTDSEDELPPGWEERTTKDWWYYANHTEEKTOWEHKTCRKRJAGDLPYGEQETDENQFFDHIN		
consensus/70%			MAALRYAGLDDTDSEDELPPGWEERTTKDWWYYANHTEEKTOWEHKTCRKRJAGDLPYGEQETDENQFFDHIN		
1	Macaca	100.0%	100.0%	KRTTYLDRLAFTDDNPTKPTTQRYDSTTAMEILQGRDFTCKVVVTGANSIGFETKSFAHLGAHVILACRNMR	160
2	Nomascus	100.0%	97.8%	KRTTYLDRLAFTDDNPTKPTTQRYDSTTAMEILQGRDFTCKVVVTGANSIGFETKSFAHLGAHVILACRNMR	
3	Pan	100.0%	97.8%	KRTTYLDRLAFTDDNPTKPTTQRYDSTTAMEILQGRDFTCKVVVTGANSIGFETKSFAHLGAHVILACRNMR	
4	Homo	100.0%	98.1%	KRTTYLDRLAFTDDNPTKPTTQRYDSTTAMEILQGRDFTCKVVVTGANSIGFETKSFAHLGAHVILACRNMR	
5	Gorilla	100.0%	98.1%	KRTTYLDRLAFTDDNPTKPTTQRYDSTTAMEILQGRDFTCKVVVTGANSIGFETKSFAHLGAHVILACRNMR	
6	Pongo	100.0%	98.3%	KRTTYLDRLAFTDDNPTKPTTQRYDSTTAMEILQGRDFTCKVVVTGANSIGFETKSFAHLGAHVILACRNMR	
consensus/100%			KRTTYLDRLAFTDDNPTKPTTQRYDSTTAMEILQGRDFTCKVVVTGANSIGFETKSFAHLGAHVILACRNMR		
consensus/90%			KRTTYLDRLAFTDDNPTKPTTQRYDSTTAMEILQGRDFTCKVVVTGANSIGFETKSFAHLGAHVILACRNMR		
consensus/80%			KRTTYLDRLAFTDDNPTKPTTQRYDSTTAMEILQGRDFTCKVVVTGANSIGFETKSFAHLGAHVILACRNMR		
consensus/70%			KRTTYLDRLAFTDDNPTKPTTQRYDSTTAMEILQGRDFTCKVVVTGANSIGFETKSFAHLGAHVILACRNMR		
1	Macaca	100.0%	100.0%	ASEAVSRILEEWHKKVEMAMIDLALLRSQHFAEFKAKNVPULHVLVCNAAFALPNSLTKDGLETTFQVNLGHFYLV	240
2	Nomascus	100.0%	97.8%	ASEAVSRILEEWHKKVEMAMIDLALLRSQHFAEFKAKNVPULHVLVCNAATFALPNSLTKDGLETTFQVNLGHFYLV	
3	Pan	100.0%	97.8%	ASEAVSRILEEWHKKVEMAMIDLALLRSQHFAEFKAKNVPULHVLVCNAATFALPNSLTKDGLETTFQVNLGHFYLV	
4	Homo	100.0%	98.1%	ASEAVSRILEEWHKKVEMAMIDLALLRSQHFAEFKAKNVPULHVLVCNAATFALPNSLTKDGLETTFQVNLGHFYLV	
5	Gorilla	100.0%	98.1%	ASEAVSRILEEWHKKVEMAMIDLALLRSQHFAEFKAKNVPULHVLVCNAATFALPNSLTKDGLETTFQVNLGHFYLV	
6	Pongo	100.0%	98.3%	ASEAVSRILEEWHKKVEMAMIDLALLRSQHFAEFKAKNVPULHVLVCNAATFALPNSLTKDGLETTFQVNLGHFYLV	
consensus/100%			ASEAVSRILEEWHKKVEMAMIDLALLRSQHFAEFKAKNVPULHVLVCNAATFALPNSLTKDGLETTFQVNLGHFYLV		
consensus/90%			ASEAVSRILEEWHKKVEMAMIDLALLRSQHFAEFKAKNVPULHVLVCNAATFALPNSLTKDGLETTFQVNLGHFYLV		
consensus/80%			ASEAVSRILEEWHKKVEMAMIDLALLRSQHFAEFKAKNVPULHVLVCNAATFALPNSLTKDGLETTFQVNLGHFYLV		
consensus/70%			ASEAVSRILEEWHKKVEMAMIDLALLRSQHFAEFKAKNVPULHVLVCNAATFALPNSLTKDGLETTFQVNLGHFYLV		
1	Macaca	100.0%	100.0%	QLIQDVLCRSAPARVIVISSESHRTTDINDSLKLDFSRSPSKNDYWAMLAYNRSKCNVLFSENHRRLSIRGTSN	320

PDB stryctural similarity (Screenshot – 1 Descriptions)

Database pdb See details ▾

Query ID lcl|Query_46844

Description None

Molecule type amino acid

Query Length 414

Other reports Distance tree of results Multiple alignment MSA viewer ?

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 Manage Columns Show 100 ?

select all 100 sequences selected GenPept Graphics Distance tree of results Multiple alignment

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/>	Chain A_Mypaa_01249.c [Mycobacterium avium subsp. paratuberculosis]	134	134	68%	9e-36	36.27%	3RD5_A
<input checked="" type="checkbox"/>	Chain A_WW domain containing oxidoreductase [Homo sapiens]	113	113	12%	8e-31	96.23%	1WMV_A
<input checked="" type="checkbox"/>	Chain A_NADPH-protochlorophyllide oxidoreductase [Thermosynechococcus elongatus]	117	117	73%	3e-29	31.81%	6RNW_A
<input checked="" type="checkbox"/>	Chain A_Thermosynechococcus elongatus protochlorophyllide oxidoreductase (POR) [Thermosynechococcus elongatus]	115	115	67%	2e-28	32.20%	6RN

Screenshot -2 (Taxonomy – Organism)

