Karen Guerrero

[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.

If you do not have a favorite protein, select human RBP4 or KIF11. Do not use beta globin as this is in the worked example report that I provide you with online.

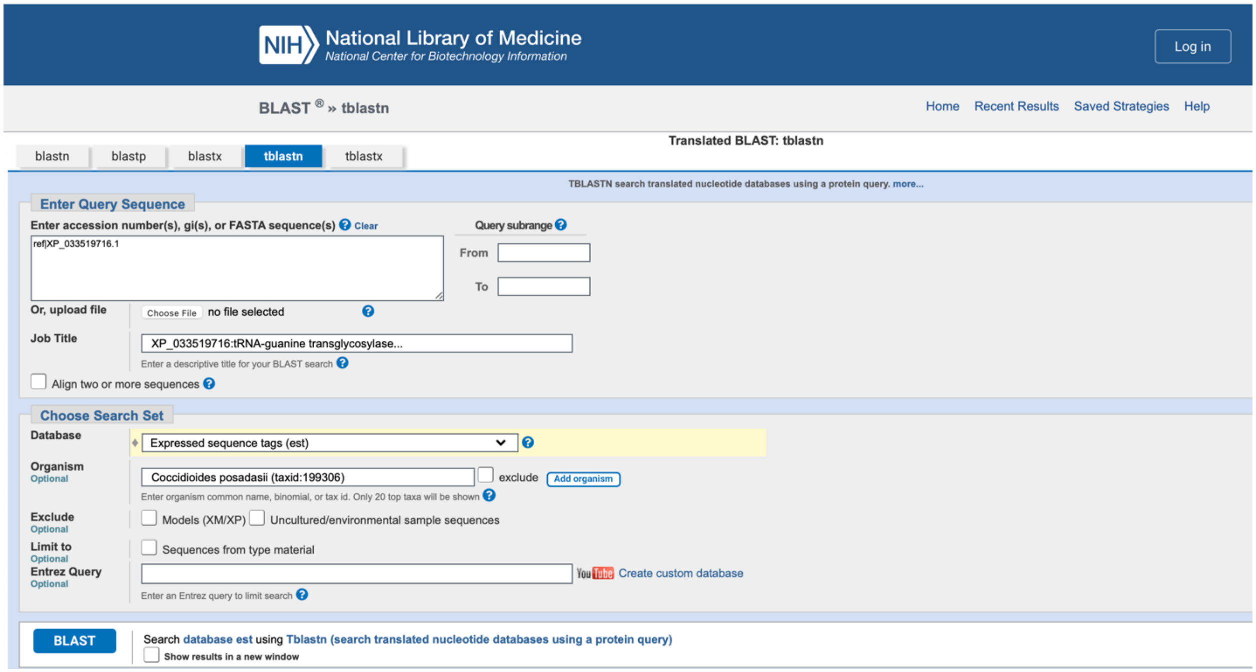
Protein Name: tRNA-guanine transglycosylase
Species: *Dothidotthia symphoricarpi CBS 119687*
Accession Number: XP_033519716
Function Known: Exchanges guanine for queuine in the following tRNA (HIS, TYR, ASP or ASN)

[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). On the BLAST results, clearly indicate a match that represents a protein sequence, encoded from some DNA sequence, that is homologous to your query protein. I need to be able to inspect the pairwise alignment you have selected, including the E value and score. It should be labeled a "genomic clone" or "mRNA sequence", etc. - but include no functional annotation.

BLAST Method: TBLASTN (2.13.0+)
Database searched: Expressed sequence tags (est)
Limits applied (Organism): Coccidioides posadasii (taxid:199306)

Chosen match: G874P535RD8.T0 C. posadasii Silveira, 72HR SPHERULE_NORMALIZED Coccidioides posadasii cDNA, mRNA sequence
Accession Number: GH437857
Organism: Coccidioides posadasii
E Value: 2e-85
Percentage Identity: 47.19%

Karen Guerrero



NIH National Library of Medicine
National Center for Biotechnology Information
Log in

BLAST ® » tblastn

Home Recent Results Saved Strategies Help
blastn blastp blastx tblastn tblastx

Translated BLAST: tblastn

TBLASTN search translated nucleotide databases using a protein query. more...

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) ? Clear

ref|XP_033519716.1

Query subrange ?

From

To

Or, upload file: Choose File no file selected ?

Job Title: XP_033519716:tRNA-guanine transglycosylase...

Enter a descriptive title for your BLAST search ?

□ Align two or more sequences ?

Choose Search Set

Database: Expressed sequence tags (est) ?

Organism (Optional): Coccidioides posadasii (taxid:199306) □ exclude Add organism

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown ?

Exclude (Optional): □ Models (XM/XP) □ Uncultured/environmental sample sequences

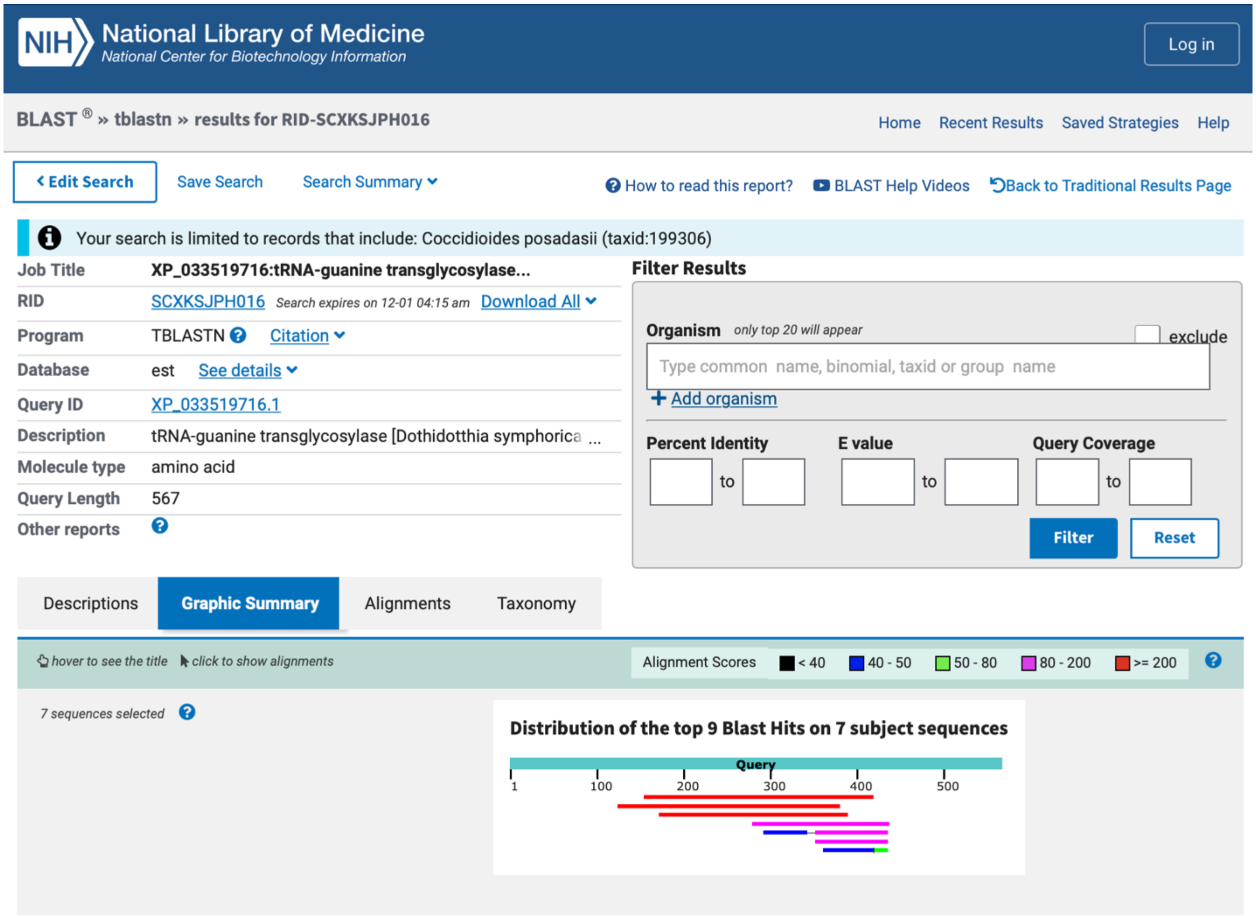
Limit to (Optional): □ Sequences from type material

Entrez Query (Optional): YouTube Create custom database

Enter an Entrez query to limit search ?

BLAST — Search database est using Tblastn (search translated nucleotide databases using a protein query)

□ Show results in a new window



NIH National Library of Medicine
National Center for Biotechnology Information
Log in

BLAST ® » tblastn » results for RID-SCXKSJPH016

Home Recent Results Saved Strategies Help
‹ Edit Search
Save Search
Search Summary
How to read this report?
BLAST Help Videos
Back to Traditional Results Page

Your search is limited to records that include: Coccidioides posadasii (taxid:199306)

Job Title	XP_033519716:tRNA-guanine transglycosylase...
RID	SCXKSJPH016 Search expires on 12-01 04:15 am Download All
Program	TBLASTN ? Citation
Database	est See details
Query ID	XP_033519716.1
Description	tRNA-guanine transglycosylase [Dothidotthia symphorica ...
Molecule type	amino acid
Query Length	567
Other reports	?

