UniProt Hints

http://www.uniprot.org/

Part I: Searching UniProt

(Also look at http://www.uniprot.org/help/text-search)

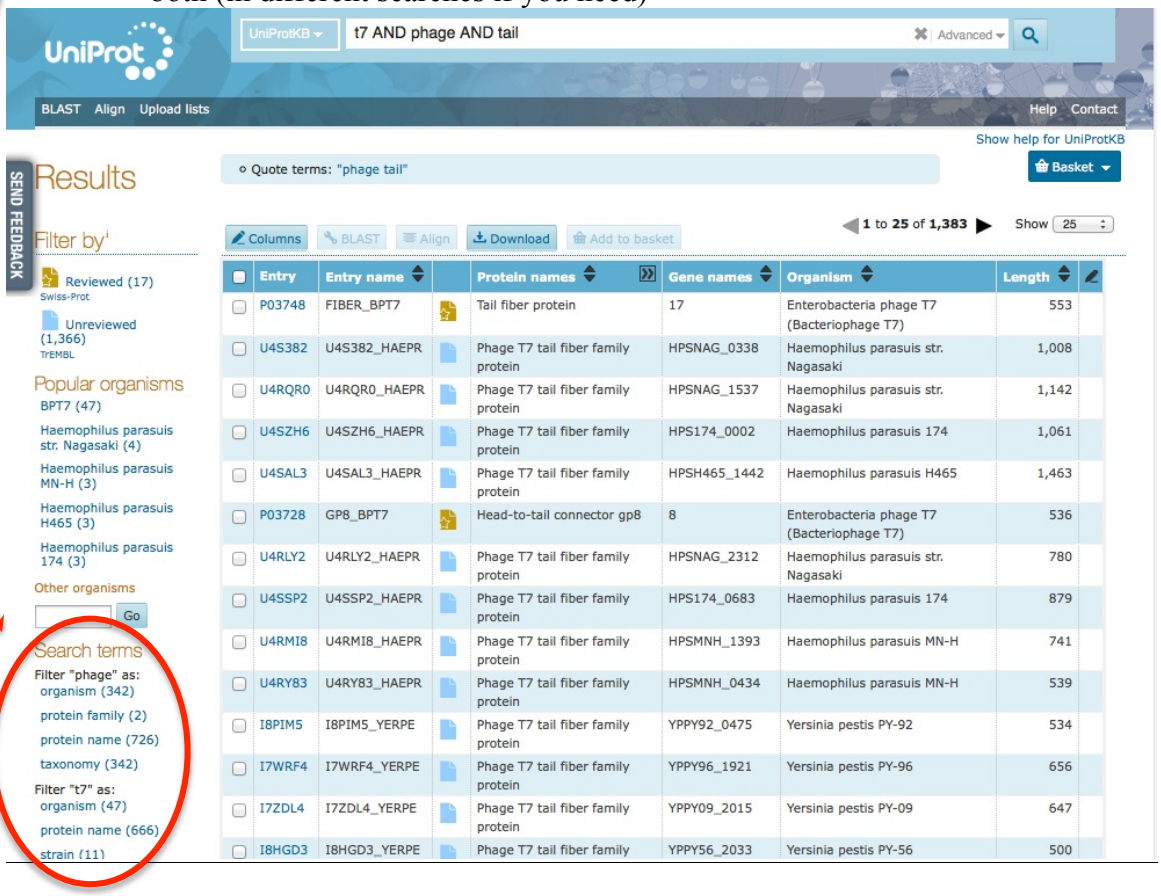
BASIC SEARCHES:

- Need gene or protein name
- Need organism name
  - → coli tolB
  - → tyrosine kinase human

USUALLY SEARCHES WILL NEED MODIFICATIONS:

Use the *Restrictions* at the top of results to refine searches you've done

- If your original search consisted of multiple terms, you will be given the option to apply that term to a specific field
  - Apply "cucumber", "coli", "Bacillus", etc. to the "organism" field
  - Apply your protein/gene name to the appropriate field- if you're not sure, try both (in different searches if you need)



ADVANCED SEARCHES (NOT ESSENTIAL IF YOU MODIFY AS SHOWN ABOVE):

- You can streamline your searches by entering the restrictions directly into the query bar. The commonly used options include protein family (FAMILY:), gene name (GENE:), protein name (NAME:), organism (ORGANISM:), strain (STRAIN:).
- AND/OR/NOT can also be represented by &&, ||, and -/! respectively:
  - → cancer AND gene OR genetic NOT human
  - → cancer && gene || genetic !human
  - → ORGANISM:phage AND NAME:"tail protein"

Part II: How do we select a protein from the search results?

Typically there will be several, to hundreds or thousands, of results. To select the appropriate accession (the "Entry" number in the far left column):

- Make sure the Species/Organism matches what is discussed in the paper
- Make sure the Strain matches what is discussed in the paper
  - "Escherichia coli K12" IS VERY DIFFERENT FROM "Escherichia coli O8"
- Consider the lengths of the potential proteins. Sometimes this info is presented in the paper
- Compare the sequences, if the information is given in the paper
- Sometimes the UniProt, or occasionally a NCBI reference is given in the paper
- If you still have several similar possibilities, **ask for help if you need it**!

Part III: What do the stars (reviewed/unreviewed) mean?

UniProt IDs are assigned to many, many sequences submitted by many, many researchers. Some are actually proteins (which is what they want), but some are fragments, repeats of previously known sequences, etc. The entries that have NOT been professionally reviewed have grey stars- this kind of means "this MIGHT be a protein, but we haven't checked it yet". If you don't have any gold-starred entries, it is perfectly acceptable to use these grey ones- especially with rarer, non-model organisms including many phage.

The gold stars mean the professional reviewers (Swiss-Prot) have verified the entry. If you have a gold-starred option, it will GENERALLY be the correct one; if in doubt, it's usually safe to choose these. If you have multiple gold-starred options that seem to be the same length, same organism & strain, same exact gene, ask for help- it probably won't matter which one you choose though.