**CME Assignment 2**

Drashti Mehta

ID: 801262877

Q1. What does your resulting phylogeny look like? (Include an image of the tree as your answer). Where is the platypus in relation to the other mammals on the tree? Where is the chicken (which is supposed to be the outgroup, i.e. the most phylogenetically distant species)? (3 pts)

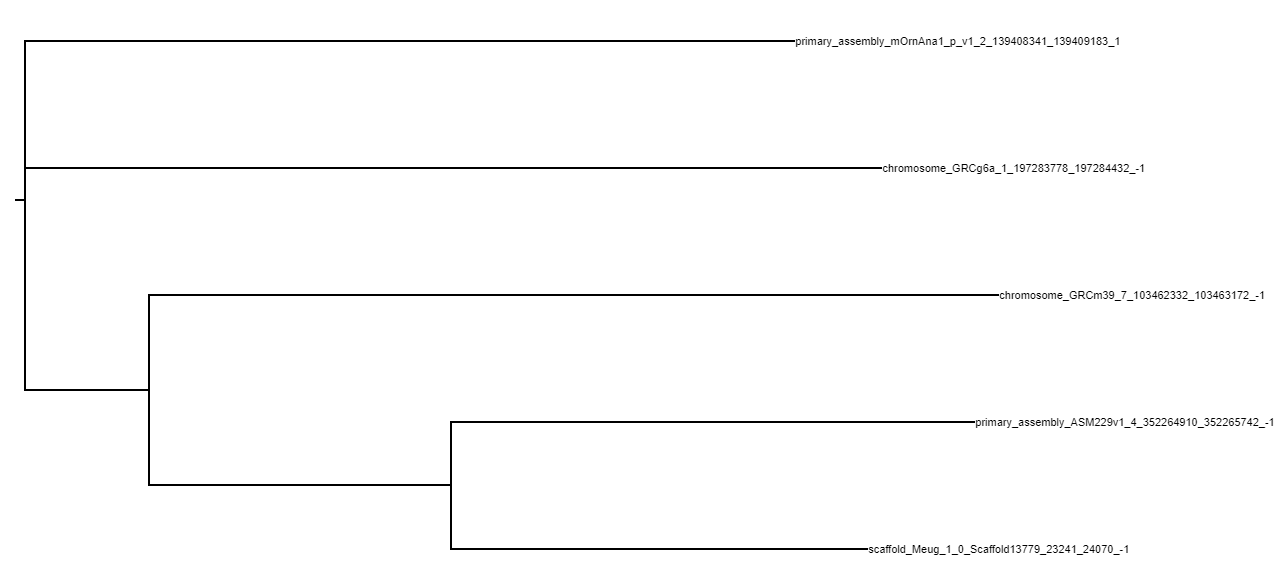


Fig 1 A: Phylogenetic tree (Without humans)

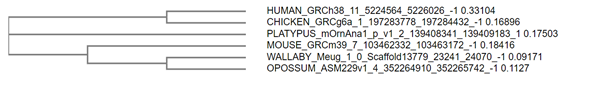


Fig 1 B: Phylogenetic tree (With humans)

A) So as per the image shown above, the platypus is located in the exact middle of the phylogenetic tree . It does not share relations with any other mammals on the tree. Also, chicken shares relation with humans

Q2. After you filtered out all of the other species, did any of our 6 species of interest have an ortholog of human HBB? Which one(s)? (2 pts)

A) Chicken (Gallus gallus) has two genes which are orthologous to the human HBB gene, namely GBE and HBAD and both have a many : many relationship to the HBB gene.

