

Multivariate Networks in *Ecology and Evolution* Design Study

Process and
Development

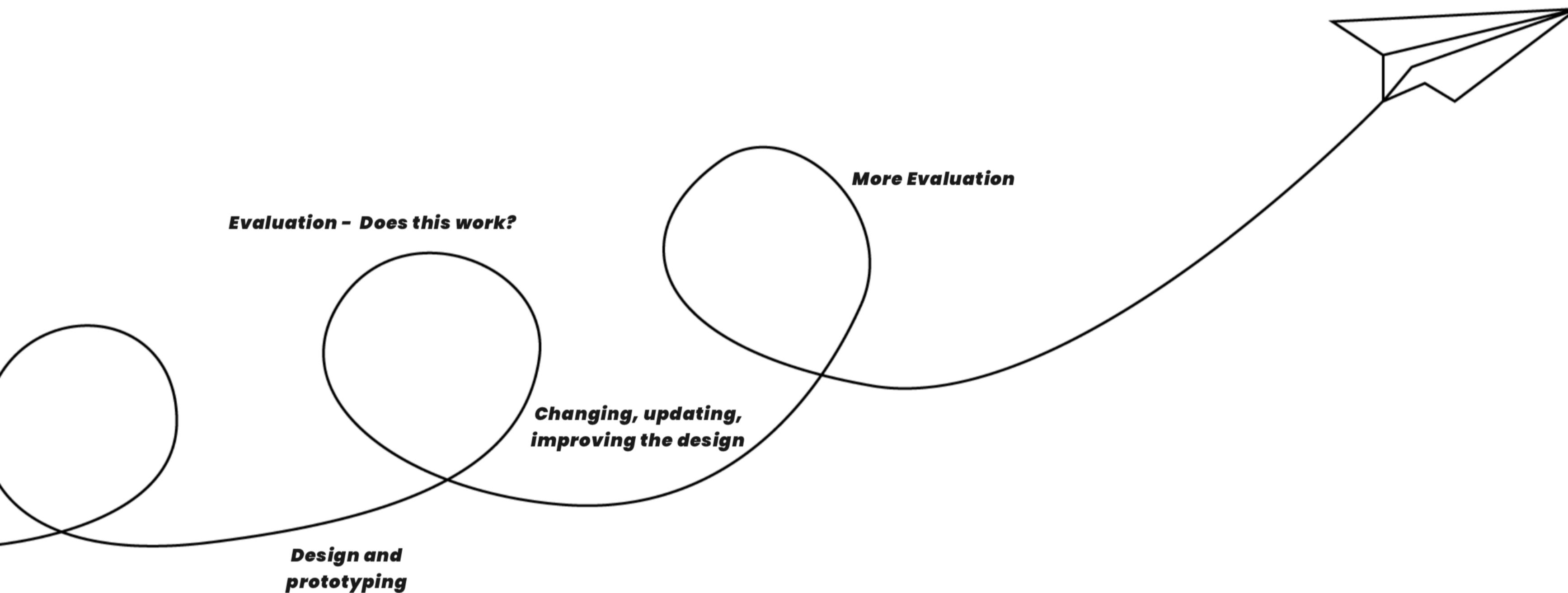
8. 30. 2019



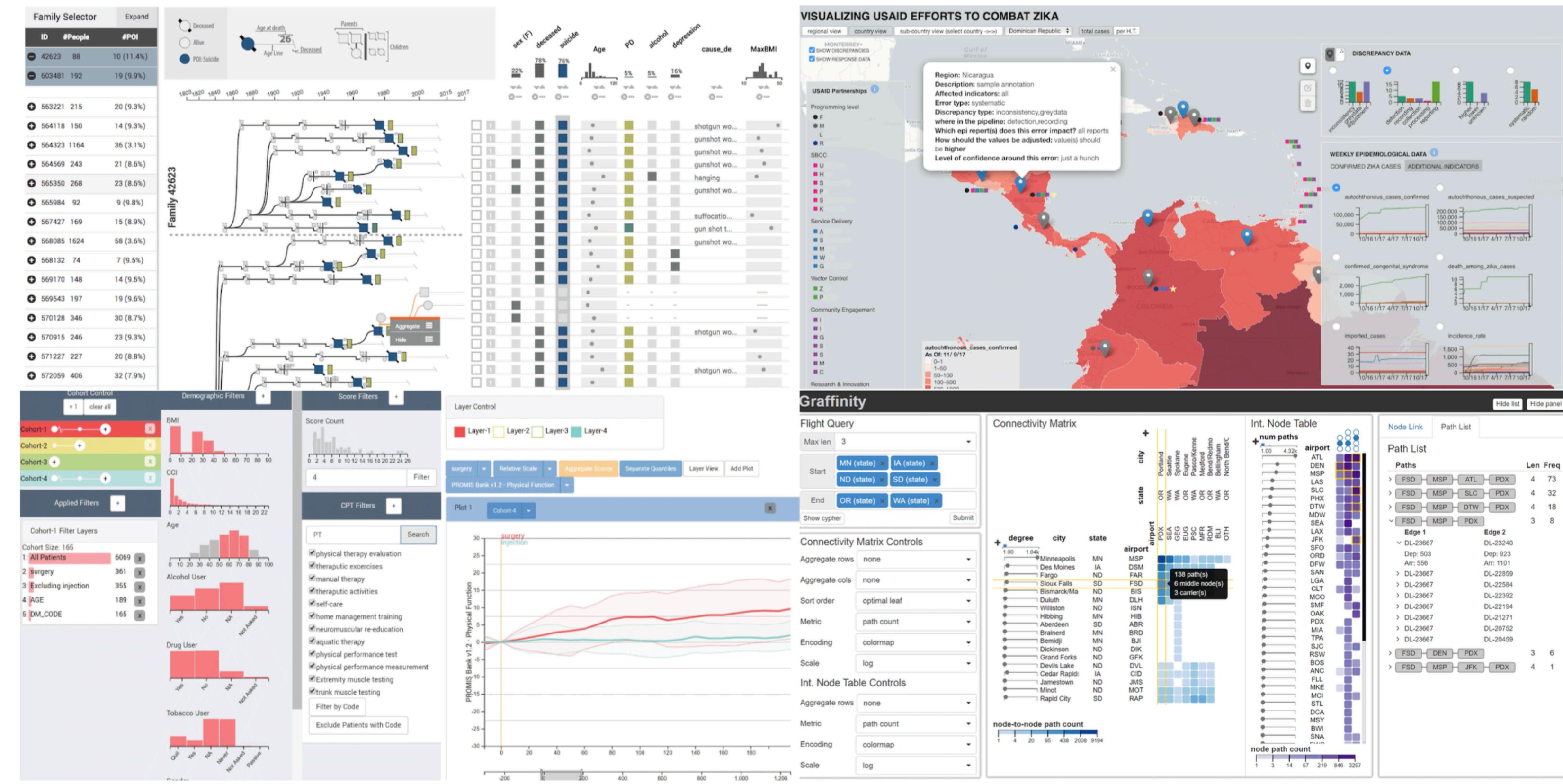
Photo by [Edi Libedinsky](#) on [Unsplash](#)



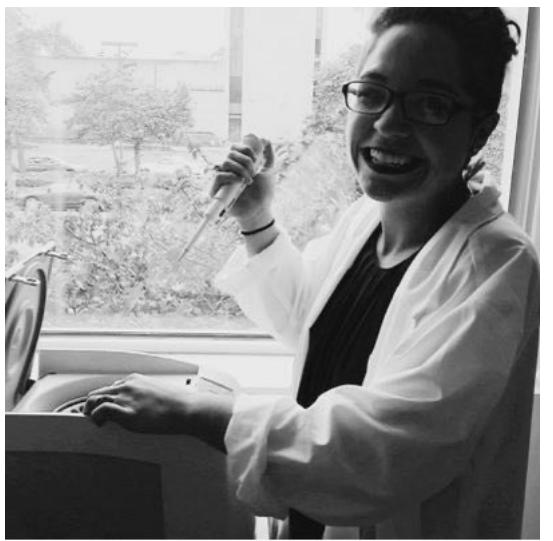
**Met with Luke the first day.
Got a desk.
Got an email.
Joined the herd.**



