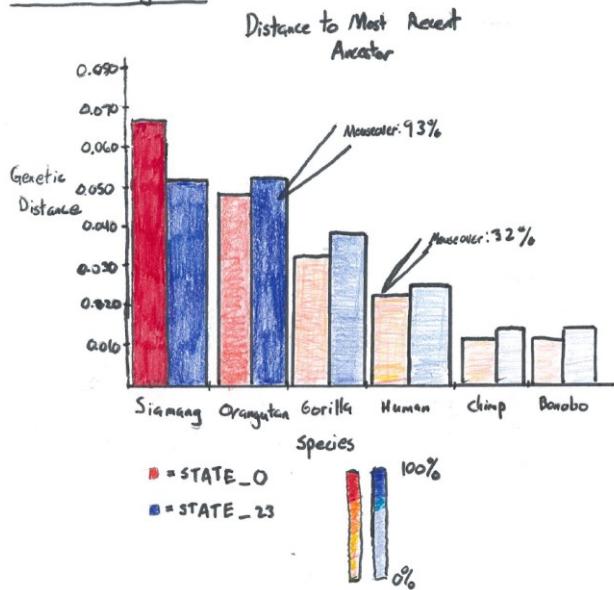


Design

Task 1: Visualization that shows the difference between “STATE\_0” tree and “STATE\_23000” tree.

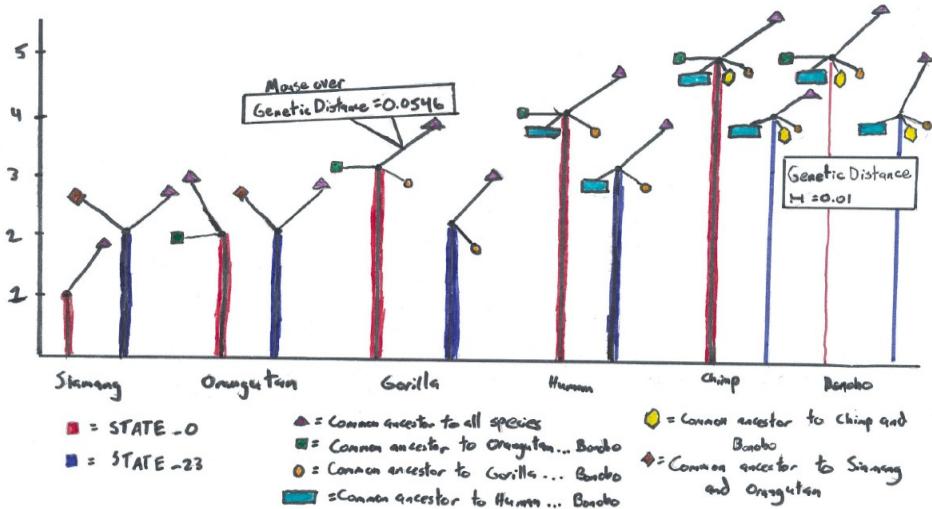
Task 1: Design 1

The first design is a graph that shows the distance to the most recent ancestor of each species for both trees. The darkness of each bar represents the percentage of the total genetic distance, and the exact percentage can be seen by putting the cursor over the specific bar. This graph shows that the most recent common ancestor for all the species in the “STATE\_0” tree has a greater genetic distance than the same ancestor in the “STATE\_23000” tree. Also, it is apparent that there is a common ancestor for siamang and orangutan in the “STATE\_23000” tree that is not present in the other tree, and thus the siamang species in the “STATE\_0” tree is older than in the “STATE\_23000” tree. The coloration of the bars helps show that the speciation events in the “STATE\_23000” tree are more bunched up

than the speciation events in the “STATE\_0” tree. The “STATE\_23000” tree has a shorter total genetic distance, but the last speciation event occurs earlier in this tree than the “STATE\_0” tree.