BLAST® » blastp suite » results for RID-7XSTJ64B014

Home Recent Results Saved Strategies Help

[Edit Search](#) Save Search Search Summary

Job Title **Protein Sequence**

RID **7XSTJ64B014** Search expires on 03-29 13:40 pm Download All

Program **BLASTP** Citation

Database **pdb** See details

Query ID **lcl|Query_83553**

Description **None**

Molecule type **amino acid**

Query Length **414**

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

Filter Reset

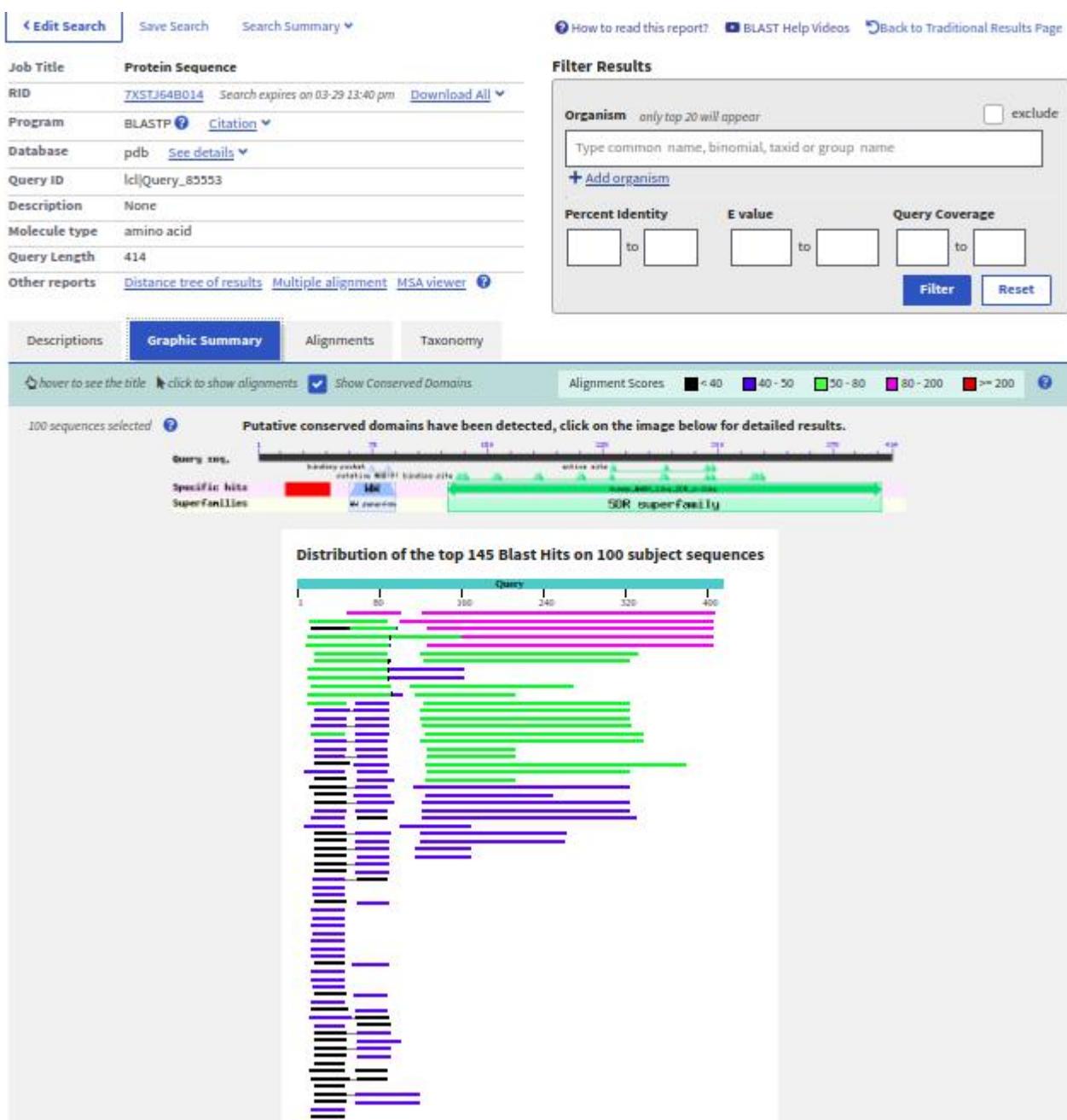
Descriptions Graphic Summary Alignments **Taxonomy**

Reports Lineage **Organism** Taxonomy

100 sequences selected

Description	Score	E value	Accession
Mycobacterium avium subsp. paratuberculosis [high GC Gram+]	▼ Next	▲ Previous	◀ First
Chain A, Mycaa01249.c [Mycobacterium avium subsp. paratuberculosis]	134	9e-38	3R05_A
Homo sapiens (human) [primates]	▼ Next	▲ Previous	◀ First
Chain A, WW domain containing oxidoreductase [Homo sapiens]	113	8e-31	3WMV_A
Chain A, E3 Ubiquitin-protein Ligase Itchy Homolog [Homo sapiens]	71.6	3e-15	3CQ2_A
Chain A, NEDD4-like E3 ubiquitin-protein ligase WWPL [Homo sapiens]	66.2	3e-11	6J1Y_A
Chain B, NEDD4-like E3 ubiquitin-protein ligase WWPL [Homo sapiens]	66.2	3e-11	6J1Y_B
Chain B, NEDD4-like E3 ubiquitin-protein ligase WWPL [Homo sapiens]	65.9	4e-11	6J1X_B
Chain A, E3 ubiquitin-protein ligase SMURF2 [Homo sapiens]	59.7	6e-11	2KKQ_A
Chain A, Dehydrogenase/reductase SDR family member 4 [Homo sapiens]	52.8	2e-07	3O4R_A
Chain B, Dehydrogenase/reductase SDR family member 4 [Homo sapiens]	52.8	2e-07	3O4R_B
Chain C, Dehydrogenase/reductase SDR family member 4 [Homo sapiens]	52.8	2e-07	3O4R_C
Chain D, Dehydrogenase/reductase SDR family member 4 [Homo sapiens]	52.8	2e-07	3O4R_D
Chain A, WW2 domain and PPxY motif complex [Homo sapiens]	47.8	3e-07	3YDY_A
Chain A, Yorkie homolog [Homo sapiens]	47.0	4e-07	3LTV_A
Chain A, Yorkie homolog [Homo sapiens]	47.0	4e-07	3LAW_A
Chain A, Helical Box Domain And Second Ww Domain Of The Human E3 Ubiquitin-Protein Ligase Hecw1 [Homo sapiens]	48.9	5e-07	3LAH_A
Chain A, Peptidyl-prolyl Cis-trans Isomerase Nims-Interacting 1 [Homo sapiens]	50.1	6e-07	3F21_A
Chain A, E3 ubiquitin-protein ligase Itchy homolog [Homo sapiens]	45.8	1e-06	4RDF_A
Chain B, E3 ubiquitin-protein ligase Itchy homolog [Homo sapiens]	45.8	1e-06	4RDF_B
Chain A, E3 ubiquitin-protein ligase Itchy homolog [Homo sapiens]	45.8	1e-06	5DWS_A
Chain C, E3 ubiquitin-protein ligase Itchy homolog [Homo sapiens]	45.8	1e-06	5DWS_C
Chain E, E3 ubiquitin-protein ligase Itchy homolog [Homo sapiens]	45.8	1e-06	5DWS_E
Chain G, E3 ubiquitin-protein ligase Itchy homolog [Homo sapiens]	45.8	1e-06	5DWS_G
Chain A, Itchy homolog E3 ubiquitin protein ligase [Homo sapiens]	45.4	2e-06	2DMV_A

Screenshot -3 (Graphical Summary)



- > The structural similarity has quite low percent identity, than its nt and peptide sequence counterparts.
- > The above fact may be because that one can never, see the structure or convert the sequence to the 2nd ary structure using any algorithm. One has to depend on predictions only.

BioInformatics Lab-8

Ashish Kumar

2017B1A70854P

1) Screenshot – 1 (Nucleotide sequence MSA)

DNA Sequences	Translated Protein Sequences
Species/Abbrv	
1. Homo_sapiens	- - - - G A C G G C T G G G T T A C T A C
2. House_mouse	- - - - G A C G G C T G G G T G T A C T A C
3. Zebrafish	- - - - G A C G G A T G G G T T T A T T A C
4. Chicken	- - - - G A C G G C T G G G T T A C T A C
5. Wolf	- - - - G A C G G C T G G G T T T A C T A T
6. Cattle	- - - - G A C G G C T G G G T T T A C T A T
7. Rhesus_macaque	- - - - G A C G G C T G G G T T A C T A C
8. Clawed_frog	- - - - G A T G G C T G G G T G T A C T A T
9. Wild_boar	- - - - G A C G G C T G G G T T A C T A T
10. Cat	- - - - G A C G G C T G G G T T A C T A T
11. Rhinoceros	- - - - G A C G G C T G G G T T T A C T A T
12. Armadillo	- - - - G A T G G C T G G G T T A C T A T
13. Goat	- - - - G A C G G C T G G G T T A C T A T
14. Brandts_bat	- - - - G A C G G C T G G G T G T A C T A T
15. Alligator	- - - - T A T T I C C T C A C T C T C T T A T
16. Turtle	- - - - C T T T T T T A A T A G T C T T T T T T A A T T T T T T T T A A T
17. Camel	- - - - G A C G G C T G G G T T T A C T A T
18. Burmese_python	- - - - G A C G G C T G G G T C T A C T A C
19. Australian_ghostshark	- - - - G A T G G C T G G G A T C T A C T A T
20. Polar_Bear	- - - - G A C G G C T G G G T T T A C T A T
21. Penguin	- - - -
22. Pangolin	- - - - G A C G G C T G G G T T T A C T A T
23. Electric_eel	- - - - G A T G G T T G G G T T T A C T A T