**Gave a presentation:
Explained what the VDL does
Showed examples of work
Set expectations for my timeline there**

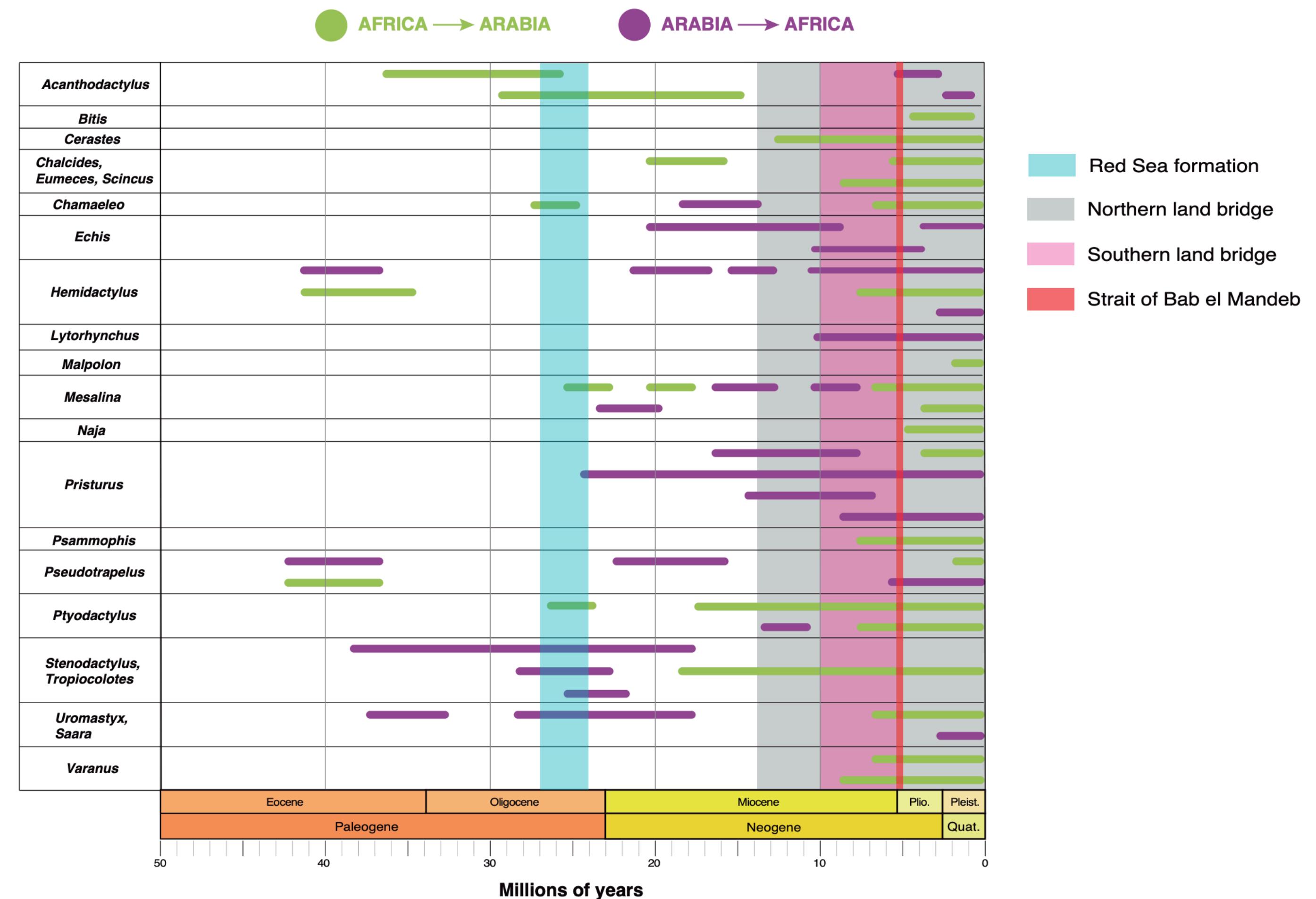


Conducted One-on-One Interviews

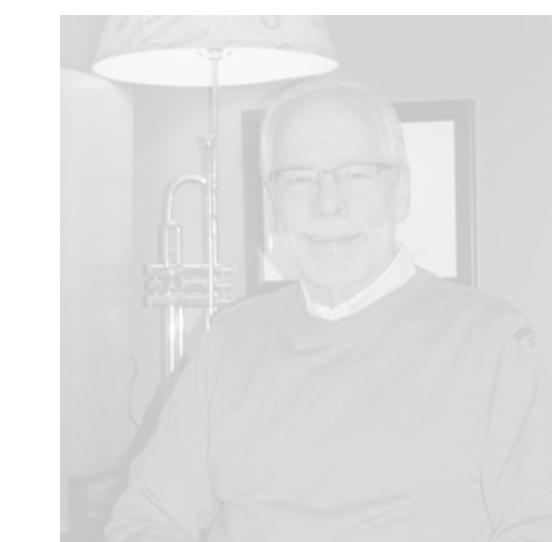
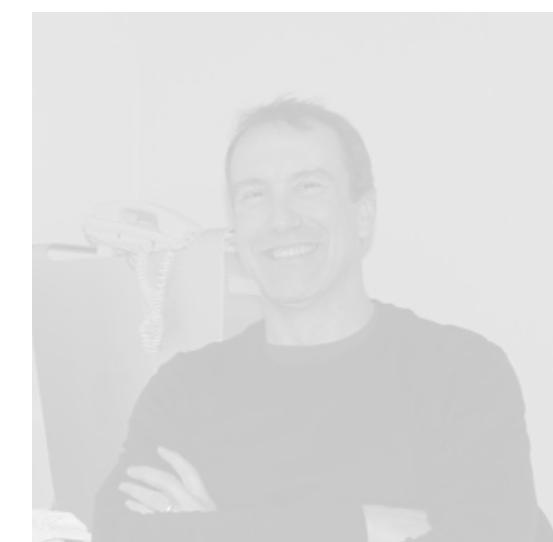
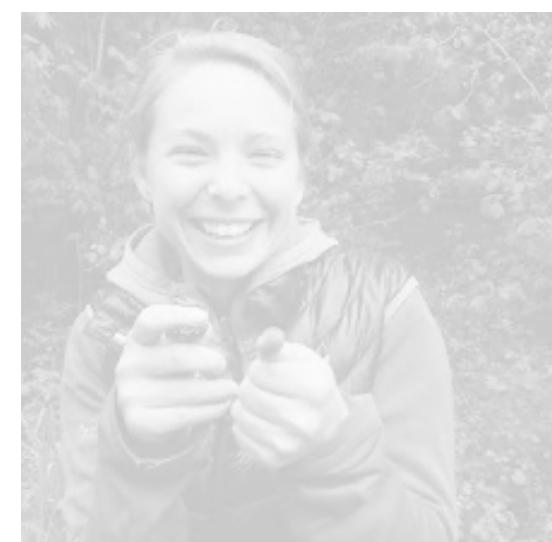
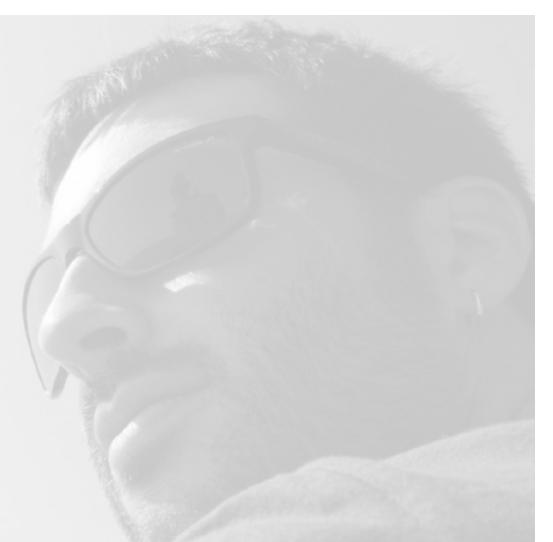
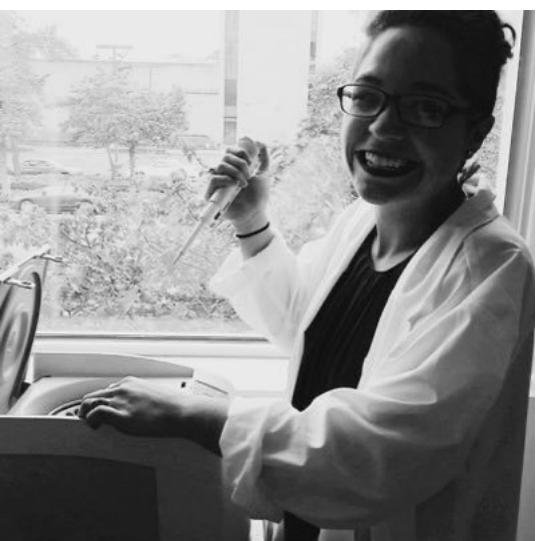


Gathered
examples from
their work

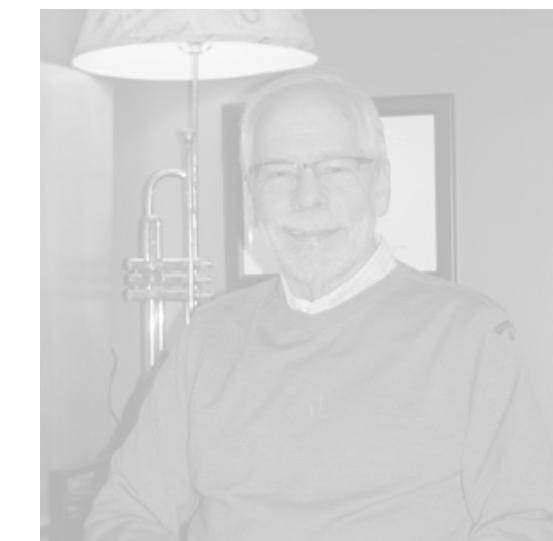
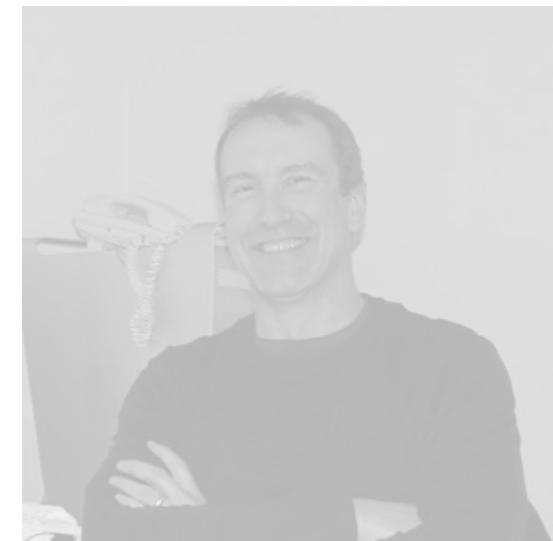
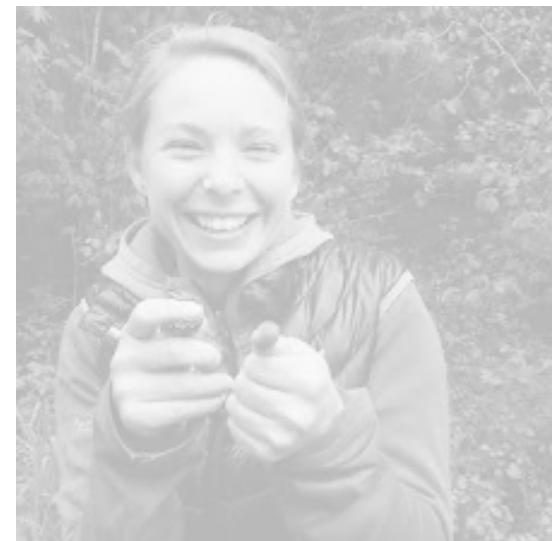
DISPERSAL AND VICARIANCE OF AFRO-ARABIAN SQUAMATE REPTILES



