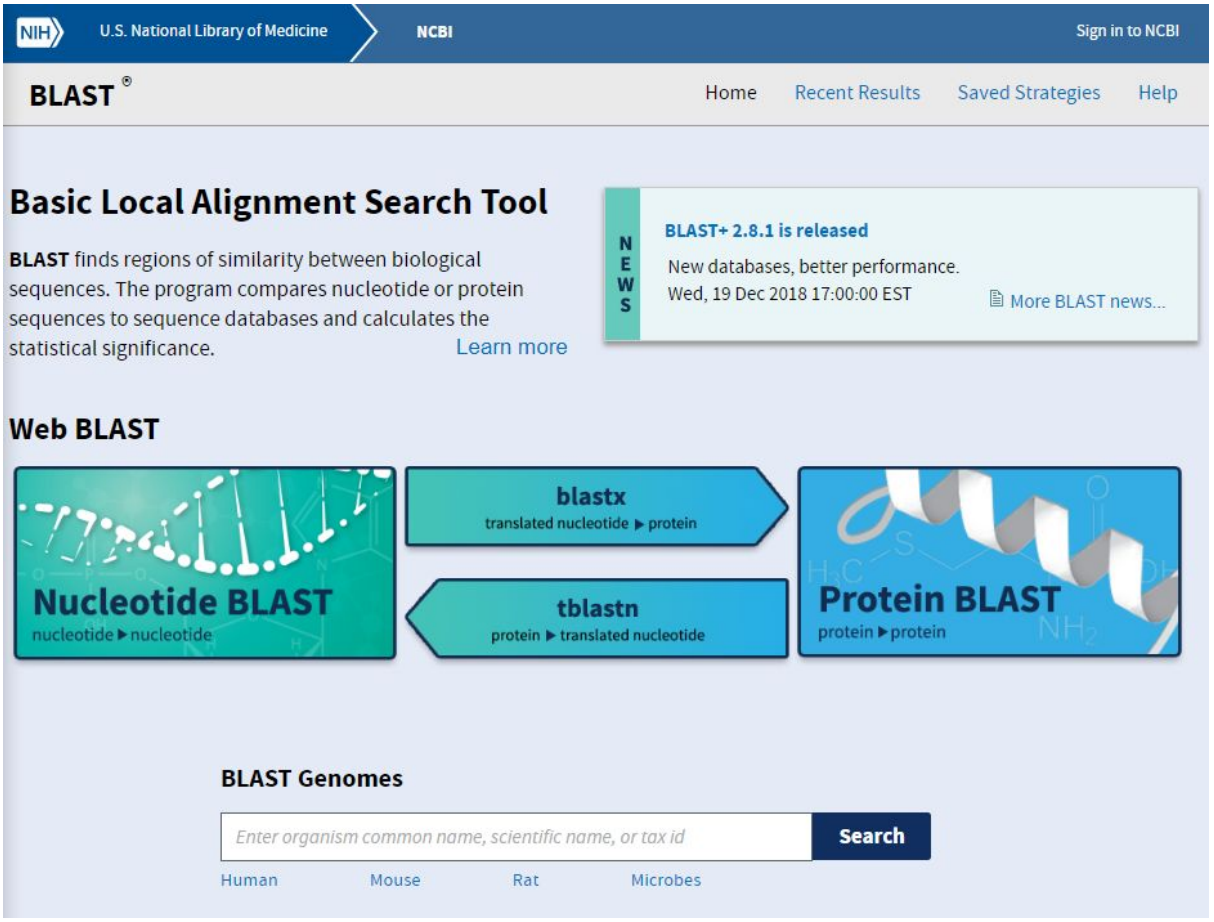
BLAST Guidebook

BLAST, which stands for Basic Local Alignment Search Tool, is a tool that looks for regions of similarity between nucleotide or protein sequences. It allows for researchers to look for other people or databases that have used same or similar sequences in their work.