Task 1: Design 2

Ancestors and Their Genetic Distance

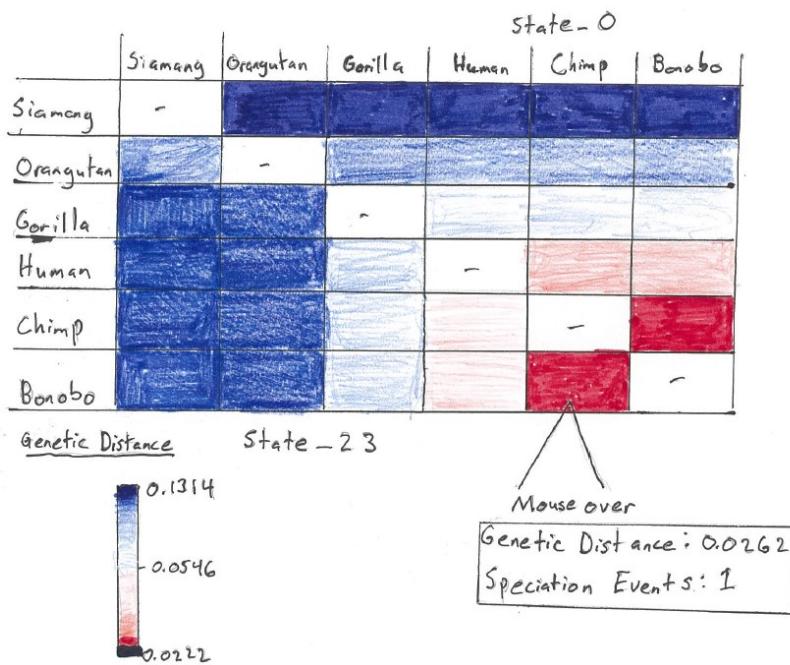


My second design shows the total number of ancestors for each species in both trees, and then the genetic distance from this species to each of its ancestors. Each of the shapes represents a different ancestor, and the length of the line to the shape is the genetic distance to this ancestor. To make comparisons between

species and the two trees easier to visualize, the similar shapes should be in the same position for each species in each tree, unlike in the orangutan for “STATE\_0” (i.e. the triangle for the common ancestor to all species should come out of the center at 60 degrees every time). Plotting the total number of ancestors makes it is easy to see the overall speciation structure of the trees. The genetic distances to the ancestors gives comparisons of relative similarity between species, and the shapes show what ancestors are similar and different for both trees. For example, the speciation event leading to chimps and bonobos happened last in both trees. For bonobos and chimps, the length of the line to the hexagon is shorter than the length of the line to the circle, so it is clearly seen that chimps and bonobos are more closely related to each other than they are gorillas. Also, by looking at the shapes it can be easily seen that the “STATE\_0” tree has a common ancestor to orangutan, gorilla, human, chimp, and bonobo that is unique to this tree, while the “STATE\_23” tree has a unique common ancestor to siamang and orangutan.