Filter Results

Organism *only top 20 will appear* □ exclude

Type common name, binomial, taxid or group name

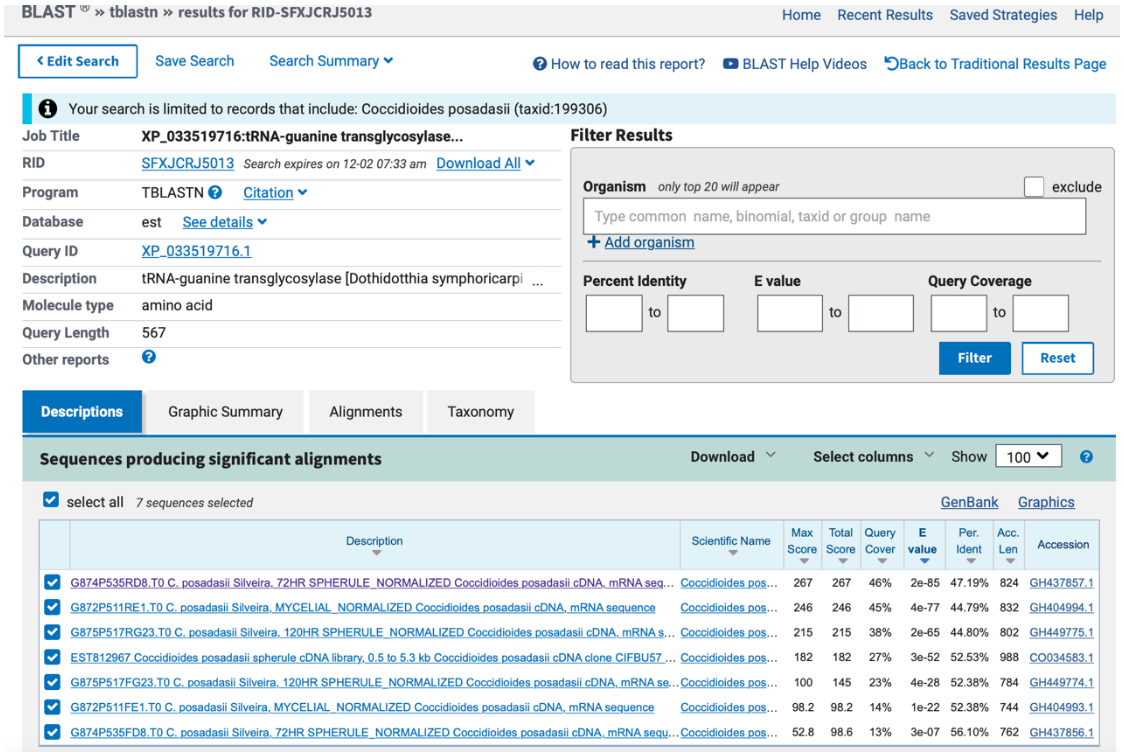
+ Add organism

Percent Identity ___ to ___   E value ___ to ___   Query Coverage ___ to ___

Alignment Scores ■ < 40 🟦 40 - 50 🟩 50 - 80 🟪 80 - 200 🟥 >= 200

7 sequences selected ?

Distribution of the top 9 Blast Hits on 7 subject sequences



BLAST® » tblastn » results for RID-SFXJCRJ5013

Your search is limited to records that include: Coccidioides posadasii (taxid:199306)

Job Title	XP_033519716:tRNA-guanine transglycosylase...
RID	SFXJCRJ5013 *Search expires on 12-02 07:33 am* Download All ⌄
Program	TBLASTN Citation ⌄
Database	est See details ⌄
Query ID	XP_033519716.1
Description	tRNA-guanine transglycosylase [Dothidotthia symphoricarpi ...
Molecule type	amino acid
Query Length	567
Other reports	

Filter Results

Organism *only top 20 will appear* □ exclude

Type common name, binomial, taxid or group name

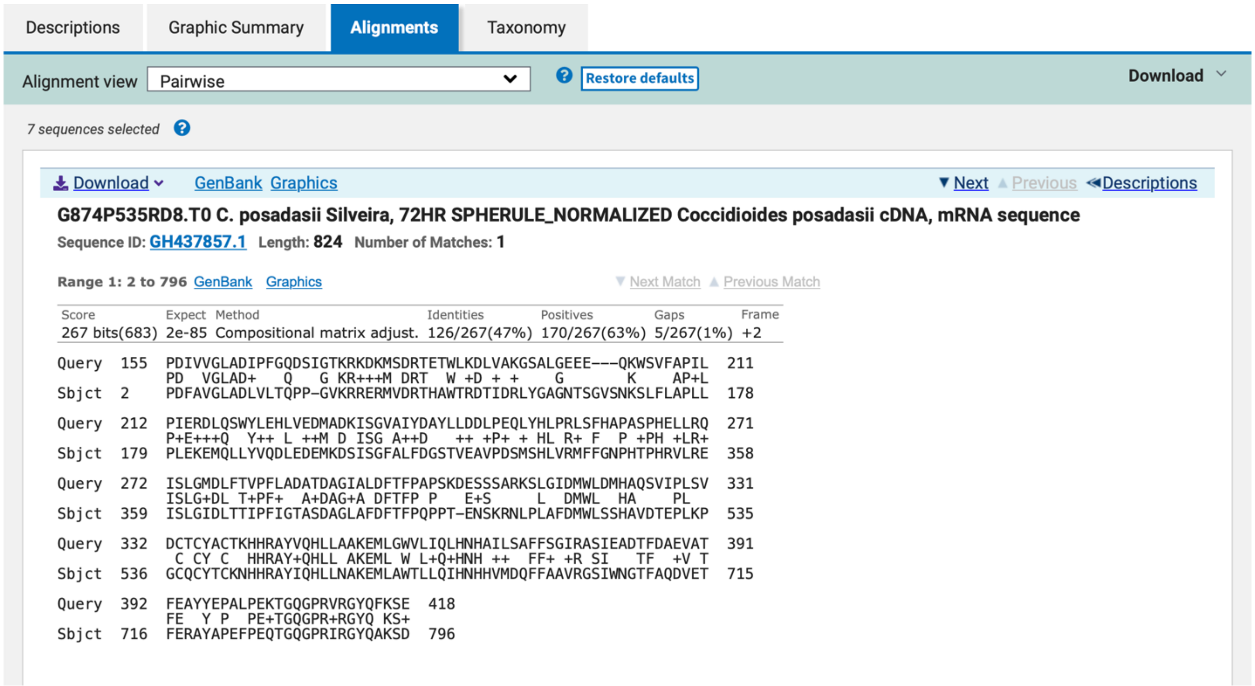
+ Add organism

Percent Identity ___ to ___ E value ___ to ___ Query Coverage ___ to ___

Sequences producing significant alignments

☑ select all *7 sequences selected*

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	G874P535RD8.T0 C. posadasii Silveira, 72HR SPHERULE_NORMALIZED Coccidioides posadasii cDNA, mRNA seq...	Coccidioides pos...	267	267	46%	2e-85	47.19%	824	GH437857.1
☑	G872P511RE1.T0 C. posadasii Silveira, MYCELIAL_NORMALIZED Coccidioides posadasii cDNA, mRNA sequence	Coccidioides pos...	246	246	45%	4e-77	44.79%	832	GH404994.1
☑	G875P517RG23.T0 C. posadasii Silveira, 120HR SPHERULE_NORMALIZED Coccidioides posadasii cDNA, mRNA s...	Coccidioides pos...	215	215	38%	2e-65	44.80%	802	GH449775.1
☑	EST812967 Coccidioides posadasii spherule cDNA library, 0.5 to 5.3 kb Coccidioides posadasii cDNA clone CIFBU57 ...	Coccidioides pos...	182	182	27%	3e-52	52.53%	988	CO034583.1
☑	G875P517FG23.T0 C. posadasii Silveira, 120HR SPHERULE_NORMALIZED Coccidioides posadasii cDNA, mRNA se...	Coccidioides pos...	100	145	23%	4e-28	52.38%	784	GH449774.1
☑	G872P511FE1.T0 C. posadasii Silveira, MYCELIAL_NORMALIZED Coccidioides posadasii cDNA, mRNA sequence	Coccidioides pos...	98.2	98.2	14%	1e-22	52.38%	744	GH404993.1
☑	G874P535FD8.T0 C. posadasii Silveira, 72HR SPHERULE_NORMALIZED Coccidioides posadasii cDNA, mRNA sequ...	Coccidioides pos...	52.8	98.6	13%	3e-07	56.10%	762	GH437856.1



