

Bioinformatics project

Submitted to: Dr. Samina Shakeel

Submitted by: Sayeda Zahra Batool

Reg No: 04032213005

Department: Biochemistry Bs 7th Morning

Task No: 01

DNA Sequence Analysis:

1. DNA Sequence:

CTACAACTCACAAACGAGAGGCCACGTTGTCAACACACAGCAGCTACTACTAACAGACA
GACAGACAGACAG

ACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGGGCTGAGTGGTTCACTGCA
GGACGCACGGATGCG

CAATTACGCACGAGACTAGAGGAATGTCACCGTCGGCGCGCTGAGCTGTGGTCACCT
GTTGAAACGACAT

CACCTGTGCGTCAGTATAGTCTCTCCCAGTCCTCCAGAACACTTGTAGCAGAACATGGG
ATGCTCCACCAG

TACTCACACCTCCGGAGTGGATCCGAGCAGACCGAGCACCAAACCGGAGGAGAGCA
ACGGGGCCAGCACC

ACAGGGGCGGCCACGAGAACGGAAATGTAGCTGAGGACACTGAAACCGCTCCCCGA
TGAGACACCTACTG

GAGGCACTGTGGCAGAACCTGCAGCAGATGCAACATCTGCACAGCCCACAGCAGGG
GTGGAACCTCCACC

TGCAGCTGCAGCAGCACCTGTAGAGGCAACACCACCTGCAGAACAGGCAACGCC
CTGCAGAAGAAGCA

GCTCCACCTGCAGAAGAGGTAACACCACCTGCAGAAGAGGCAGCACCACCTGCAGA
AGAACGAGCACCAAG

CAGCAGAAAAGCCTTCAGAGGCTGCAACACAAGAACAGCACCAGCAGCTGCAGAGCC
AGGGAGCGAGGCCAC

AGCAGCTGACTCAGCAGAACAGCACCAGCTCCCAGTGAGTGAAGAGCAGCACCATAAT
GCTAACGCTGTTG

TATGACTGAGTAACAGTACGAATGTCAGTATGTGTGTAAATTGCCTGTTGCCAATT
AGCAGACTATG

GTCACCACTGTGCCTTGATTCC

2. NCBI BLASTn Result:

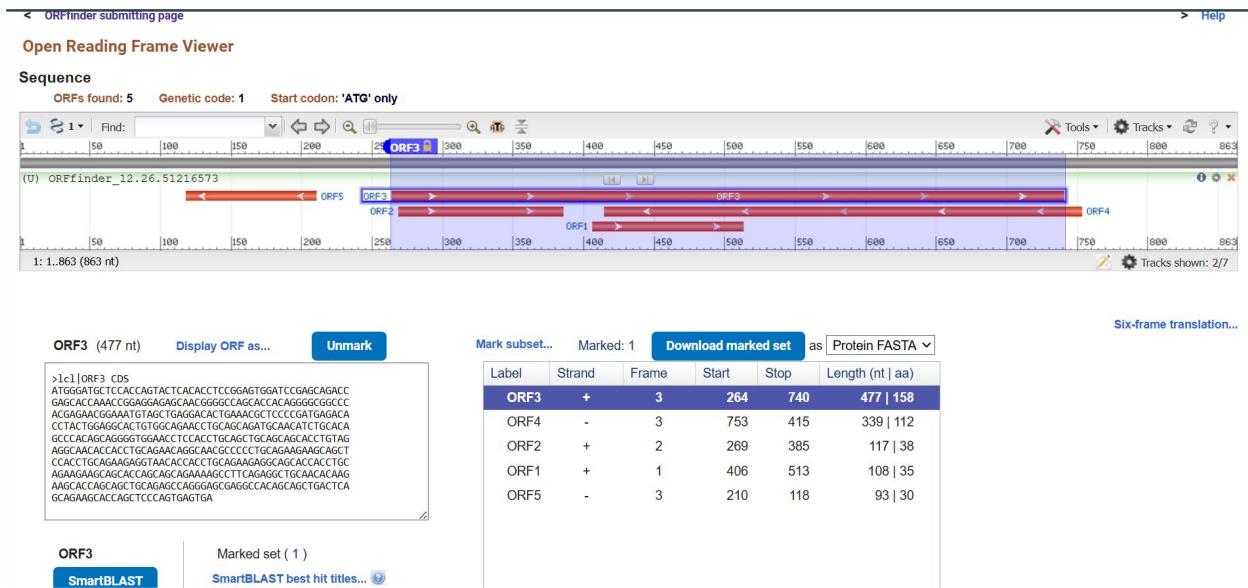
The screenshot shows the NCBI BLASTn search results for the provided sequence. The top navigation bar includes tabs for Descriptions, Graphic Summary, Alignments, and Taxonomy. The main content area is titled "Sequences producing significant alignments". It displays a table with the following columns: Description, Scientific Name, Max Score, Total Score, Query Cover, E value, Per. Ident, Acc. Len, and Accession. There are 6 sequences selected, all of which are PREDICTED entries from Gouania willdenowi. The first two results have 100% identity across the entire length (1594/1594), while the third result has only 8% identity over a small portion. The table also includes download and selection options like GenBank, Graphics, Distance tree of results, and MSA Viewer.