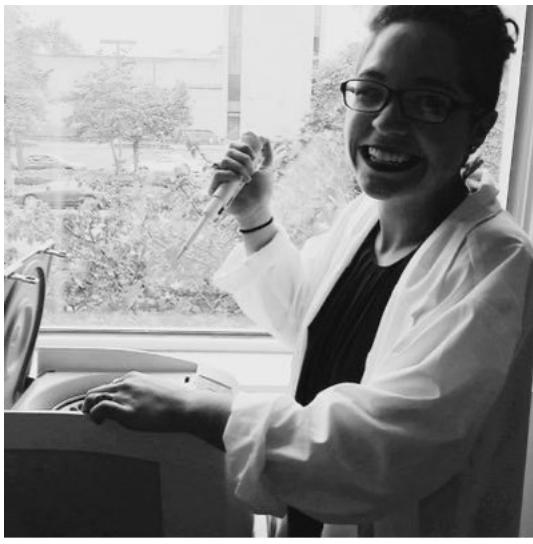
Main Contacts in Moscow



Secondary Interviews and Continuous Input on Design







Participants in the Workshop

Workshop Structure:

2:30 -2:40 *Workshop Opening*

2:40 - 3:00 *Walk-through of Analysis*

3:00 - 3:30 *What is your data?*

3:30 - 4:05 *Wishful Thinking Activity*

4:05 - 4:10 *5 Minute Break*

4:10 - 4:20 *Task Prioritization*

4:20 - 5:20 *Storyboarding Activity*

5:20 - 5:30 *Closing*

Workshop Structure:

2:30 -2:40 Workshop Opening

2:40 - 3:00 Walk-through of Analysis

3:00 - 3:30 What is your data?

3:30 - 4:05 Wishful Thinking Activity

4:05 - 4:10 5 Minute Break

4:10 - 4:20 Task Prioritization

4:20 - 5:20 Storyboarding Activity

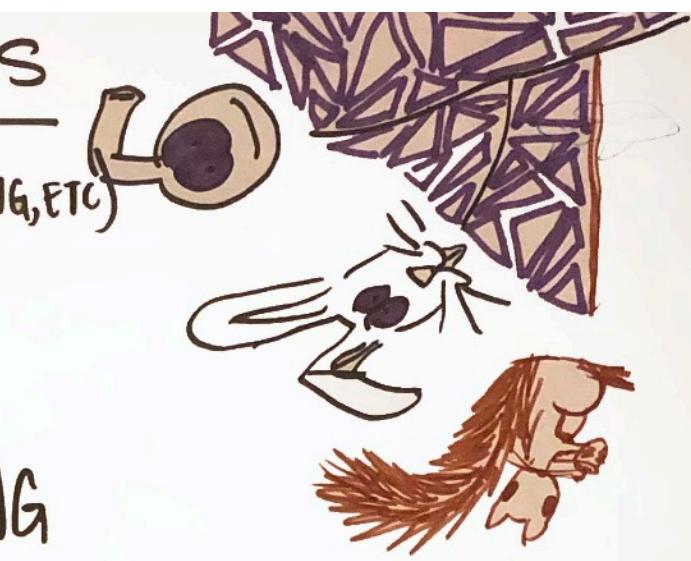
5:20 - 5:30 Closing

“I think we did it wrong,”

*“It’s funny that this is revealing
some deep insecurities in you guys
- I don’t know why but it’s good,”*

*IMPORTANT TASKS

- DATA CURATION (WRANGLING, ETC)
- READ PAPERS
- MAKE PROTOCOLS
- WRITING CODE/DEBUGGING
- INTERPRET ANALYSES
- WRITE PAPERS
- DISCUSSING RESULTS w/ COLLABORATORS



Questions?

Is it working? (software)

Am I right? (hypothesis testing)

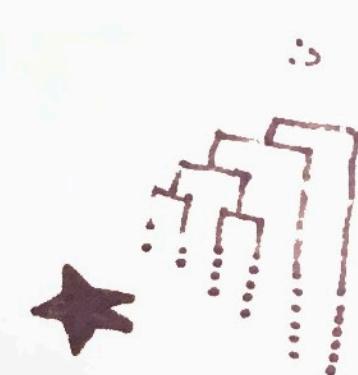
How are x + y related?

Does mite domatia increase diversification rates in Trichomyces?

What influences community structure?

VISUALIZATION

- PLANT STRUCTURES
(morphological)
- PHYLOGENIES
- CONCEPTUAL FIGURE
FOR SIMULATIONS,
REVIEW PAPERS, ETC.



Individual Walkthrough

To get the gears turning, we are going to spend a little time thinking about your workflow, important tasks involved, and the data you depend on to answer questions in your research

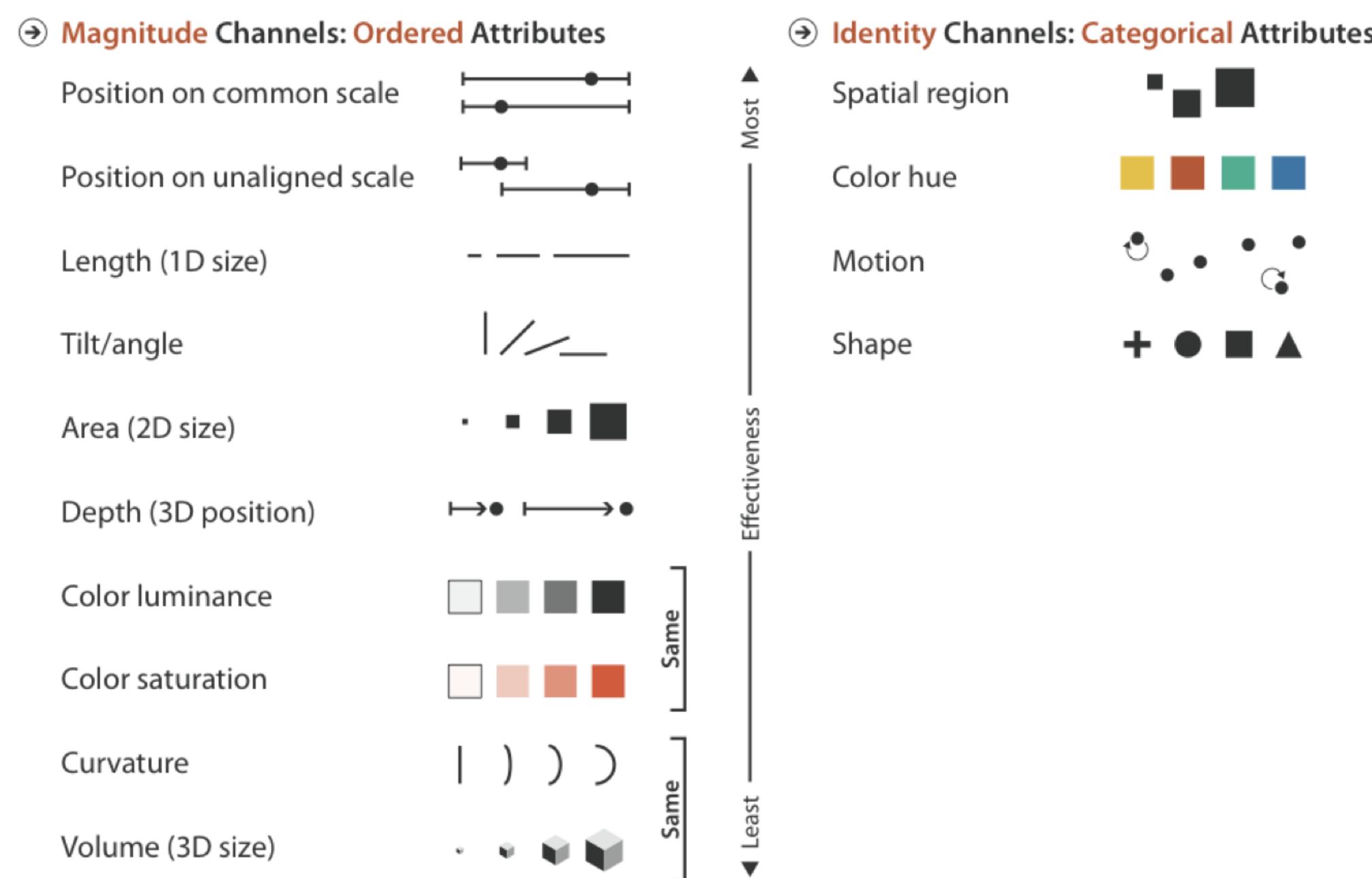
Main Workflow Points:

Generation of tree topology

Mapping of data to topology

Comparison of trees and trait distributions

What is Your Data?