Access BLAST at
https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastHome

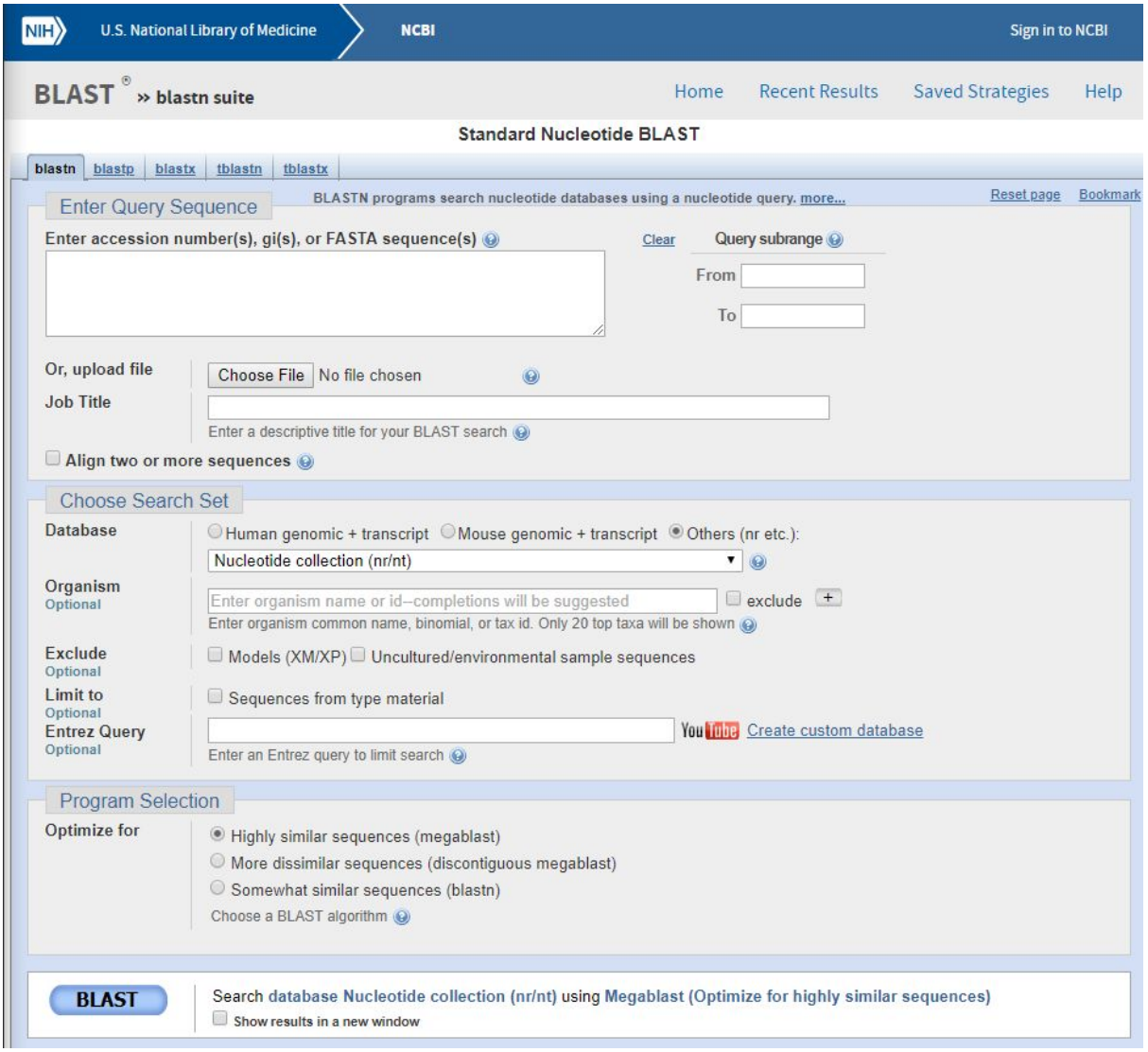
The home page looks like this:



1. Nucleotide BLAST

Nucleotide BLAST compares a nucleotide sequence of your choice to other nucleotide sequences found on databases on the internet.

