

San Luc  
PID: A50910657

### Questions:

[Q1] Tell me the name of a protein you are interested in. Include the species and the accession number. This can be a human protein or a protein from any other species as long as it's function is known.

#### Glutathione S-transferase 4 [Caenorhabditis elegans]

NCBI Reference Sequence: NP\_501848.1

[Identical Proteins](#) [FASTA](#) [Graphics](#)

Go to: ☒

LOCUS NP\_501848 207 aa linear INV 09-AUG-2021  
DEFINITION Glutathione S-transferase 4 [Caenorhabditis elegans].  
ACCESSION NP\_501848  
VERSION NP\_501848.1  
DBLINK BioProject: [PRJNA158](#)  
BioSample: [SAMEA3138177](#)  
DBSOURCE REFSEQ: accession [NM\\_069447.8](#)  
KEYWORDS RefSeq.  
SOURCE Caenorhabditis elegans  
ORGANISM [Caenorhabditis elegans](#)  
Eukaryota; Metazoa; Ecdysozoa; Nematoda; Chromadorea; Rhabditida;  
Rhabditina; Rhabditomorpha; Rhabditoidea; Rhabditidae; Peloderinae;  
Caenorhabditis.  
REFERENCE 1 (residues 1 to 207)  
AUTHORS Sulson,J.E. and Waterston,R.  
CONSRM Caenorhabditis elegans Sequencing Consortium  
TITLE Genome sequence of the nematode C. elegans: a platform for  
investigating biology  
JOURNAL Science 282 (5396), 2012-2018 (1998)  
PUBMED [9851916](#)  
REMARK Erratum:[Science 1999 Jan 1;283(5398):35]  
REFERENCE 2 (residues 1 to 207)  
CONSRM NCBI Genome Project  
TITLE Direct Submission  
JOURNAL Submitted (09-AUG-2021) National Center for Biotechnology  
Information, NIH, Bethesda, MD 20894, USA

