

Task 1. Demonstrate Tree0 and Tree23



Problems of this common phylogenetic tree (generated by R)

- Hard to compare length (i.e. evolutionary time) between two trees
- Hard to compare evolutionary relationship, as the topological structure difference doesn't mean relationship difference. And if use different node as root, the topological structure will be different.

Design 1. Mirror trick (interactive) – Processing

Demonstrate two trees at mirror position, for human eyes, it will be clearer to compare things if they are

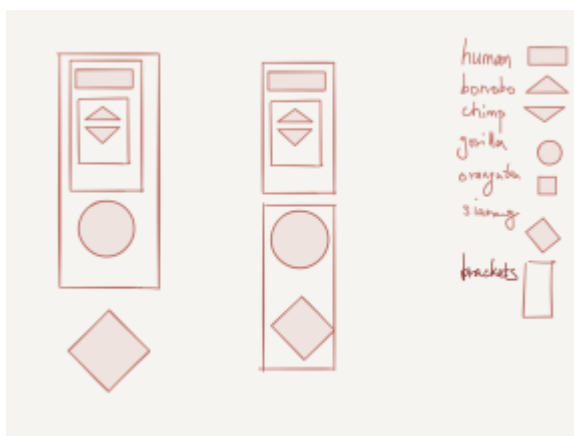


symmetrical or not.

Interactive: By moving mouse, the angle of each bunch will change, so user can use mouse to see more detail or compare them in general. Same species always have same angle

From this figure, we can easily see the main difference is the position of orangutan. Better with more details like label.

Design 2. Gestalt Law



Law of Proximity: Objects that are near one another in space or time are perceived as belonging together.

Law of Similarity: Objects with similar attributes (such as shape, color, size, or brightness) are perceived as belonging together.

- Use different shape to represent different species
- Use rectangle to function as brackets.

The function of closed rectangle make user to think the shape in rectangle are closer.

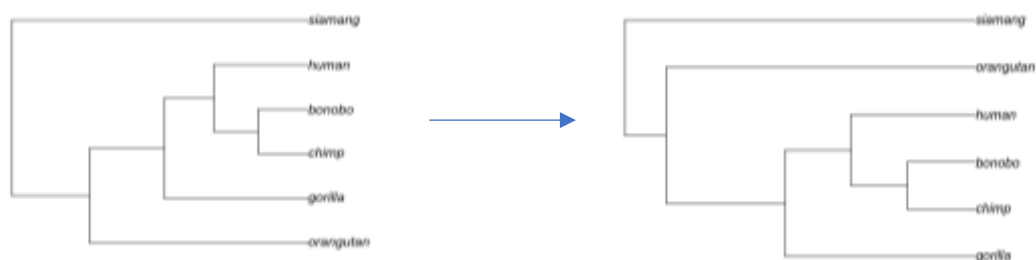
From this figure, we can easily see the only difference between tree0 and tree23 is the relationship between gorilla and siamang.

Pros and Cons: Easier to see difference and similarity but neglect evolutionary time.

Design 3. Root change and structure rebuild

As root will influence topological structure of the tree a lot, design an interactive tree which can interact by changing roots will be much useful. Comparing trees have same root will be clearer. To illustrate, by clicking node, the tree will change to new tree based on new root. Also, if we can order species in same order and in the meantime, keep relationship, will be easier to compare.

For example, tree0 and tree1, their evolutionary relationships are same, even if they look different. After root change and structure rebuild will solve this problem.



Task 2. Multiple trees

Compare all 101 phylogenetic trees in one figure.

Design 1. Color

User Assumption 1: assume that users are human beings, then users want to use human as a point of origin, they want to see the relationship between other species and themselves.

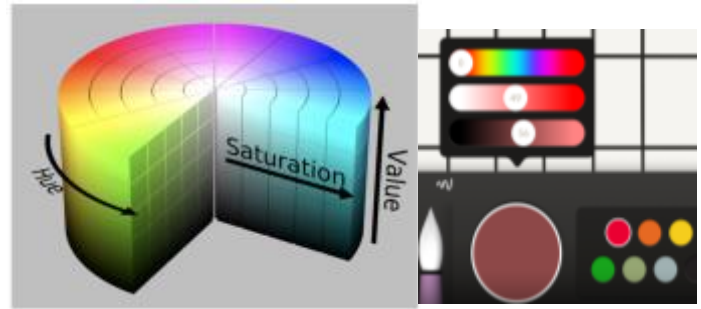
User Assumption 2: assume that users without scientific training, then users may not know the concept of tree and numbers' meanings.

Based above assumptions, we use color instead topological structure and numbers to represent evolutionary relationship.

1. Set 'human' as center (allow to pick other species as center).
2. Select center color.
3. Relationship calculation.

a) Method 1:

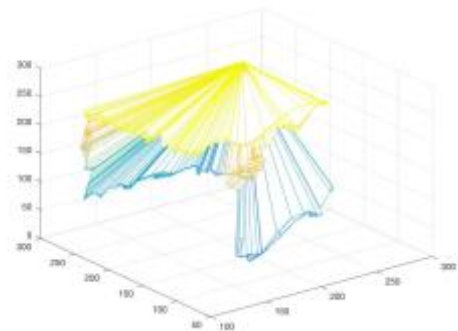
- i. According to the length of evolutionary time, calculate saturation of color. Evolutionary time longer, saturation higher.
- ii. According to the distance of relationship to human, calculate hue of color. Farther the distance from human, hue difference bigger.



