**4-26-20 Final Project Planning — revised notes from the project planning proposal**

>What is the biological question you are trying to answer with this project?

Taste Receptors (TRs) are a group of chemoreceptor proteins, which in mammals have been widely identified as key to taste sensation. TRs belong to the A class of GCPRs (a massive gene superfamily), and are further broken down into two families containing multiple T1R and T2R taste receptors, respectively.

The identification and number of TRs is less known outside of Mammalia. The general sense I’ve seen in the literature is that they are found in numerous other vertebrate lineages, but not in protostomes.

Given the known instances of gene duplication in vertebrates and the apparent lack of homologs in protostomes, I would like to investigate the following:

What is the gene structure (coding and non-coding sequences) of TRs amongst different deuterstome taxa? (Non-chordata vs chordata, and mammal vs non-mammal)

How many copies of each taste receptor do these different taxa have?

Are there any similar sequences or homologs in some protostomes? (Personal communications with Dave Plachetski suggests some cnidarians may have them).

>What dataset(s) will you use in to your analysis? What is the accession number?

Reference purple sea urchin genome:

ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/002/235/GCF\_000002235.5\_Spur\_5.0/GCF\_000002235.5\_Spur\_5.0\_genomic.fna.gz

Reference Mus Musculus genome:

https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/001/635/GCF\_000001635.26\_GRCm38.p6/GCF\_000001635.26\_GRCm38.p6\_genomic.fna.gz

Reference Zebrafish genome:

https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/002/035/GCF\_000002035.6\_GRCz11/GCF\_000002035.6\_GRCz11\_genomic.fna.gz

Reference Drosophila genome:

https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/001/215/GCF\_000001215.4\_Release\_6\_plus\_ISO1\_MT/GCF\_000001215.4\_Release\_6\_plus\_ISO1\_MT\_genomic.fna.gz

Reference Cnidarian (Hydra vulgaris) genome:

https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/004/095/GCF\_000004095.1\_Hydra\_RP\_1.0/GCF\_000004095.1\_Hydra\_RP\_1.0\_genomic.fna.gz

>What software will you need to complete your analysis?

For reference genomes:

- BLAST Suite (used to blast T1R and T2R homologs from mice and Drosophila in order to ID these genes in the other assemblies.)

- MSA (MAFFT) (Compare exons/introns and functional regions of genes identified

>What are your expected results? What do you hypothesize that you will find?

Based in part on previously published knowledge, I tentatively expect that Mammals will have the largest number of Taste 1 and 2 Receptors; however, I should be able to identify them in any other chordate, as well as deuterostomes in general.

In regard to Protostomes, I anticipate finding potentially similar sequences; whether they will be true homologs or (maybe?) structurally similar but unrelated remains to be seen.

>What are the steps (as far as you can tell) to your analysis?

Download all genome data via wget

Download T1R and T2R homologs from Mus musculus and Drosophila melanogaster

ID full genes of interest (T1R T2R, and similar sequences / homologs) from genomes

Use MSA to compare gene sequences