Computational Analysis of Unknown DNA Sequences Using Integrated Bioinformatics Tools

DNA SEQUENCE

ATGTCAGATACGGATAGCAGGCGGCTGGAACAGCACGAGCTGCAGGTGGTGCGCATTGGCCAAGGCAACGGCAGCTCCAGTGATAATGCCAGTCTGGACACCACGCCGA
AGCTCTCCAAGCGCAACAGCTCAGAGCGCAGCCTGCCGCTGCGCAGCTACAGCAAATGGTCGCCCGTGGAGCAGGGCGCGACGCTCGTTTGGAGGGACTTGTGTGTCTA
CACGAATGTGGGCTCGGGGCGTGGCATGAAGCGGATCATTAACAACTCAACGGGCGCGGTGCAGCCCGGTTCGCTGATGGCTCTGATGGGCGCCAGTGGCTCGGGCAAG
ACGACGCTAATGTCAACACTGGGCTACAGGCAGCCAGCTGGCACTGTGGTTCAAGGTGATATACTGATCAATGGCCGGCGCGTTGGTCCCTTTATGCATCGCATCAGCGGC
TATGTCTATCAGGACGATCTCTTCATAGGCTCGCTCACTGTGCTGGAGCATCTGAACTTTATGGCCCATCTGCGCTTGGATCGGCGCGTTTCCAGCGAAGAGCGTCGCTTGAT
CATCAAGGAGCTGCTGGAGCGCACGGGTCTGCTTTCCGCCTCCCACACCGCGCATTGGCAGCGGAGACGATGAGAAGGTTCTGTCTGGCGGTGAACGCAAGCGTCTGGCC
TTTGCCGTCGAGCTGCTCAACAATCCGGTGATTCTCTTCTGCGATGAGCCCACAACGGGACTCGATTCGTATAGCGCCCAGCAGCTGGTGGCTACGCTGCACGAGCTGGCA
CAGAAGGGCACCACCATCCTATGCACCATTCATCAGCCCAGCTCTCAGCTCTTCGACAACTTCAACAATGTGATGCTGCTGGCGGATGGGCGTGTCGCCTTCACGGGATCG
CCGCAGCATGCGCTGAGCTTCTTTGCGGACCATGGCTACTATTGCCCGGAGGCGTACAATCCGGCGGACTTTCTTATTGGCGTGCTGGCCACAGATCCCGGCTACGAGCAG
GCCTCCCAGCGCTCCGCGCAGCATCTGTGCGATCAGTTCGCCGTCAGCTCGGCGGCCAAGCAGCGCGATATGTTGGTCAACTTGGAGATTCACATGGCGCAGTCGGGCAA
CTTTCCCTACGATACGGAACTGGACTCCTTCCAGGGCGTCGCCTGGTACAAACGCTTCCATGTCGTCTGGCTGCGTGCCTCGCTGACGCTGCTACGCGACCCGACCATACA
GTGGATGCGCTTTCTGCAGAAGATGGCCATGGCCTTCATCATTGGCGCCTGCTTTGCGGGCACCACGGAGCTAACGCAGCTGGGCGTGCAGGCGGTGCAGGGCACCATCT
TCATCATGATATCGGAGAACACCTACCATCCCATGTACTCCGCGCTGAACATCTTCCCTCAGGGCTTTCCGCTCTTCATGCGTGAAACCCGCTCCGGCTTGTACACCACGGC
TCAGTACTACGTCGCCAACATTTTGGCCATGCTGCCTGGCATGATTATCGAGCCGCTGCTGTTCGTCGTCATCTGCTATTGGCTAACAGGATTGCGTGCCACCTTGTATGCCT
TCGGCATTACGGCCATTTGCGTTGTCCTGGTGATGAATGTGGCCACCGCCTGCGGCTGCTTCTTCTCCACGGCCTTCAACTCGGTGCCGCTGGCGATGGCGTATCTCGTGCC
GCTTGATTACATTATCATGATTACGTCGGGCATCTTCATCAAAATTAGCTCGCTGCCTGTGGCATTCTATTGGACACAGTTTCTATCCTGGATGCTCTATGCGAACGAGGCCAT
GACAATTGCCCAATGGTCAGGCGTGCAGAACATAACCTGCTTTGGAGAGAGCGCTGATTTGCCCTGCTTCCACACCGGGCAGGATGTGCTGGACAAGTATAGCTTCATAG
AGGCGAACCTCTATCGCAATCTCTTGGCCATAGTCGGCATCTACTTCGGCTTCCATCTGCTGGGCTACTACTTCCTCTGGCGCCGCGCGCGCAAGATCTAGGCAGAAATTGG
CGCCGCTTC

INTRODUCTION

The scarlet (st) gene is a well-characterized gene in the fruit fly *Drosophila melanogaster* that plays a crucial role in eye pigmentation. It encodes an ATP-binding cassette (ABC) transporter protein, which is involved in the transport of pigment precursors into pigment granules during eye development.