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
PREDICTED: Gouania willdenowi tropomyosin-1 isoforms 33/34 (LOC114471988), transcript variant X2... <i>Gouania willde...</i>	<i>Gouania willde...</i>	1594	1594	100%	0.0	100.00%	863	XM_028461004.1
PREDICTED: Gouania willdenowi tropomyosin-1 isoforms 33/34 (LOC114471988), transcript variant X1... <i>Gouania willde...</i>	<i>Gouania willde...</i>	1491	1688	100%	0.0	99.75%	965	XM_028461003.1
PREDICTED: Gouania willdenowi uncharacterized LOC114471991 (LOC114471991). ncRNA	<i>Gouania willde...</i>	122	122	8%	8e-23	100.00%	908	XR_003675055.1
PREDICTED: Thunnus albacares skin secretory protein xp2-like (LOC122969450), mRNA	<i>Thunnus albac...</i>	58.4	58.4	4%	0.002	100.00%	1416	XM_044335117.1
PREDICTED: Thunnus thynnus uncharacterized protein protein homolog (s1dkey-284p5.3), transcript varia...	<i>Thunnus thynnus</i>	58.4	58.4	4%	0.002	100.00%	1277	XM_067612771.1
PREDICTED: Thunnus maccoyii brain acid soluble protein 1 homolog (LOC121913660), transcript varia...	<i>Thunnus maccoyii</i>	58.4	58.4	4%	0.002	100.00%	1456	XM_042436385.1

Summary:

The results show that my sequence is from a fish called *Gouania willdenowi* (Clingfish), specifically the Tropomyosin 1 gene. The first two results are the best because they match my sequence almost 100% across the entire length. The third result also matches well but only for a very small part (8%) of the sequence. Since the E-value is 0.0, it proves that this identification is correct and not by chance.

3. ORF Map:



Selection of ORF:

After analyzing the ORF Finder results, I selected ORF3 as the final sequence for my project.

Reasoning:

I chose this ORF because it is the longest open reading frame available.

It starts with the 'ATG' (Methionine) codon, which is the standard start signal.

4. Expasy Translate Tool:

Results of translation

- Open reading frames are highlighted in red
- Select your initiator on one of the following frames to retrieve your amino acid sequence

[Download all the translated frames](#)

5'3' Frame 1

MGCSTSTHTSGVDPSPRSTKPEESNGASTTGAAHENGNVAEDTETLPDETPTGGTVAEPAADATSAQPTAGVEPPPAAAAAPVEATP
PAEQATPPAEEAAPAEEVTTPAEEAAPAEEAAPAAEKPSEAATQEAPAAAEPGSEATAADSAEAPAPSE-

5'3' Frame 2

WDAPPVLTPPEWIRADRAPNRRATGPAPQGRPTRTEM-IRTLKRSPMRHILLEALWQNLQQMQHLHSPOQGWNLHLQLQQHL-RQHH
LQRQRPLQKKQLHLQKR-HHLQKRQHHLQKKHQQQQKSLQRLQHKKHQQLQSQGARPQQLTQQKHQLPVS

5'3' Frame 3

GMLHQYSHLRSGSEQTEHQTGGEQRGQHHRGGPRERKCS-GH-NAPR-DTYWRHCGRTCSRNCNICTAHSRGGTSTCSCSSTCRGNTT
CRTGNAPCRSSSTCRRGNTTCRRGTTCRSSSTSSRKAFRCNTRSTSSCRARERGHSS-LSRSTSSQ-V

3'5' Frame 1

Protein translation:

After selecting ORF3, I translated the nucleotide sequence into its corresponding protein sequence. This step is important to understand the functional role of the gene.

GC Content:

Total count, all bases:	477
Adenine (A) count:	146
Thymine (T) count:	44
Guanine (G) count:	135
Cytosine (C) count:	152
%G~C content:	60.2

Longest repeat length: 17

ATGGGATGCTCCACCACTACTCACACCTCCGG
AGTGGATCCGAGCAGACCGAGCACCAAACCGG
AGGAGAGCAACGGGGCCAGCACCACAGGGGCG
GCCACGAGAACGGAAATGTAGCTGAGGACAC
TGAAACGCTCCCCGATGAGACACCTACTGGAG
GCACTGTGGCAGAACCTGCAGCAGATGCAACA
TCTGCACAGCCCACAGCAGGGTGGAACCTCC
ACCTGCAGCTGCAGCAGCACCTGTAGAGGCAA
CACCACCTGCAGAACAGGCAACGCCCTGCA
GAAGAAGCAGCTCCACCTGCAGAAGAGGTAAC
ACCACCTGCAGAACAGAGGCAGCACCACTGCAG
AAGAACGAGCACCAAGCAGCAGAAAAGCCTCA
GAGGCTGCAACACAAGAACAGCAGCTGC
AGAGCCAGGGAGCGAGGCCACAGCAGCTGACT
CAGCAGAACGACCACTCCCAGTGAGTGA

Repeats found: 17