Final Report for Bioinformatics Course

Mitochondrial Comparison and Phylogenetic Tree Construction

Moein Derakhshan

In this project, we examined mitochondrial DNA from various animal species, including birds, insects, and mammals. We aligned these sequences using multiple methods and then constructed a phylogenetic tree based on the nucleotide sequences.

Initially, we retrieved data from the NCBI database and included 10 species in the given dataset. The original dataset contained mostly mammalian and insect species, so we attempted to balance the selection. The mitochondrial sequences were filtered using the 'mitochondrion' keyword to ensure genetic relevance. The final dataset contains approximately 16,000 base pairs and is included in the species.csv file.

We downloaded mitochondrial sequences from the NCBI database, processed them using Python scripts, and saved them in "concatenated.fasta" file. We used the Muscle tool for sequence alignment. Due to compatibility issues of Muscle's latest version with Python, we downloaded and ran Muscle 3.8 exe manually. We used the presented code in the notebook for running the program. The execution took about 100 minutes on a Linux-based system. The aligned sequences were saved in muscle output.fasta, and we used three different methods to construct phylogenetic trees:

1. Bayesian Inference Trees

We used the BEAST software along with BEAUti for data processing. The fasta file was converted into xml format, and multithreading was enabled for performance

optimization. The analysis ran for approximately one million steps, and it resulted in 10017 trees which are all in the beauti-output.trees. By using the 'tracer' software, introduced on the BEAST website, we analyzed these trees. Through tracer we ran a similarity test on all the trees, and as can be concluded in Figure 1, we can see that the results have reached a common ground and constructed trees after one million steps are almost the same. The final tree was visualized using FigTree software.
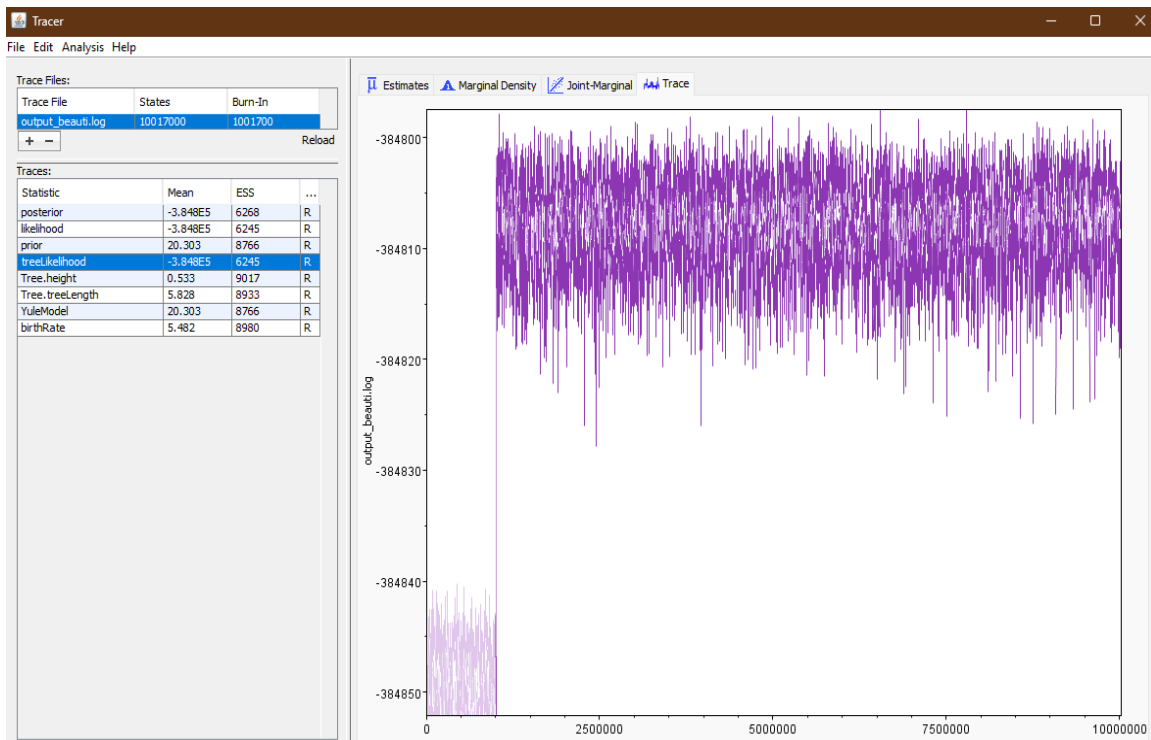

Figure 1 – Similarity of Constructed Trees vs the running steps

2. Neighbor-Joining Trees

We used the MEGA software to generate a Neighbor-Joining tree. The software itself had visualization tools. The resulting tree image is provided in Figure 3.