Task 1: Design 3

Heat Map of Pairwise Genetic Distance



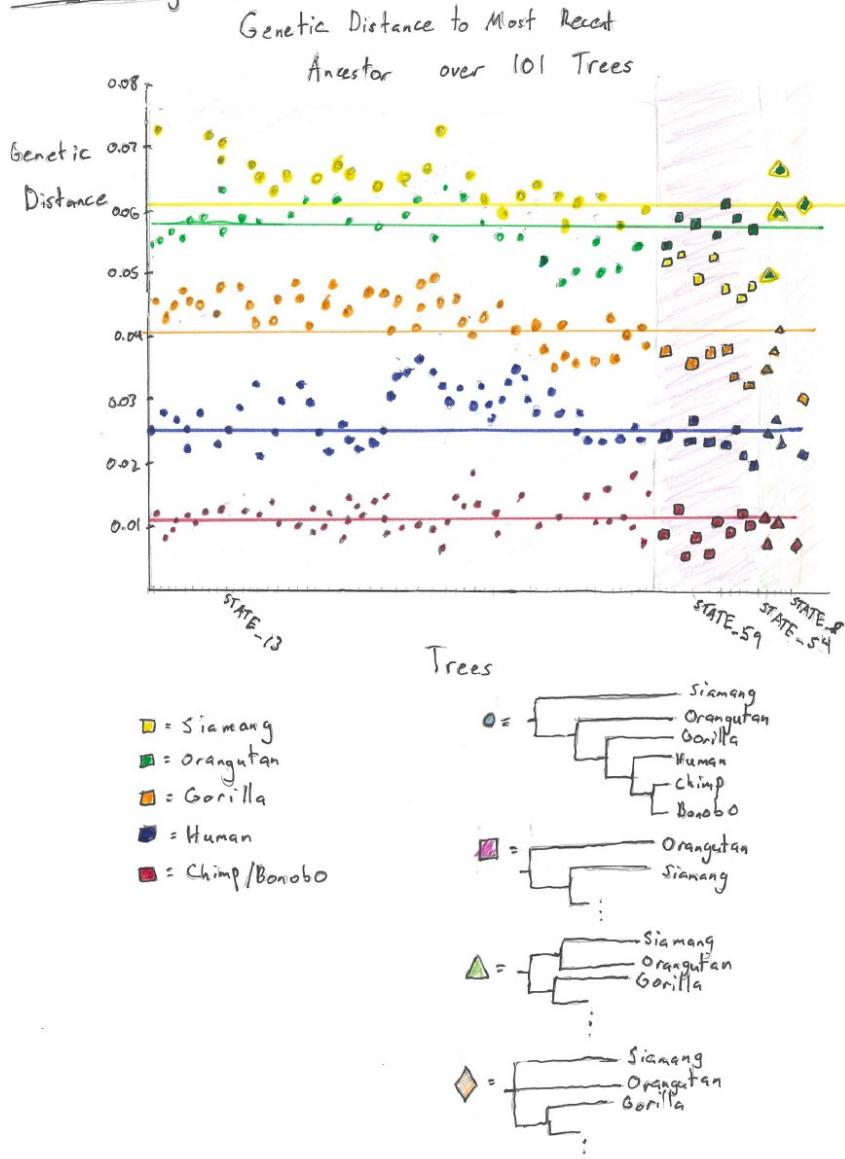
My third design for Task 1 represents pairwise genetic distances as a heatmap, with the “STATE\_0” tree in the upper right triangle of the matrix, and the “STATE\_23000” tree in the lower left triangle of the matrix. The dark blue represents the maximum genetic distance possible in the two trees, while the dark red represents the minimum genetic distance in the two trees. Putting the cursor over specific positions in the matrix shows the exact genetic distance and the number of speciation events that separates the two species.

The number of speciation events is included because it gives greater insight into the similarity of the two species that are being compared in that specific square in the matrix. Pairwise genetic distance was found by adding the total genetic distance between each of the species in both trees, and a greater distance means greater differentiation. Visualizing the pairwise genetic distances as a heat map makes it obvious that the main difference between the two trees is in the orangutan and siamang species: in the “STATE\_0” tree siamang and orangutan are separated while in the “STATE\_23” tree they are grouped together and are similarly differentiated from the rest of the species in the tree.

(When doing my implementation, I realized I made a calculation error in my design: the middle genetic distance should be 0.0768, not 0.0546. This means the gorilla column for “STATE\_23” and the gorilla row for “STATE\_0” should both be shaded slightly red, not slightly blue.)

