

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:

CTACAACTCACAACGAGAGCCACGTTTGTCAACACAGCAGCTACTACTCAACAGACA
GACAGACAGACAG

ACAGACAGACAGACAGACAGACAGACAGACAGACAGGCTGAGTGGTTTCACTGCA
GGACGCACGGATGCG

CAATTACGCACGAGACTAGAGGAATGTCACCGTCGGCGCGCTGAGCTGTGGTCACTT
GTTGAAACGACAT

CACCTGTGCGTCAGTATAGTCTCTCCCAGTCCTCCCAGAACACTTGTAGCAGAATGGG
ATGCTCCACCAG

TACTCACACCTCCGGAGTGGATCCGAGCAGACCGAGCACCAAACCGGAGGAGAGCA
ACGGGGCCAGCACC

ACAGGGGCGGCCCACGAGAACGGAAATGTAGCTGAGGACACTGAAACGCTCCCCGA
TGAGACACCTACTG

GAGGCACTGTGGCAGAACCTGCAGCAGATGCAACATCTGCACAGCCCACAGCAGGG
GTGGAACCTCCACC

TGCAGCTGCAGCAGCACCTGTAGAGGCAACACCACCTGCAGAACAGGCAACGCCCC
CTGCAGAAGAAGCA

GCTCCACCTGCAGAAGAGGTAACACCACCTGCAGAAGAGGCAGCACCCACCTGCAGA
AGAAGCAGCACCAG

CAGCAGAAAAGCCTTCAGAGGCTGCAACACAAGAAGCACCAGCAGCTGCAGAGCC
AGGGAGCGAGGCCAC

AGCAGCTGACTCAGCAGAAGCACCAGCTCCCAGTGAGTGAAGAGCAGCACCATAAT
GCTAACGCTGTTTG

TATGACTGAGTAACAGTACGAATGTCAGTATGTGTGTGTAAATTGCCTTGTTGCCAATT
AGCAGACTATG

GTCACCACTGTGCCTTTGATTCC

2. NCBI BLASTn Result:

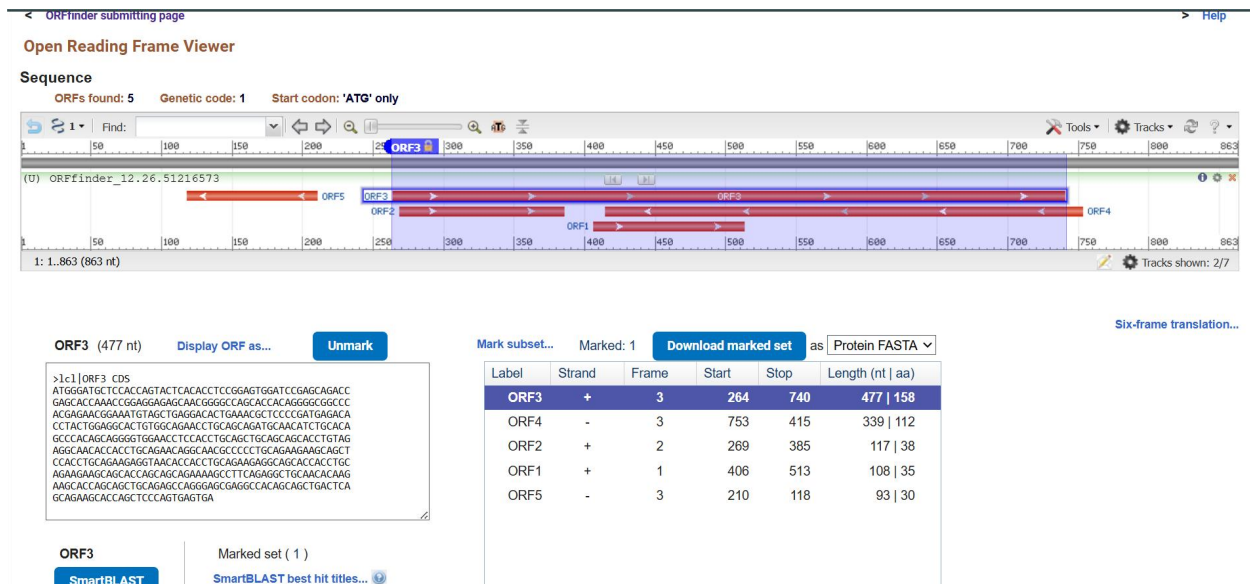
Descriptions	Graphic Summary	Alignments	Taxonomy					
Sequences producing significant alignments								
Download Select columns Show 100								
select all 6 sequences selected								
GenBank Graphics Distance tree of results MSA Viewer								
Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
PREDICTED: Gouania wilidenowi tropomyosin-1 isoforms 33/34 (LOC114471988), transcript variant X2...Gouania wilde...		1594	1594	100%	0.0	100.00%	863	XM_028461004.1
PREDICTED: Gouania wilidenowi tropomyosin-1 isoforms 33/34 (LOC114471988), transcript variant X1...Gouania wilde...		1491	1688	100%	0.0	99.75%	965	XM_028461003.1
PREDICTED: Gouania wilidenowi uncharacterized LOC114471991 (LOC114471991), ncRNA	Gouania wilde...	122	122	8%	8e-23	100.00%	908	XR_003675055.1
PREDICTED: Thunnus albacares skin secretory protein xP2-like (LOC122969450), mRNA	Thunnus albac...	58.4	58.4	4%	0.002	100.00%	1416	XM_044335117.1
PREDICTED: Thunnus thynnus uncharacterized protein protein homolog (si:dkey-284p5.3), transcript v...	Thunnus thynnus	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...	Thunnus maccoyii	58.4	58.4	4%	0.002	100.00%	1456	XM_042436385.1