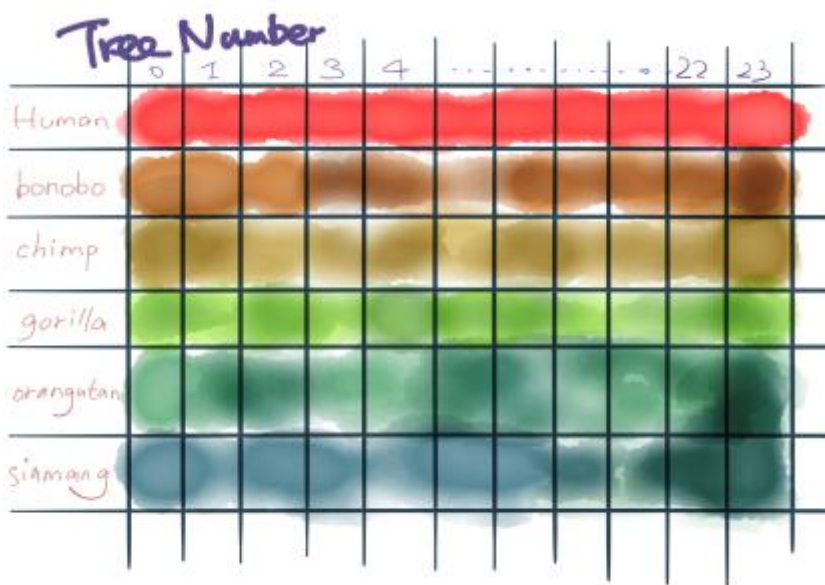
b) Method 2: (Matlab)

- i. Build RGB 3D coordinates, and common phylogenetic tree (scaled).
- ii. Map tree into 3D coordinates, set 1D coordinate manually, according to the position in the coordinates, get color value.

Example 1: use Matlab transform a figure to 3D coordinates (RGB), then It is viable to get position in 3D coordinates then generate color chart.



Example 2: Use app on iPad make a sketch



From this figure, we can see the evolutionary relationship by Hue, for example, red is closer to orange, purple is closer to blue, we can get evolutionary time by saturation, for example, some boxes have low saturation, some have high saturation, indicates after separating from node, species undergo different evolutionary time. Specifically, tree 23 is very different as color of orangutan and color of siamang are more similar, which means they are closer to each other than others.

siamang are more similar, which means they are closer to each other than others.

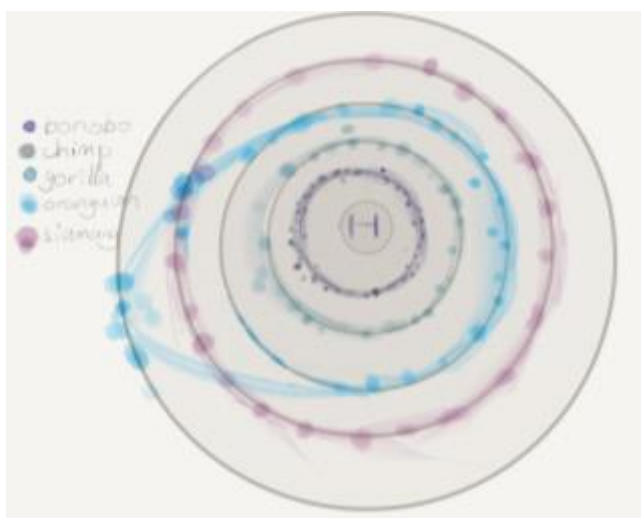
In general, the difference between trees is small. (This sketch is very rough, but proves it is a viable method.

Next, we can use processing to implement on all trees)

Pros and Cons:

- Color easy to catch;
- Need good scaled, otherwise the difference between saturation is too slight to be felt;
- Wrong color design may mislead.

Design 2 Solar system



User Assumption same as above. Use human as center, according to evolutionary distance, plot dots, use different colored closed circle to represent different species.

We can see the circle of orangutan and siamang have overlap, which represents they split up from same time, some orangutan dots even are outside of circle of siamang, which means who is more ancient on evolutionary history is controversial.

Design 3 Quantification

User Assumption: Assume those 101 phylogenetic trees come from different tree building algorithms or datasets, then our users are scientist who try to figure out the overall trend and summarize all research results.

Scheme : use statistic and demonstrate data in statistical way. Tool : Tableau

For example, frequency of evolutionary time and frequency of specific relationship.

Insights

As we can see, most of trees' relationships are similar, difference in two points:

- Differences of evolutionary time are most prevalent differences among trees, so if I need to generate a tree as reference, I would use statistical method to get a tree with mean value and majority relationship.
- The position of orangutan and siamang is most controversial, which is worth further research.

Video: <https://youtu.be/xBSyiStj4iQ>