No. of conserved regions = 275

2) Screenshot – 2 (Protein sequence MSA)

DNA Sequences	Translated Protein Sequences
Species/Abbrv	
1. Homo_sapiens	- - - - R T A G F T T P - - - I T P - - - R R R - - - L S G - - - N I Q K L E K E N E W Q - - - E I C H - T D G N K K L L M R T D K C F L L T
2. House_mouse	- - - - S E D E L P P G - - - W E E - - - R T T - - - K D G - - - W V Y Y A N H T E E K T - - - Q W E H P K T G K R K R V A G D L P Y G W E
3. Zebrafish	- - - - K T D G F I T L - - - I M R - - - R X R - - - P S G - - - S I P K P A R R N A V L - - - E L C H T D G S R R P M I K A K S S M L I T S
4. Chicken	- - - - S E E E L P P G - - - W E E - - - R T T - - - K D G - - - W V Y Y A N H L E E K T - - - Q W E H P K T G K R K R V A G D L P Y G W E
5. Wolf	- - - - G R A A P G L G - - - A E N - - - H C G - - - R L G - - - L L C Q S R X G E D P V G T S K N W K K K T N S R R F A L W M G T R N X X E
6. Cattle	- - - - S E D E L P P G - - - W E E - - - R T T - - - K D G - - - W V Y Y A N H L E E K T - - - Q W E H P K T G K R K R V A G D L P Y G W E
7. Rhesus_macaque	- - - - S E D E L P P G - - - W E E - - - R T T - - - K D G - - - W V Y Y A N H T E E K T - - - Q W E H P K T G K R K R V A G D L P Y G W E
8. Clawed_frog	- - - - R M A G C T M Q - - - T T L - - - M K K - - - P S G - - - N I Q K L E S V K E L L - - - E I - F P T A G N K K Q M K M A K S T L L I
9. Wild_boar	- - - - G R A A A G L G - - - A E N - - - H C G - - - R L G - - - L L C Q S H G G E D A M G T S E N W E K K A N S R R F A L W M G A R N R X E F
10. Cat	- - - - S E D E L P P G - - - W E Q - - - R T T - - - K D G - - - W V Y Y A N H T E E K T - - - Q W E H P K T G K R K R V A G D L P Y G W E
11. Rhinoceros	- - - - R T A G F I M P - - - I T L - - - R R R - - - L N G - - - N I R K Q E N E N E X Q - - - E I C R - M D G N K K L M R T D R C F L S T
12. Armadillo	- - - - S E D E L P P G - - - W E E - - - R T T - - - K D G - - - W V Y Y A N H T P Q T S - - - Q W E H P K T G K R K R V A G D L P Y G W E
13. Goat	- - - - S E D E L P P G - - - W E Q - - - R T T - - - K D G - - - W V Y Y A N H T E E K T - - - Q W E H P K T G K R K R V A G D L P Y G W E
14. Brandts_bat	- - - - R T A G C T M P - - - I T L - - - R R R - - - L N G - - - N I R K L E K E N E X Q - - - E I C R - M D G N R K R M R M D K C I L S T
15. Alligator	- - - - V I E K C A I C - - - W - P - - - M P K - - - Y S S - - - L S Y V C N H L E E K T - - - Q W E H P K S G K R K R V A G D L P Y G W E
16. Turtle	- - - - L X A T F X X S F L I F F X T V I L - - - K R K - - - P S G - - - S I L N L E R E N A L L - - - E I C H - M G G S K K Q M K T D K S T L W T
17. Camel	- - - - G R A A P G L G - - - A E N - - - H C G - - - R L G - - - L L C Q S Y R G E D S V G T S E N W K K K E N S R R F A L R M G A R N R X E F
18. Burmese_python	- - - - S E E E L P P G - - - W E E - - - R T T - - - K D G - - - W V Y Y A N H L E E F T - - - Q W D H P K S G K R K R V A G D L P Y G W E
19. Australian_ghostshark	- - - - S E D E L P P G - - - W E E - - - R S T - - - K D G - - - W I Y Y A N H V E L K T - - - Q W D H P K T G K K K R I A G D L P Y G W E
20. Polar_Bear	- - - - H D K V Y S E D E L P P G - - - W E Q - - - R T T - - - K D G - - - W V Y Y A N H A E E K T - - - Q W E H P K T G K R K R V A G D L P Y G W E
21. Penguin	- - - - T K G L P T L - - - T Q D W P L - - - O L K I I Q X N - - - L L L D K N M M E T V L - - - Q W K Y S R V V T X V E K W L - - - X S P
22. Pangolin	- - - - S E D E L P P G - - - W E Q - - - R T T - - - K D G - - - W V Y Y A N H T E E K T - - - Q W E H P K T G K R K R V A G D L P Y G W E
23. Electric_eel	- - - - R R D E D P V G A S X N W E E E A L C R R S T L R V G A G D R X O G T N L L C X P H Q X E E D I - - - L X S S A S V H C R G C A G E A Q T L X R