Alignment view Pairwise

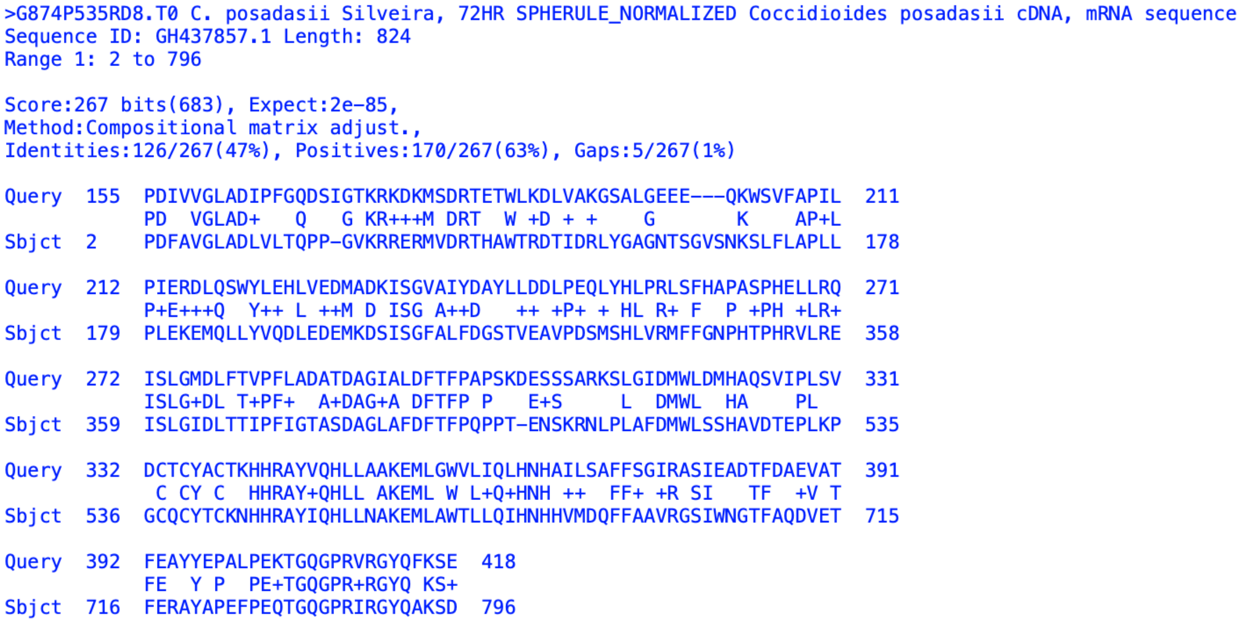
7 sequences selected

G874P535RD8.T0 C. posadasii Silveira, 72HR SPHERULE_NORMALIZED Coccidioides posadasii cDNA, mRNA sequence

Sequence ID: GH437857.1 Length: 824 Number of Matches: 1

Range 1: 2 to 796 GenBank Graphics

```
Score        Expect  Method                        Identities     Positives      Gaps       Frame
267 bits(683) 2e-85  Compositional matrix adjust.  126/267(47%)  170/267(63%)  5/267(1%)  +2

Query  155  PDIVVGLADIPFGQDSIGTKRKDKMSDRTETWLKDLVAKGSALGEEE---QKWSVFAPIL  211
            PD  VGLAD+   Q   G KR+++M DRT  W +D + +    G       K    AP+L
Sbjct  2    PDFAVGLADLVLTQPP-GVKRRERMVDRTHAWTRDTIDRLYGAGNTSGVSNKSLFLAPLL  178

Query  212  PIERDLQSWYLEHLVEDMADKISGVAIYDAYLLDDLPEQLYHLPRLSFHAPASPHELLRQ  271
            P+E+++Q  Y++ L ++M D ISG A++D   ++ +P+ + HL R+ F  P +PH +LR+
Sbjct  179  PLEKEMQLLYVQDLEDEMKDSISGFALFDGSTVEAVPDSMSHLVRMFFGNPHTPHRVLRE  358

Query  272  ISLGMDLFTVPFLADATDAGIALDFTFPAPSKDESSSARKSLGIDMWLDMHAQSVIPLSV  331
            ISLG+DL T+PF+  A+DAG+A DFTFP P   E+S     L  DMWL  HA    PL
Sbjct  359  ISLGIDLTTIPFIGTASDAGLAFDFTFPQPPT-ENSKRNLPLAFDMWLSSHAVDTEPLKP  535

Query  332  DCTCYACTKHHRAYVQHLLAAKEMLGWVLIQLHNHAILSAFFSGIRASIEADTFDAEVAT  391
             C CY C  HHRAY+QHLL AKEML W L+Q+HNH ++  FF+ +R SI   TF  +V T
Sbjct  536  GCQCYTCKNHHRAYIQHLLNAKEMLAWTLLQIHNHHVMDQFFAAVRGSIWNGTFAQDVET  715

Query  392  FEAYYEPALPEKTGQGPRVRGYQFKSE  418
            FE  Y P  PE+TGQGPR+RGYQ KS+
Sbjct  716  FERAYAPEFPEQTGQGPRIRGYQAKSD  796
```



```
>G874P535RD8.T0 C. posadasii Silveira, 72HR SPHERULE_NORMALIZED Coccidioides posadasii cDNA, mRNA sequence
Sequence ID: GH437857.1 Length: 824
Range 1: 2 to 796

Score:267 bits(683), Expect:2e-85,
Method:Compositional matrix adjust.,
Identities:126/267(47%), Positives:170/267(63%), Gaps:5/267(1%)

Query  155  PDIVVGLADIPFGQDSIGTKRKDKMSDRTETWLKDLVAKGSALGEEE---QKWSVFAPIL  211
            PD  VGLAD+   Q   G KR+++M DRT  W +D + +    G        K    AP+L
Sbjct  2    PDFAVGLADLVLTQPP-GVKRRERMVDRTHAWTRDTIDRLYGAGNTSGVSNKSLFLAPLL  178

Query  212  PIERDLQSWYLEHLVEDMADKISGVAIYDAYLLDDLPEQLYHLPRLSFHAPASPHELLRQ  271
            P+E+++Q  Y++ L ++M D ISG A++D   ++ +P+ + HL R+ F  P +PH +LR+
Sbjct  179  PLEKEMQLLYVQDLEDEMKDSISGFALFDGSTVEAVPDSMSHLVRMFFGNPHTPHRVLRE  358

Query  272  ISLGMDLFTVPFLADATDAGIALDFTFPAPSKDESSSARKSLGIDMWLDMHAQSVIPLSV  331
            ISLG+DL T+PF+  A+DAG+A DFTFP P   E+S     L  DMWL  HA    PL
Sbjct  359  ISLGIDLTTIPFIGTASDAGLAFDFTFPQPPT-ENSKRNLPLAFDMWLSSHAVDTEPLKP  535

Query  332  DCTCYACTKHHRAYVQHLLAAKEMLGWVLIQLHNHAILSAFFSGIRASIEADTFDAEVAT  391
             C CY C  HHRAY+QHLL AKEML W L+Q+HNH ++  FF+ +R SI   TF  +V T
Sbjct  536  GCQCYTCKNHHRAYIQHLLNAKEMLAWTLLQIHNHHVMDQFFAAVRGSIWNGTFAQDVET  715

Query  392  FEAYYEPALPEKTGQGPRVRGYQFKSE  418
            FE  Y P  PE+TGQGPR+RGYQ KS+
Sbjct  716  FERAYAPEFPEQTGQGPRIRGYQAKSD  796
```

[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. Here, tell me the name of the novel protein, and the species from which it derives. It is very unlikely (but still definitely possible) that you will find a novel gene from an organism such as S. cerevisiae, human or mouse, because those genomes have already been thoroughly annotated. It is more likely that you will discover a new gene in a genome that is currently being sequenced, such as bacteria or plants or protozoa.

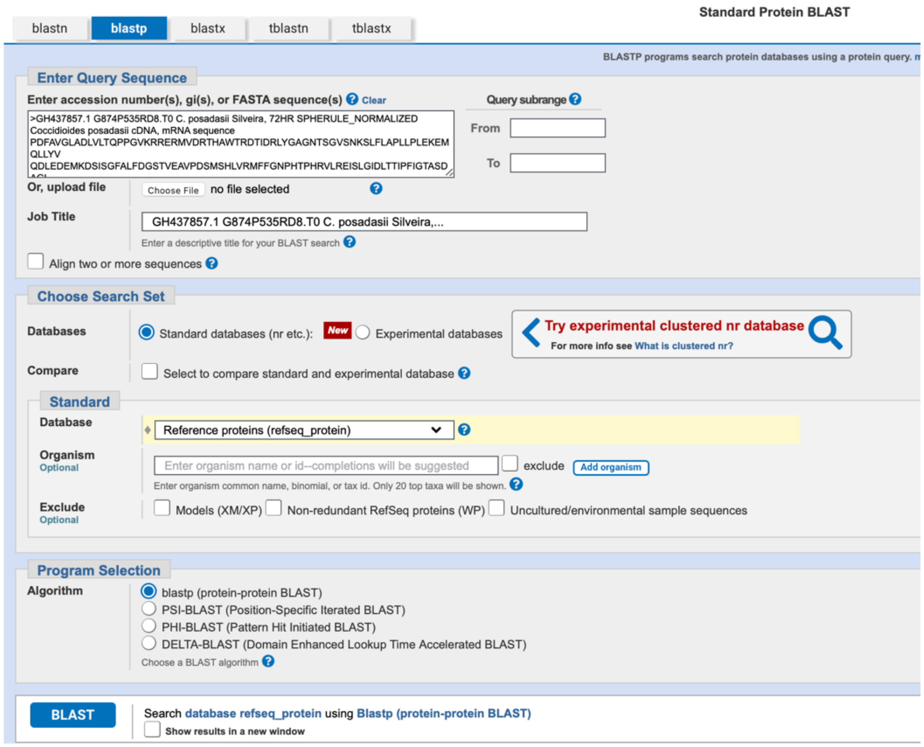
Chosen Sequence:

> G874P535RD8.T0 C. posadasii Silveira, 72HR SPHERULE_NORMALIZED Coccidioides posadasii cDNA, mRNA sequence, Coccidioides posadasii tRNA-guanine transglycosylase
PDFAVGLADLVLTQPPGVKRRERMVDRTHAWTRDTIDRLYGAGNTSGVSNKSLFLAPL
LPLEKEMQLLYVQDLEDEMKDSISGFALFDGSTVEAVPDSMSHLVRMFFGNPHTPHRVL
REISLGIDLTTIPFIGTASDAGLAFDFTFPQPPTENSKRNLPLAFDMWLSSHAVDTEPLKPG
CQCYTCKNHHRAYIQHLLNAKEMLAWTLLQIHNHHVMDQFFAAVRGSIWNGTFAQDV
ETFERAYAPEFPEQTGQGPRIRGYQAKSD

Name: Coccidioides posadasii tRNA-guanine transglycosylase
Species: Coccidioides posadasii

[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.