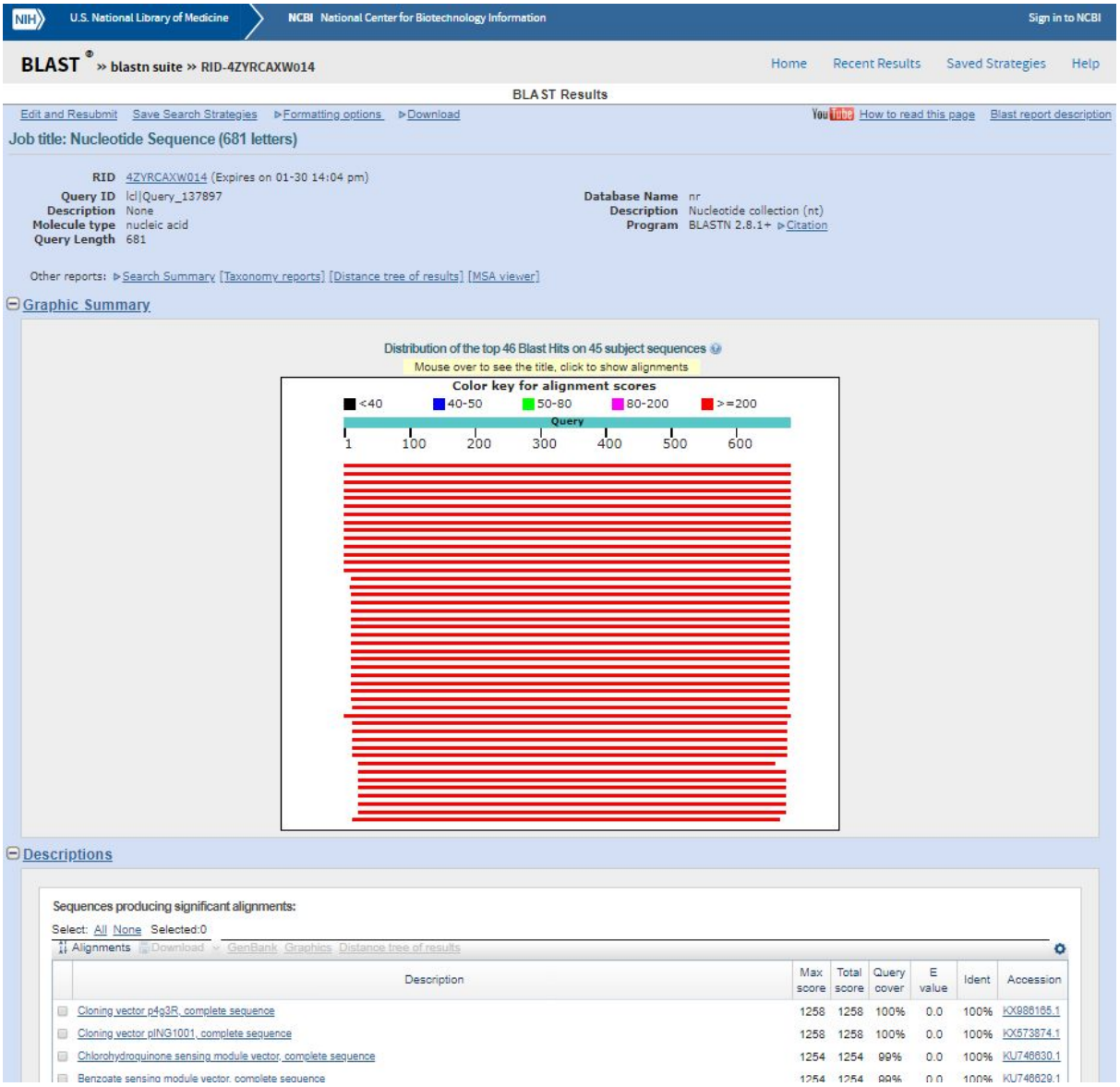
When you click on “Nucleotide BLAST” on the BLAST home page, it will take you to this page:



Enter your sequence of interest in FASTA format in the large box on the top left. FASTA format refers to a method of representing nucleotide sequences where you give only the bases on one strand of DNA. Due to base complementarity, the other strand can be derived. You can set the subrange from your input sequence from which to blast search.