Gene: Scarlet
Organism: Drosophila *melanogaster* (fruit fly)
Amino Acids: 2006
Location: Chromosome 3 (Autosome)
Accession Number:P45843
Protein: Eye Protein Granule Membrane
No. of matches: 1

DNA ANALYSIS
PROTEIN ANALYSIS
PHYLOGENTIC ANALYSIS
RNA ANALYSIS
SUBCELLULAR LOCALIZATION
PROTEIN INTERACTION

DNA ANALYSIS

1. ALLIGNMENT

Score	Expect	Identities	Gaps	Strand
3705 bits(2006)	0.0	2006/2006(100%)	0/2006(0%)	Plus/Plus

```
Query  1    ATGTCAGATACGGATAGCAGGCGGCTGGAACAGCACGAGCTGCAGGTGGTGCGCATTGGC  60
Sbjct  1    ATGTCAGATACGGATAGCAGGCGGCTGGAACAGCACGAGCTGCAGGTGGTGCGCATTGGC  60

Query  61   CAAGGCAACGGCAGCTCCAGTGATAATGCCAGTCTGGACACCACGCCGAAGCTCTCCAAG  120
Sbjct  61   CAAGGCAACGGCAGCTCCAGTGATAATGCCAGTCTGGACACCACGCCGAAGCTCTCCAAG  120

Query  121  CGCAACAGCTCAGAGCGCAGCCTGCCGCTGCGCAGCTACAGCAAATGGTCGCCCGTGGAG  180
Sbjct  121  CGCAACAGCTCAGAGCGCAGCCTGCCGCTGCGCAGCTACAGCAAATGGTCGCCCGTGGAG  180

Query  181  CAGGGCGCGACGCTCGTTTGGAGGGACTTGTGTGTCTACACGAATGTGGGCTCGGGGCGT  240
Sbjct  181  CAGGGCGCGACGCTCGTTTGGAGGGACTTGTGTGTCTACACGAATGTGGGCTCGGGGCGT  240

Query  241  GGCATGAAGCGGATCATTAACAACTCAACGGGCGCGGTGCAGCCCGGTTCGCTGATGGCT  300
Sbjct  241  GGCATGAAGCGGATCATTAACAACTCAACGGGCGCGGTGCAGCCCGGTTCGCTGATGGCT  300

Query  301  CTGATGGGCGCCAGTGGCTCGGGCAAGACGACGCTAATGTCAACACTGGGCTACAGGCAG  360
Sbjct  301  CTGATGGGCGCCAGTGGCTCGGGCAAGACGACGCTAATGTCAACACTGGGCTACAGGCAG  360
```

