

HOW DO I RUN a BLAST SEARCH?

**Note: This is only a set of directions (i.e. what to copy/paste). There are no explanations; if you feel confused and want to understand the principles behind BLASTs, take a look at the transcript/notes and the accompanying powerpoint.

At the end of the BLAST directions, we will help you answer the other questions found on DSAP.

Blastn (nr/nt):

1. On the DSAP page, click the “retrieve sequence” button.
2. Copy the sequence.
3. Click on the “Blastn” link. Paste the sequence into the box labeled “Enter accession number(s), gi(s), or FASTA sequence(s)”.
4. Scroll to the bottom, and click the big blue BLAST button.
5. After the results page loads, scroll until you see the list of results. Click on the first one.
6. Copy and paste the necessary information, highlighted below (the organism name has two boxes around it, as it is part of the description):

PREDICTED Setaria italica dynein light chain LC6, flagellar outer arm-like (LOC101770043), mRNA
Sequence ID: ref|XM_004969977.1| Length: 791 Number of Matches: 1

Range 1: 222 to 499 [GenBank](#) [Graphics](#) ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
244 bits(270)	8e-61	221/278(79%)	0/278(0%)	Plus/Plus

Query 11 ATGTTGGAAGGGAGGGCGAGAGTAGAAGACACCGACATGCCGAGGAAGATGCAGGCGGAG 70
Sbjct 222 ATGTTGGAAGGGAAGGCGATGGTGGAGGACACCGACATGCCGGCGAAGATGCAGGCGCAG 281

Query 71 GCCATGAACGCCGCTCTCACGGCTCGATCTGTTCGACGTCCGGACTGCAAGAGCCTC 130
Sbjct 282 GCGATGGCGGCGGCGTCCAGGGCCCTCGACCGCTTCGACGTCTCGACTGCCGGAGCATC 341

Query 131 GCCGCGCATATCAAGAAGGAATTTGATAAGATCTACGGTCCGGGATGGCAGTGCCTCGTC 190
Sbjct 342 GCGGCCACATCAAGAAGGAGTTTGACACGATCCATGGACCGGATGGCAATGCGTGGTG 401

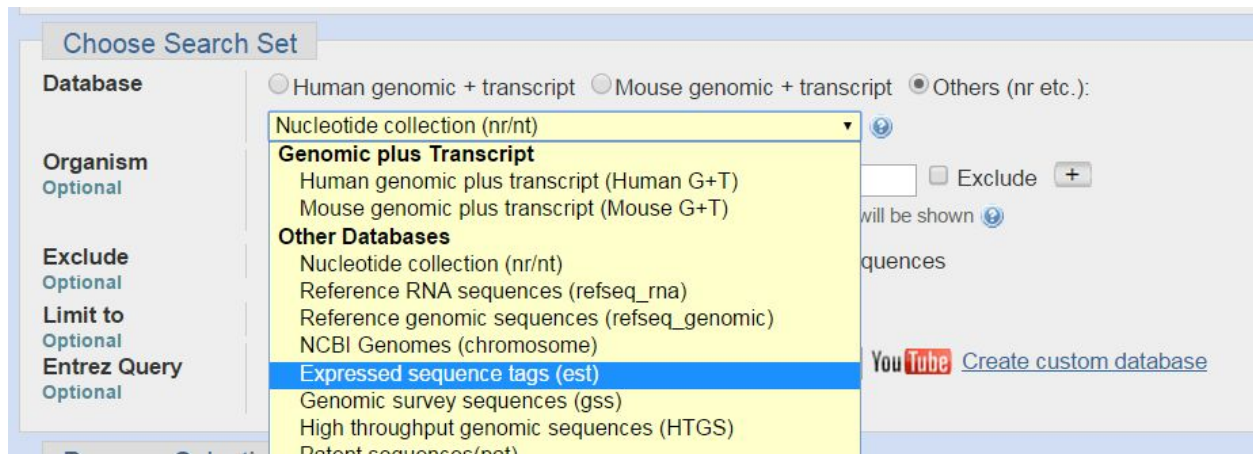
Query 191 GGCTCCAGCTTCGGCTGTTTCTTCACTCACAAGAAAGGCAGCTTCATCTACTTCCGCCTG 250
Sbjct 402 GGCTCCAGCTTCGGCTGCTACTTCACTCACAAGAAAGGCAGCTTCATCTACTTCCGCCTG 461

Query 251 GAGACGCTCCACTTCCCTCATCTTCAAAGGCGCGGCC 288
Sbjct 462 GAGTCCGCTCAGGTTCTCGTCTTCAAAGGGCGGCAC 499

7. Repeat this for two DIFFERENT organisms.

Blastn (est):

1. Repeat steps 1-3 from previous BLAST.
2. Before clicking the big blue BLAST button, find the box that says “Choose Search Set”. Next to “database”, click the drop-down menu and select “expressed sequence tags (est)”
3. Hit the big blue BLAST button.



4. Repeat steps 5-6

Blastx:

1. After finishing the Blastns, repeat steps 1-6 for Blastx.

Toolbox: Take a look at the Frame given to you in Blastx. After copying and pasting your DNA sequence into toolbox, select that frame. Your ORF is USUALLY the bolded, underlined region (but not always). Your ORF begins at the first green M after a red asterisk and stops at the next red asterisk. There are a number of scenarios that can happen:

-If your first M is NOT after a red asterisk, try two scenarios: First, copy starting from the very beginning, BEFORE the M. If those results do not match in Blastp, then copy starting from the first M and ignore the section before it.

-If you copied the underlined portion and returned poor results, then re-try Blastp and copy other sequences from toolbox.

Blastp:

1. After copying your ORF from toolbox, repeat steps 1-6, from the first Blastn step, for Blastp.

Blastx v.s. Blastp: Compare your results from Blastx and Blastp. Your accession numbers should be the same and your e-values should be similar.

Other questions:

Is your edited sequence significantly similar to sequences found in other organisms?

Answer: Almost always YES, unless if you received very poor results.

Look down the list of Blastn/Blastx/ results. is your edited sequence significantly similar to sequences found in different kingdoms?

Answer: 99% of the time NO: our results are almost always plants. You can always double check by scrolling through the results.

Does your DNA code for the likely start of the protein?

The start of the protein begins with an M. If you highlighted an M in toolbox and there was a match in Blastp, then YES. If not, then NO. (You may highlight an M in toolbox, but there may be no match to that in Blastp. Further details in the BLAST transcript/notes).

You can also confirm this by looking at the graphic organizer: if it codes for the start, the protein will have a colored band starting from the beginning until the end.

Does your DNA code for the likely end of the protein?

The end is the asterisk in Toolbox. You can also confirm this in the graphic organizer, as stated above (but for the end).

Compare e-values of Blastx and Blastp: E-values should all be similar (the negative exponent should be close to each other).

Compare alignments: You have to manually go back to the Blastx and Blastp sections of DSAP. Blastp should be only one-third the length of Blastx.

THAT IS THE END OF BLAST!!!! CONGRATS! IF THERE ARE QUESTIONS, FEEL FREE TO ASK.