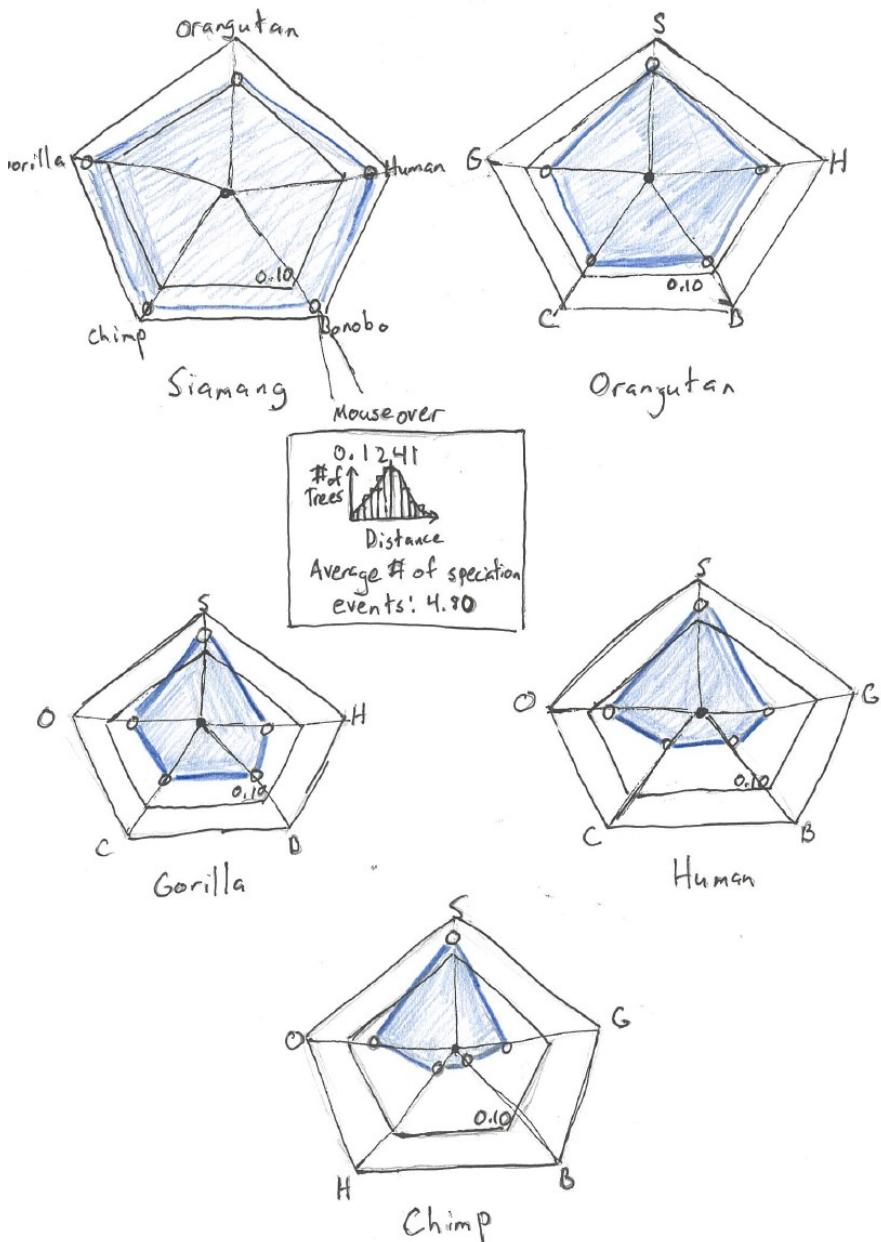
Task 2: Visualization containing all trees and how they are similar/different.

Task 2: Design 1



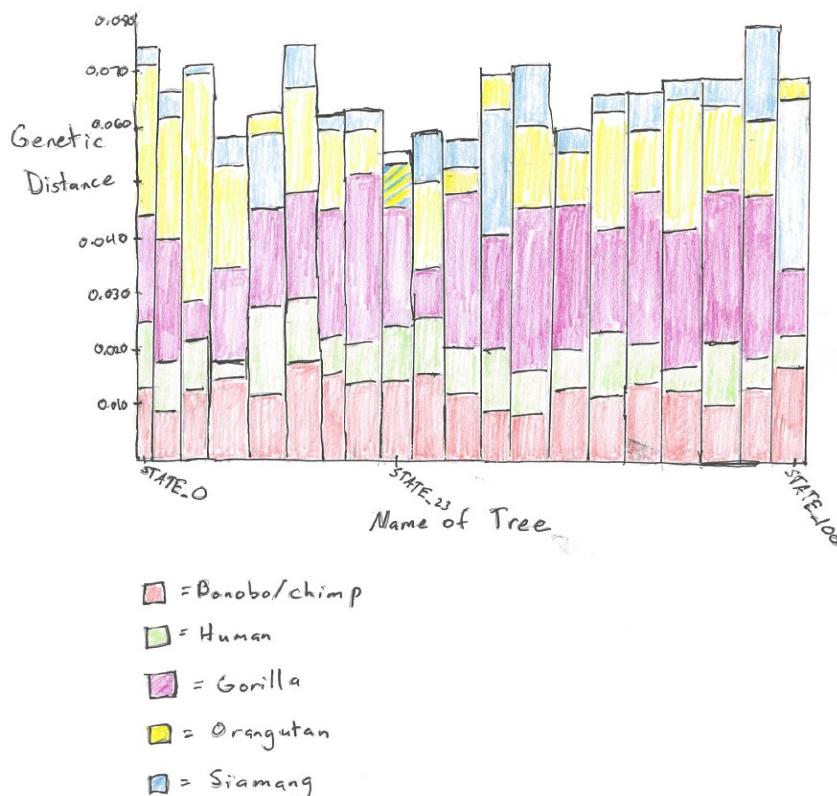
This is a graph of the genetic distance to the most recent ancestor for all the species in all trees. Each color corresponds to a different species, with chimp and bonobo grouped together because their most recent ancestor is common in all the 101 trees. The line through the points is the average genetic distance for each of the species to their most recent ancestor. The shape of the dot corresponds to the shape of the tree, and all the trees of the same shape are grouped together along the x-axis. It is difficult to see in this scan, but the background of the graph is shaded in different colors corresponding to the shape of the tree. This clearly shows that there are four different shapes of trees that show four different ways that these six species have evolved. The distribution of the different shape of trees, and thus the support values for each speciation event can be visualized. The differences in speciation events between the four shapes of trees can also be seen in the graph where the green dots jump above the yellow dots, meaning in this tree shape orangutans have the greatest