**Name:** Glutathione S-Transferase

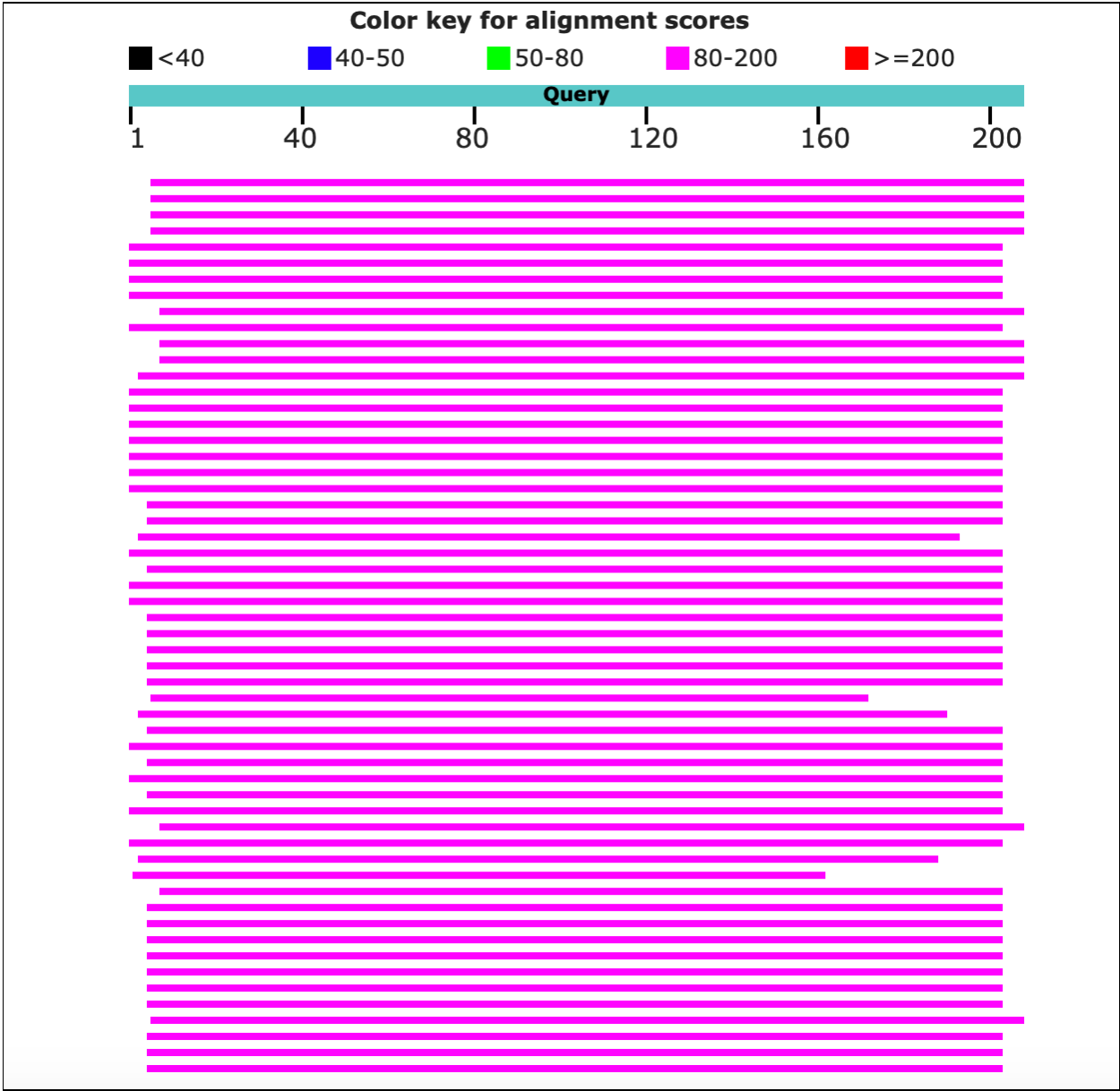
**Accession:** NP\_501848.1

[Q2] Perform a BLAST search against a DNA database, such as a database consisting of genomic DNA or ESTs. The BLAST server can be at NCBI or elsewhere. Include details of the BLAST method used, database searched, and any limits applied (e.g. Organism).

Method: TBLASTN (2.7.1) search against flatworms ESTs

Database: Expressed Sequence Tags (est)

Organism: flatworms (taxid:6157)



Download
GenBank
Graphics
Next
Previous
Descriptions

FY942128 planarian head cDNA library Dugesia japonica cDNA clone Dj\_aH\_304\_P22, mRNA sequence  
Sequence ID: [FY942128.1](#) Length: 716 Number of Matches: 1

Range 1: 7 to 597
GenBank
Graphics
Next Match
Previous Match

| Score         | Expect   | Method                       | Identities  | Positives    | Gaps      | Frame |
|---------------|--|------------------------------|-------------|--------------|-----------|-------|
| 102 bits(255) | 6e-26  | Compositional matrix adjust. | 66/204(32%) | 103/204(50%) | 9/204(4%) | +1    |
| Query 6       | LLYFDARALAEPIRIMFAMLNVPYEDYRVSVEEWSKLKPTTPFGQLPILQVD-GEQFGQS |                              |             |              |           | 64    |
| Sbjct 7       | L YF+AR AE IR + + +V +ED R+ EEW +LKPT P GQLPI+Q+ G +S        |                              |             |              |           | 186   |
| Query 65      | MSITRYLARKFGLAGKTAEEEEAYADSIVDQYRDFIFFFRQFTSSVFYGSADHINKVRFE |                              |             |              |           | 124   |
| Sbjct 187     | MAIARYFAKKYHLTGSNENEEYKVDRVVCTLDD---LFNKVI-DVFHEKDEGKKETLKHE |                              |             |              |           | 354   |
| Query 125     | VVEPARDDFLAIINKFLAKSKSGFLVGDSLWADIADNLTSLKNGFLDFNKEKKLEE     |                              |             |              |           | 184   |
| Sbjct 355     | LNETHLPAFLDRLDYLLKDKNGDFFLGDHPSLADLQLVNVMDHFEES---QYQSHPKLVH |                              |             |              |           | 525   |
| Query 185     | FYNKI-HSIPEIKNYVATRKDSIV                                     |                              |             |              |           | 207   |
| Sbjct 526     | CYQKVLHYPKLKHYKDNQKSII                                       |                              |             |              |           | 597   |

Chosen sequence: FY942128 planarian head cDNA library Dugesia japonica cDNA clone Dj\_aH\_304\_P22, mRNA sequence.

| Score         | Expect | Identities  | Positives    | Gaps      |
|---------------|--------|-------------|--------------|-----------|
| 102 bits(255) | 6e-26  | 66/204(32%) | 103/204(50%) | 9/204(4%) |

Query6

LLYFDARALAEPIRIMFAMLNVPYEDYRVSVEEWSKLKPTTPFGQLPILQVD-GEQFGQS

64

L YF+AR AE IR + + +V +ED R+ EEW +LKPT P GQLPI+Q+ G +S

Sbjct7

LTYFNARGKAELIRFVLIVSDVEFEDKRIEFEEWPQLKPTIPTGQLPIVQLSCGGIINES

186

Query65

MSITRYLARKFGLAGKTAEEEEAYADSIVDQYRDFIFFFRQFTSSVFYGSADHINKVRFE

124

M+I RY A+K+ L G EE D +V D F + VF+ D ++ E

Sbjct187

MAIARYFAKKYHLTGSNENEEYKVDRVVCTLDD---LFNKVI-DVFHEKDEGKKETLKHE

354

Query125

VVEPARDDFLAIINKFLAKSKSGFLVGDSLWADIADNLTSLKNGFLDFNKEKKLEE

184

+ E FL ++ +L F +GD + AD+ + + + ++ + KL

Sbjct355

LNETHLPAFLDRLDYLLKDKNGDFFLGDHPSLADLQLVNVMDHFEES---QYQSHPKLVH

525

Query185

FYNKI-HSIPEIKNYVATRKDSIV

207

Y K+ P++K+Y R+ SI+

Sbjct526

CYQKVLHYPKLKHYKDNQKSII

597

[Q3] Gather information about this “novel” protein. At a minimum, show me the protein sequence of the “novel” protein as displayed in your BLAST results from [Q2] as FASTA format (you can copy and paste the aligned sequence subject lines from your BLAST result page if necessary) or translate your novel DNA sequence using a tool called EMBOSS Transeq at the EBI. Don’t forget to translate all six reading frames; the ORF (open reading frame) is likely to be

the longest sequence without a stop codon. It may not start with a methionine if you don't have the complete coding region. Make sure the sequence you provide includes a header/subject line and is in traditional FASTA format.