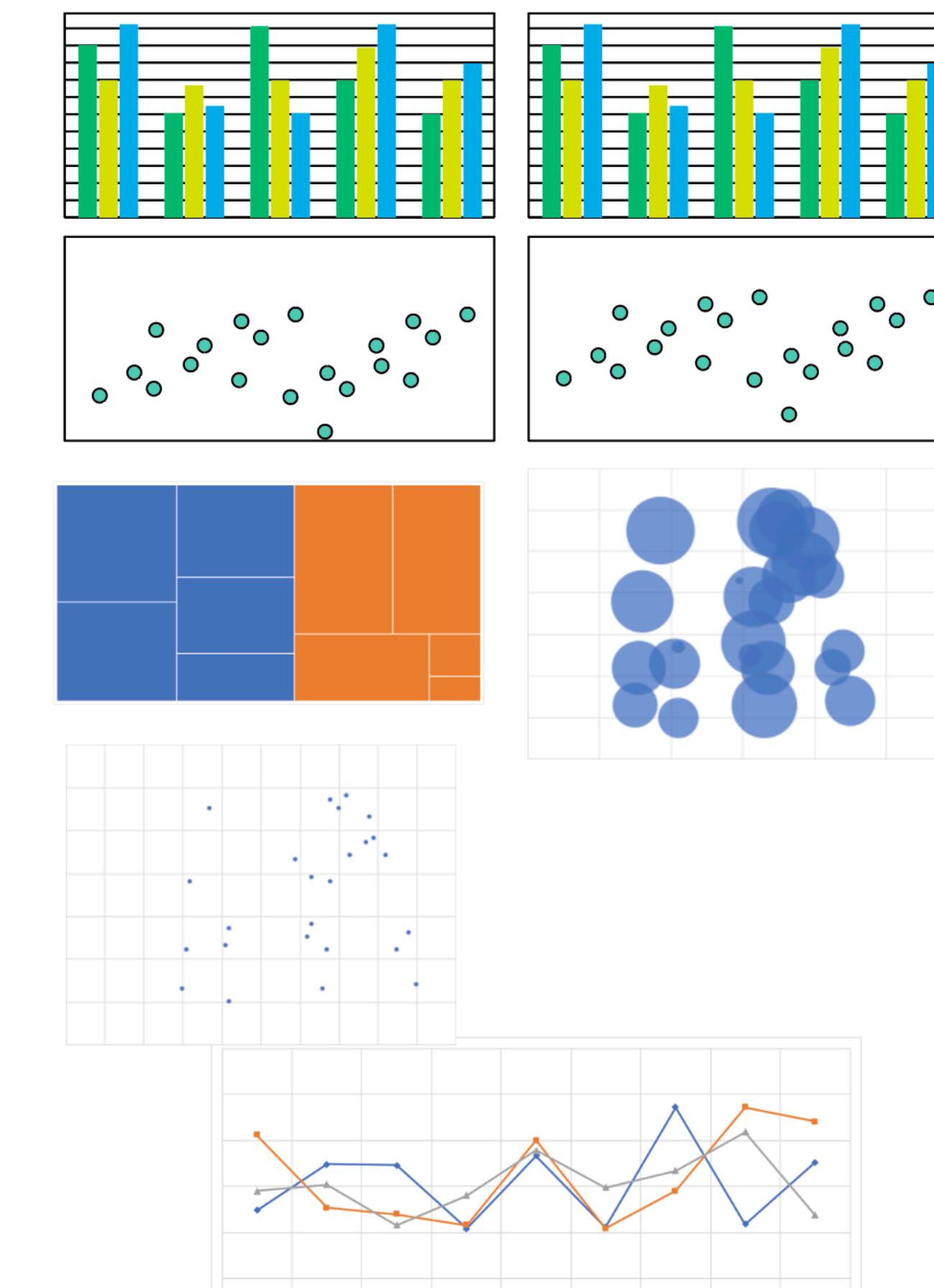


Prompt: We are going to start by diagramming your data. The main purpose of this is to show how you currently represent your data and how you are using these encodings to answer research questions. For this, you can directly sketch the representations or data encodings, you can use a combination of sketching and pieces from the “vis toolkit”, or you can map the data types and relationships with post-it notes.

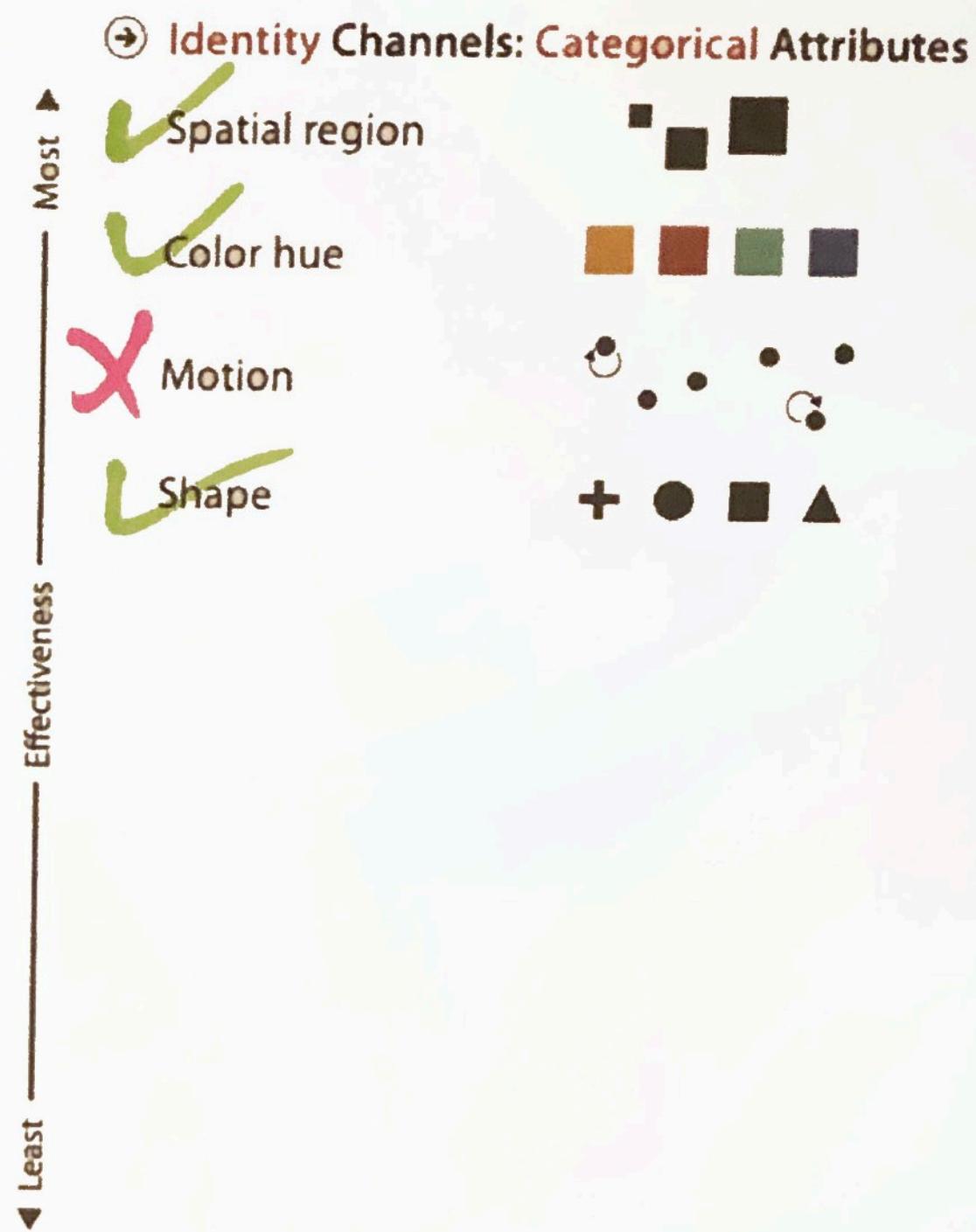
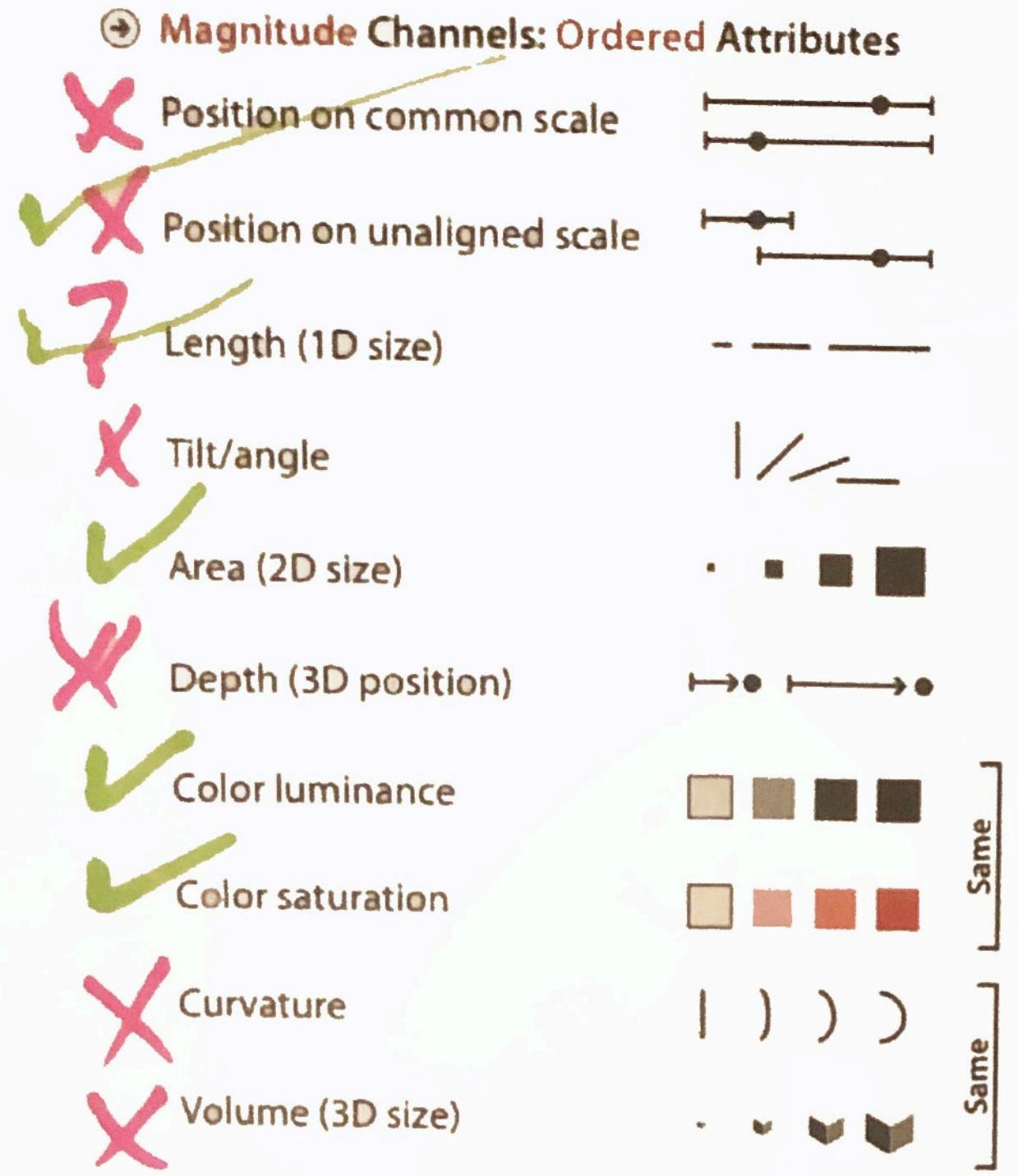
Please use the questions below to inform your work:

- What do your data consist of?
- What encodings are you using to represent your data?
- How are different types of data related?
- How do you combine them to answer questions?