Method: BLASTP (2.13.0+)
Database: Reference proteins (refseq_protein)
Screenshot of Protein BLAST set up page:



Enter accession number(s), gi(s), or FASTA sequence(s)

```
>GH437857.1 G874P535RD8.T0 C. posadasii Silveira, 72HR SPHERULE_NORMALIZED
Coccidioides posadasii cDNA, mRNA sequence
PDFAVGLADLVLTQPPGVKRRERMVDRTHAWTRDTIDRLYGAGNTSGVSNKSLFLAPLLPLEKEM
QLLYV
QDLEDEMKDSISGFALFDGSTVEAVPDSMSHLVRMFFGNPHTPHRVLREISLGIDLTTIPFIGTASD
```

Query subrange: From: To:

Or, upload file: no file selected

Job Title: GH437857.1 G874P535RD8.T0 C. posadasii Silveira,...
Enter a descriptive title for your BLAST search

□ Align two or more sequences

Databases: ● Standard databases (nr etc.): ○ Experimental databases

Compare: □ Select to compare standard and experimental database

Database: Reference proteins (refseq_protein)

Organism (Optional): Enter organism name or id--completions will be suggested □ exclude
Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown.

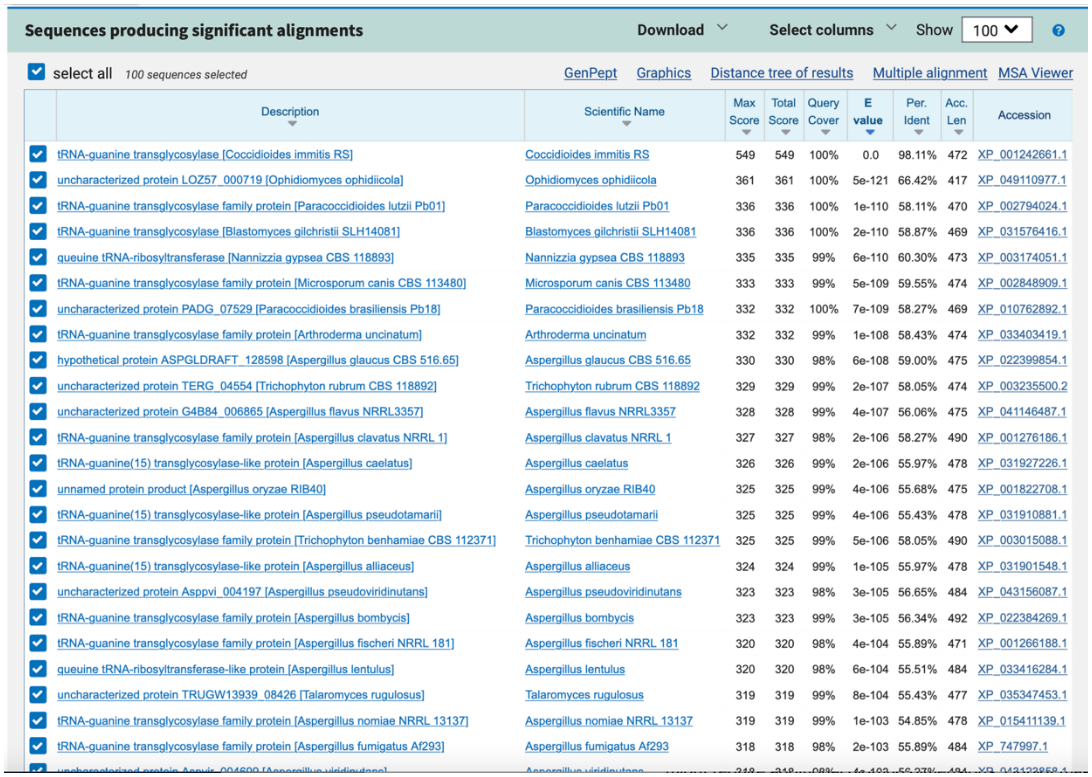
Exclude (Optional): □ Models (XM/XP) □ Non-redundant RefSeq proteins (WP) □ Uncultured/environmental sample sequences

Algorithm:
● blastp (protein-protein BLAST)
○ PSI-BLAST (Position-Specific Iterated BLAST)
○ PHI-BLAST (Pattern Hit Initiated BLAST)
○ DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)
Choose a BLAST algorithm

Search database refseq_protein using Blastp (protein-protein BLAST)
□ Show results in a new window

Karen Guerrero

Screenshot of top hits (results of protein BLAST):

