

Visualization Assignment of Management of Large-Scale Omics Data

(IOU19a)

Hai LYU(r0605528)

Data Description

The object file contains 101 trees specifying the evolutionary models for several selected animals in Hominidae. Throughout the trees, the subtrees involving the human beings, the chimpanzees, the gorillas and the bonobos are nearly the same with differences in evolutionary distances (the length of branches). This complies with the knowledge regarding the evolutionary of these four Hominidae animals. The target of the visualization of these task is to represent the differences between two ‘states’ and similarities and difference among all trees.

Design

Task I

The differences between the state_0 and the state_23000 mainly concern the topology of the tree and the length of the branches.

The first design to demonstrate the differences is that merging the two trees into a complex. The subtree regarding the 4 species which is only different in branch lengths will be shared by the two trees. By clicking to switch between the 2 trees, the tree of selected state will be emphasized in bold or dark colour and the rest part which definitely belongs to the non-selected tree will be shown in a distinct way. As the state have been selected, the branch lengths will be shown in figures.

The second design will try to represent the evolutionary processes by animation of the development of the two trees. The trees will grow separately at the same development speed for instance 0.005 branch length per second. The main benefit of the animation is to reconstruct the evolutionary process modelled by the trees.

The third design will try to directly compare the two states. The species will be listed at a certain sequence in the central. Then the first state will be shown at the left side and the other tree will be shown at the right side. The direction of the two trees will be demonstrated in opposite. By clicking to select any two of the species, the branches in each tree connecting the two species will be bolded and the distances by each tree will be shown at the bottom of the tree.

Task II

As the design will involve all the 101 trees, the problems will rise in quite many aspects. Fortunately, the subtree shared by all trees offered the opportunity to decrease the

complexity in tree topology.

The first design aims at the tree topology. The subtree will be taken as the whole and there can only be 3 possibilities in the tree topology regardless to the branch lengths. Users will be able to select among the 3 possibilities and numbers of tree states complying with the select topology will be shown and a list of these states will be given. If the specific state is not selected, a tree based on statistical result will be shown. Once a specific state is selected, the tree will be constructed with its detailed branch lengths.

The second design will give an alternation in tree topology selection. Users may drag the species of orangutan and siamang and the node of the subtree to rearrange their position in the tree. Once the positions of the 3 parts get set, the supporting states will be listed for selection. The rest part is the same with the first design after tree topology selection.

The third design will try to figure out the distances between any two species. By selection through two dropdown menus, users can select the specific state or statistical result to be represented. By default, the species will be ranked by the evolutionary distances to human beings ascendingly. The figures will be arranged in the form of matrix. If users select a specific species at the rows, the columns will be rearranged according to the distances to the selected species from left to right ascendingly. Of course, the relative constant relationship between humans, gorillas, bonobos and chimpanzee will be treated distinctly. To be the same with the first design, if the state is not specified, the statistical result will be represented as summary.

Implement of Design

Task I

By p5.js, the first design has been implemented. The basic idea is to rise direct comparison between the two states and emphasize the common subtree shared. Since the difference in branch lengths is not very significant, the trees are constructed regardless to the length differences which would be shown by figures near the branch.
<https://www.youtube.com/watch?v=qfOZwuMblbE>

At first, you will see the complex with different parts represented in distinct colours and weight. Then, you would be able to switch between the two states. Once you select a state, the tree demonstrating the state will be shown in bold and the part completely belonging to the another state will be shown in a minor way. Also, the lengths of each branch of the selected state will be given with state_0 figures shown at the left side of the corresponding branch and state_23000 figures shown at the right side of the corresponding branch.

Task II

By p5.js and java script, the first design has been implemented. The basic idea is to summary the 3 tree topologies together with basic statistics results.
<https://youtu.be/IT3OUZfFSxQ>

At first, when the selection is not specified, the notes will be represented to ask the selection to be made by dropdown menus. When a topology is selected and the state is not specified, the statistical result will be given with a tree constructed according to average of the lengths of each branch. When the state is selected, the tree will be constructed according to the branch lengths of the state.

Since the time is limited for me to fulfil the assignment, the refinement would not be undertaken and I will try to specify some here. For both tasks, images that could show a portrait or special trait of the corresponding species will be represented aside with the name of the species. The lengths of the tree in task I can be refined to give slightly differences in evolutionary distances between the two states. For task II, perhaps the distribution of the lengths of each branch given a selected tree topology can be shown when users clicking at any of the branches. If the state is not specified, a summary of distribution will be given. Once the state is selected, the position where the state's length lies in the total distribution will be shown.

Insights

Briefly, except for the common subtree shared by all, the most important insight is that siamang is more evolutionarily distant to the 4 species forming the subtree than orangutan. 78 out of 101 trees support it. Also, 17 trees support that orangutan is more evolutionarily distant to humans than siamang. 6 trees support that siamang and orangutan is equally evolutionarily distant to human beings. Since we have no idea how the distances of each state measured, we can't move further to imply more with respect to the detailed branch lengths.