Run web-based BLAST

Query: human brain type creatine kinase, *NP_001814*
Program: BLASTP
Database: refseq_protein
Goals:

- Identify members of this protein family in mammals.
- Use taxonomy report, formatting options, TreeView, and links to explore results.

Procedure:

1. Retrieve *NP_001814* from the Entrez protein service **http://www.ncbi.nlm.nih.gov/protein/**.
2. Click **"BLAST"** , then Click on **"Protein BLAST"**
3. Select **refseq_protein** as the database.
4. Enter "mammals" in the Organism input box, select from the suggested list
5. Click **Algorithm Parameters** and increase the **Max target sequences** to 1000.

   Q1. Please list the default word size, threshold, scoring matrix and gap penalties.

6. Click **BLAST** to submit the search
   * The matches have different types of RefSeq accessions with *XP* entries representing proteins from gene models.
7. Click **Taxonomy report** to see the organism distribution and examine the matches for dog `(dog genome: Canis lupus familiaris)`

NCBI Reference Sequences with *XP* style accessions are predicted using gene models. Some of them maybe incomplete due to missing data in the genome or may represent potential but unsupported splice variants.

8. From the results page, click **Edit and resubmit**
9. Check the Exclude **Models (XM/XP)** checkbox
10. Click **BLAST** to submit the search.
11. The results now contain only *NP_* style accessions, experimentally supported gene products. Click **Taxonomy report**, to see the different creatine kinase products found in humans and other mammals.
12. Return to the BLAST results, click on the Distance tree of results and set Max Seq. Difference to 0.5. The resulting figure demonstrates relationship between the different proteins. The mitochondrial and cytosolic isoforms are two distinct clusters.

    Q2. Save the distance tree in a pdf file and submit.

Align two Sequences

Query 1: Human Albumin, *NP_000468*
Query 2: Human GC, *NP_000574*
Program: BLASTP

Procedure:

1. Retrieve *NP_000468* from the Entrez protein system **http://www.ncbi.nlm.nih.gov/protein/**.
2. Click **"BLAST"** , then Click on **"Protein BLAST"**

3. Check the box that reads **Align 2 or more sequences**.
4. Enter *NP_000574* in the subject sequence box.
5. Click **BLAST**
6. Expand and examine the **Dot Matrix View**

Q3. Submit the dot plot matrix you obtained.

Needleman Wunsch Global Sequence Alignment

Query 1: Human Albumin, NP_000468
Query 2: Human GC, NP_000574
Program: Protein

Procedure:

8. Click on the **Global Align** in the **Specialized searches** section of the **BLAST homepage**.
9. Click the **Protein tab** over the Query sequence text area.
10. Click the **Align** button

Q4. Submit the dot plot of the alignment.

The "Align 2 (or more) sequences" service is now combined with Basic BLAST. Checking the "Align two or more sequences" on the BLAST form will transform the BLAST form to allow direct comparison of two input sequences. This service produces only local alignments since this is BLAST. In cases such as the albumin family used here – where there is a set of repeated domains, more than one alignment is found. This is easily seen in the dot matrix graphic of the alignments found between albumin and the vitamin D binding protein. The new Needleman–Wunsch alignment tool allows a global comparison of albumin and the vitamin D binding protein and produces the single best alignment that includes all residues.