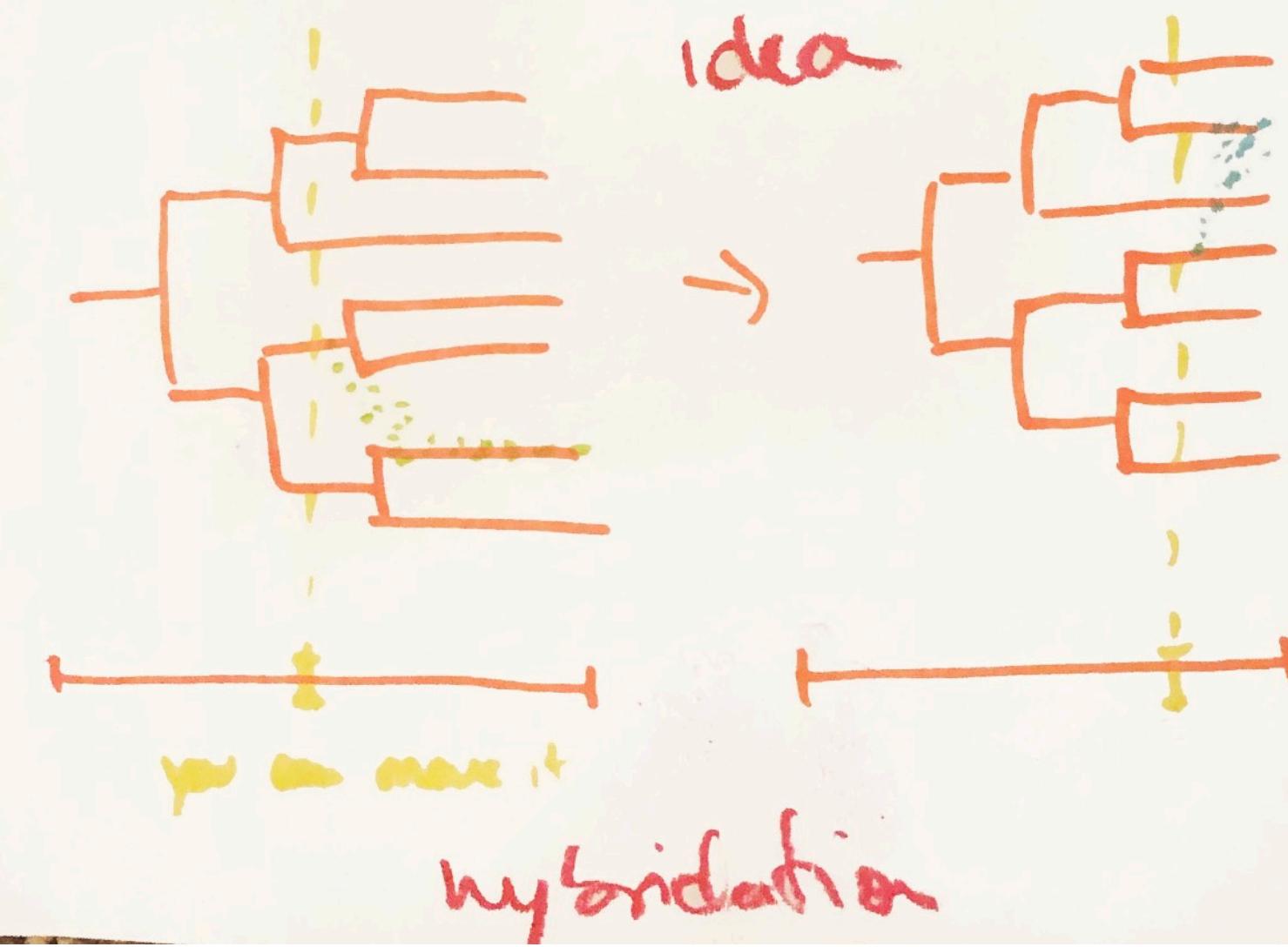
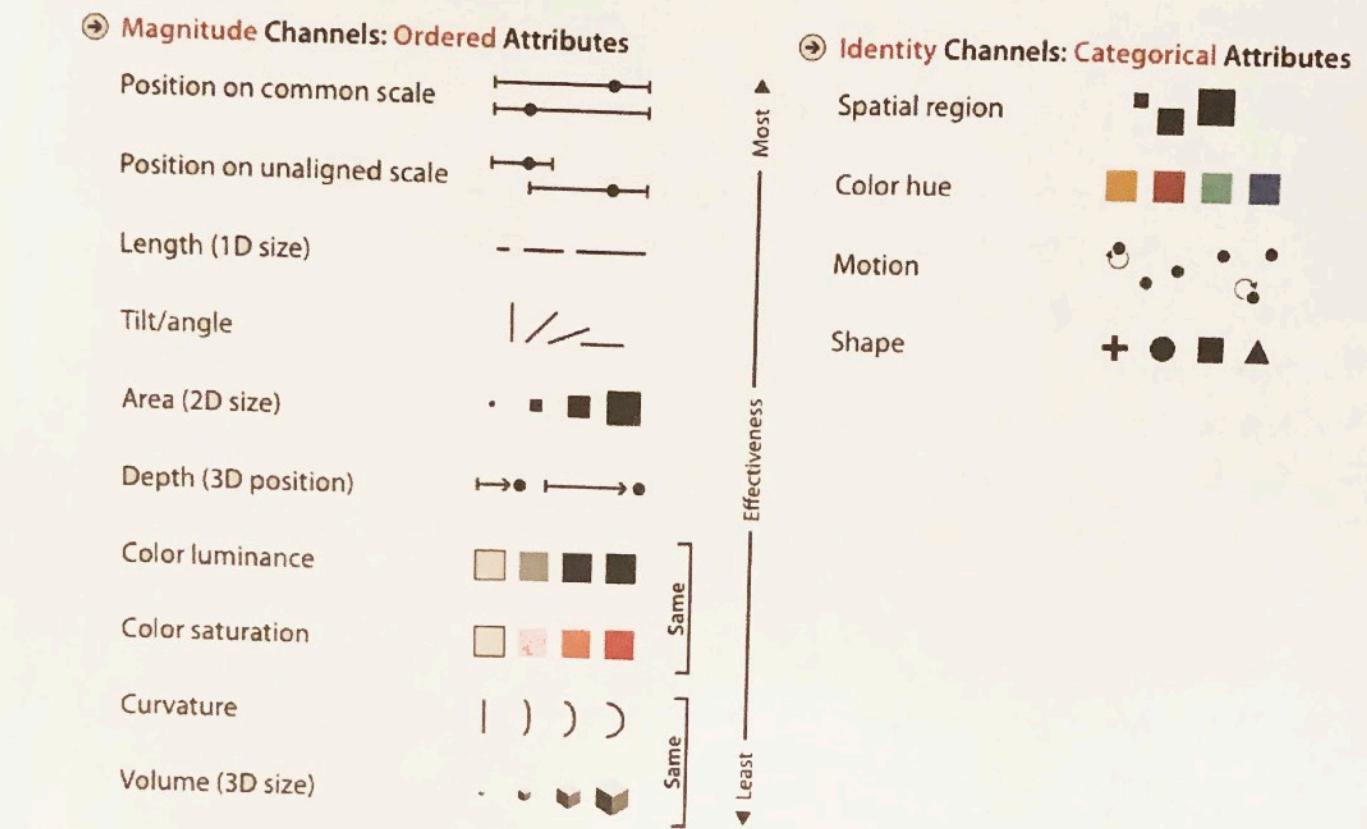
Vis toolkit:

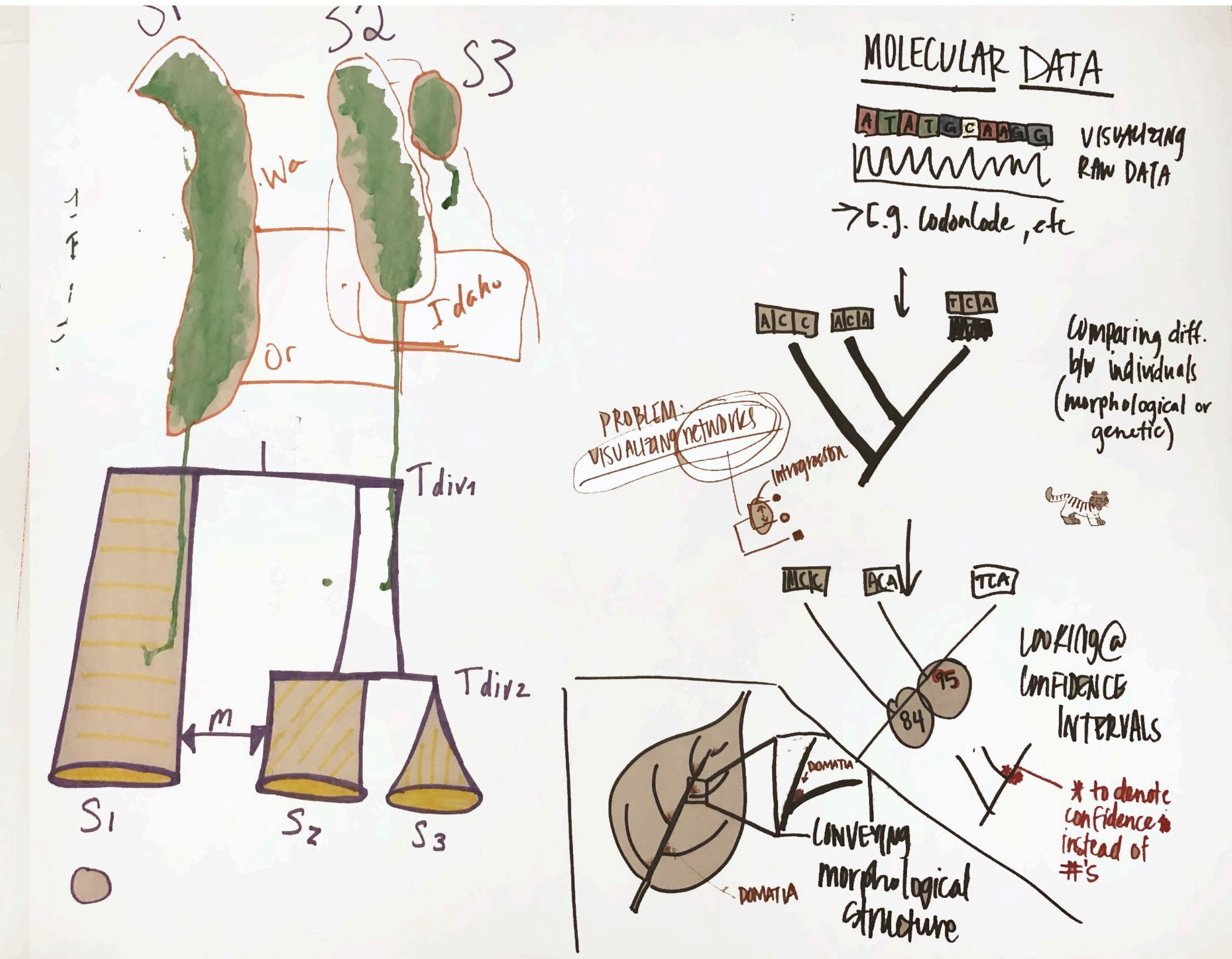
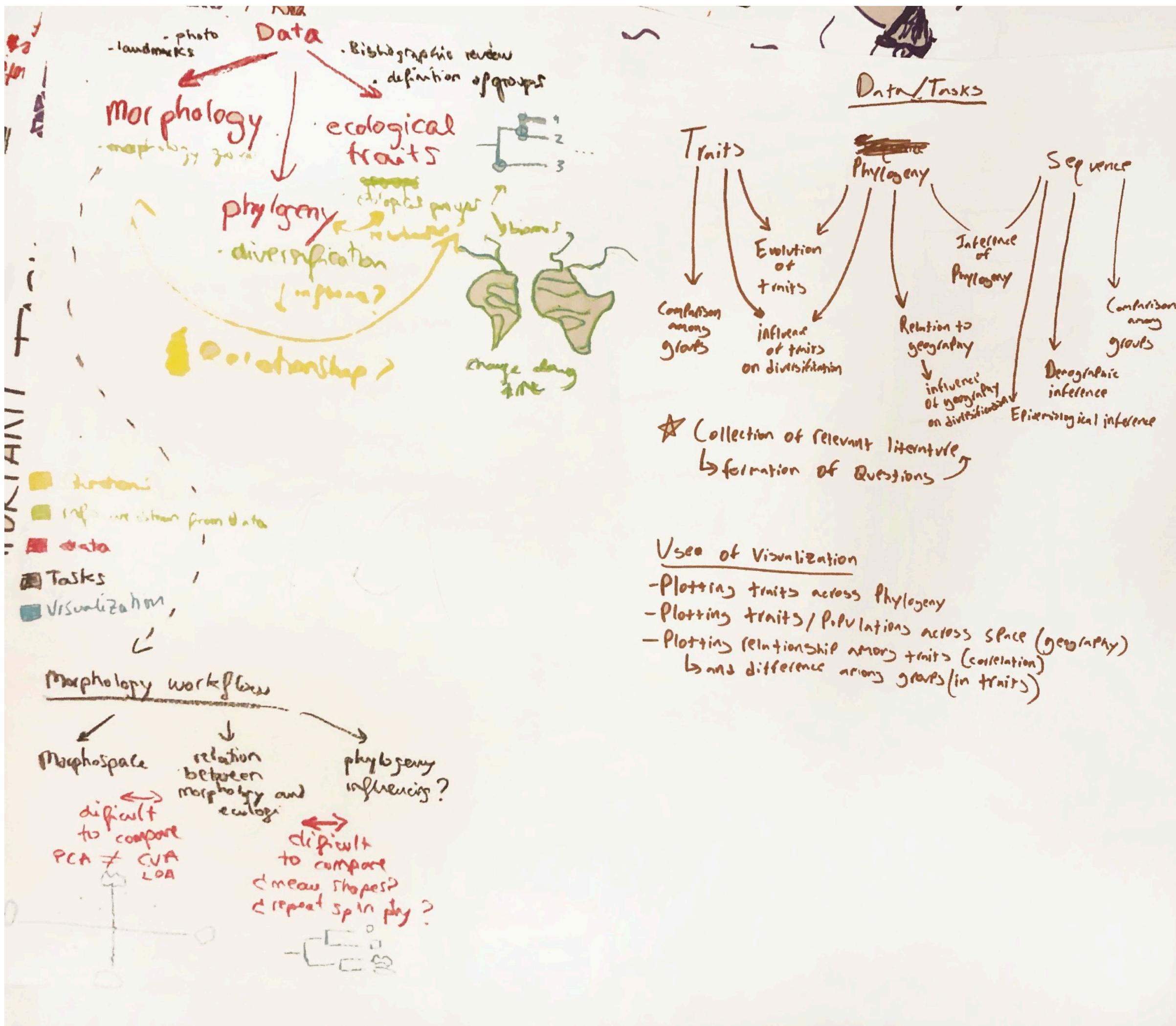


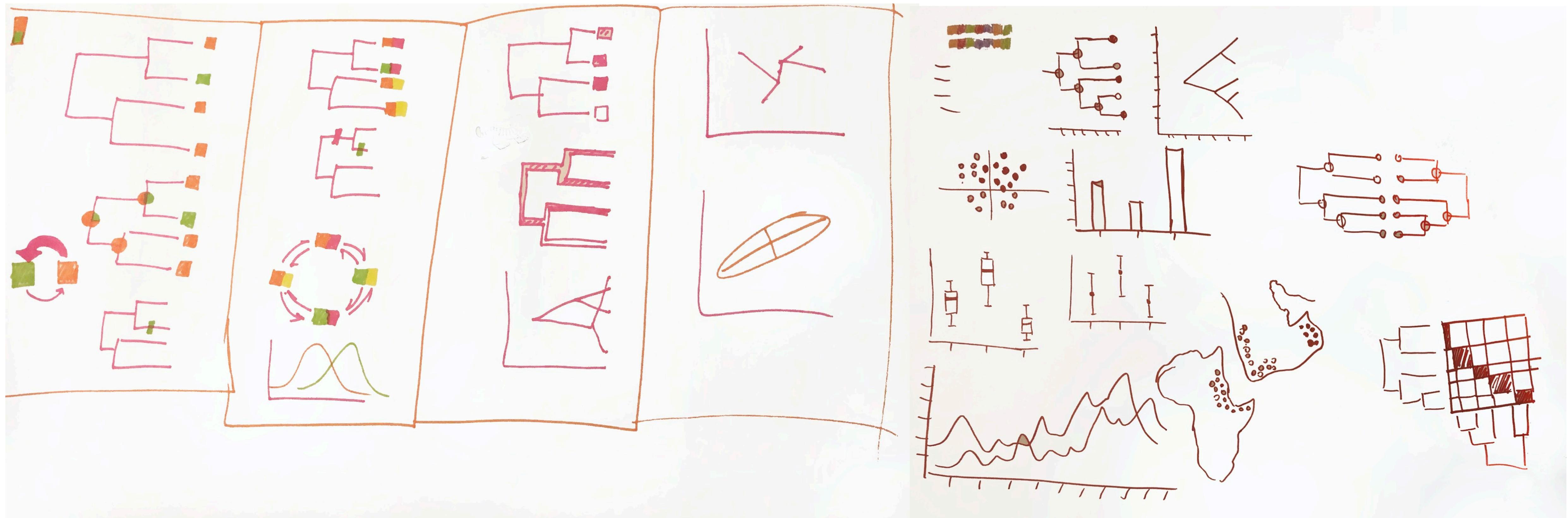
Evolutionary Biology CVO Workshop, May 22, 2019