Feedback

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

MGCSTSTHTSGVDPSPSTKPEESNGASTTGAAHENGVAEDTETLPDETPTGGTVAEPAADATSAQPTAGVEPPPPAAAAAPVEATP
PAEQATPPAEEAAPAEVTPPAEEAAPAEAAAPAEKPSAATQEAPAAEPGSEATAADSAEAPAPSE-

5'3' Frame 2

WDAPPVLTPEWIRADRAPNRRRATGPAPQGRPTRTEM-LRTLKRSPMRHLLLEALWQNLQQMQHLHSPQQGWNLHLQLQQHL-RQHH
LQNRQRPLQKKQLHLQKR-HHLQKRQHHLQKKQHQQKSLQRLQHKKHQQLQSQGARPQQLTQQKHQLPVS

5'3' Frame 3


GMLHQYSHLRSGSEQTEHQTGGEQRGQHHRGGPRERKCS-GH-NAPR-DTYWRHCGRTC SRCNICTAHSRGGTSTCSCSSTCRGNTT
CRTGNAPCRSSSTCRRGNTTCRRGSTTCRRSSTSSRKAFRGCNTRSTSSCRARERGHSS-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

ATGGGATGCTCCACCAGTACTCACACCTCCGG
AGTGGATCCGAGCAGACCGAGCACCAAACCGG
AGGAGAGCAACGGGGGCCAGCACACAGGGGGCG
GCCCACGAGAACGGAAATGTAGCTGAGGACAC
TGAAACGCTCCCCGATGAGACACCTACTGGAG
GCACTGTGGCAGAACCTGCAGCAGATGCAACA
TCTGCACAGCCCACAGCAGGGGTGGAACCTCC
ACCTGCAGCTGCAGCAGCACCTGTAGAGGCAA
CACCACCTGCAGAACAGGCAACGCCCCCTGCA
GAAGAAGCAGCTCCACCTGCAGAAGAGGTAAC
ACCACCTGCAGAAGAGGCAGCACCACCTGCAG
AAGAAGCAGCACCAGCAGCAGAAAAGCCTTCA
GAGGCTGCAACACAAGAAGCACCAGCAGCTGC
AGAGCCAGGGAGCGAGGCCACAGCAGCTGACT
CAGCAGAAGCACCAGCTCCCAGTGAGTGA

Repeats found: 17