You can also indicate which database to search from - NCBI has lots of organism or species-specific sequences that you can search from.

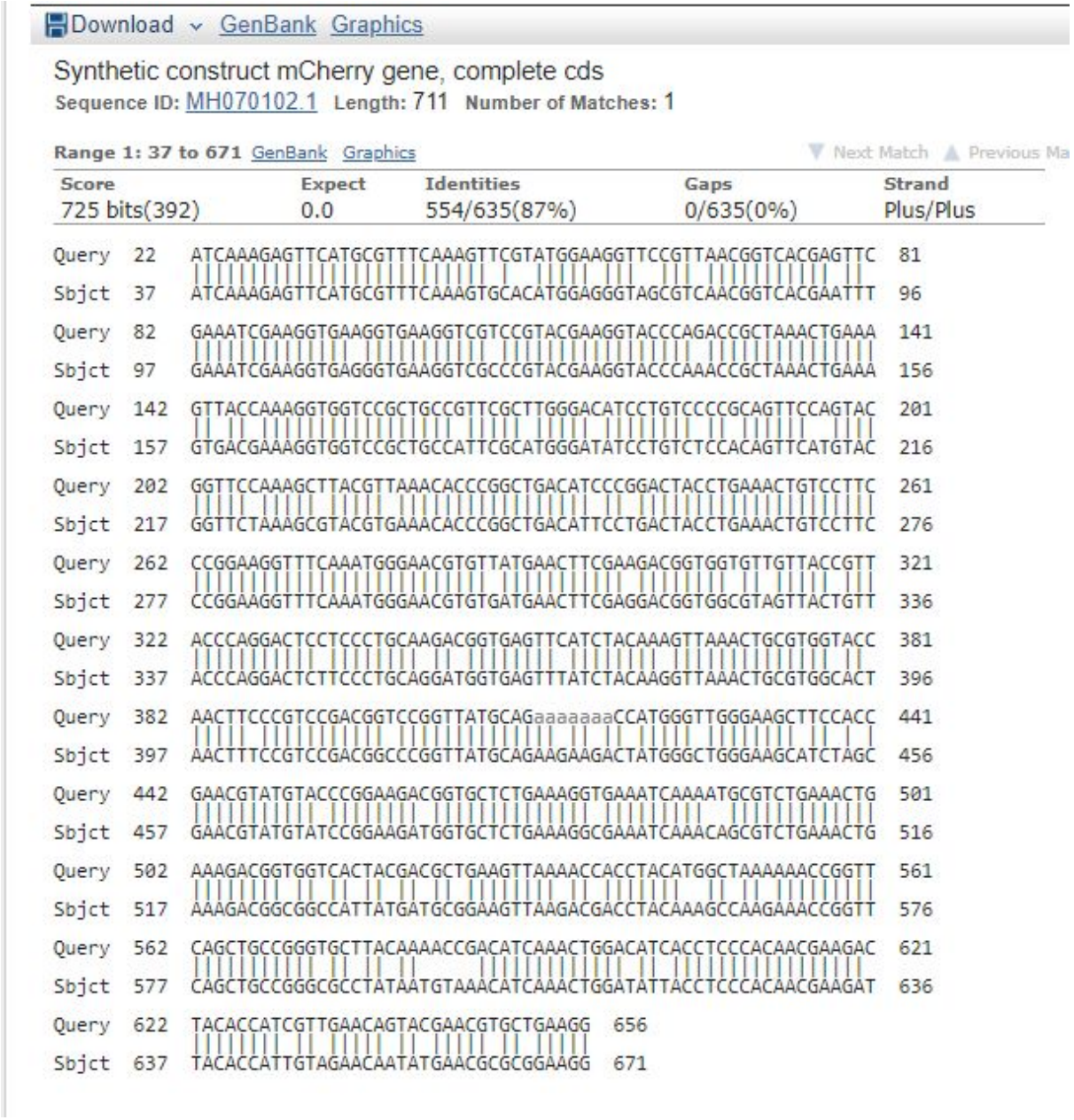
When you are done inputting all the necessary information above, click “BLAST” to begin your search. Your search output should look something like this:



The graph gives us a visual representation of how closely the inputted sequence aligns with sequences found in various databases.