Evolutionary Biology CVO Workshop, May 22, 2019

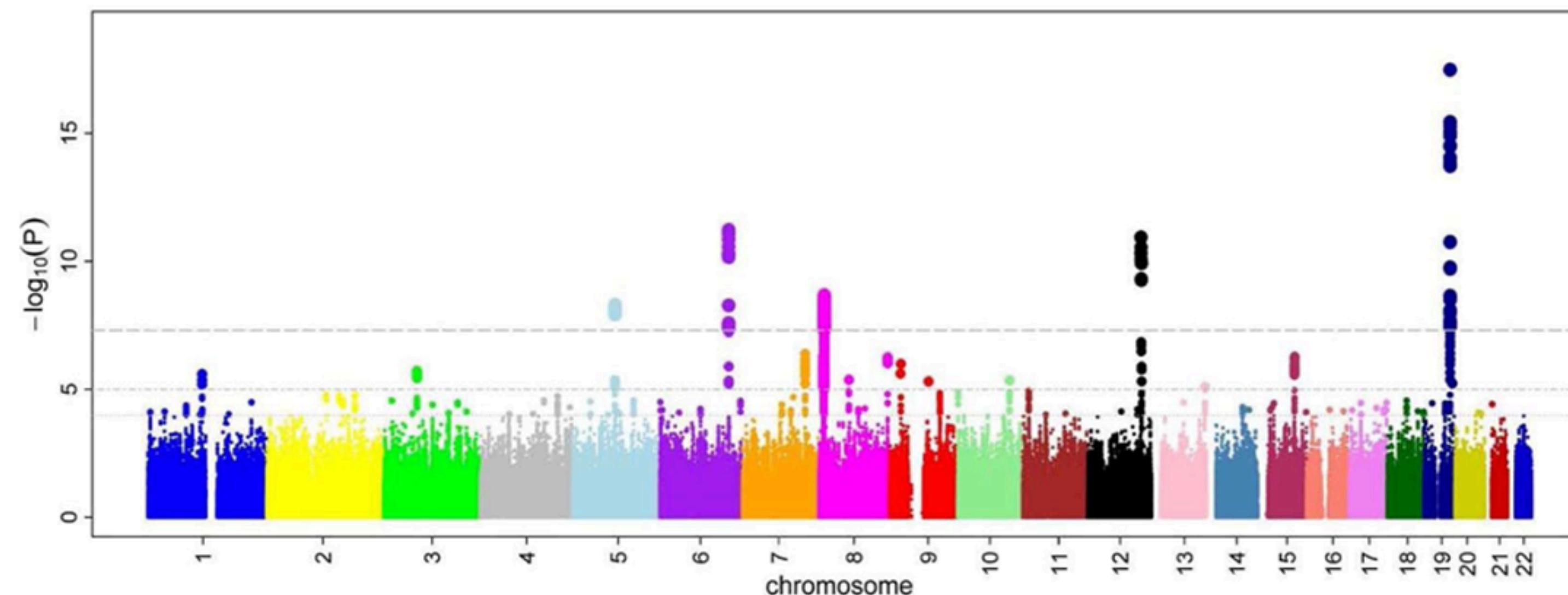






Did “What is your data?” encourage people to think about their own data representations in a critical way?

“I don't understand why people make those plots when they always look the same. You can draw it with a marker with no data it looks perfectly the same”.



“It’s what you do. It’s not necessarily what’s right- not necessarily what’s useful”





Wishful-Thinking Exercise

- What would you like to be able to do?
- What would like to know?
- What would you like to see?

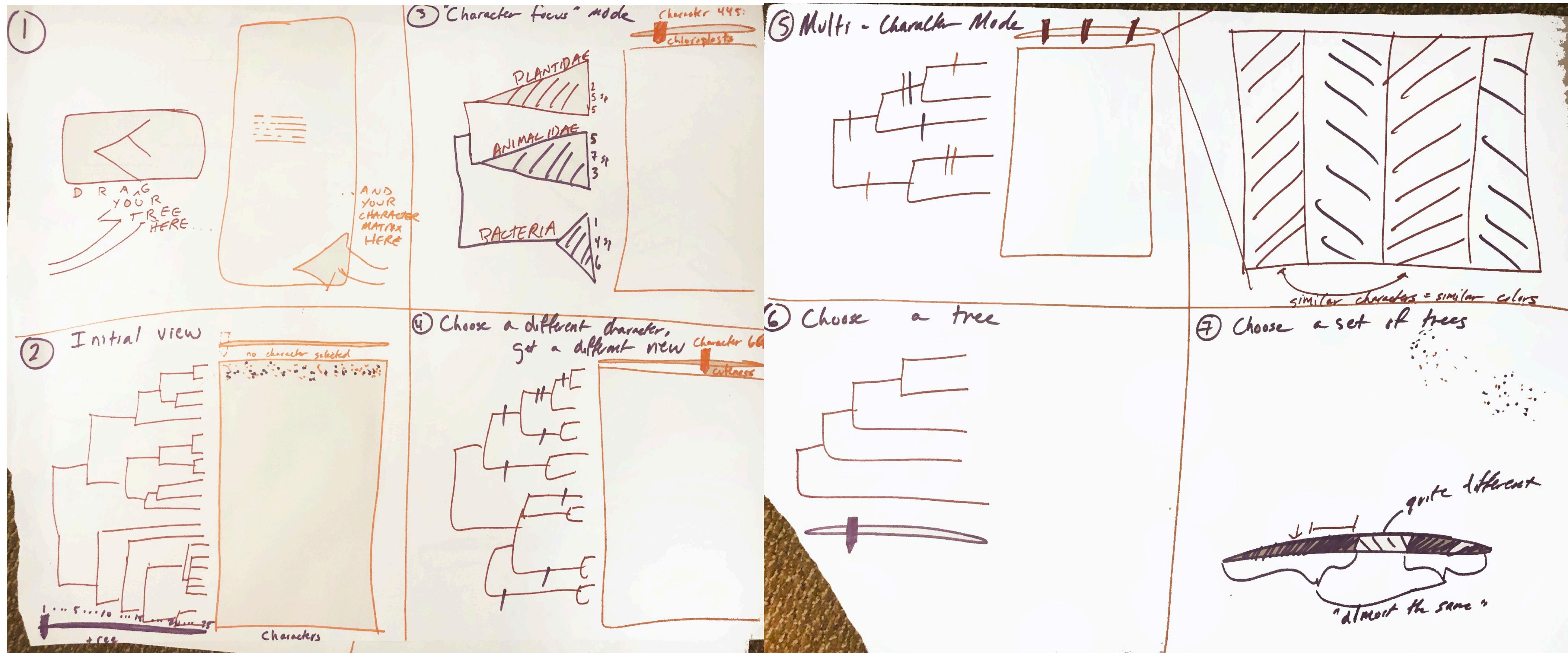
***The wishful thinking conversation
gravitated toward limitations in current
visualizations..***

How could a visualization of your data look like?

Sketch a visualization or a feature that addresses the most important ideas we generated

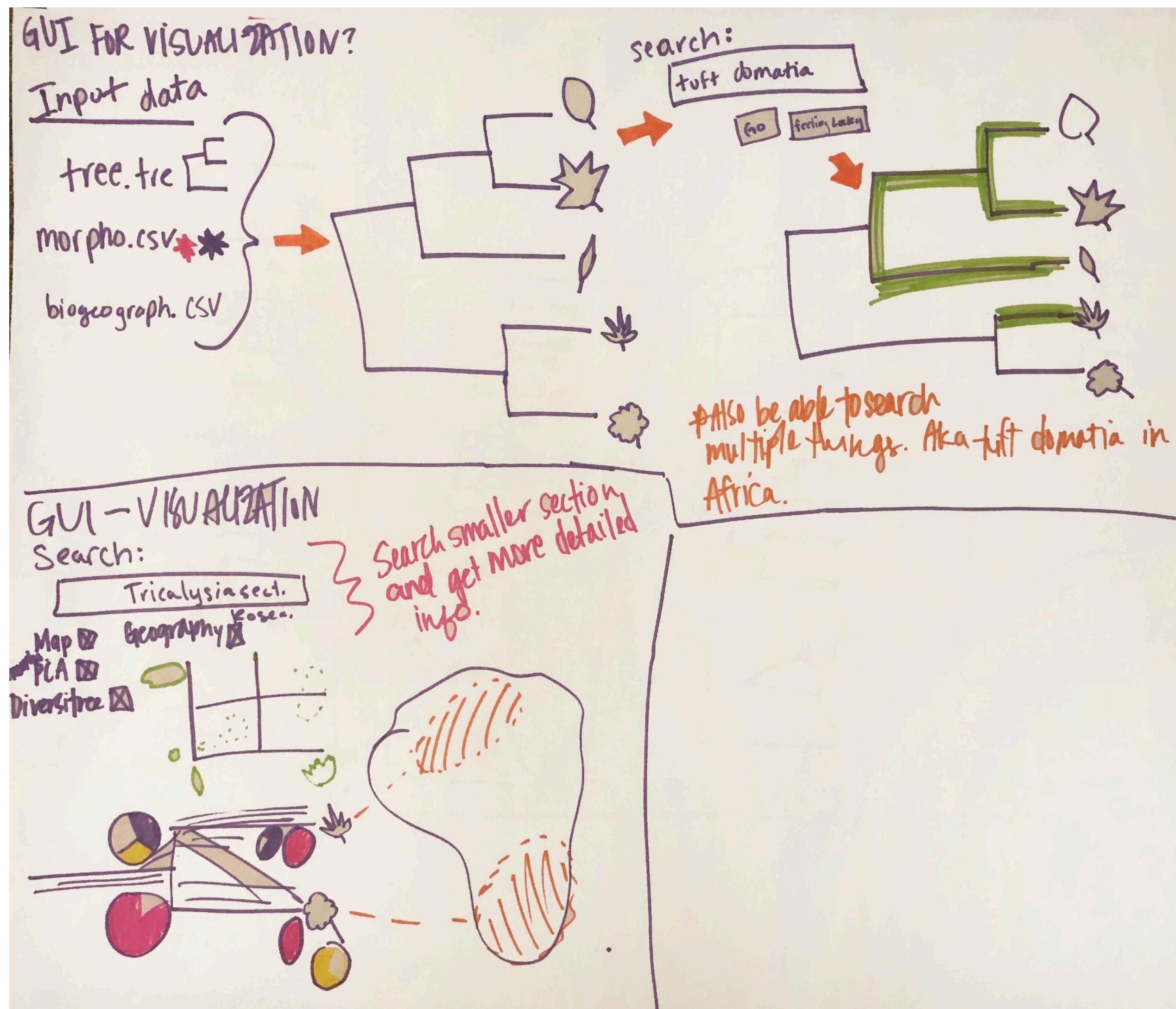
“Complex ideas begin as simple drawings,”
-Giorgia Lupi

Focused on interactivity.