**Chosen sequence:**

```
>FY942128.1_1 planarian head cDNA library Dugesia japonica cDNA clone Dj_aH_304_P22,  
mRNA sequence
```

```
IILTYFNARGKAELIRFVLIVSDVEFEDKRIEFEEWPQLKPTIPTGQLPIVQLSCGGIIN  
ESMAIARYFAKKYHLTGSNENEYKVDREVCTLDDLFNKVIDVFHEKDEGKKETLKHELN  
ETHLPAFLDRLDYLLKDKNGDFFLGDHPSLADLQLVNVMDHFEESQYQSHPKLVHCYQKV  
LEHYPKLKHYKDNRQKSII*KNSFTVSEYL*KLMKLF*LFQKLMIINLLLIVEKKKKKX
```

```
Eukaryota; Metazoa; Spiralia; Lophotrochozoa; Platyhelminthes;  
Rhabditophora; Seriata; Tricladida; Continenticola; Geoplanioidea;  
Dugesidae; Dugesia.
```

**[Q4]** Prove that this gene, and its corresponding protein, are novel. For the purposes of this project, “novel” is defined as follows. Take the protein sequence (your answer to [Q3]), and use it as a query in a blastp search of the nr database at NCBI.

- If there is a match with 100% amino acid identity to a protein in the database, from the same species, then your protein is NOT novel (even if the match is to a protein with a name such as “unknown”). Someone has already found and annotated this sequence, and assigned it an accession number.
- If the top match reported has less than 100% identity, then it is likely that your protein is novel, and you have succeeded.
- If there is a match with 100% identity, but to a different species than the one you started with, then you have likely succeeded in finding a novel gene.
- If there are no database matches to the original query from [Q1], this indicates that you have partially succeeded: yes, you may have found a new gene, but no, it is not actually homologous to the original query. You should probably start over.

RID

PZZB63XM01N

Search expires on 10-21 19:28 pm

Download All

Program

BLASTP

Citation

Database

nr

See details

Query ID

lcl|Query\_153764

Description

FY942128.1\_1 planarian head cDNA library Dugesia japoni ...

Molecule type

amino acid

Query Length

239

Other reports

Distance tree of results

Multiple alignment

MSA viewer

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

Sequences producing significant alignments

Download

New

Select columns

Show

100

select all

100 sequences selected

GenPept

Graphics

Distance tree of results

Multiple alignment

New

MSA Viewer

|                                     | Description   | Scientific Name      | Max Score | Total Score | Query Cover | E value | Per. Ident | Acc. Len | Accession      |
|-------------------------------------|---|----------------------|-----------|-------------|-------------|---------|------------|----------|----------------|
| <input checked="" type="checkbox"/> | LOW QUALITY PROTEIN: glutathione S-transferase-like [Crassostrea virginica] | Crassostrea virgi... | 148       | 148         | 80%         | 2e-40   | 36.22%     | 203      | XP_022319415.1 |
| <input checked="" type="checkbox"/> | glutathione S-transferase-like [Crassostrea virginica]                      | Crassostrea virgi... | 147       | 147         | 80%         | 3e-40   | 36.73%     | 203      | XP_022316933.1 |
| <input checked="" type="checkbox"/> | hypothetical protein P879_00235 [Paragonimus westermani]                    | Paragonimus we...    | 145       | 145         | 81%         | 2e-39   | 39.41%     | 206      | KAF8572165.1   |
| <input checked="" type="checkbox"/> | glutathione S-transferase [Crassostrea gigas]                               | Crassostrea gigas    | 141       | 141         | 80%         | 1e-37   | 34.69%     | 203      | XP_011444849.1 |
| <input checked="" type="checkbox"/> | glutathione S-transferase-like [Gigantopelta aegis]                         | Gigantopelta aegis   | 140       | 140         | 80%         | 2e-37   | 36.55%     | 204      | XP_041346664.1 |
| <input checked="" type="checkbox"/> | S-crystallin SL11 [Crassostrea gigas]                                       | Crassostrea gigas    | 138       | 138         | 81%         | 1e-36   | 39.90%     | 201      | XP_034305104.1 |
| <input checked="" type="checkbox"/> | Glutathione S-transferase [Fasciola hepatica]                               | Fasciola hepatica    | 138       | 138         | 82%         | 3e-36   | 35.61%     | 226      | THD22549.1     |
| <input checked="" type="checkbox"/> | S-crystallin SL11-like [Crassostrea gigas]                                  | Crassostrea gigas    | 137       | 137         | 81%         | 3e-36   | 38.89%     | 201      | XP_034305105.1 |