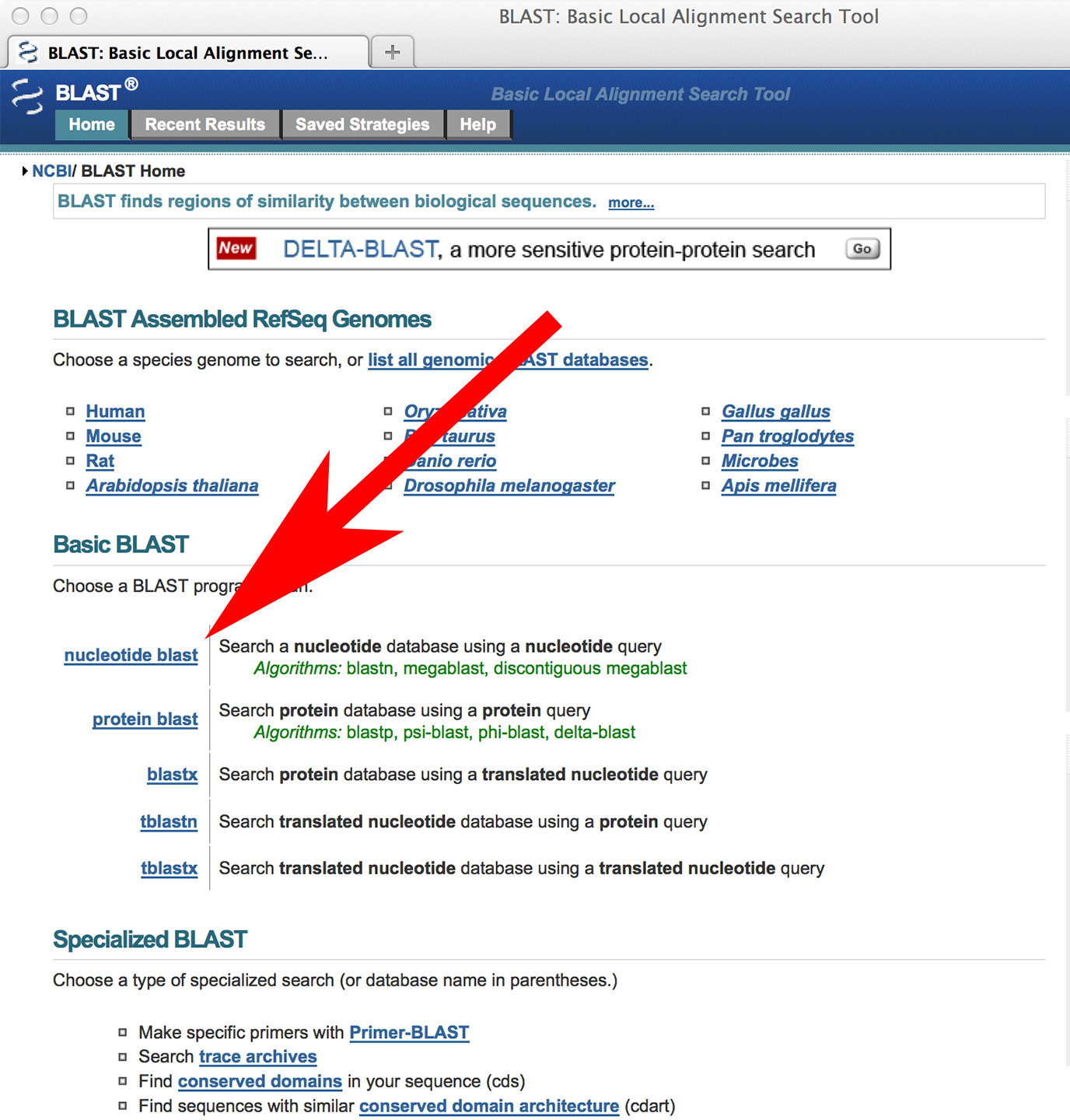
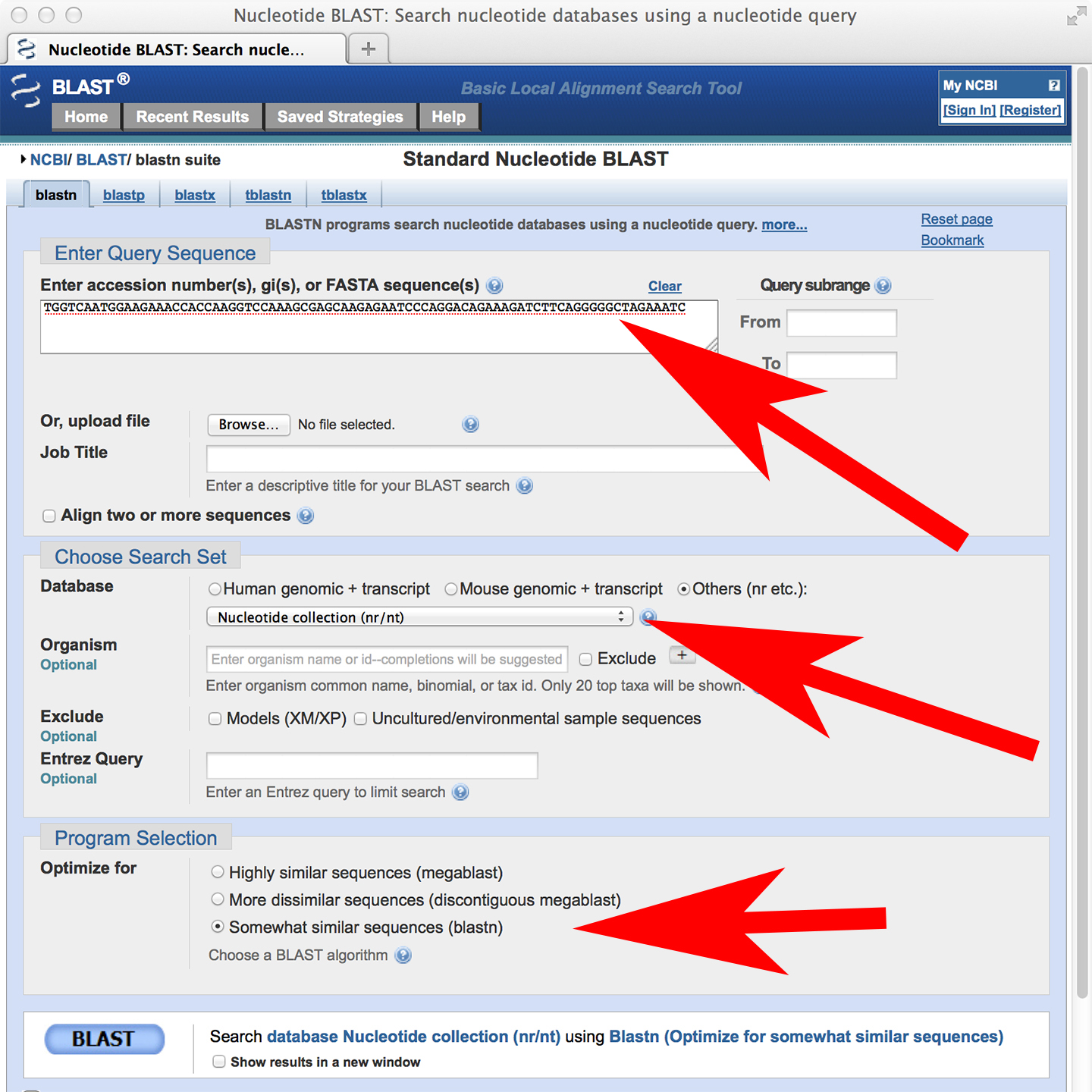
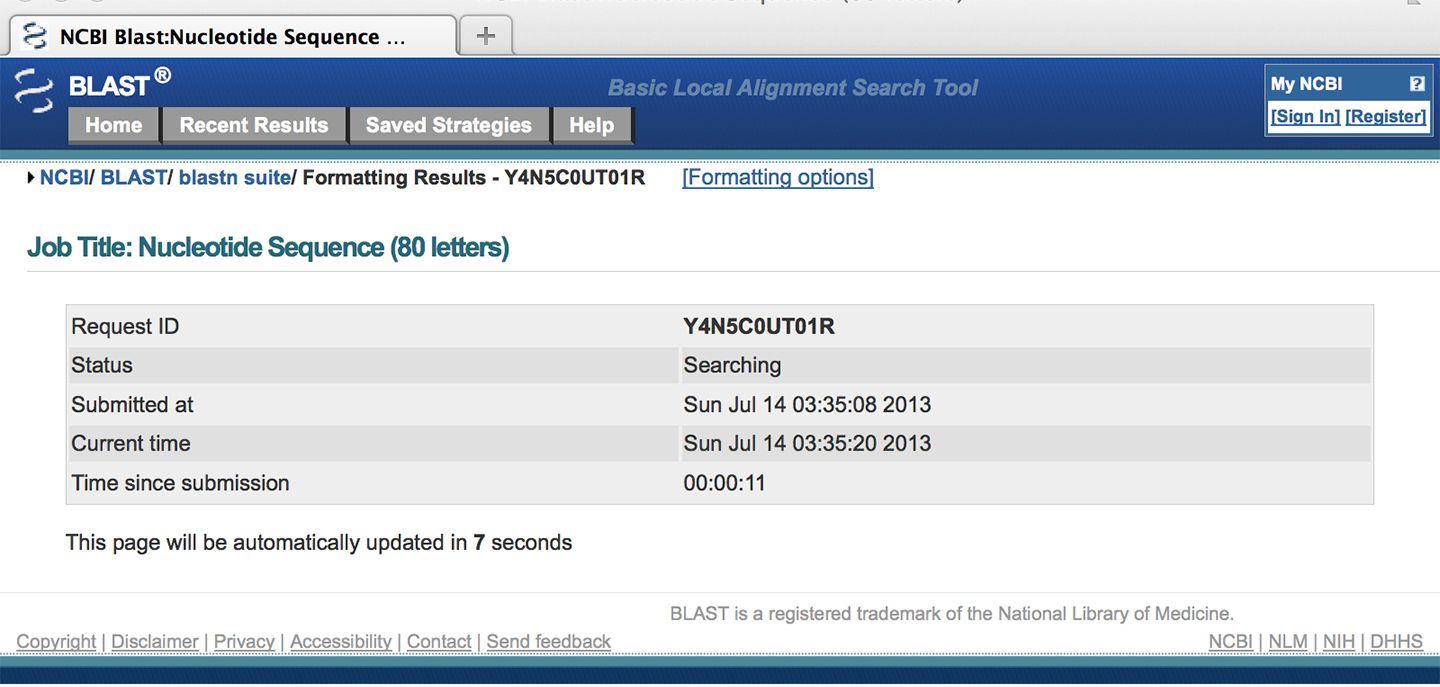
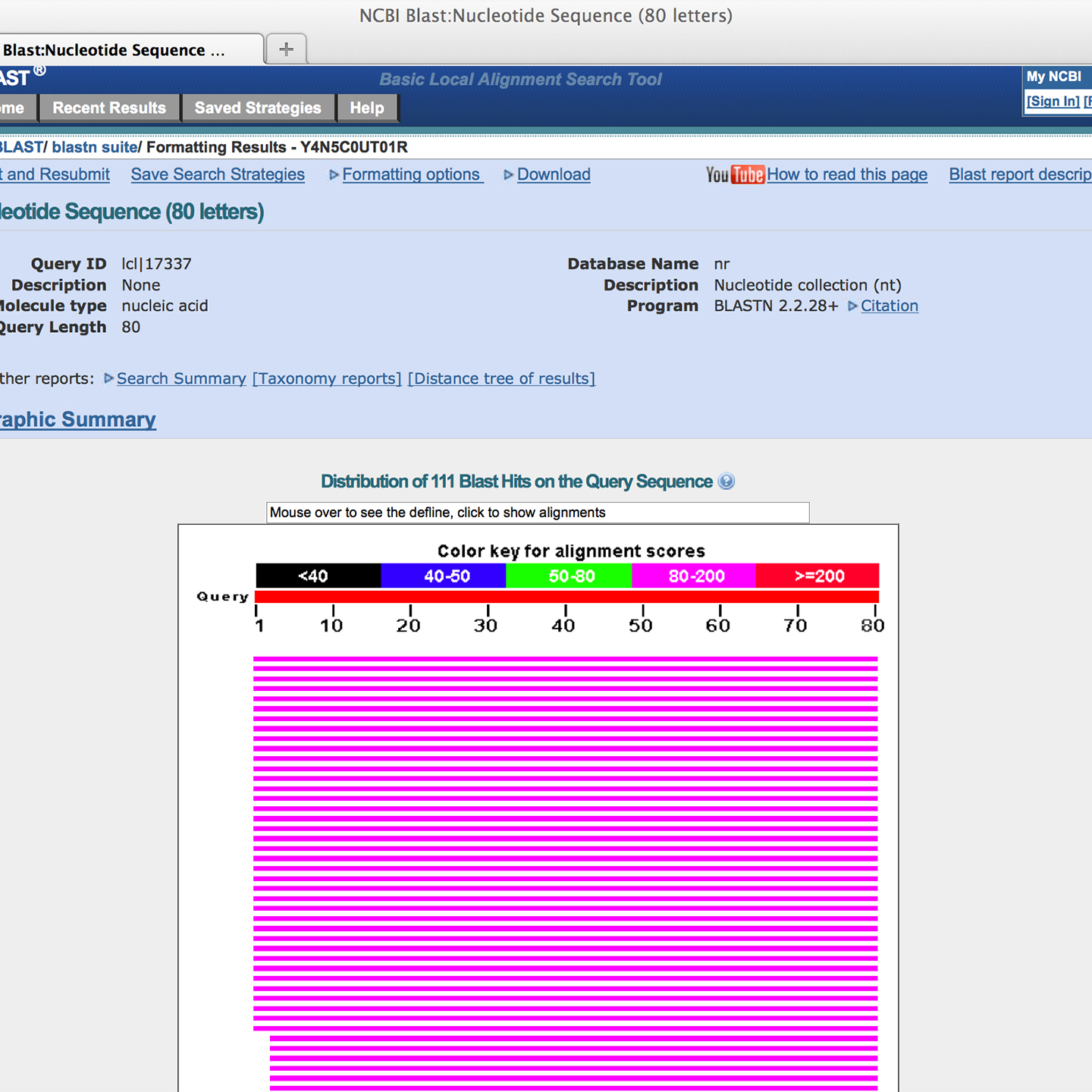