3. Maximum Likelihood Trees

The muscle output.fasta file was converted to phylip format using Biopython.:

```python
from Bio import AlignIO

AlignIO.convert("muscle output.fasta", "fasta", "file.phy", "phylip")
```

We used PhyML to construct the tree, and the results were visualized in FigTree. (Fig4)
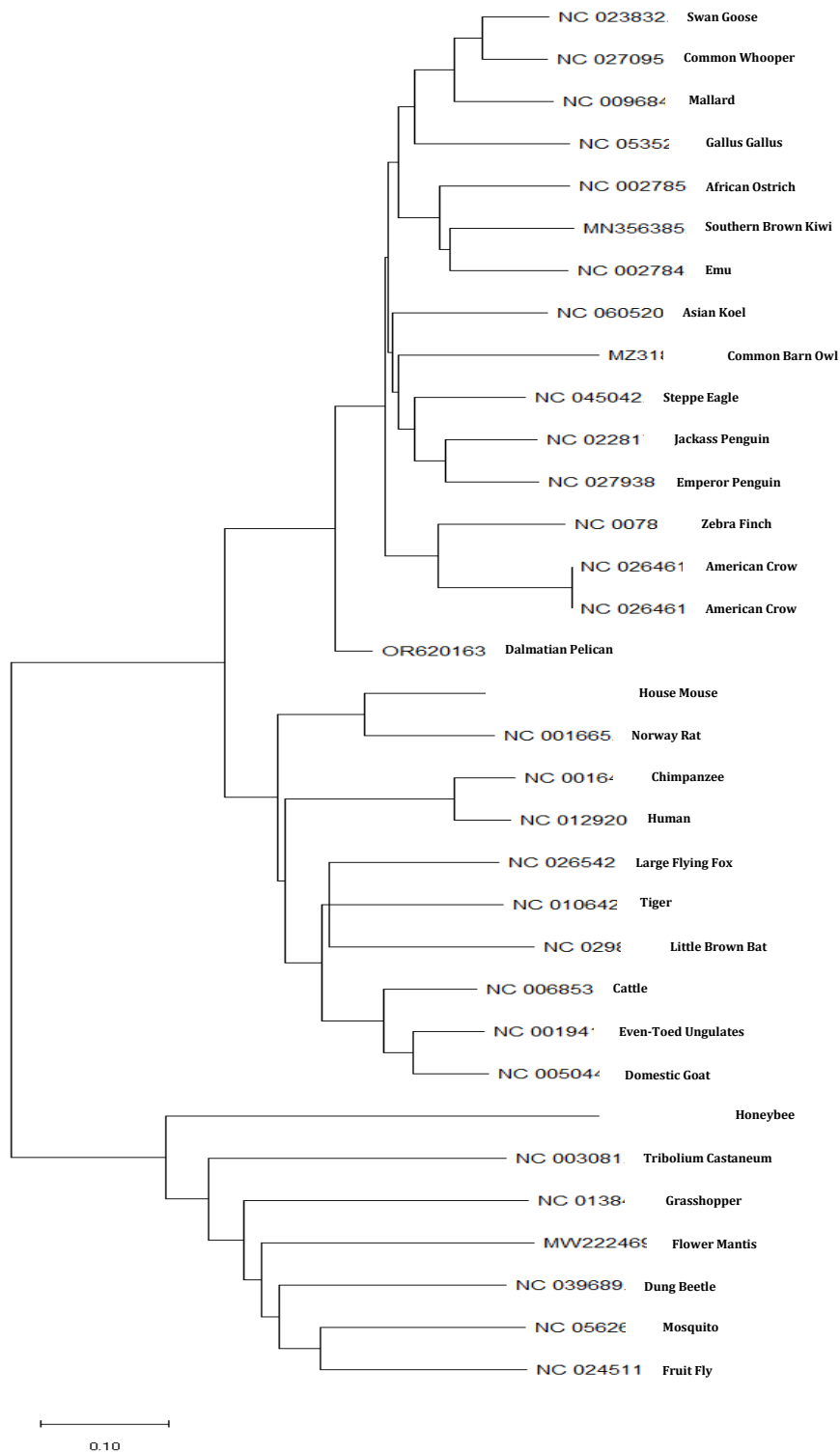
Figure 2 - Bayesian Inference Tree