Number of conserved regions = 10

3) Screenshot – 3 (Distance matrix for nucleotides)

```
[ 1]      1     2     3     4     5     6     7     8     9    10    11    12    13    14    15    16    17    18    19    20    21    22
[ 2] 0.3053
[ 3] 0.8145 0.8011
[ 4] 0.5742 0.6559 0.9559
[ 5] 0.2227 0.3714 0.7987 0.5484
[ 6] 0.3535 0.4266 0.8083 0.6564 0.2978
[ 7] 0.0511 0.3190 0.8261 0.5759 0.2278 0.3629
[ 8] 0.7762 0.8885 1.1050 0.7305 0.7655 0.9468 0.8141
[ 9] 0.2359 0.3568 0.7893 0.5862 0.2290 0.2864 0.2854 0.8682
[10] 0.1864 0.2921 0.7082 0.4504 0.1175 0.2481 0.1961 0.6250 0.1989
[11] 0.1915 0.3377 0.7949 0.5409 0.1622 0.2990 0.2202 0.8073 0.1851 0.1517
[12] 0.1617 0.2434 0.6490 0.3551 0.1656 0.2189 0.2257 0.5333 0.2497 0.1995 0.1834
[13] 0.3574 0.4241 0.8444 0.6956 0.2938 0.0719 0.3840 0.9963 0.2943 0.2457 0.3121 0.2490
[14] 0.2340 0.3385 0.8315 0.5793 0.2058 0.3103 0.2651 0.8263 0.2244 0.1781 0.1648 0.2098 0.3387
[15] 0.8861 0.8495 1.1995 0.7020 0.8117 0.8955 0.9144 1.1953 0.9403 0.6397 0.8708 0.6195 0.9113 0.9136
[16] 0.8699 1.0299 1.4053 0.7580 0.8881 1.0767 0.9133 1.1545 1.0114 0.7161 0.9270 0.6066 1.1074 0.9262 0.8690
[17] 0.4416 0.5053 0.9699 0.7778 0.4178 0.4416 0.4676 1.0741 0.4239 0.2531 0.3870 0.3042 0.4292 0.4194 0.9478 1.0333
[18] 0.5575 0.5877 0.8854 0.5096 0.5251 0.5992 0.5658 0.6739 0.6042 0.5169 0.5542 0.4831 0.6206 0.5942 0.6246 0.5739 0.5781
[19] 0.8392 0.9330 0.9882 0.9096 0.8334 0.8993 0.8695 1.1646 0.9131 0.6994 0.9054 0.5188 0.8923 0.8662 1.0750 1.2945 1.0840 0.7152
[20] 0.4127 0.4941 0.8918 0.6799 0.3167 0.4478 0.7317 0.9221 0.5344 0.3104 0.4376 0.6610 0.5538 0.4514 0.8612 0.8996 0.5415 0.5924 0.8764
[21] 0.5650 0.6587 0.9625 0.1218 0.5436 0.7168 0.5668 0.6798 0.5688 0.4256 0.5278 0.3291 0.7334 0.5686 0.6071 0.6359 0.8239 0.4638 0.9535 0.6730
[22] 0.2340 0.3467 0.8353 0.6082 0.2062 0.3397 0.2513 0.8313 0.2339 0.1664 0.1771 0.2007 0.3636 0.2078 0.8906 0.9692 0.4026 0.5776 0.9298 0.4347 0.5748
[23] 0.6395 0.6351 0.3975 0.6677 0.5631 0.5621 0.6722 0.8444 0.6436 0.6801 0.6842 0.6634 0.5955 0.6899 0.9780 1.0269 0.7218 0.7832 0.6487 0.8347 0.6527 0.6154
```

```
[ 1] #Homo_sapiens
[ 2] #House_mouse
[ 3] #Zebrafish
[ 4] #Chicken
[ 5] #Wolf
[ 6] #Cattle
[ 7] #Rhesus_macaque
[ 8] #Clawed_frog
[ 9] #Wild_boar
[10] #Cat
[11] #Rhinoceros
[12] #Armadillo
[13] #Goat
[14] #Brandts_bat
[15] #Alligator
[16] #Turtle
[17] #Camel
[18] #Burmese_python
[19] #Australian_ghostshark
[20] #Polar_Bear
[21] #Penguin
[22] #Pangolin
[23] #Electric_eel
```

Species with the least distance are → (1,7) i.e. Homo sapiens and Rhesus macaque

4) Screenshot – 4 (Distance matrix for proteins)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1]																						
2]	5.000																					
3]	5.034	7.500																				
4]	7.052	0.800	8.963																			
5]	5.258	4.933	5.615	6.987																		
6]	6.675	0.621	7.714	0.790	6.000																	
7]	4.870	0.651	8.492	0.670	5.388	0.647																
8]	2.206	8.041	7.672	10.228	10.962	9.203	8.280															
9]	6.011	5.299	5.460	5.588	0.928	5.471	4.034	9.463														
10]	5.145	0.364	6.915	0.486	4.831	0.292	0.317	8.830	5.200													
11]	0.807	6.196	5.276	7.182	4.759	5.288	5.068	2.343	3.293	6.329												
12]	8.766	0.266	7.809	0.327	5.879	0.209	0.524	8.938	4.725	0.266	8.740											
13]	5.694	0.625	7.561	0.787	5.367	0.344	0.719	8.230	4.488	0.286	5.324	0.363										
14]	0.898	6.154	5.734	8.104	6.818	7.554	4.669	2.356	4.656	4.574	0.872	7.492	5.647									
15]	7.822	1.102	8.982	0.957	7.219	1.058	1.216	9.460	7.160	0.904	6.878	0.728	1.117	8.246								
16]	2.699	9.897	6.687	10.824	13.095	11.735	9.141	2.758	9.820	10.163	2.740	10.000	11.294	2.602	9.632							
17]	8.484	5.750	6.632	6.930	1.068	6.000	5.286	13.524	1.000	5.506	8.984	4.247	6.091	9.356	8.091	11.213						
18]	8.022	0.561	9.333	0.465	6.179	0.556	0.566	9.641	6.203	0.444	7.809	0.376	0.557	8.244	0.647	9.757	6.000					
19]	7.662	0.982	9.089	0.874	7.931	0.925	1.003	11.127	6.207	0.606	7.123	0.384	0.907	8.456	1.253	7.914	7.000	0.606				
20]	8.207	0.575	8.894	0.698	6.269	0.507	0.853	10.286	5.495	0.484	7.773	0.748	0.664	8.175	0.956	9.875	7.712	0.553	0.855			
21]	2.783	7.750	8.333	3.814	10.286	8.156	6.272	3.482	7.514	6.737	3.033	6.708	7.268	3.496	7.797	3.820	10.976	8.026	8.391	6.754		
22]	4.521	0.673	7.582	0.802	4.732	0.599	0.713	10.459	3.918	0.296	3.664	0.379	0.681	5.052	1.224	10.796	6.838	0.589	1.053	0.676	7.162	
23]	9.111	1.662	6.714	1.739	3.592	1.546	1.849	7.533	3.636	1.560	7.591	1.866	1.714	8.625	3.137	9.618	4.472	2.140	1.760	2.785	6.902	1.797

```

1] #Homo_sapiens
2] #House_mouse
3] #Zebrafish
4] #Chicken
5] #Wolf
6] #Cattle
7] #Rhesus_macaque
8] #Clawed_frog
9] #Wild_boar
10] #Cat
11] #Rhinoceros
12] #Armadillo
13] #Goat
14] #Brandts_bat
15] #Alligator
16] #Turtle
17] #Camel
18] #Burmese_python
19] #Australian_ghostshark
20] #Polar_Bear
21] #Penguin
22] #Pangolin
23] #Electric_eel