**BLAST Tutorial, step by step directions to search publically known DNA sequences from all organism.**

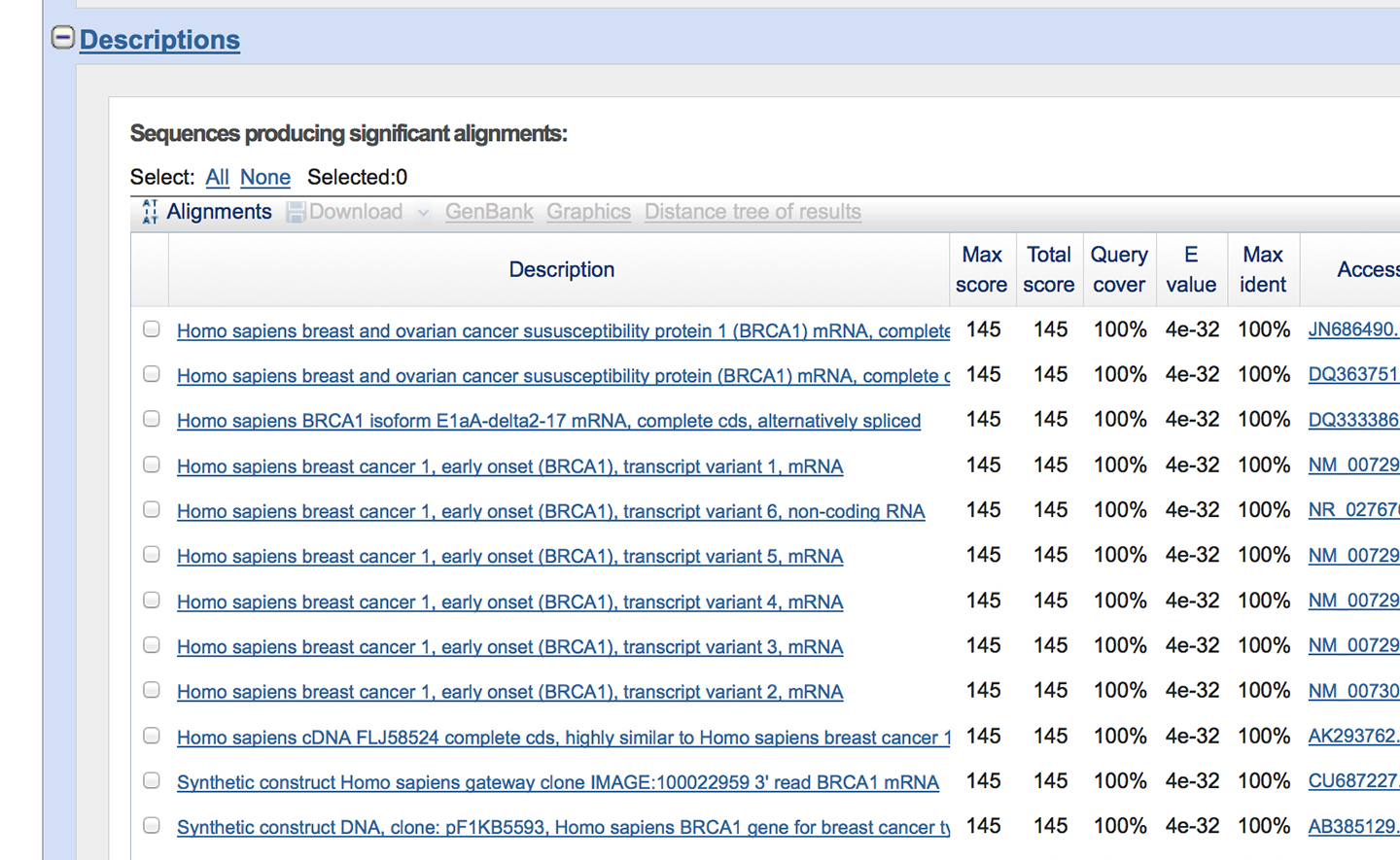