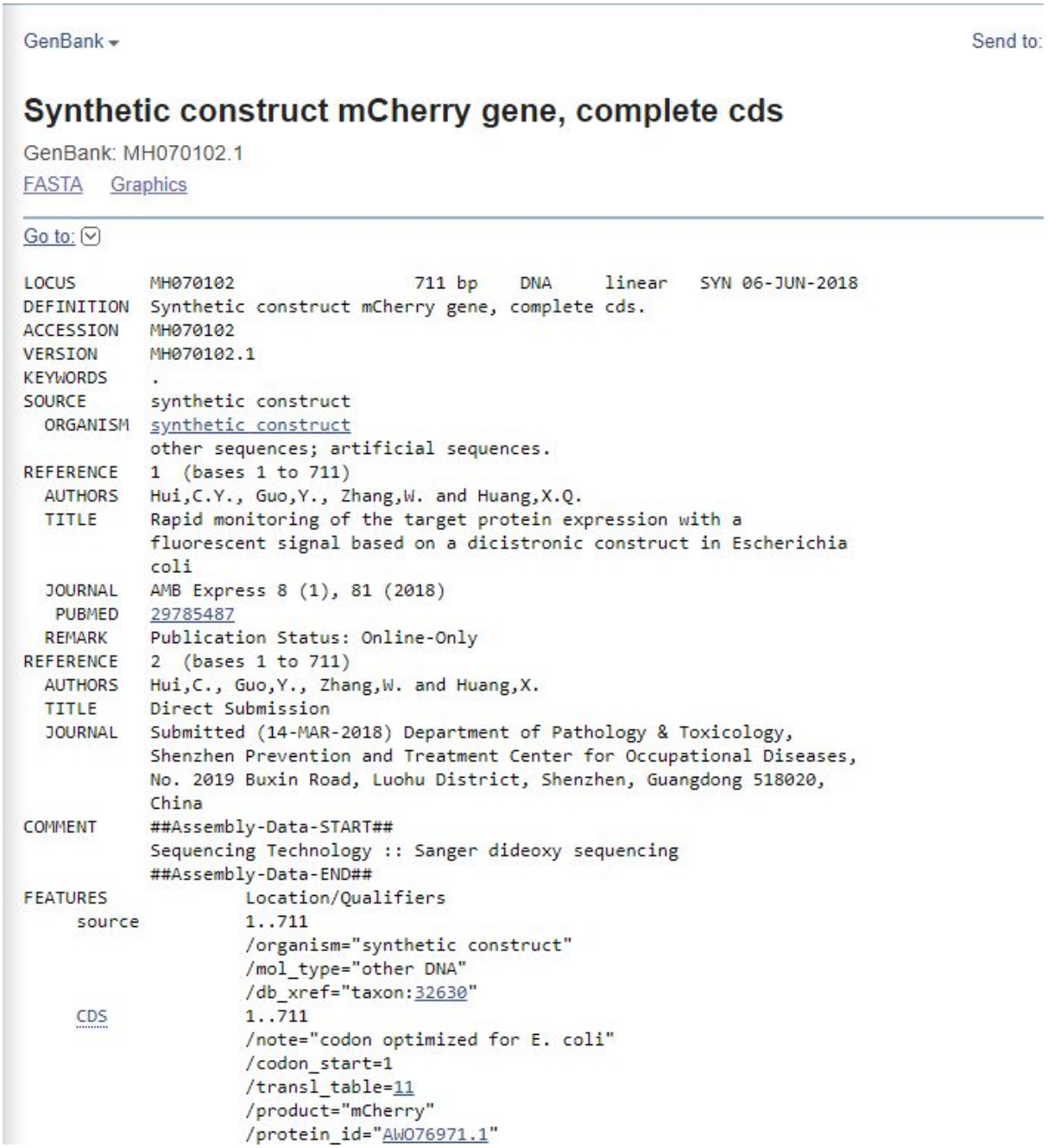
The Descriptions box below provides more specific information about the search results. The most important value in this table is the “Query cover” value, which tells us what percent of a match the database sequence is with our sequence of interest.

