**Week 4**

**DAY 6: Find orthologs in NCBI**

***General instruction***

***This activity is an extension of the Day 5 activity. You will work in parallel with your group. Make one copy of this document that you add your (own) answers to. Each group member will pick one of the human genes from the previous activity and add it to the table below. You don’t need to submit this activity, but it would be nice if you did. Save the resulting files for later.***

|  |  |
| --- | --- |
| Gene ID | Gene |
|  | PAH |
|  | TPH1 |
|  | TPH2 |
|  | TH |

The teaching team will circle through the different breakout rooms. If you need assistance, click the *Ask for help* icon and we will come to you as soon as possible.

It is highly recommended to keep your cameras on and be unmuted so you can chat with each other while you complete the assignment.

**A.** **For your gene, go to its Gene page (hints in figure). Here we will use p53 from human (GeneID 7157) as an example. Then click on Orthologs as shown by the arrow:**

****

**B.** **Select from the following groups at least 10 mammals, 3 birds, 1 amphibian, 3 fish, and 1 shark and in total at least 20 sequence and add to cart.**

****

**C. Click on your Cart and then on Download. Select Refseq proteins, followed by One sequence per gene, and Download. This will result in one representative protein from each gene, if the gene is alternatively spliced (resulting in protein isoforms). For your data, what may pros and cons of only selecting one sequence be?**

****

**E. Use NotePad++ or BBEdit to count how many sequences you have in the dataset.**

**F. Build a PhyML tree (SH-like) for your combined sequences using Muscle with default as the alignment. Remember to assign short names to your sequences and discuss in the group how will rename the sequence**

**G. Make a tree figure that nicely shows where the 5 groups of your species are.**

**H. Build a species tree using NIH Common Tree.**

**I. Do your sequences (species and groups) follow the species tree?**

**J. Does you tree contain sequences that are related to each other through orthology, paralogy or both?**