2. DESCRIPTION

Reports | Lineage | Organism | Taxonomy

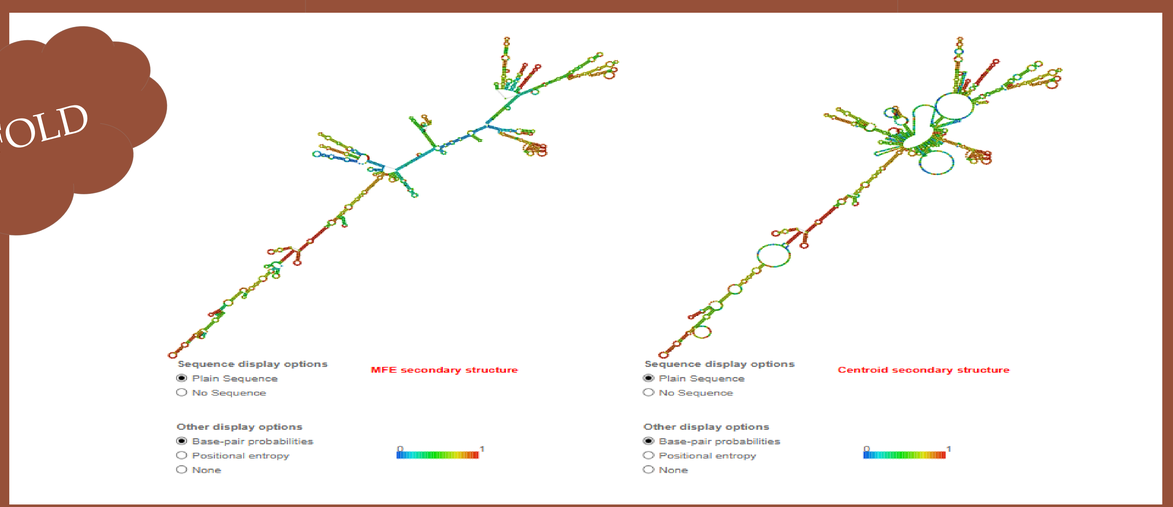
49 sequences selected

Organism	Blast Name	Score	Number of Hits	Description
root			49	
. Acalyptratae	flies		47	
. . Drosophilinae	flies		45	
. . . Drosophila	flies		44	
. . . . Drosophila	flies		16	
. . . . . repleta group	flies		7	
. . . . . . Drosophila hydei	flies	3705	3	Drosophila hydei hits
. . . . . Drosophila arizonae	flies	2915	2	Drosophila arizonae hits

Feedback

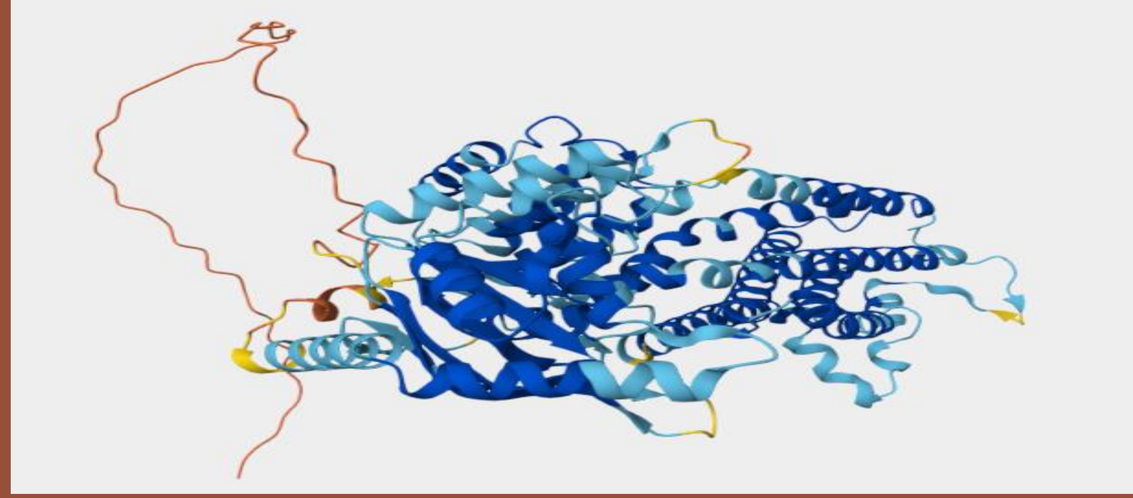
RNA ANALYSIS

RNAFOLD



The selected DNA coding region was transcribed into mRNA by replacing thymine (T) with uracil (U). RNAfold predicted the secondary structure of mRNA based on minimum free energy (MFE), indicating thermodynamically stable folding.

PROTEIN ANALYSIS



ORF VIEWER

ORF MAP

Translate tool

Translate is a tool which allows the translation of a nucleotide (DNA/RNA) sequence to a protein sequence.

DNA or RNA sequence

```
atgtcagatacggatagcaggcggctggaacagcacgagctgcag
gtggtgcgcattggccaaggcaacggcagctccagtgataatgccag
tctggacaccacgccgaagctctccaagcgcaacagctcagagcgc
agcctgccgctgcgcagctacagcaaatggtcgcccgtggagcagg
gcgcgacgctcgtttggagggacttgtgtgtctacacgaatgtgggctc
ggggcgtggcatgaagcggatcattaacaactcaacgggcgcggtg
cagcccggttcgctgatggctctgatgggcgccagtggctcgggcaa
gacgacgctaatgtcaacactgggctacaggcagccagctggcact
gtggttcaaggtgatatactgatcaatggccggcgcgttggtccctttat
gcatcgcatcagcggctatgtctatcaggacgatctcttcataggctcg
```