```

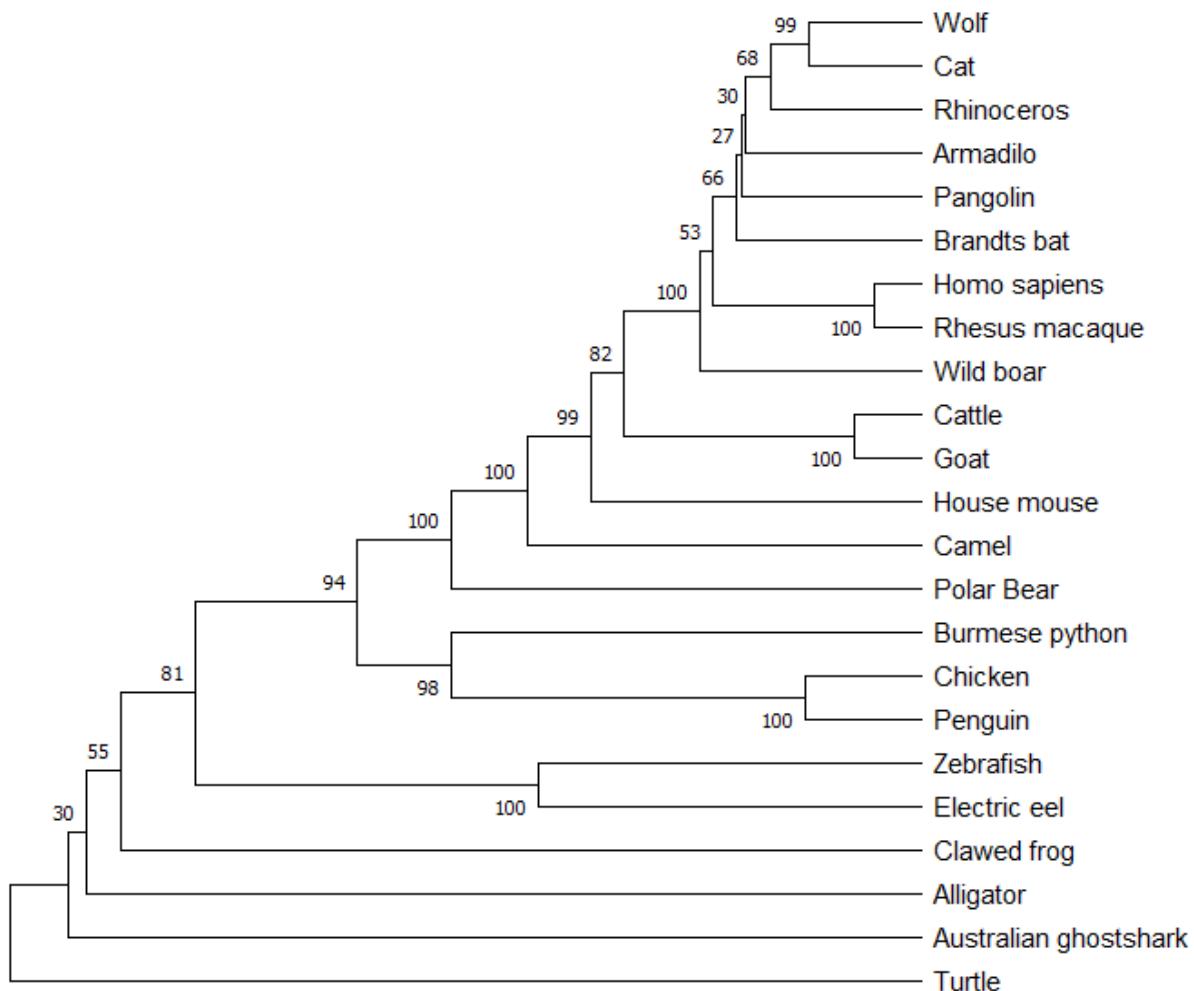
Species with least distance-

a) Cattle, Armadillo = 0.209

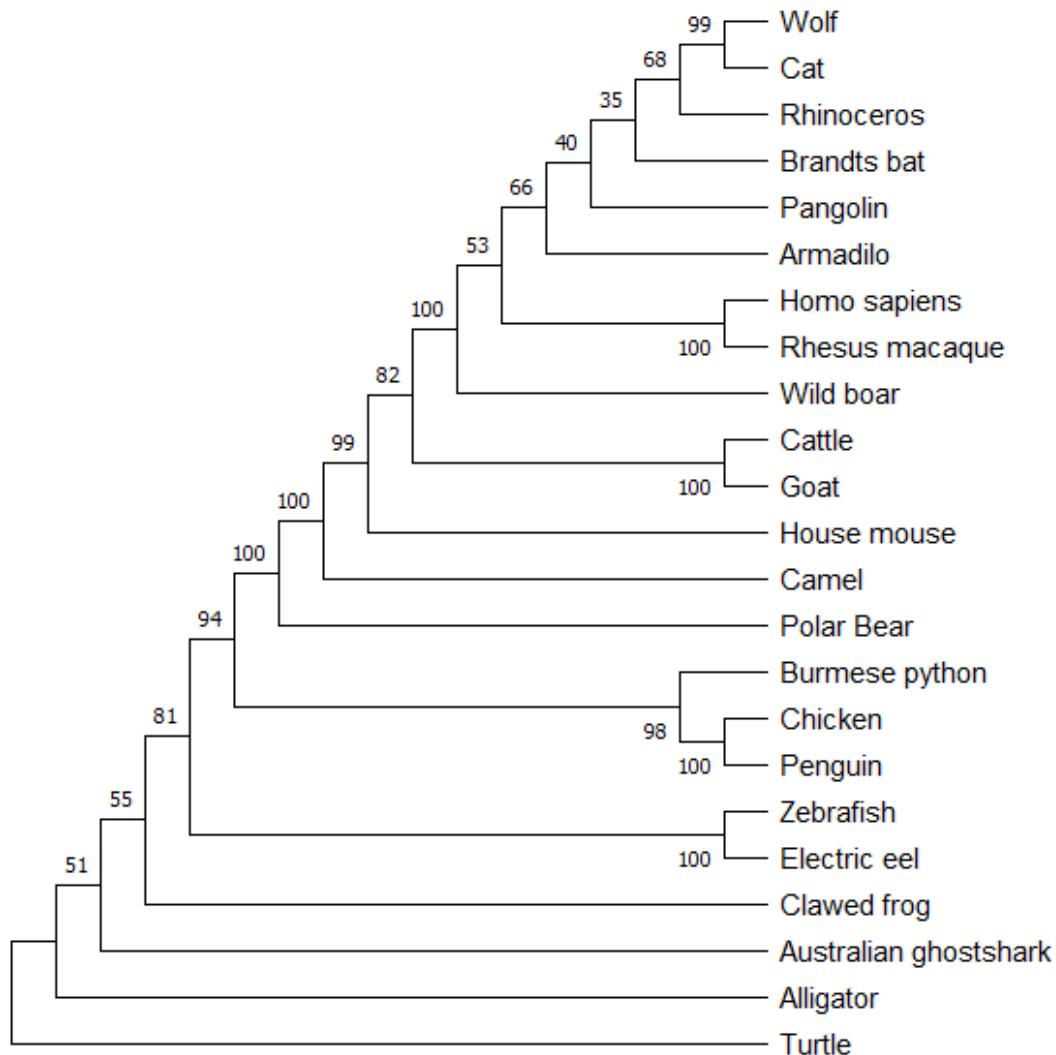
The interpretation from this is that the distance matrix for sequence MSA gives us a better evolutionary close species approximation than protein MSA distance matrix.

As, Humans and macaques are primates, they must be close evolutionary as shown by nucleotide matrix than Humans and armadillos.

5) Screenshot – 5 (Nucleotide phylogenetic tree)



