

Are the proteins from the Uniprot activity related?

- These proteins are not related to one another because they each produce a different structure/effect.
- These proteins are related as they all have the same catalytic domain (biopterin).
- These proteins are most likely related to each other since they share a majority of features, such as being found in the same organism, having similar subcellular locations and sharing a majority of pfam domains.

| Database | For the last hit | | | Taxonomic range of hits | Number of hits in human? | BLAST RID – last character | Number of sequences in database |
|--------------------------|------------------|---------|-------|-------------------------|--------------------------|----------------------------|---------------------------------|
| | %QC | e-value | % ID | | | | |
| Nr | 100 | 0.0 | 92 | Placentals | 20 | 6 | 340795435 |
| Refseq_proteins | 100 | 0.0 | 91.37 | Placentals | 2 | 6 | 185399577 |
| Modelorganism (landmark) | 11 | 0.006 | 44 | Eukaryotes | 9 | 6 | 438637 |
| Uniprot | 15 | 4e-04 | 40 | Eukaryotes | 4 | 6 | 475372 |

Many very similar hits, many from human

Many very similar hits, but few from human

Many hits from human, even some bad hits. This is good!

Fewish hits from human, even some bad hits. Is this good or bad?

nr

Refseq_protein

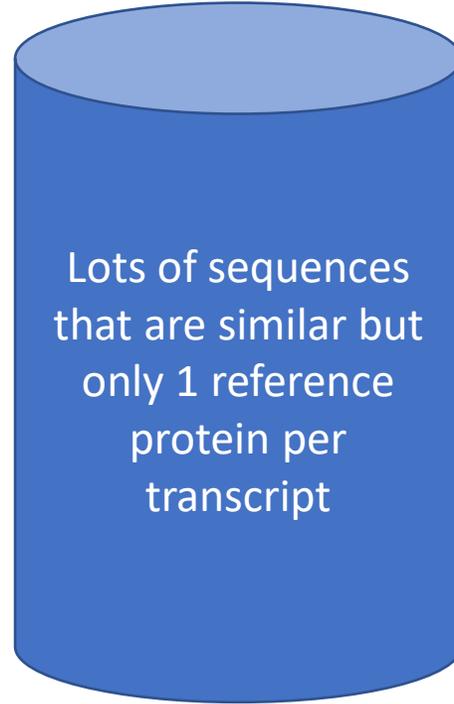
Landmark

Uniprot



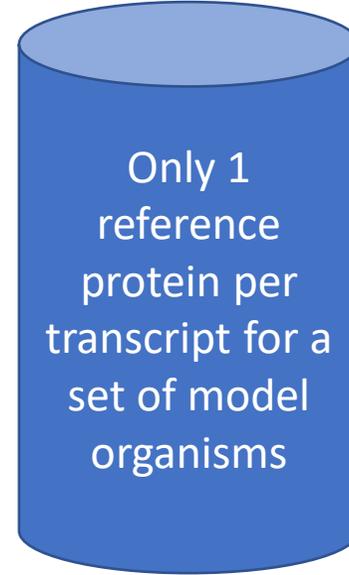
Lots of sequences that are very similar. Many proteins are represented multiple times because they may differ on the nucleotide level but not on the amino acid level.

Search with default setting yields very similar sequences.



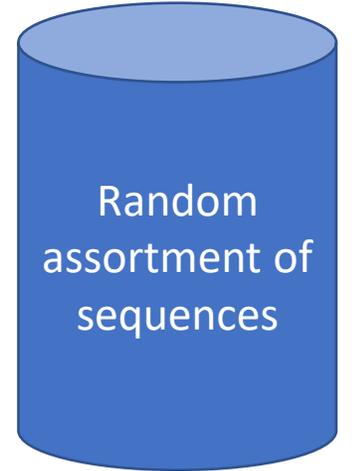
Lots of sequences that are similar but only 1 reference protein per transcript

Search with default setting yields similar sequences.



Only 1 reference protein per transcript for a set of model organisms

Search with default setting yields more diverse sequences.



Random assortment of sequences

It depends! But mostly diverse. The least inconclusive of a species has a homolog or not...

| Name | Length | Protein accession | GeneID |
|----------------|--------|-------------------|--------|
| PAH | 452 aa | NP_000268.1 | 5053 |
| PAH | 452 aa | NP_001341233.1 | 5053 |
| PAH isoform X1 | 240 aa | XP_016874859.1 | 5053 |
| TPH2 | 490 aa | NP_775489.2 | 121278 |
| TPH1 | 444 aa | NP_004170.1 | 7166 |
| TH isoform b | 497 aa | NP_000351.2 | 7054 |
| TH isoform X1 | 501 aa | XP_011518637.1 | 7054 |
| TH isoform a | 528 aa | NP_954986.2 | 7054 |
| TH isoform c | 524 aa | NP_954987.2 | 7054 |

Many proteins, but only 4 genes:

Alternative splicing -> protein isoforms

Isoforms may be very similar, may just differ by one or two amino acids in length, but can also be very different in length.

Are there paralogs in this table?

Yes, different gene ids in the same species

Are there orthologs?

No, all are from the same species. Why can there not be orthologs from the same species?

In class today:

- Complete the computer set up
- Complete Day 4 activity
 - remember to submit to Canvas by Wed 1 PM
- Documents including sequence dataset for Day 4 activity are on the class website.
- The quiz from last week is open until 11:59 tonight...