Output format

- Verbose: Met, Stop, spaces between residues
- Compact: M - no spaces

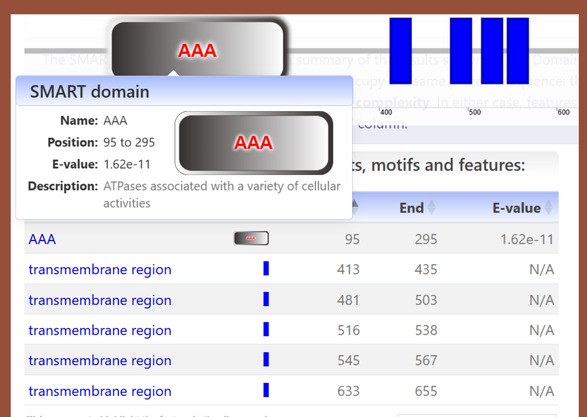
5'3' Frame 1

```
MSDTDSRRLEQHELQVVRIGQGNGSSSDNASLDTTPKLSKRNSSERSLPLRSYSKWSPVEQGATLVWRDLCVYTNVGSGRGMKRIIN
NSTGAVQPGSLMALMGASGSGKTTLMSTLGYRQPAGTVVQGDILINGRRVGPFMHRISGYVYQDDLFIGSLTVLEHLNFMAHLRLDR
RVSSEERRLIIKELLERTGLLSASHTRIGSGDDEKVLSGGERKRLAFAVELLNNPVILFCDEPTTGLDSYSAQQLVATLHELAQKGT
TILCTIHQPSSQLFDNFNNVMLLADGRVAFTGSPQHALSFFADHGYYCPEAYNPADFLIGVLATDPGYEQASQRSAQHLCDQFAVSS
AAKQRDMLVNLEIHMAQSGNFPYDTELDSFQGVAWYKRFHVVWLRASLTLLRDPTIQWMRFLQKMAMAFIIGACFAGTTELTQLGVQ
AVQGTIFIMISENTYHPMYSALNIFFPQGFPLFMRETRSGLYTTAQYYVANILAMLPGMIIEPLLFVVICYWLTGLRATLYAFGITAI
CVVLVMNVATACGCFFSTAFNSVPLAMAYLVPLDYIIMITSGIFIKISSLPVAFYWTQFLSWMLYANEAMTIAQWSGVQNITCFGES
ADLPCFHTGQDVLDKYSFIEANLYRNLLAIVGIYFGFHLLGYYFLWRRARKI-AEIGAA
```

reset | TRANSLATE!

PROPERTIES

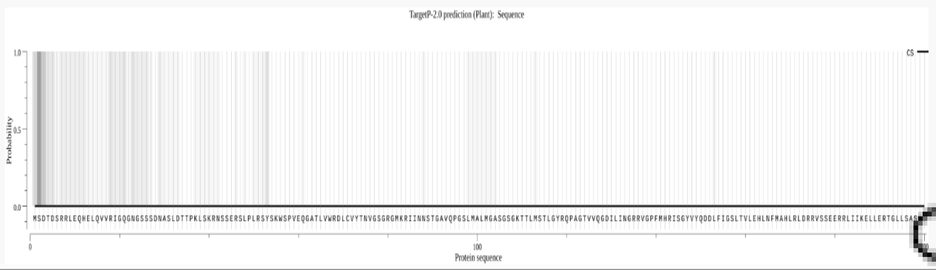
Number of amino acids	667
Theoretical pI	6.73
Molecular weight	74222.47
Instability Index	94.20
Aliphatic Index	94.20



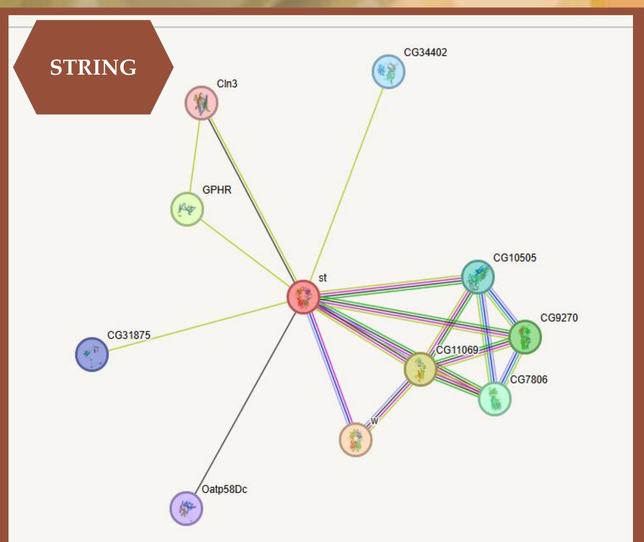
TARGET P

Protein type	Other	Signal peptide	Mitochondrial transfer peptide	Chloroplast transfer peptide	Thylakoid luminal transfer peptide
Likelihood	1	0	0	0	0