1. The BLAST home page directs you to many types of analysis and DNA sequence databases that are maintained by the NCBI. For our lab only the "nucleotide blast" option is needed.

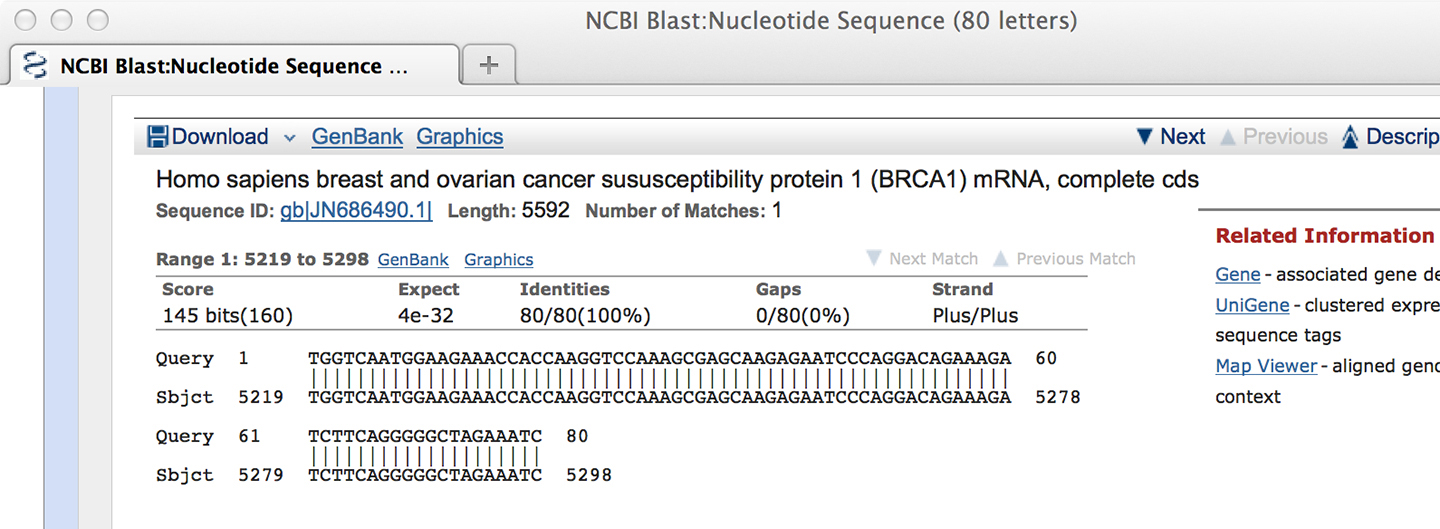
2. The Nucleotide BLAST page (also called BLASTN) is where you control the DNA search. In the upper input box paste in an unknown DNA sequence. The Database should be "Others" and "Nucleotide collection (nr/nt)" selected. The Program Selection should be "Somewhat similar sequences".