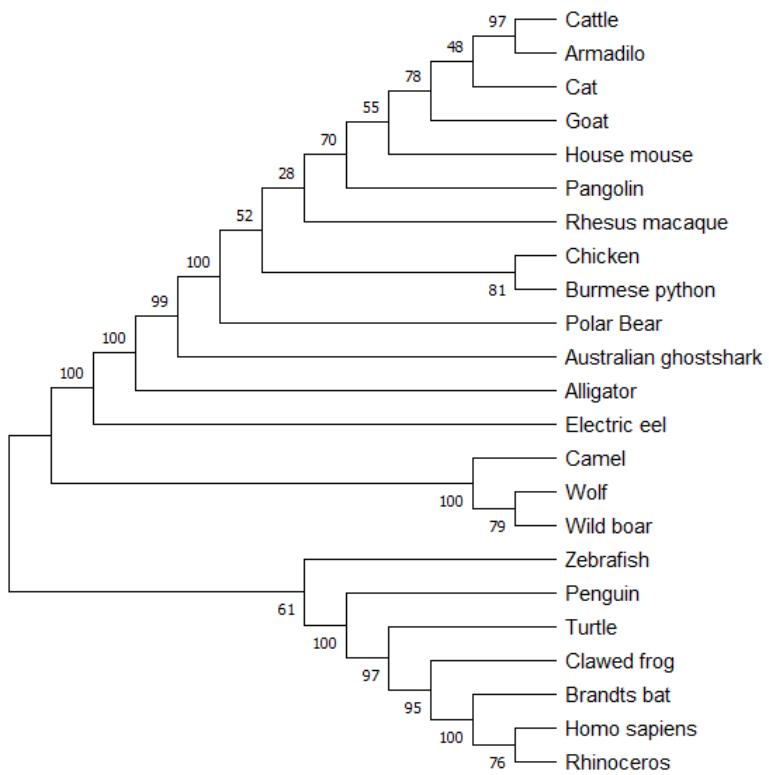
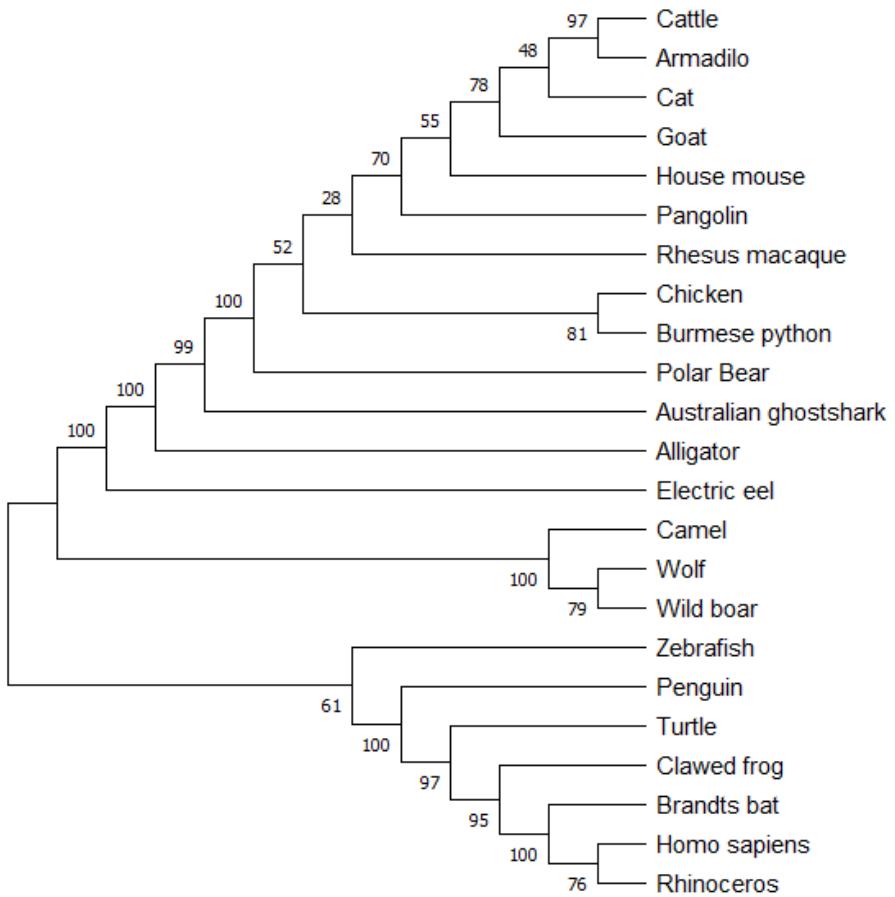
Below, is the tree after clicking bootstrap option.



The following are closely related organisms.

- a) Zebrafish – Eel → both are marine species
- b) Chicken – Penguin → Both evolved as birds
- c) Cattle – Goat → Both are primary consumers and 4 legged.
- d) Cat – Wolf

6) Screenshot – 6 (Protein phylogenetic tree)



The tree shows cattle armadillo as the most closely related according to their protein sequence, even more than humans and rhesus macaque.

1) Screenshot 1 (GOR-IV structure prediction)



--> GOR IV predicts 0% beta secondary structure present in my protein.

--> GOR IV predicts that the protein secondary structure is made completely of helixes and random turns.

2) Screenshot - 2(JPred E-value + Structure Prediction)

--> Jpred predicts beta strands to be present at specified ranges in the figure.

Match found in PDB

The sequence you submitted is similar to those with known structure. These may provide a more accurate secondary structure assignment than a JPred prediction.

If you still want to carry out a Jpred prediction click [continue](#)

Hits found

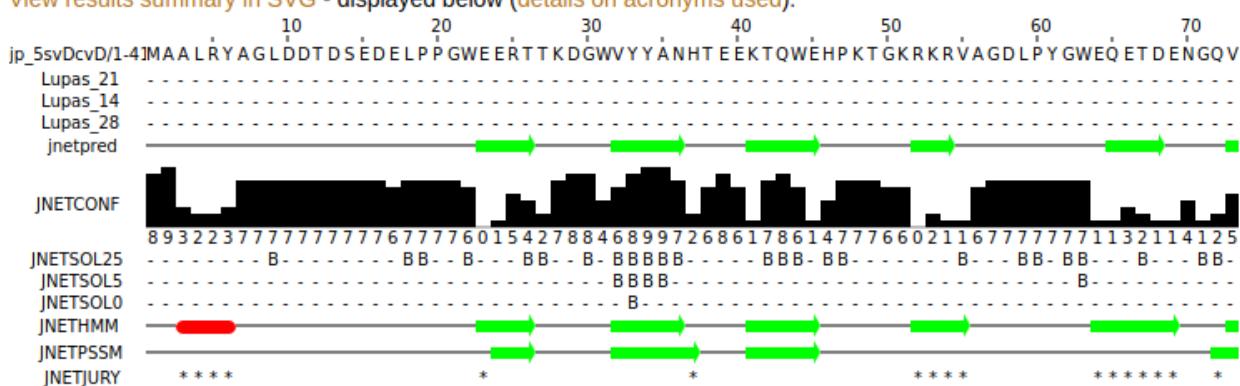
Show **25** entries

PDB	Chain	Description	Blast E-value
3rd5	A	MYPAA.01249.C	4e-29
6rnw	A	NADPH-protochlorophyllide oxidore...	2e-27
6rnv	A	Thermosynechococcus elongatus pro...	2e-26
6r46	A	NADPH-protochlorophyllide oxidore...	2e-26
6l1h	B	NADPH-protochlorophyllide oxidore...	2e-26
6l1h	A	NADPH-protochlorophyllide oxidore...	2e-26
1wmv	A	WW domain containing oxidoreductase	2e-24
6r48	B	Light-dependent protochlorophylli...	9e-22
6r48	A	Light-dependent protochlorophylli...	9e-22
6l1g	B	Light-dependent protochlorophylli...	9e-22
6l1g	A	Light-dependent protochlorophylli...	9e-22
5xmc	A	E3 ubiquitin-protein ligase Itchy	2e-13

Results

After much trouble and strife, Bob the scheduling penguin has retrieved your results! Rejoice. For your pleasure the following viewing options are available. You may bookmark this page for future reference although data is not kept on the server for more than two days.