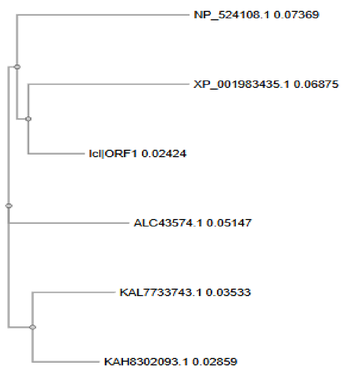
Download: PNG / EPS / Tabular



STRING



PHYLOGENTIC TREE



PROTEIN ALLIGMENT

```
NP_524108.1       -MSDSDSK---RIDVEAPERVEQHELQVMPVGSTI----EVPSLDSTPKLSKRNSSERSL   52
XP_001983435.1    MSSDSSDSRQLLADVPQQQEKEQHELQVVVSRGSGT-----GSADTTPKLSKRNSSERSL   55
ALC43574.1        -MSDEDSSKQLLLTEVQTTRPVSQELQMVTIPPESNTVNEQPSWDNTPKLSKRNSSERSL   59
lcl|ORF1          -MSDTDS-R----------RLEQHELQVVRIGQGNGSSSDNASLDTTPKLSKRNSSERSL   48
KAL7733743.1      -MSDEDS-RQ---------LLDQHELEVVSIGQGSAEKLP-SVETTPSRLSKRNSSERSL   48
KAH8302093.1      -MSDTDG-RQLLGDMGGVVPLDQHELQLVSVGGSDLN---SVDS-TPKLSKRNSSERSL   53
                    **  ..            .:**:::             :***********

NP_524108.1       PLRSYSKWSPTEQGATLVWRDLCVYTNVGGSGQRMKRIINNSTGAIQPGTLMALMGSSGS  112
XP_001983435.1    PLRSYSKWSPTEQGATLVWRDLCVYTNVG-SGRGMKRIINNSTGAVQPGSLMALMGASGS  114
ALC43574.1        PLRSYNKWAPTEHGATLVWRDLCVYTNVG-SGKHMKRIINNSTGAVQPGSLMALMGASGS  118
lcl|ORF1          PLRSYSKWSPVEQGATLVWRDLCVYTNVG-SGRGMKRIINNSTGAVQPGSLMALMGASGS  107
KAL7733743.1      PLRSYSKWSPTEQGATLVWRDLCVYTNVG-SGNKMKRIINNSTGAVQPGSLMALMGASGS  107
KAH8302093.1      PLRSYSKWAPTEHGATLVWRDLCVYTNVG-SGSKMKRIINNSTGAVQPGSLMALMGASGS  112
                  *****.**:*.*:***************** **   **********:***:*******:***

NP_524108.1       GKTTLMSTLAFRQPAGTVVQGDILINGRRIGPFMHRISGYVYQDDLFLGSLTVLEHLNFM  172
XP_001983435.1    GKTTLMSTLAYRQPAGTVVQGDILINGRRVGPFMHRISGYVYQDDLFIGALTVLEHLNFM  174
ALC43574.1        GKTTLMSTLAYRQPAGTVVQGDILINGRRVGPFMHRISGYVYQDDLFIGALTVLEHLNFM  178
lcl|ORF1          GKTTLMSTLGYRQPAGTVVQGDILINGRRVGPFMHRISGYVYQDDLFIGSLTVLEHLNFM  167
KAL7733743.1      GKTTLMSTLGYRQPAGTVVQGDILINGRRVGPFMHRISGYVYQDDLFIGALTVLEHLNFM  167
KAH8302093.1      GKTTLMSTLAYRQPAGTVVQGDILINGRRVGPFMHRISGYVYQDDLFIGALTVLEHLNFM  172
                  *********.:*******************:****************:*:**********
```

CELLO