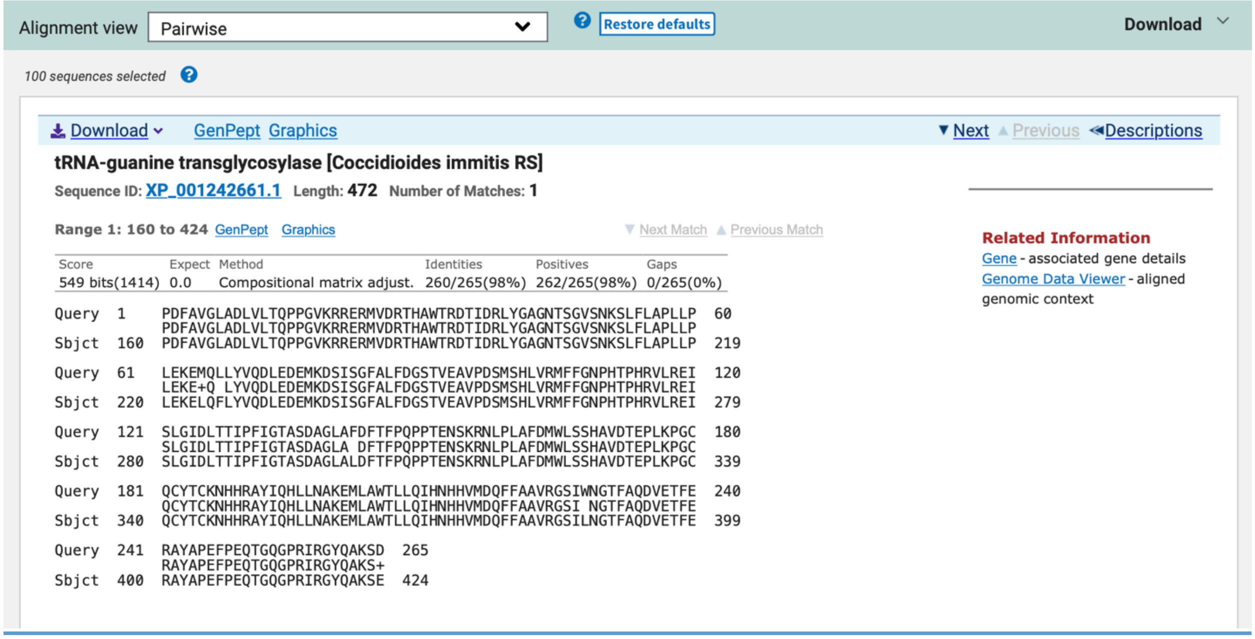


Sequences producing significant alignments

☑ select all *100 sequences selected*

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	tRNA-guanine transglycosylase [Coccidioides immitis RS]	Coccidioides immitis RS	549	549	100%	0.0	98.11%	472	XP_001242661.1
☑	uncharacterized protein LOZ57_000719 [Ophidiomyces ophidiicola]	Ophidiomyces ophidiicola	361	361	100%	5e-121	66.42%	417	XP_049110977.1
☑	tRNA-guanine transglycosylase family protein [Paracoccidioides lutzii Pb01]	Paracoccidioides lutzii Pb01	336	336	100%	1e-110	58.11%	470	XP_002794024.1
☑	tRNA-guanine transglycosylase [Blastomyces gilchristii SLH14081]	Blastomyces gilchristii SLH14081	336	336	100%	2e-110	58.87%	469	XP_031576416.1
☑	queuine tRNA-ribosyltransferase [Nannizzia gypsea CBS 118893]	Nannizzia gypsea CBS 118893	335	335	99%	6e-110	60.30%	473	XP_003174051.1
☑	tRNA-guanine transglycosylase family protein [Microsporum canis CBS 113480]	Microsporum canis CBS 113480	333	333	99%	5e-109	59.55%	474	XP_002848909.1
☑	uncharacterized protein PADG_07529 [Paracoccidioides brasiliensis Pb18]	Paracoccidioides brasiliensis Pb18	332	332	100%	7e-109	58.27%	469	XP_010762892.1
☑	tRNA-guanine transglycosylase family protein [Arthroderma uncinatum]	Arthroderma uncinatum	332	332	99%	1e-108	58.43%	474	XP_033403419.1
☑	hypothetical protein ASPGLDRAFT_128598 [Aspergillus glaucus CBS 516.65]	Aspergillus glaucus CBS 516.65	330	330	98%	6e-108	59.00%	475	XP_022399854.1
☑	uncharacterized protein TERG_04554 [Trichophyton rubrum CBS 118892]	Trichophyton rubrum CBS 118892	329	329	99%	2e-107	58.05%	474	XP_003235500.2
☑	uncharacterized protein G4B84_006865 [Aspergillus flavus NRRL3357]	Aspergillus flavus NRRL3357	328	328	99%	4e-107	56.06%	475	XP_041146487.1
☑	tRNA-guanine transglycosylase family protein [Aspergillus clavatus NRRL 1]	Aspergillus clavatus NRRL 1	327	327	98%	2e-106	58.27%	490	XP_001276186.1
☑	tRNA-guanine(15) transglycosylase-like protein [Aspergillus caelatus]	Aspergillus caelatus	326	326	99%	2e-106	55.97%	478	XP_031927226.1
☑	unnamed protein product [Aspergillus oryzae RIB40]	Aspergillus oryzae RIB40	325	325	99%	4e-106	55.68%	475	XP_001822708.1
☑	tRNA-guanine(15) transglycosylase-like protein [Aspergillus pseudotamarii]	Aspergillus pseudotamarii	325	325	99%	4e-106	55.43%	478	XP_031910881.1
☑	tRNA-guanine transglycosylase family protein [Trichophyton benhamiae CBS 112371]	Trichophyton benhamiae CBS 112371	325	325	99%	5e-106	58.05%	490	XP_003015088.1
☑	tRNA-guanine(15) transglycosylase-like protein [Aspergillus alliaceus]	Aspergillus alliaceus	324	324	99%	1e-105	55.97%	478	XP_031901548.1
☑	uncharacterized protein Asppvi_004197 [Aspergillus pseudoviridinutans]	Aspergillus pseudoviridinutans	323	323	98%	3e-105	56.65%	484	XP_043156087.1
☑	tRNA-guanine transglycosylase family protein [Aspergillus bombycis]	Aspergillus bombycis	323	323	99%	3e-105	56.34%	492	XP_022384269.1
☑	tRNA-guanine transglycosylase family protein [Aspergillus fischeri NRRL 181]	Aspergillus fischeri NRRL 181	320	320	98%	4e-104	55.89%	471	XP_001266188.1
☑	queuine tRNA-ribosyltransferase-like protein [Aspergillus lentulus]	Aspergillus lentulus	320	320	98%	6e-104	55.51%	484	XP_033416284.1
☑	uncharacterized protein TRUGW13939_08426 [Talaromyces rugulosus]	Talaromyces rugulosus	319	319	99%	8e-104	55.43%	477	XP_035347453.1
☑	tRNA-guanine transglycosylase family protein [Aspergillus nomiae NRRL 13137]	Aspergillus nomiae NRRL 13137	319	319	99%	1e-103	54.85%	478	XP_015411139.1
☑	tRNA-guanine transglycosylase family protein [Aspergillus fumigatus Af293]	Aspergillus fumigatus Af293	318	318	98%	2e-103	55.89%	484	XP_747997.1
☑	uncharacterized protein Aspvir_004699 [Aspergillus viridinutans]	Aspergillus viridinutans	318	318	98%	4e-103	56.27%	484	XP_043123858.1

Screenshot of alignment details:



Alignment view: Pairwise

100 sequences selected

tRNA-guanine transglycosylase [Coccidioides immitis RS]

Sequence ID: XP_001242661.1 Length: 472 Number of Matches: 1

Range 1: 160 to 424 GenPept Graphics

```
Score           Expect  Method                        Identities     Positives      Gaps
549 bits(1414)  0.0     Compositional matrix adjust.  260/265(98%)   262/265(98%)   0/265(0%)

Query  1    PDFAVGLADLVLTQPPGVKRRERMVDRTHAWTRDTIDRLYGAGNTSGVSNKSLFLAPLLP  60
            PDFAVGLADLVLTQPPGVKRRERMVDRTHAWTRDTIDRLYGAGNTSGVSNKSLFLAPLLP
Sbjct  160  PDFAVGLADLVLTQPPGVKRRERMVDRTHAWTRDTIDRLYGAGNTSGVSNKSLFLAPLLP  219

Query  61   LEKEMQLLYVQDLEDEMKDSISGFALFDGSTVEAVPDSMSHLVRMFFGNPHTPHRVLREI  120
            LEKE+Q LYVQDLEDEMKDSISGFALFDGSTVEAVPDSMSHLVRMFFGNPHTPHRVLREI
Sbjct  220  LEKELQFLYVQDLEDEMKDSISGFALFDGSTVEAVPDSMSHLVRMFFGNPHTPHRVLREI  279

Query  121  SLGIDLTTIPFIGTASDAGLAFDFTFPQPPTENSKRNLPLAFDMWLSSHAVDTEPLKPGC  180
            SLGIDLTTIPFIGTASDAGLA DFTFPQPPTENSKRNLPLAFDMWLSSHAVDTEPLKPGC
Sbjct  280  SLGIDLTTIPFIGTASDAGLALDFTFPQPPTENSKRNLPLAFDMWLSSHAVDTEPLKPGC  339

Query  181  QCYTCKNHHRAYIQHLLNAKEMLAWTLLQIHNHHVMDQFFAAVRGSIWNGTFAQDVETFE  240
            QCYTCKNHHRAYIQHLLNAKEMLAWTLLQIHNHHVMDQFFAAVRGSI NGTFAQDVETFE
Sbjct  340  QCYTCKNHHRAYIQHLLNAKEMLAWTLLQIHNHHVMDQFFAAVRGSILNGTFAQDVETFE  399

Query  241  RAYAPEFPEQTGQGPRIRGYQAKSD  265
            RAYAPEFPEQTGQGPRIRGYQAKS+
Sbjct  400  RAYAPEFPEQTGQGPRIRGYQAKSE  424
```

Related Information

Gene - associated gene details

Genome Data Viewer - aligned genomic context

[Q5] Generate a multiple sequence alignment with your novel protein, your original query protein, and a group of other members of this family from different species. A typical number of proteins to use in a multiple sequence alignment for this assignment purpose is a minimum of 5 and a maximum of 20 - although the exact number is up to you. Include the multiple sequence alignment in your report. Use Courier font with a size appropriate to fit page width.
Side-note: Indicate your sequence in the alignment by choosing an appropriate name for each sequence in the input unaligned sequence file (i.e. edit the sequence file so that the species, or short common, names (rather than accession numbers) display in the output alignment and in the subsequent answers below). The goal in this step is to create an interesting an alignment for building a phylogenetic tree that illustrates species divergence.

```
> Dothidotthia symphoricarpi | tRNA-guanine transglycosylase [Dothidotthia symphoricarpi
CBS 119687] XP_033519716.1
MAQKLDQLPPEMLDFTLLKTAGALTPRLGRLAVPGRKTLLTPDFLGNTSRGAIPHLSQD
NYRKSVDINGVYIALEDFVEKYPAKTPPVLSYDVPEPLRQFIALPHDTLVVLGARRNPPIP
CPSANTNTAISLLTSVGFRSVSSEYYAAAIQKLKPDIVVGLADIPFGQDSIGTKRKDKMSD
RTETWLKDLVAKGSALGEEEQKWSVFAPILPIERDLQSWYLEHLVEDMADKISGVAIYD
AYLLDDLPEQLYHLPRLSFHAPASPHELLRQISLGMDLFTVPFLADATDAGIALDFTFPAP
SKDESSSARKSLGIDMWLDMHAQSVIPLSVDCTCYACTKHHRAYVQHLLAAKEMLGW
VLIQLHNHAILSAFFSGIRASIEADTFDAEVATFEAYYEPALPEKTGQGPRVRGYQFKSEE
HAKREKKNPKAFTKFDEEQIAELKNASELQKDRKLPASNVVDDEALMGLVGLNGVNFN
ADPVEGLTIEDDKKTTYPYVRCTLRTVQKAPNKLVGVLRCCAISVENAKTDGVKALDQ
CTYQKRTTTPQNKDTVGHEDSGSSQLSEKERV
```

```
> Coccidioides posadasii | GH437857.1 G874P535RD8.T0 C. posadasii Silveira, 72HR
SPHERULE_NORMALIZED Coccidioides posadasii cDNA, mRNA sequence
PDFAVGLADLVLTQPPGVKRRERMVDRTHAWTRDTIDRLYGAGNTSGVSNKSLFLAPL
LPLEKEMQLLYVQDLEDEMKDSISGFALFDGSTVEAVPDSMSHLVRMFFGNPHTPHRVL
REISLGIDLTTIPFIGTASDAGLAFDFTFPQPPTENSKRNLPLAFDMWLSSHAVDTEPLKPG
CQCYTCKNHHRAYIQHLLNAKEMLAWTLLQIHNHHVMDQFFAAVRGSIWNGTFAQDV
ETFERAYAPEFPEQTGQGPRIRGYQAKSD
```

```
> Escherichia coli (strain K12) | TGT_ECOLI Queuine tRNA-ribosyltransferase OS=Escherichia
coli (strain K12) OX=83333 GN=tgt PE=1 SV=1|P0A847|
MKFELDTTDGRARRGRLVFDRGVVETPCFMPVGTYGTVKGMTPEEVEATGAQIILGNT
FHLWLRPGQEIMKLHGDLHDFMQWKGPILTDSGGFQVFSLGDIRKITEQGVHFRNPING
DPIFLDPEKSMEIQYDLGSDIVMIFDECTPYPADWDYAKRSMEMSLRWAKRSRERFDSL
GNKNALFGIIQGSVYEDLRDISVKGLVDIGFDGYAVGGLAVGEPKADMHRILEHVCPQIP
ADKPRYLMGVGKPEDLVEGVRRGIDMFDCVMPTRNARNGHLFVTDGVVKIRNAKYKS
DTGPLDPECDCYTCRNYSRAYLHHLDRCNEILGARLNTIHNLRYYQRLMAGLRKAIEEG
KLESFVTDFYQRQGREVPPLNVD
```