genetic distance to their most recent ancestor and therefore diverged first. The two-colored dots in the last few graphs show that in a few trees, the siamang and orangutan diverged and formed a separate clade.

Task 2: Design 2 - Average Substitutions per site Between Species

My next design are spider graphs showing genetic distance between species as average substitutions per site between species over all trees. The inner ring represents 0.10 base substitutions per site. These graphs clearly show which species is overall most similar/different from the other species, and from this we know the order that these species evolved in all the trees on average. The siamang graph has the most blue-shaded area, so this species has the highest average substitutions per site compared to all the other species. This means that it is the most different to all other species, and evolved first in the majority of the trees. The chimp graph has the least blue-shaded area, meaning it is the species most genetically similar to the other species and most likely evolved last. In the design, I should have noted that the bonobo would look exactly like the chimp graph, just with the bonobo node replaced with chimp. Putting the mouse cursor over each of the nodes brings up a window that shows a histogram with the exact

average substitutions per site between the two species and the average number of speciation events that separate the two species. The x-axis of the histogram are intervals of genetic distance, and the y-axis is the number trees that fall into each interval. The histogram allows for the visualization of the variation in the genetic distances between the two species that comes from each individual tree.

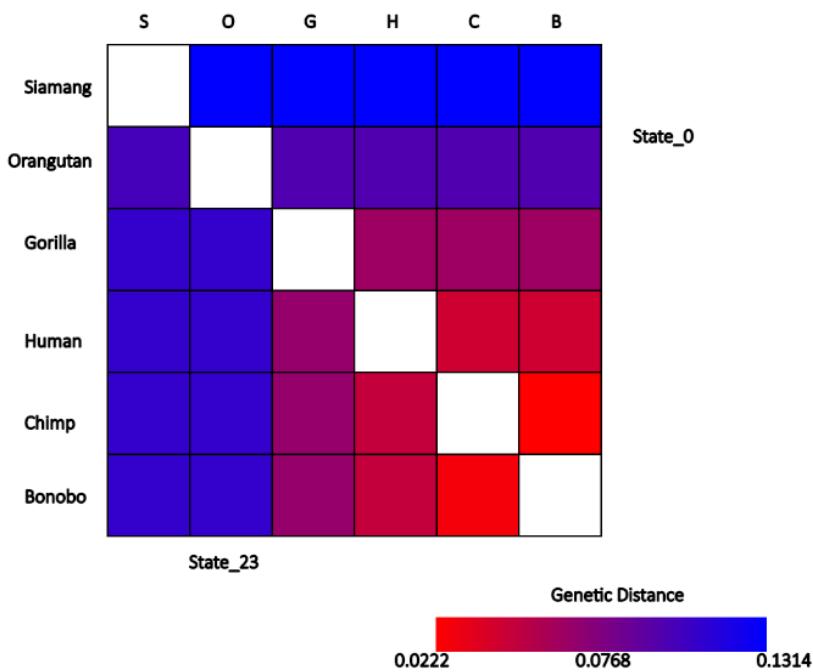
Task 2: Design 3

The last design shows the total genetic distance of each tree, and the distance to each speciation event that led to one of the final 6 species. This visualization highlights that most of the trees have the siamang as the species with the greatest genetic distance to its most recent ancestor, while for a few trees this distinction goes to the orangutan. Also, this graph shows how soon the speciation events happened to each other, which tells us how long each ancestor was living. The distance between Human-chimp/bonobo MRCA speciation event and the chimp-bonobo speciation event is small, so this ancestor was likely living longer than the MRCA of gorillas and orangutans whose speciation events are separated by a greater distance.