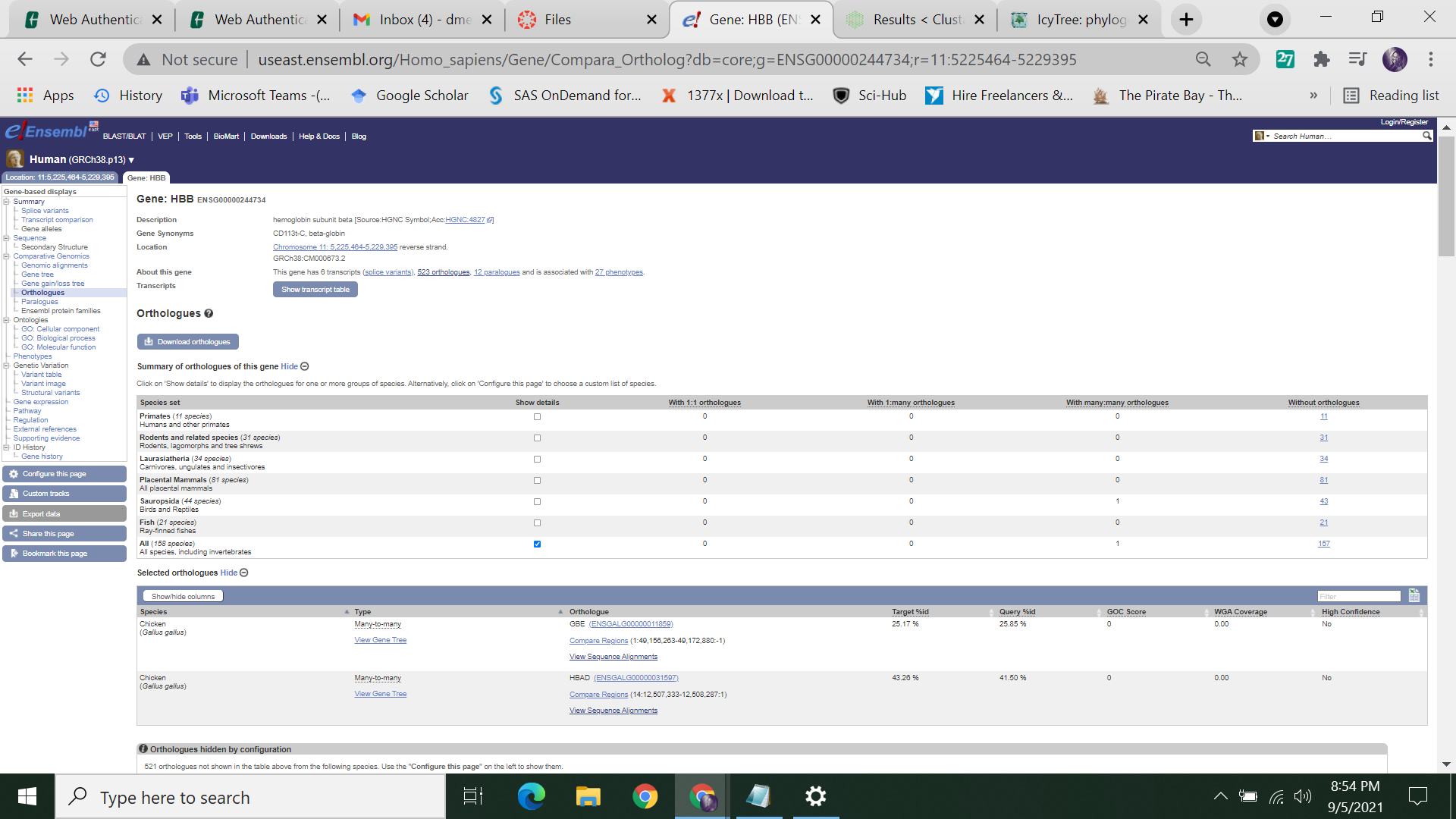


Fig 2: Orthologues to HBB

**Q3.** Based on your results from looking at the list of orthologs in the database, what kinds of comparisons do you think you were actually making in Part I? Do you think the comparison in part I is likely to reflect the real evolutionary history of these species? Why or why not? (4pts)

A) The comparisons that we made in part 1 are mainly depending on different species with respect to the human HBB gene. The phylogenetic tree that we see at the end depicts the evolutionary history of species but lacks to explain the lineage and their actual homological type of relation, orthologs or paralogs. Here we see that all the mammals are related while chicken is at a distance from remaining, which seems correct. So, moving further in the assignment we come across the solution for the same issues by understanding the orthologs and paralogs of the genes based on other hemoglobin genes too apart from HBB.

For this, click on the “Paralogues” option on the left side of the page which will result in 12 paralogs: 11 paralogs and 1 duplication. For each of them, we have to click the “Orthologues” option from the left side of the page to understand the result. It will be depicted further

**Q4.**  Fill in the table below to indicate which species have one or more orthologs for all 12 genes in our table. Indicate if the orthologous relationship is 1-to-1, 1-to-many, or Many-to-Many by putting either "1-1", "1-M", or "M-M" in each box. If a species has no orthologs of a particular gene, you can either leave the box blank, right in "NA" or write "0". (6pts)

A)

| Gene Name | Mouse | Wallaby | Opossum | Platypus | Chicken |
| --- | --- | --- | --- | --- | --- |
| MB | 1-1 | 1-1 | 0 | 1-1 | 1-1 |
| HBG1 | 0 | 1-M | 1-M | 0 | M-M (2 genes) |
| HBG2 | 0 | 1-M | 1-M | 0 | M-M (2 genes) |
| Novel Ensemble prediction | 0 | 1-M | 1-M | 0 | M-M (2 genes) |
| HBQ1 | 1-M (2 genes) | 1-1 | 0 | 0 | 0 |
| CYGB | 1-1 | 1-1 | 1-1 | 1-1 | 1-1 |
| HBA1 | M-M (2 genes) | 1-M | 1-M | M-M (2 genes) | 1-M |
| HBA2 | M-M (2 genes) | 1-M | 1-M | M-M (2 genes) | 1-M |
| HBE1 | 1-1 | 1-1 | 1-1 | 1-1 | 1-M (4 genes) |
| HBM | 0 | 0 | 0 | 1-1 | 0 |
| HBZ | 1-1 | 1-M (2 genes) | 0 | 1-M (2 genes) | 1-1 |
| HBD | 0 | 0 | 0 | 0 | M-M (2 genes) |

**Q5.** Based on the results in the table, which gene do you think might work best for the phylogeny that we want to get? Explain your answer (2pts)

A) The gene that might work the best for phylogeny that we desire is the CYGB gene because it has a consistent 1:1 orthologous relationship with all the species. It shows the ancestral lineage of all the species. Hence, considering it into account will help us to find the lineage in the most clear way.

**Q6.** Download the orthologous sequences for the gene that you picked in question 5 (hint: there is a download orthologs button that will let you get all fasta seqs for the displayed orthologs), and re-do the sequence alignment and phylogeny. **Paste a copy of your phylogeny below. Be sure to specify which gene you used.** Note that you may or may not get a phylogeny that looks correct for the species, but you should definitely get a phylogeny that looks different from the one you obtained in Part I. (3pts)

A) As I considered the gene CYGB as the best for our current use, I am going forward with the same.