4

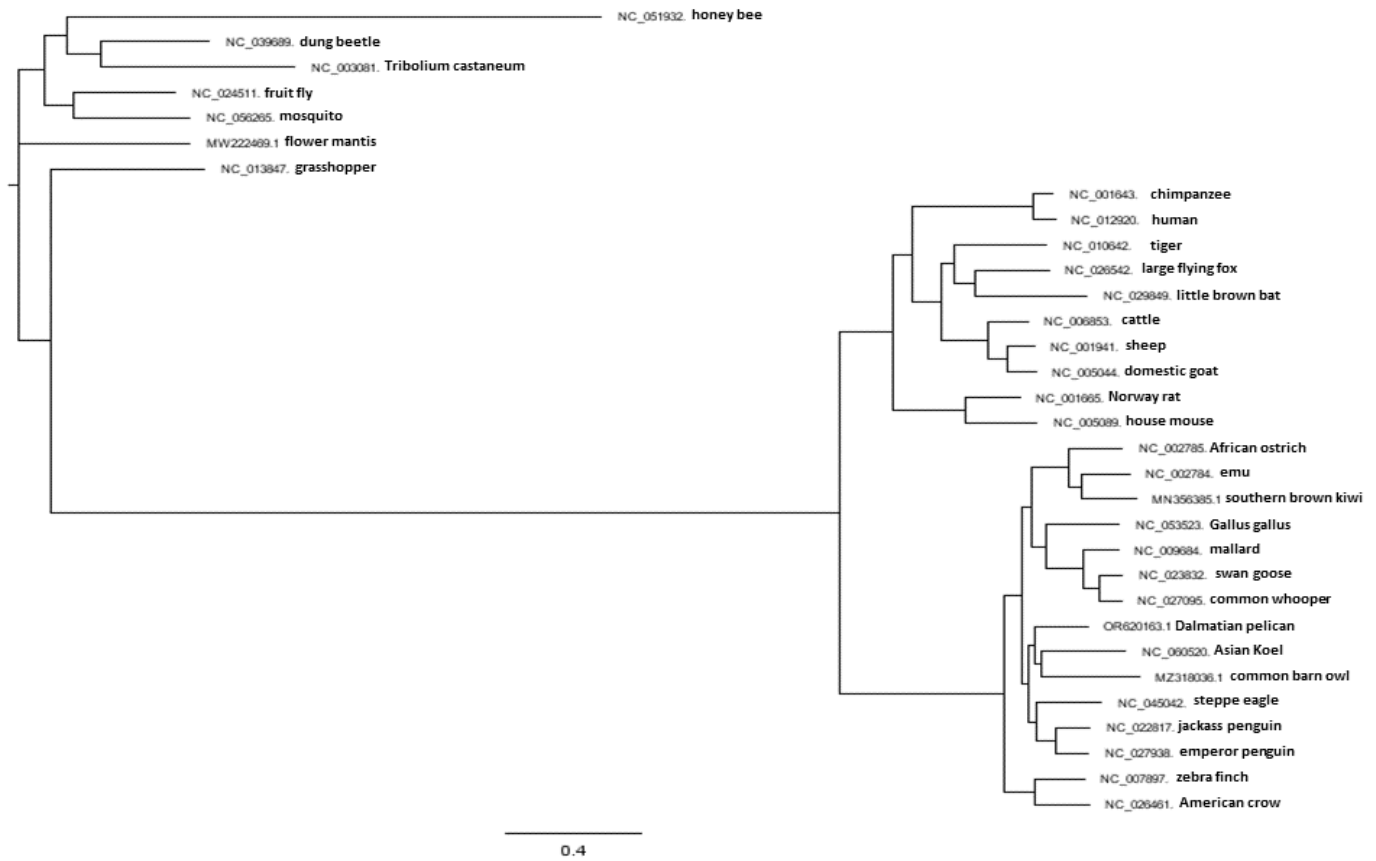**Figure 3 - Neighbor-Joining Tree**

Figure 4 - Maximum Likelihood Tree

Conclusion

In this project, we managed to analyze mitochondrial DNA from from mammals, birds, and insects to perform a phylogenetic analysis via three different methods: Bayesian Inference, Neighbor-Joining and Maximum Likelihood. Our results reflect on species divergence, phylogenetic relationships, and possible conservation considerations.

