Will use protein this narwhal FASTA file RefSeq\_Narwhal\_100\_GCF\_005190385.1\_NGI\_Narwhal\_1\_protein.fasta

(*https://github.com/pwilmart/PAW\_BLAST*)

Make subset FASTA database (make\_subset\_DB\_from\_list\_3.py)

Use identifications.txt and Narwhal fasta above.

Will make \*NGI\_Narwhal\_1\_protein\_subset.fasta

For the target we will use a human, but this can be any other fasta on compareWeb

Ex. 2019.06\_UP000005640\_9606\_Homo\_sapiens\_canonical.fasta

Run BLAST ortholog script (db\_to\_db\_blaster.py)

Select Narwhal DB first, then human

BLAST script makes a tab-delimited results file

RefSeq\_Narwhal\_100\_GCF\_005190385.1\_NGI\_Narwhal\_1\_protein\_subset\_vs\_2019.06\_UP000005640\_9606\_Homo\_sapiens\_canonical.txt

\*the db\_to\_db\_blaster script is looking for Blast to be in a certain location, so be sure that is correct and correct name to correct version (if diff). [Here](ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/) is blast download and [here](https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastDocs&DOC_TYPE=Download) is info.

**Stop here unless you’re target is human.**

Select hit\_acc column (without header) from the \*.txt above and copy to clipboard

*What is about to happen is within a GUI but you could hard code in the sprot.dat and keywlist.txt and settings.*

*(https://github.com/pwilmart/annotations)*

Run annotation script (add\_protein\_annotations\_6.py)

Human species should be default radio button

Click “Get accessions” button

Click “Parse DAT file” button

Need to browse to a human Swiss-Prot flat text file (sending separately since is too big for GitHub)

GitHub\compare\_web\testing\_files\Sprot\_dat\keywlist.txt file should also be present

Click “Add annotations” button

Have to select location for output files to be written

Annotations will be written to clipboard

At this point we need to merge the original identifications.txt file with the .txt output from db\_to\_db\_blaster and from add\_protein\_annotations.

Check that “Index” column and “hit\_acc” column match row-by-row, or more easily create an index to re-merge.

Save annotated results as a tsv or whatever for user