[View results summary in SVG](#) - displayed below (details on acronyms used):



[View full results in HTML](#)

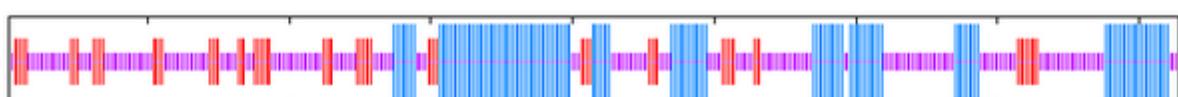
3) Screenshot - 3(Predator proteins structure prediction)

PREDATOR result for : UNK_541090

Abstract Frishman D., Argos P., Incorporation of non-local interactions in protein secondary structure prediction from the amino acid sequence. View PREDATOR in: [AnTheProt \(PC\)](#) , [Download...](#)] [[HELP](#)]

Sequence length : 414

PREDATOR :				
Alpha helix	(Hh) :	130	is	31.40%
3 ₁₀ helix	(Gg) :	0	is	0.00%
Pi helix	(Ii) :	0	is	0.00%
Beta bridge	(Bb) :	0	is	0.00%
Extended strand	(Ee) :	67	is	16.18%
Beta turn	(Tt) :	0	is	0.00%
Bend region	(Ss) :	0	is	0.00%
Random coil	(Cc) :	217	is	52.42%
Ambiguous states (?)	:	0	is	0.00%
Other states	:	0	is	0.00%



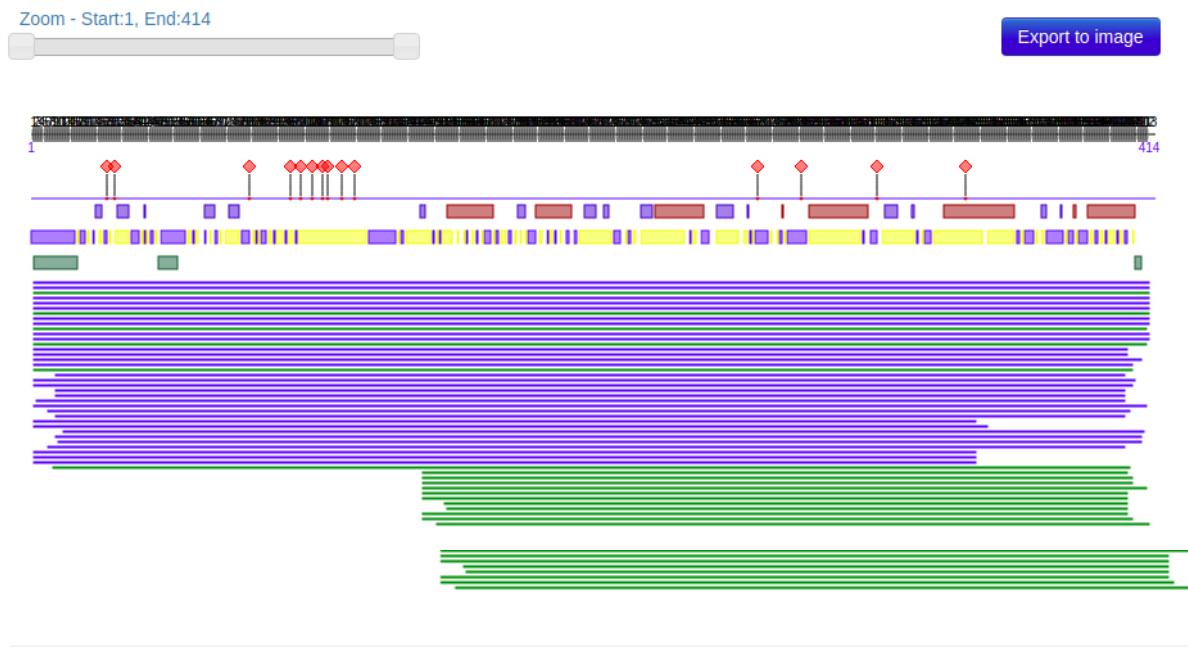
--> The prediction is very similar to GOR IV, and similar inference can be drawn - protein secondary structure comprises entirely of helixes and turns and no beta bridges.

4)Screenshot - 4(Phd protein prediction)



--> The sequence comprises only of helices and turns without any beta bridges.

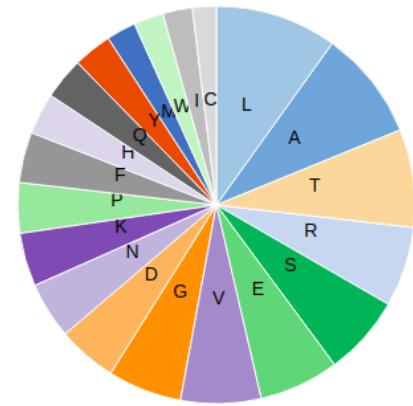
5)Screenshot - 5(Predict Protein secondary structure)



Summary

Sequence Length	414
Number of Aligned Proteins	48
Likely Organism	HUMAN

Amino Acid composition



--> Lysine is the most common amino acid in the sequence and cysteine is the least common amino acid in the sequence.

--> The predicted subcellular localization of the protein is the cytoplasm.

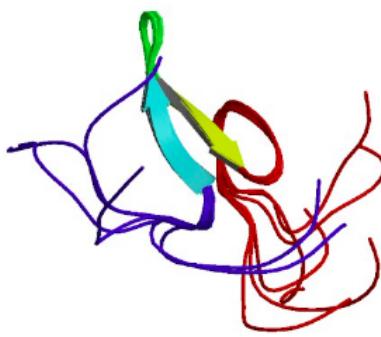
6) Screenshot - 6(Psipred secondary structure prediction)



--> This predicts that the sequence secondary structure comprises of mostly coils(grey), followed by helix(pink), and strand(yellow) the least.

7)Screenshot - 7 (icn3d)

NMR Ensemble



1WMV

Solution structure of the second WW domain of WWOX

DOI: [10.2210/pdb1WMV/pdb](https://doi.org/10.2210/pdb1WMV/pdb)

Classification: OXIDOREDUCTASE, APOPTOSIS
Organism(s): Homo sapiens
Expression System: Escherichia coli BL21

Deposited: 2004-07-21 Released: 2005-08-09
Deposition Author(s): Kowalski, K., Merkel, A.L., Colella, A., Richards, F.

Experimental Data Snapshot

Method: SOLUTION NMR
Conformers Calculated: 200
Conformers Submitted: 20
Selection Criteria: structures with the lowest energy

wwPDB Validation

Metric

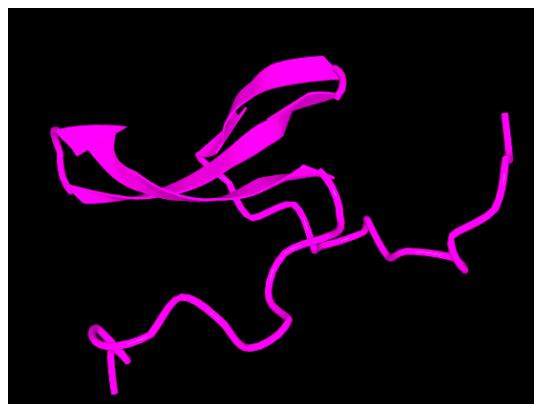
Clashscore | Ramachandran outliers | Sidechain outliers |

3D View: Structure

Standalone Viewers

- My protein name was 1WMV according to PDB.