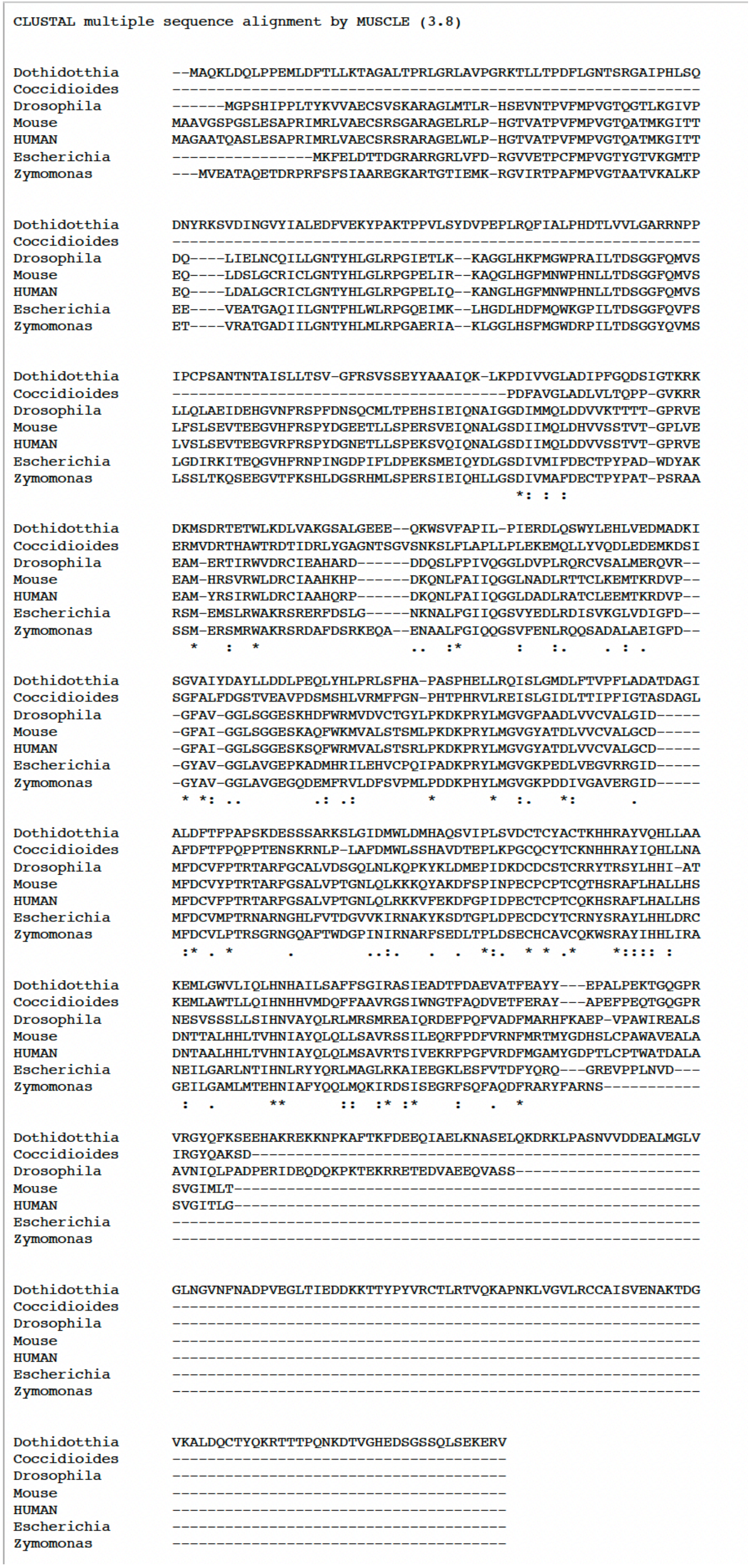
```
> Zymomonas mobilis subsp. mobilis (strain ATCC 31821 / ZM4 / CP4) | TGT_ZYMMO
Queuine tRNA-ribosyltransferase OS=Zymomonas mobilis subsp. mobilis (strain ATCC 31821 /
ZM4 / CP4) OX=264203 GN=tgt PE=1 SV=4|P28720|
MVEATAQETDRPRFSFSIAAREGKARTGTIEMKRGVIRTPAFMPVGTAATVKALKPETV
RATGADIILGNTYHLMLRPGAERIAKLGGLHSFMGWDRPILTDSGGYQVMSLSSLTKQS
EEGVTFKSHLDGSRHMLSPERSIEIQHLLGSDIVMAFDECTPYPATPSRAASSMERSMRW
AKRSRDAFDSRKEQAENAALFGIQQGSVFENLRQQSADALAEIGFDGYAVGGLAVGEG
QDEMFRVLDFSVPMLPDDKPHYLMGVGKPDDIVGAVERGIDMFDCVLPTRSGRNGQAF
TWDGPINIRNARFSEDLTPLDSECHCAVCQKWSRAYIHHLIRAGEILGAMLMTEHNIAFY
QQLMQKIRDSISEGRFSQFAQDFRARYFARNS
```

```
> Mouse | TGT_MOUSE Queuine tRNA-ribosyltransferase catalytic subunit 1 OS=Mus
musculus OX=10090 GN=Qtrt1 PE=1 SV=2|Q9JMA2|
MAAVGSPGSLESAPRIMRLVAECSRSGARAGELRLPHGTVATPVFMPVGTQATMKGITT
EQLDSLGCRICLGNTYHLGLRPGPELIRKAQGLHGFMNWPHNLLTDSGGFQMVSLFSLS
EVTEEGVHFRSPYDGEETLLSPERSVEIQNALGSDIIMQLDHVVSSTVTGPLVEEAMHRS
VRWLDRCIAAHKHPDKQNLFAIIQGGLNADLRTTCLKEMTKRDVPGFAIGGLSGGESKA
QFWKMVALSTSMLPKDKPRYLMGVGYATDLVVCVALGCDMFDCVYPTRTARFGSAL
VPTGNLQLKKKQYAKDFSPINPECPCPTCQTHSRAFLHALLHSDNTTALHHLTVHNIAY
QLQLLSAVRSSILEQRFPDFVRNFMRTMYGDHSLCPAWAVEALASVGIMLT
```

```
> Drosophila melanogaster | TGT_DROME Queuine tRNA-ribosyltransferase catalytic subunit
OS=Drosophila melanogaster OX=7227 GN=Tgt PE=2 SV=1|Q9VPY8|
MGPSHIPPLTYKVVAECSVSKARAGLMTLRHSEVNTPVFMPVGTQGTLKGIVPDQLIEL
NCQILLGNTYHLGLRPGIETLKKAGGLHKFMGWPRAILTDSGGFQMVSLLQLAEIDEHG
VNFRSPFDNSQCMLTPEHSIEIQNAIGGDIMMQLDDVVKTTTTGPRVEEAMERTIRWVD
RCIEAHARDDDQSLFPIVQGGLDVPLRQRCVSALMERQVRGFAVGGLSGGESKHDFWR
MVDVCTGYLPKDKPRYLMGVGFAADLVVCVALGIDMFDCVFPTRTARFGCALVDSGQ
LNLKQPKYKLDMEPIDKDCDCSTCRRYTRSYLHHIATNESVSSSLLSIHNVAYQLRLMRS
MREAIQRDEFPQFVADFMARHFKAEPVPAWIREALSAVNIQLPADPERIDEQDQKPKTE
KRRETEDVAEEQVASS
```

```
> HUMAN | Queuine tRNA-ribosyltransferase catalytic subunit 1 OS=Homo sapiens OX=9606
GN=QTRT1 PE=2 SV=1|B2RAR3|
MAGAATQASLESAPRIMRLVAECSRSRARAGELWLPHGTVATPVFMPVGTQATMKGIT
TEQLDALGCRICLGNTYHLGLRPGPELIQKANGLHGFMNWPHNLLTDSGGFQMVSLVS
LSEVTEEGVRFRSPYDGNETLLSPEKSVQIQNALGSDIIMQLDDVVSSTVTGPRVEEAMY
RSIRWLDRCIAAHQRPDKQNLFAIIQGGLDADLRATCLEEMTKRDVPGFAIGGLSGGESK
SQFWRMVALSTSRLPKDKPRYLMGVGYATDLVVCVALGCDMFDCVFPTRTARFGSAL
VPTGNLQLRKKVFEKDFGPIDPECTCPTCQKHSRAFLHALLHSDNTAALHHLTVHNIAY
QLQLMSAVRTSIVEKRFPGFVRDFMGAMYGDPTLCPTWATDALASVGITLG
```

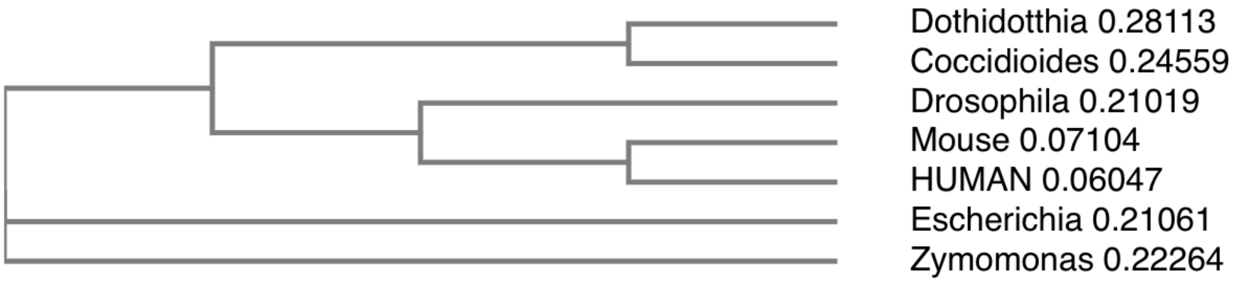
Alignment Results: Used MUSCLE (version 3.8) at EBI:



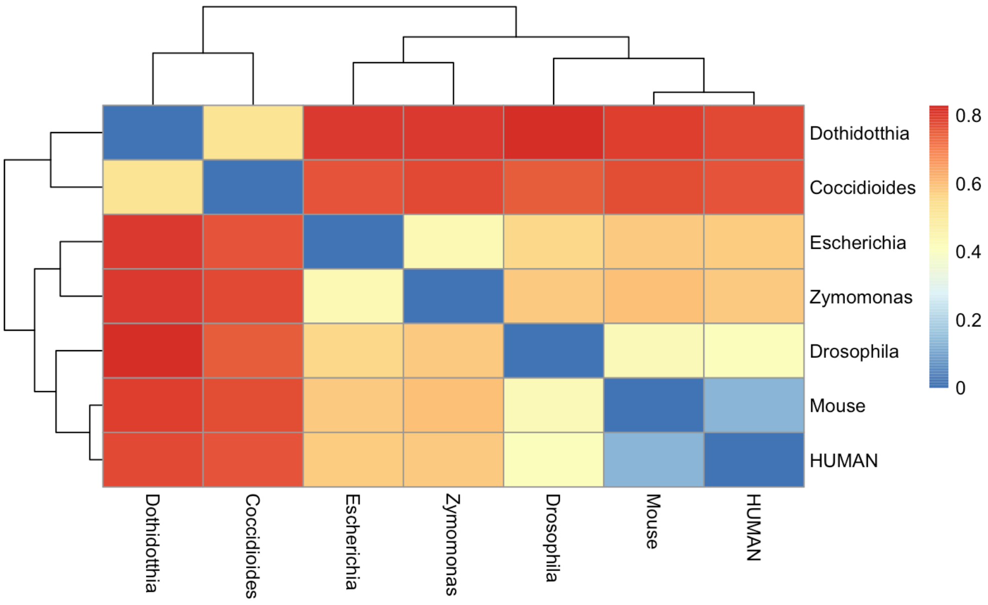
```
CLUSTAL multiple sequence alignment by MUSCLE (3.8)


Dothidotthia      --MAQKLDQLPPEMLDFTLLKTAGALTPRLGRLAVPGRKTLLTPDFLGNTSRGAIPHLSQ
Coccidioides      ------------------------------------------------------------
Drosophila        ------MGPSHIPPLTYKVVAECSVSKARAGLMTLR-HSEVNTPVFMPVGTQGTLKGIVP
Mouse             MAAVGSPGSLESAPRIMRLVAECSRSGARAGELRLP-HGTVATPVFMPVGTQATMKGITT
HUMAN             MAGAATQASLESAPRIMRLVAECSRSRARAGELWLP-HGTVATPVFMPVGTQATMKGITT
Escherichia       ----------------MKFELDTTDGRARRGRLVFD-RGVVETPCFMPVGTYGTVKGMTP
Zymomonas         ---MVEATAQETDRPRFSFSIAAREGKARTGTIEMK-RGVIRTPAFMPVGTAATVKALKP


Dothidotthia      DNYRKSVDINGVYIALEDFVEKYPAKTPPVLSYDVPEPLRQFIALPHDTLVVLGARRNPP
Coccidioides      ------------------------------------------------------------
Drosophila        DQ----LIELNCQILLGNTYHLGLRPGIETLK--KAGGLHKFMGWPRAILTDSGGFQMVS
Mouse             EQ----LDSLGCRICLGNTYHLGLRPGPELIR--KAQGLHGFMNWPHNLLTDSGGFQMVS
HUMAN             EQ----LDALGCRICLGNTYHLGLRPGPELIQ--KANGLHGFMNWPHNLLTDSGGFQMVS
Escherichia       EE----VEATGAQIILGNTFHLWLRPGQEIMK--LHGDLHDFMQWKGPILTDSGGFQVFS
Zymomonas         ET----VRATGADIILGNTYHLMLRPGAERIA--KLGGLHSFMGWDRPILTDSGGYQVMS


Dothidotthia      IPCPSANTNTAISLLTSV-GFRSVSSEYYAAAIQK-LKPDIVVGLADIPFGQDSIGTKRK
Coccidioides      --------------------------------------PDFAVGLADLVLTQPP-GVKRR
Drosophila        LLQLAEIDEHGVNFRSPFDNSQCMLTPEHSIEIQNAIGGDIMMQLDDVVKTTTT-GPRVE
Mouse             LFSLSEVTEEGVHFRSPYDGEETLLSPERSVEIQNALGSDIIMQLDHVVSSTVT-GPLVE
HUMAN             LVSLSEVTEEGVRFRSPYDGNETLLSPEKSVQIQNALGSDIIMQLDDVVSSTVT-GPRVE
Escherichia       LGDIRKITEQGVHFRNPINGDPIFLDPEKSMEIQYDLGSDIVMIFDECTPYPAD-WDYAK
Zymomonas         LSSLTKQSEEGVTFKSHLDGSRHMLSPERSIEIQHLLGSDIVMAFDECTPYPAT-PSRAA
                                                         *: : :

Dothidotthia      DKMSDRTETWLKDLVAKGSALGEEE--QKWSVFAPIL-PIERDLQSWYLEHLVEDMADKI
Coccidioides      ERMVDRTHAWTRDTIDRLYGAGNTSGVSNKSLFLAPLLPLEKEMQLLYVQDLEDEMKDSI
Drosophila        EAM-ERTIRWVDRCIEAHARD------DDQSLFPIVQGGLDVPLRQRCVSALMERQVR--
Mouse             EAM-HRSVRWLDRCIAAHKHP------DKQNLFAIIQGGLNADLRTTCLKEMTKRDVP--
HUMAN             EAM-YRSIRWLDRCIAAHQRP------DKQNLFAIIQGGLDADLRATCLEEMTKRDVP--
Escherichia       RSM-EMSLRWAKRSRERFDSLG-----NKNALFGIIQGSVYEDLRDISVKGLVDIGFD--
Zymomonas         SSM-ERSMRWAKRSRDAFDSRKEQA--ENAALFGIQQGSVFENLRQQSADALAEIGFD--
                    *   :  *                 ..  :*      :   :.    . : .

Dothidotthia      SGVAIYDAYLLDDLPEQLYHLPRLSFHA-PASPHELLRQISLGMDLFTVPFLADATDAGI
Coccidioides      SGFALFDGSTVEAVPDSMSHLVRMFFGN-PHTPHRVLREISLGIDLTTIPFIGTASDAGL
Drosophila        -GFAV-GGLSGGESKHDFWRMVDVCTGYLPKDKPRYLMGVGFAADLVVCVALGID-----
Mouse             -GFAI-GGLSGGESKAQFWKMVALSTSMLPKDKPRYLMGVGYATDLVVCVALGCD-----
HUMAN             -GFAI-GGLSGGESKSQFWRMVALSTSRLPKDKPRYLMGVGYATDLVVCVALGCD-----
Escherichia       -GYAV-GGLAVGEPKADMHRILEHVCPQIPADKPRYLMGVGKPEDLVEGVRRGID-----
Zymomonas         -GYAV-GGLAVGEGQDEMFRVLDFSVPMLPDDKPHYLMGVGKPDDIVGAVERGID-----
                   * *: ..        .: .:        *      *  :.   *:      .

Dothidotthia      ALDFTFPAPSKDESSSARKSLGIDMWLDMHAQSVIPLSVDCTCYACTKHHRAYVQHLLAA
Coccidioides      AFDFTFPQPPTENSKRNLP-LAFDMWLSSHAVDTEPLKPGCQCYTCKNHHRAYIQHLLNA
Drosophila        MFDCVFPTRTARFGCALVDSGQLNLKQPKYKLDMEPIDKDCDCSTCRRYTRSYLHHI-AT
Mouse             MFDCVYPTRTARFGSALVPTGNLQLKKKQYAKDFSPINPECPCPTCQTHSRAFLHALLHS
HUMAN             MFDCVFPTRTARFGSALVPTGNLQLRKKVFEKDFGPIDPECTCPTCQKHSRAFLHALLHS
Escherichia       MFDCVMPTRNARNGHLFVTDGVVKIRNAKYKSDTGPLDPECDCYTCRNYSRAYLHHLDRC
Zymomonas         MFDCVLPTRSGRNGQAFTWDGPINIRNARFSEDLTPLDSECHCAVCQKWSRAYIHHLIRA
                   :* . *      .        ..:.   .  .  *:.  * * .*    *:::: :

Dothidotthia      KEMLGWVLIQLHNHAILSAFFSGIRASIEADTFDAEVATFEAYY---EPALPEKTGQGPR
Coccidioides      KEMLAWTLLQIHNHHVMDQFFAAVRGSIWNGTFAQDVETFERAY---APEFPEQTGQGPR
Drosophila        NESVSSSLLSIHNVAYQLRLMRSMREAIQRDEFPQFVADFMARHFKAEP-VPAWIREALS
Mouse             DNTTALHHLTVHNIAYQLQLLSAVRSSILEQRFPDFVRNFMRTMYGDHSLCPAWAVEALA
HUMAN             DNTAALHHLTVHNIAYQLQLMSAVRTSIVEKRFPGFVRDFMGAMYGDPTLCPTWATDALA
Escherichia       NEILGARLNTIHNLRYYQRLMAGLRKAIEEGKLESFVTDFYQRQ---GREVPPLNVD---
Zymomonas         GEILGAMLMTEHNIAFYQQLMQKIRDSISEGRFSQFAQDFRARYFARNS-----------
                   :  .      **      ::  :* :*    :   .  *

Dothidotthia      VRGYQFKSEEHAKREKKNPKAFTKFDEEQIAELKNASELQKDRKLPASNVVDDEALMGLV
Coccidioides      IRGYQAKSD---------------------------------------------------
Drosophila        AVNIQLPADPERIDEQDQKPKTEKRRETEDVAEEQVASS---------------------
Mouse             SVGIMLT-----------------------------------------------------
HUMAN             SVGITLG-----------------------------------------------------
Escherichia       ------------------------------------------------------------
Zymomonas         ------------------------------------------------------------


Dothidotthia      GLNGVNFNADPVEGLTIEDDKKTTYPYVRCTLRTVQKAPNKLVGVLRCCAISVENAKTDG
Coccidioides      ------------------------------------------------------------
Drosophila        ------------------------------------------------------------
Mouse             ------------------------------------------------------------
HUMAN             ------------------------------------------------------------
Escherichia       ------------------------------------------------------------
Zymomonas         ------------------------------------------------------------


Dothidotthia      VKALDQCTYQKRTTTPQNKDTVGHEDSGSSQLSEKERV
Coccidioides      --------------------------------------
Drosophila        --------------------------------------
Mouse             --------------------------------------
HUMAN             --------------------------------------
Escherichia       --------------------------------------
Zymomonas         --------------------------------------
```