Using computational tools like BEAST, MEGA and PhyML, we identified evolutionary relationships. Phylogenetic trees show a clear separation between major taxonomic groups — mammals, birds, and insects cluster separately. The findings also show expected evolutionary relationships, like the tight grouping of chimpanzees and humans, as well as surprising relationships, like the close association of bats and primates. These results highlight the role of mtDNA in reconstructing evolutionary history. The differences in tree-building methods also bring out their strengths and weaknesses. Bayesian Inference is the best supported by statistics, Neighbor-Joining has computational efficiency, and Maximum Likelihood is a good compromise. This helps in better understanding of Phylogenetic results.

From a conservation standpoint, knowing which species are more closely related evolutionarily can help with conservation efforts, particularly those at risk due to environmental pressures. The preservation of genetically related species together allows for greater ecosystem stability and effective biodiversity conservation.

Answers to Questions from the Notebook

- **Considering the evolutionary relationships revealed in your phylogenetic analysis, what insights can be gained for conservation strategies? Specifically, how could understanding the close evolutionary ties between species, which might be facing distinct environmental challenges, guide targeted conservation efforts?**

Close evolutionary relationships between species are important for conservation initiatives. "For instance, with chimpanzees being so closely related to us as species, studying their genetic make-up offers potential knowledge we don't have in relation to human diseases and genetic disorders." Likewise, the evolutionary clustering of bats and primates suggests that conservation strategies designed for one group can benefit the other. For avian species, recent studies indicate that closely related species, swan goose and whooper swan as an example, could assist habitat conservation efforts. If one species is threatened by climate change or habitat destruction, conservationists can work to preserve the ecosystems that sustain both species, ensuring both thrive over the long term.

- **Are there any unexpected pairings or groupings of species that suggest a closer evolutionary relationship than previously thought? How could this reshape our understanding of these species' evolutionary paths?**

Bats, the large flying fox and little brown bat respectively, and primates (e.g. chimpanzees and humans) are closely related. Previous analyses have hinted at a more distant evolutionary relationship, while the phylogenetic analysis describes bats and primates as being more closely related to each other than to rodents. This may revolutionize our understanding of mammalian evolution and may imply certain genetic or physiological similarities that deserve further investigation.

- **Considering the discrepancies between the trees generated by different methods, what might this tell us about the limitations and strengths of each phylogenetic analysis method?**

The trees break down into three distinct phylogenetic trees, exemplifying the strengths and weaknesses of the methods used:

- Bayesian Inference: Generates statistically solid trees, but computation heavy.
- Enter neighbor-joining, a method that is fast and more accurate but does not give statistical support for evolutionary distances.
- Max likelihood: is a good tradeoff between accuracy and computation, however this can also be sensitive to input parameters

With the different ways of doing, it each has their pros and cons, this is why using just one way might not always give you the correct evolutionary relationships. A variety of methods combined can increase confidence in its phylogenetic interpretations.

- **Based on the trees, which species appear to have the most ancient divergence? How might this information contribute to our understanding of their evolutionary history?**

The phylogenetic trees show the most ancient divergence of insects, particularly of honeybee, dung beetle, and Tribolium castaneum from the other species. This is consistent with our understanding that insects are among the earliest diverging lineages in the tree of animal life. These early divergences provide glimpses at how evolution shifted from comparatively simple-to-more-complex organ systems and reveal genetic success stories that enabled insects to invade a wide range of environments.

- **How does maternal inheritance of mitochondrial DNA simplify evolutionary lineage analysis?**

Unlike nuclear DNA, mitochondrial DNA (mtDNA) is inherited solely from the mother. This prevents recombination between maternal and paternal genetic material, making it easier to trace evolutionary lineage.

- **Why is mtDNA a more sensitive tool for detecting recent evolutionary events?**

mtDNA mutates at a faster rate than nuclear DNA. This higher mutation rate allows for a more precise resolution in studying recent evolutionary events and species differentiation.

- **Which species show the most ancient divergence in the phylogenetic trees?**

According to Figure 2, domestic goat and sheep exhibit the most ancient divergence. In Figures 3 and 4, the swan goose and whooper swan show the most significant evolutionary distance.

- **Were there any unexpected evolutionary relationships?**

One surprising finding was that bats appear more closely related to primates than to rodents in the phylogenetic trees.