Implementation

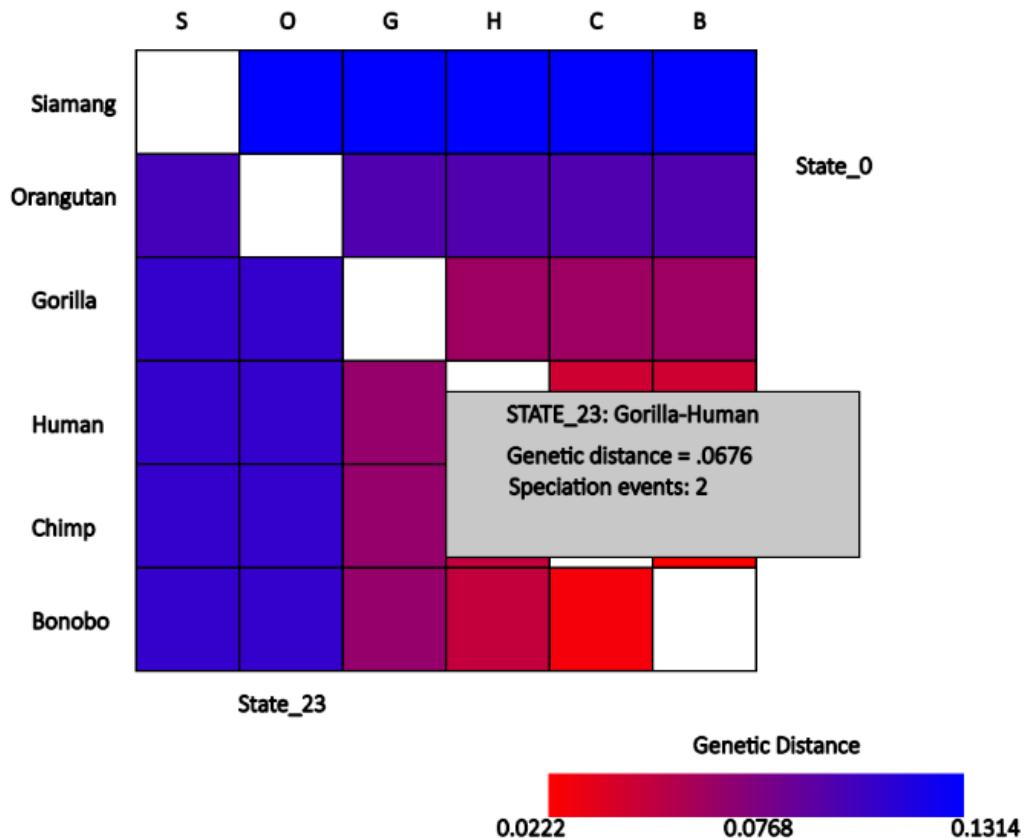
I chose to implement the heat map graph in P5.

Heat Map of Pairwise Genetic Distances



Each square in the matrix can be clicked on to show the number of speciation events between the two species for that specific tree and the exact genetic distance.

### Heat Map of Pairwise Genetic Distances



### Insights

Overall, these visualizations for Task 1 show that tree 0 and tree 23 differ most importantly in their relative relatedness for orangutans. Tree 0 shows orangutans are most similar to gorillas, while tree 0 shows them as most similar to siamang. Tree 23 has a separate clade for the orangutan and siamang species, which means they have a common ancestor that is not present in tree 0. Further investigation should focus on comparing orangutans to the siamang and gorilla species and looking for further support for this unique siamang-orangutan common ancestor.

The visualizations for Task 2 show the relative frequencies of each of the four shapes of trees, and the overall relative genetic distances for each of the 6 species. Again, these visualizations point to further research focusing on the siamang-orangutan relationship, as the difference in the shape of the trees is due to these two species. Their speciation events are closest together, so it makes sense that much of the significant variability among the 101 trees is due to slightly different hypotheses of evolution for these two species. Finally, since chimps and bonobos are always grouped together with a very low genetic distance, these two species may be further investigated to gain insights into what causes speciation events.

Screencast

<https://www.youtube.com/watch?v=3YM46kD8QUc>