If you click on any of the descriptions, it will take you to a page that looks like this:



This map shows you where the differences are between your input sequence and the result sequence. The lines indicate the same base, whereas the spaces indicate different base.

If you click Sequence ID at the top, you can be taken to the GenBank page for that sequence.



GenBank ▾ Send to:

Synthetic construct mCherry gene, complete cds

GenBank: MH070102.1

FASTA Graphics

Go to: ⊙

```
LOCUS       MH070102                711 bp    DNA     linear   SYN 06-JUN-2018
DEFINITION  Synthetic construct mCherry gene, complete cds.
ACCESSION   MH070102
VERSION     MH070102.1
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (bases 1 to 711)
  AUTHORS   Hui,C.Y., Guo,Y., Zhang,W. and Huang,X.Q.
  TITLE     Rapid monitoring of the target protein expression with a
            fluorescent signal based on a dicistronic construct in Escherichia
            coli
  JOURNAL   AMB Express 8 (1), 81 (2018)
   PUBMED   29785487
  REMARK    Publication Status: Online-Only
REFERENCE   2  (bases 1 to 711)
  AUTHORS   Hui,C., Guo,Y., Zhang,W. and Huang,X.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAR-2018) Department of Pathology & Toxicology,
            Shenzhen Prevention and Treatment Center for Occupational Diseases,
            No. 2019 Buxin Road, Luohu District, Shenzhen, Guangdong 518020,
            China
COMMENT     ##Assembly-Data-START##
            Sequencing Technology :: Sanger dideoxy sequencing
            ##Assembly-Data-END##
FEATURES             Location/Qualifiers
     source          1..711
                     /organism="synthetic construct"
                     /mol_type="other DNA"
                     /db_xref="taxon:32630"
     CDS             1..711
                     /note="codon optimized for E. coli"
                     /codon_start=1
                     /transl_table=11
                     /product="mCherry"
                     /protein_id="AWO76971.1"
```

GenBank is another tool made by NCBI that compiles DNA sequences and their origin of publication. On this page, you can find an annotated sequence, as well as the paper and authors who first published that seqence.

2. Blastx, Tblastn, pblast

BLAST is not exclusive to nucleotide sequences - it can also be used with amino acid sequences, or between nucleotides and amino acid sequences.

Blastx is a tool that allows you to search for your translated nucleotide sequence in comparison to available amino acid sequences. This is useful if you have a sequence but you do not know what it codes for.

Tblastn is a tool that matches your amino acid sequence to RNA nucleotide sequences in the database. This may be useful when you're trying to find sequences of the same protein that are codon-optimized to various organisms.

Pblast is a tool that compares amino acid sequences. You may be able to find other proteins with similar domains, or mutations of your protein with this tool.

3. Other available BLAST tools

There are many more tools that BLAST offers aside from the above four main search tools, including a conserved domain search or a search for markers. If you want to find things similar to any biological molecule, be it by function, morphology, chemical structure, etc. BLAST is probably the place to go. Find these extra functions at the bottom of the BLAST home page.

4. General Tips and tl;dr

Before any sequence is used or ordered in the lab, it must be put through BLAST. This is to make sure that we are using a sequence that has been used and verified before. The more matches you get on BLAST, the better. It might be a good idea to look through source papers on any information about that part that might help inform your experiments.

Tl;dr blast has lots of sequences, use them.

Feel free to DM us (Sara, Cassie, or Chris) if you have any questions ;)