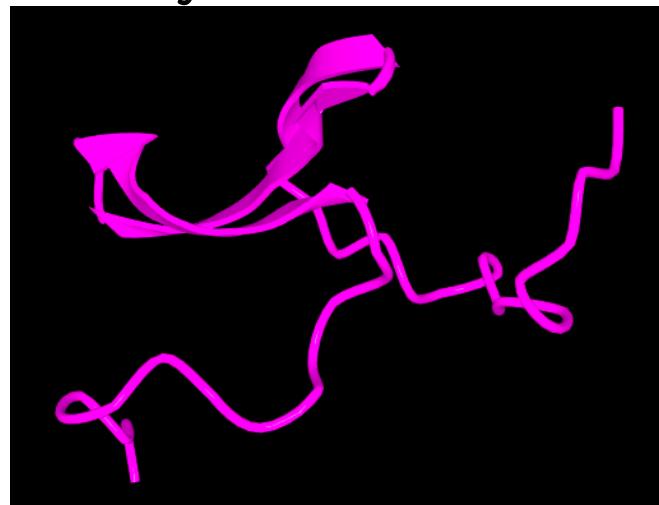
Ribbon (below)



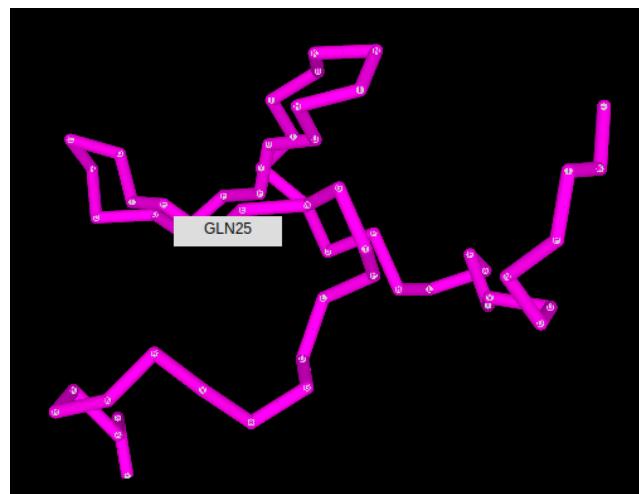
Strand



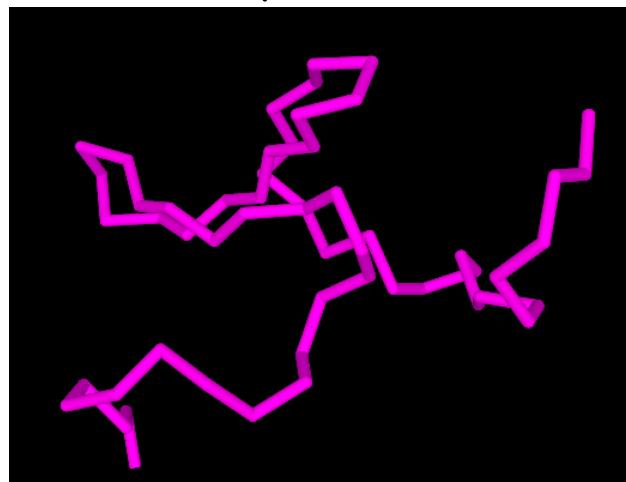
Cylinder and Plate



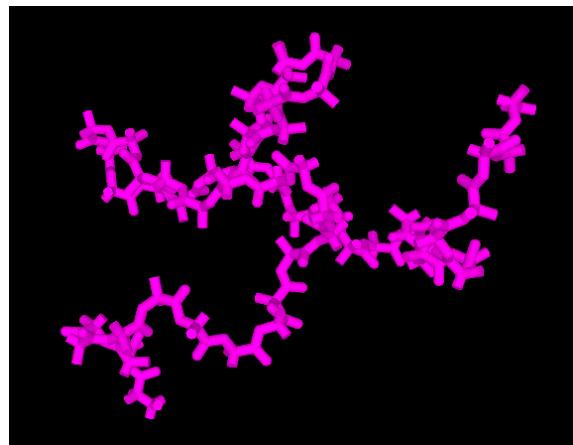
Schematic



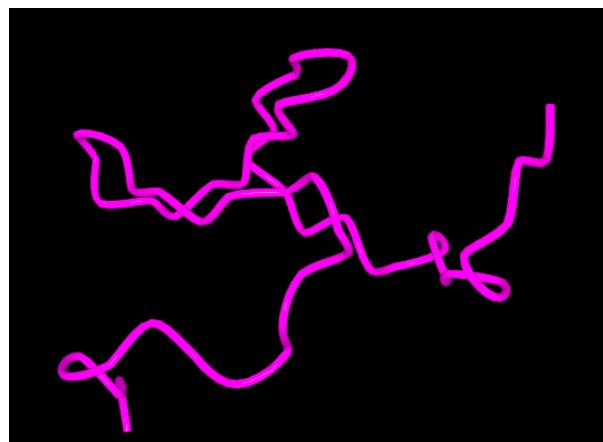
C alpha trace



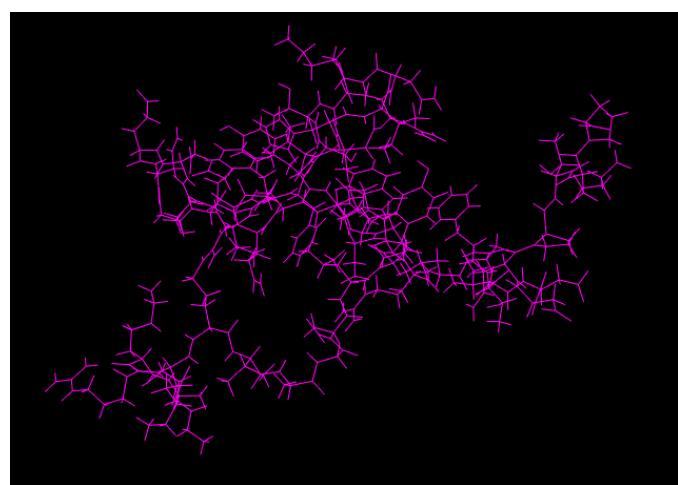
Backbone



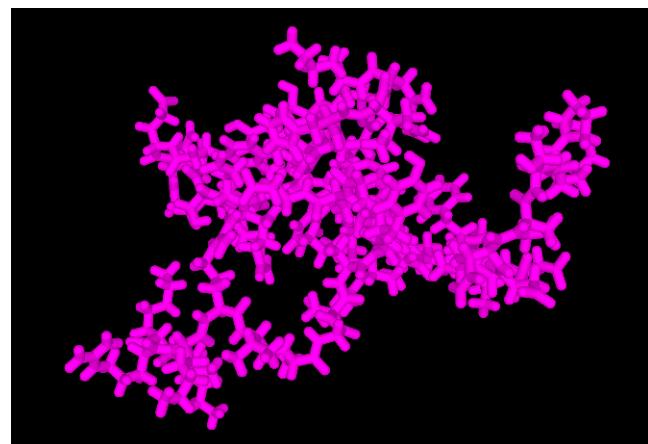
B factor tube



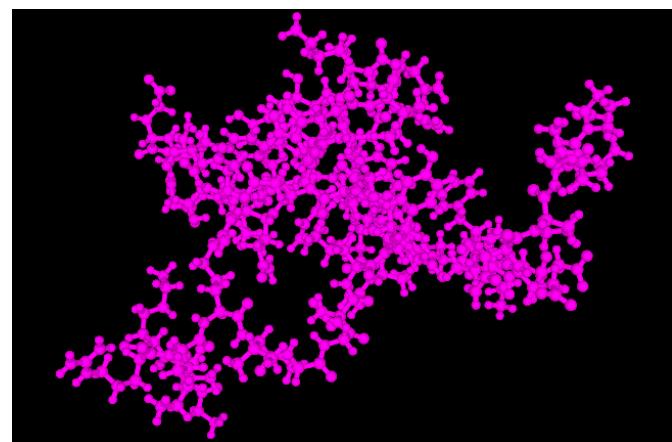
Lines



Stick



Ball and stick



Sphere