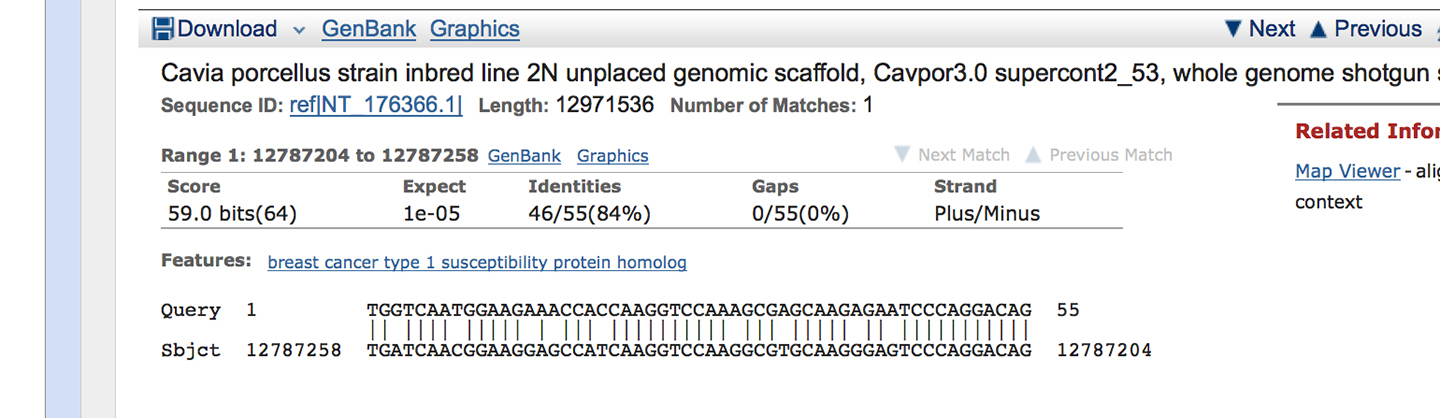


3. BLAST the DNA sequences one by one. Start with a clean BLASTN page, make sure the other options are properly set, paste in one of the unknown sequences into the top input box. Now you are ready. Click on the "BLAST" button to submit your search to the computers at the National Institutes of Health (NIH) in Bethesda, Maryland. As the search begins your webpage will change to the progress page like below. This tells you that the search is happening. The database of DNA sequences is very large and the power of the tool is great. As a result a BLAST search can sometimes take a few minutes, thus the progress page is provided. 

4. The progress page will update every few seconds. When the search is complete you will see a new page. This page provides a graphical summary of the BLAST results. For our work this summary is not very useful because there are many BRCA1 sequences in the database that are similar to the unknown sequences. 

5. The interesting results are located below the graphic and shown as a table of results. In this example the input sequence is a small portion of the human BRCA1 gene that corresponds to your animal unknowns. The table shows details about the sequences that were found to be most similar to your input sequence. In this sample there are many human BRCA1 sequences and so the matching sequences are actually identical. The table provides a short description of the sequence that was found to make your sequence entered in the input box on the form. Your input sequence is called the query sequence. The matching sequence is called the subject. The next five columns indicate the subject sequence is identical to the query, we will not go into the details of the information in these columns.

6. Click on one of the descriptions, or scroll down the page, and you'll come to the details about the match. Two sequences are show, your query sequence above the matching sequence, aka subject sequence. Here the Subject is abbreviated "Sbjct". In between the sequences a line (‘|’) represents a match. A match is called an identity. There is other information also shown for each pair of sequences. The score is a measure of the quality of the match, we will not go into how the score is calculated here. The E value (Expect value) provides an indication of the significance of the match. In the image below the E value is 4e-32, this is a notation for 4 x 10-32. The E value is the probability by chance of finding a match with a similar score or better in the database. A value of 4e-32 indicates the probability of finding a match this good by chance is 1 out of 4 x 1032. The percent similarity of the two sequences in this case is 100%, it’s a perfect match.

7. Here is an example of a more typical alignment, one that is not perfect. Notice that the E value indicates the probability of finding a match by chance that is this good or better is one out of 1 x 105. The percent identify is 84%. There are no gaps between these sequences. Gaps are spaces added in the middle of the sequence to give an optimal alignment to accommodate for an insertion or deletion. Thus with no gaps the percentage of gaps is 0%.

8. You are now ready to search your unknown DNA sequences against all known public DNA sequences using the BLAST tool.