[Q6] Create a phylogenetic tree, using either a parsimony or distance-based approach. Bootstrapping and tree rooting are optional. Use "simple phylogeny" online from the EBI or any respected phylogeny program (such as MEGA, PAUP, or Phylip). Paste an image of your Cladogram or tree output in your report.

I imported the multiple sequence alignment done with MUSCLE into Simple Phylogeny from EBI. I generated the phylogenetic tree using Simple Phylogeny from EBI.



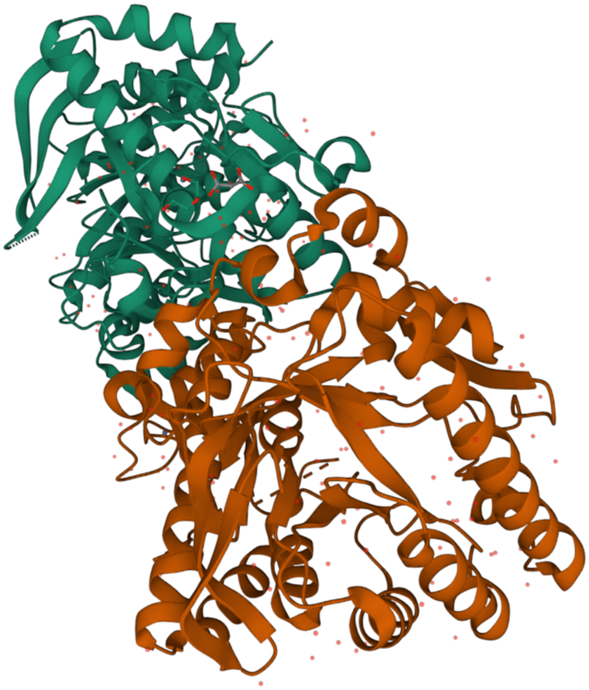
[Q7] Generate a sequence identity based heatmap of your aligned sequences using R. If necessary convert your sequence alignment to the ubiquitous FASTA format (Seaview can read in clustal format and "Save as" FASTA format for example). Read this FASTA format alignment into R with the help of functions in the Bio3D package. Calculate a sequence identity matrix (again using a function within the Bio3D package). Then generate a heatmap plot and add to your report. Do make sure your labels are visible and not cut at the figure margins.



[Q8] Using R/Bio3D (or an online blast server if you prefer), search the main protein structure database for the most similar atomic resolution structures to your aligned sequences.
List the top 3 unique hits (i.e. not hits representing different chains from the same structure) along with their Evalue and sequence identity to your query. Please also add annotation details of these structures. For example include the annotation terms PDB identifier (structureId), Method used to solve the structure (experimentalTechnique), resolution (resolution), and source organism (source).

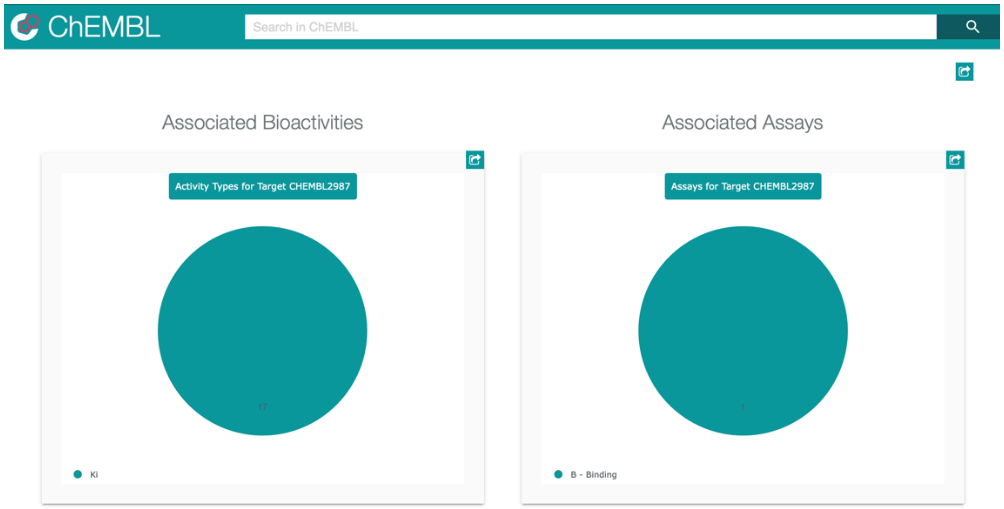
ID	Identity	E-Value	Experimental Technique	Resolution	Source
6FV5_A	28.622	6.56E-15	X-ray	2.179	Mus musculus
7B2I_A	28.622	8.13E-15	X-ray	1.65	Mus musculus
2ASH_A	33.333	3.92E-13	X-ray	1.9	Thermotoga maritima

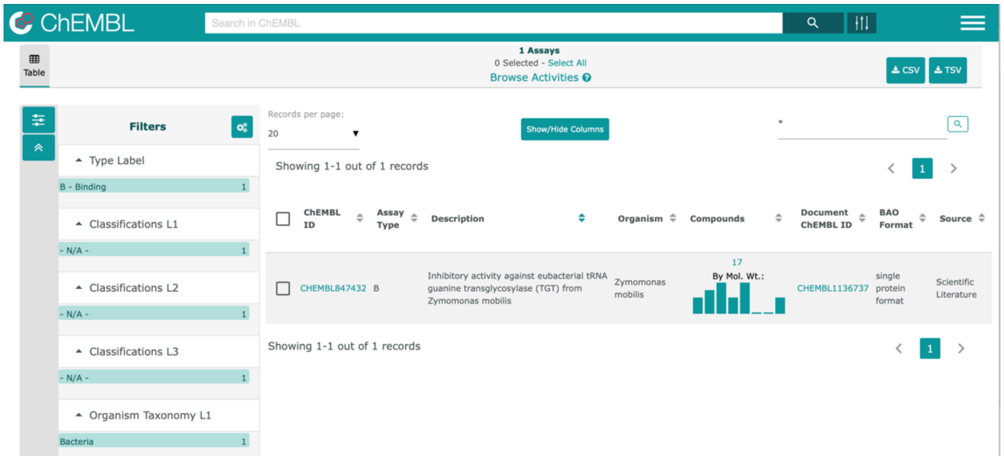
[Q9] Generate a molecular figure of one of your identified PDB structures using VMD. You can optionally highlight conserved residues that are likely to be functional. Please use a white or transparent background for your figure (i.e. not the default black). Based on sequence similarity. How likely is this structure to be similar to your "novel" protein?



[Q10] Perform a "Target" search of ChEMBEL (https://www.ebi.ac.uk/chembl/) with your novel sequence. Are there any Target Associated Assays and ligand efficiency data reported that may be useful starting points for exploring potential inhibition of your novel protein?

Based off my "Target" search of ChEMBEL of my novel sequence, there is 1 Target Associated Assay that is useful to explore the inhibition of my novel protein. This assay inhibits the activity of eubacterial tRNA guanine transglycosylase (TGT) from Zymomonas mobilis.





1 Assays
0 Selected - Select All
Browse Activities

Type Label
B - Binding 1

Classifications L1
- N/A - 1

Classifications L2
- N/A - 1

Classifications L3
- N/A - 1

Organism Taxonomy L1
Bacteria 1

Records per page: 20

Showing 1-1 out of 1 records

□	ChEMBL ID	Assay Type	Description	Organism	Compounds	Document ChEMBL ID	BAO Format	Source
□	CHEMBL847432	B	Inhibitory activity against eubacterial tRNA guanine transglycosylase (TGT) from Zymomonas mobilis	Zymomonas mobilis	17 By Mol. Wt.:	CHEMBL1136737	single protein format	Scientific Literature

Showing 1-1 out of 1 records

Citation of Scientific Literature:

Ruth Brenk, Lars Naerum, Ulrich Grädler, Hans-Dieter Gerber, George A. Garcia, Klaus Reuter, Milton T. Stubbs, and Gerhard Klebe. Virtual Screening for Submicromolar Leads of tRNA-guanine Transglycosylase Based on a New Unexpected Binding Mode Detected by Crystal Structure Analysis. *J. Med. Chem.* 46, 7, 1133-1143 (2003).