Thoughts on the process:

The value of the workshop came from the conversations sparked by the exercises, not the artifacts themselves.



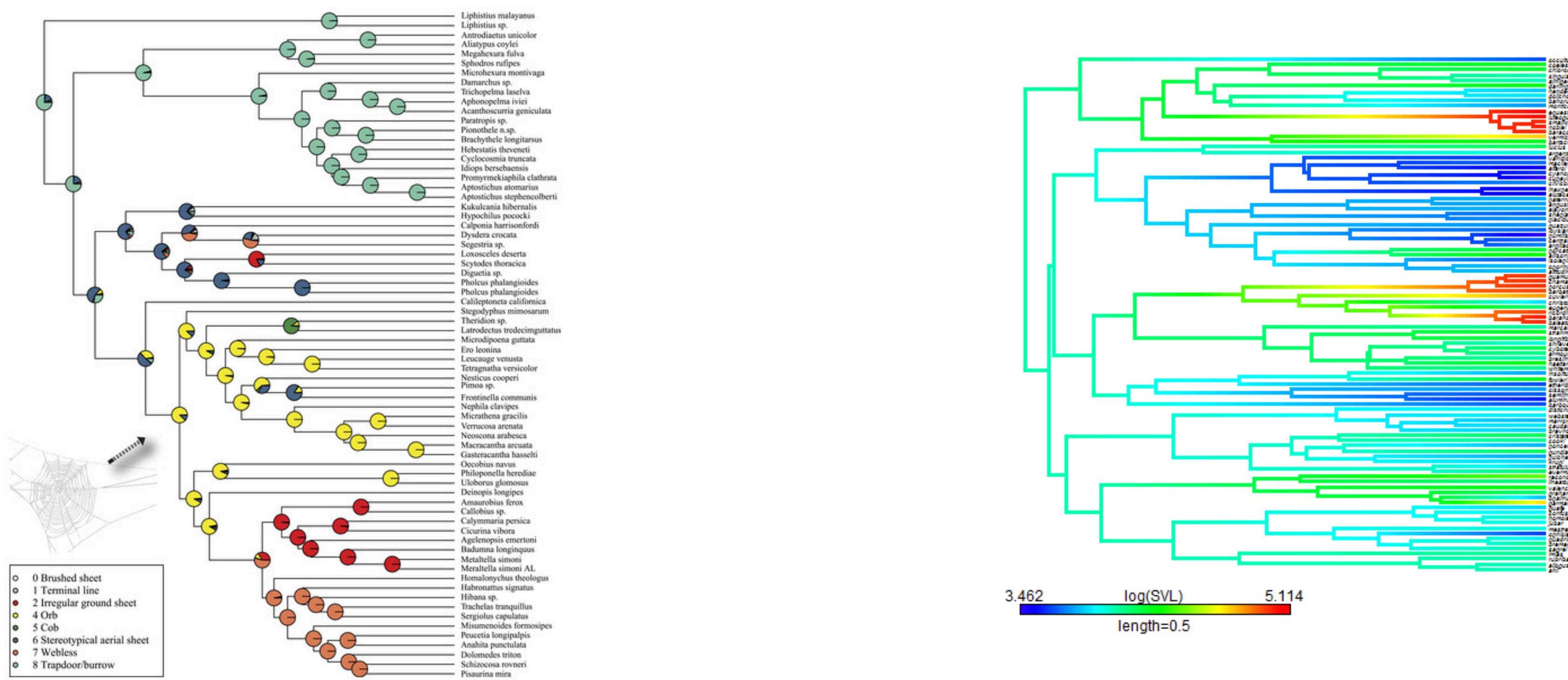
Scaling is a big issue.

Designed features that they could get more detail by drilling down.

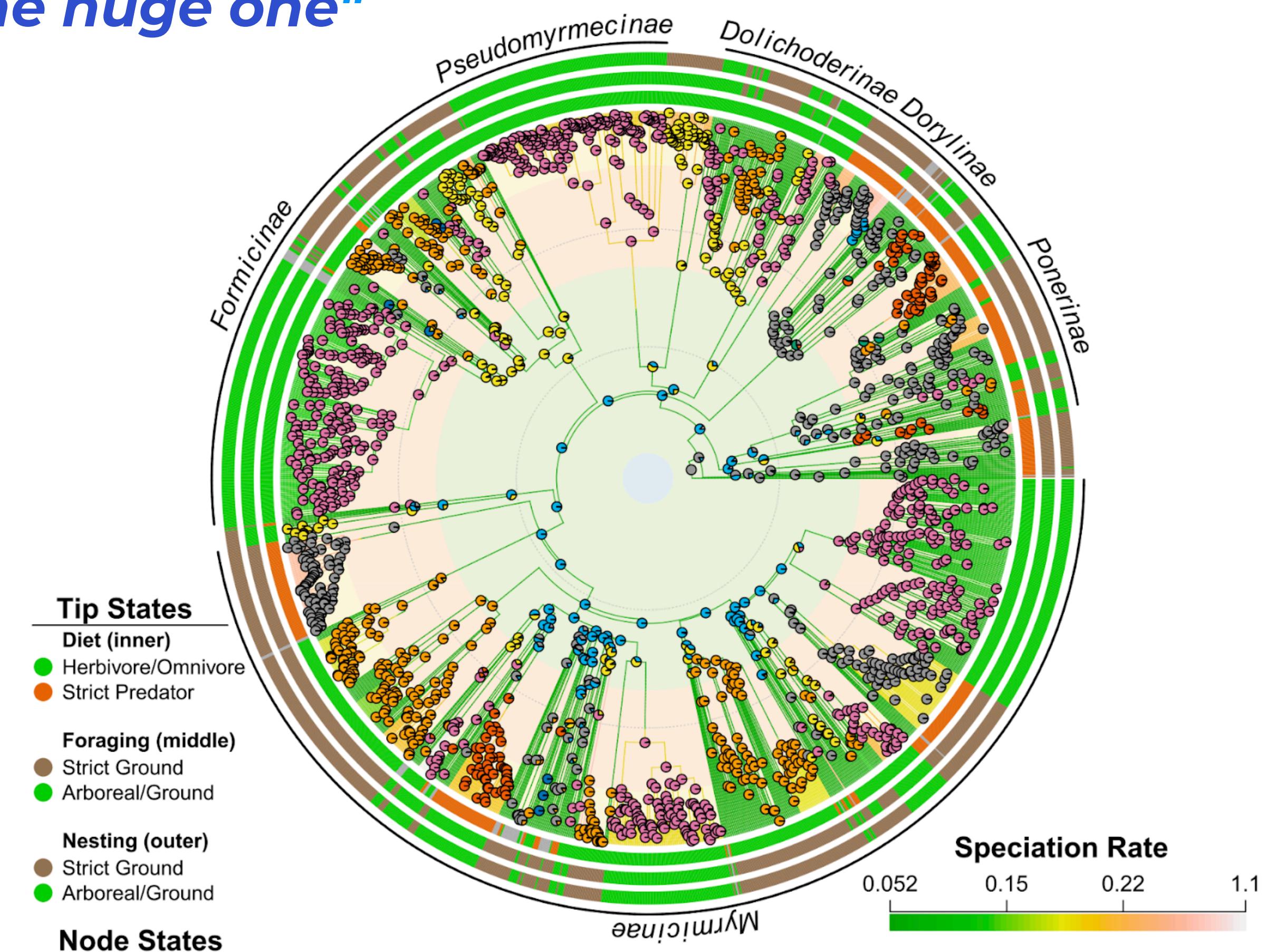
Main Points that came up in the workshop:

- **Visualizing multiple traits at once / scaling for large trees**
- Visualizing uncertainty in tree topology and estimated attributes
- Semantic zooming
- Assessment of the status quo for charts and visualizations to communicate their research

**“If you have one continuous trait you can do things.
If you two - ok. If you have three or four or five, there
is nothing really sufficient”**



“You need to invent the whole figure every time you do it because sometimes the tree has 10 species , sometimes it has 500. The space that I have is insane and I want to show as much information that I show for this (one with 10) as I show for the huge one”



Main Points that came up in the workshop:

- Visualizing multiple traits at once / scaling for large trees
- **Visualizing uncertainty in tree topology and estimated attributes**
- Semantic zooming
- Assessment of the status quo for charts and visualizations to communicate their research

Uncertainty in the topology:



***“I will use one phylogeny,
but I feel like I am lying because I used
100. How do I show that I am using 100?
How do those phylogenies differ from
one another?”***

Uncertainty in the traits:

“You loose all resolution in terms of anything [you are trying to communicate].

Some of that has to do with uncertainty, some of that has to do with the inadequacy of the approach to communicate that uncertainty”

Main Points that came up in the workshop:

- Visualizing multiple traits at once / scaling for large trees
- Visualizing uncertainty in tree topology and estimated attributes
- **Semantic zooming**
- Assessment of the status quo for charts and visualizations to communicate their research

Main Points that came up in the workshop:

- *Visualizing multiple traits at once / scaling for large trees*
- *Visualizing uncertainty in tree topology and estimated attributes*
- **Semantic zooming**
- *Assessment of the status quo for charts and visualizations to communicate their research*

“Why can’t we have interactive figures for our papers? It’s 2019”.

*“You could visualize large patterns
by collapsing what is not important”*

“Visualization [would depend] on the context”

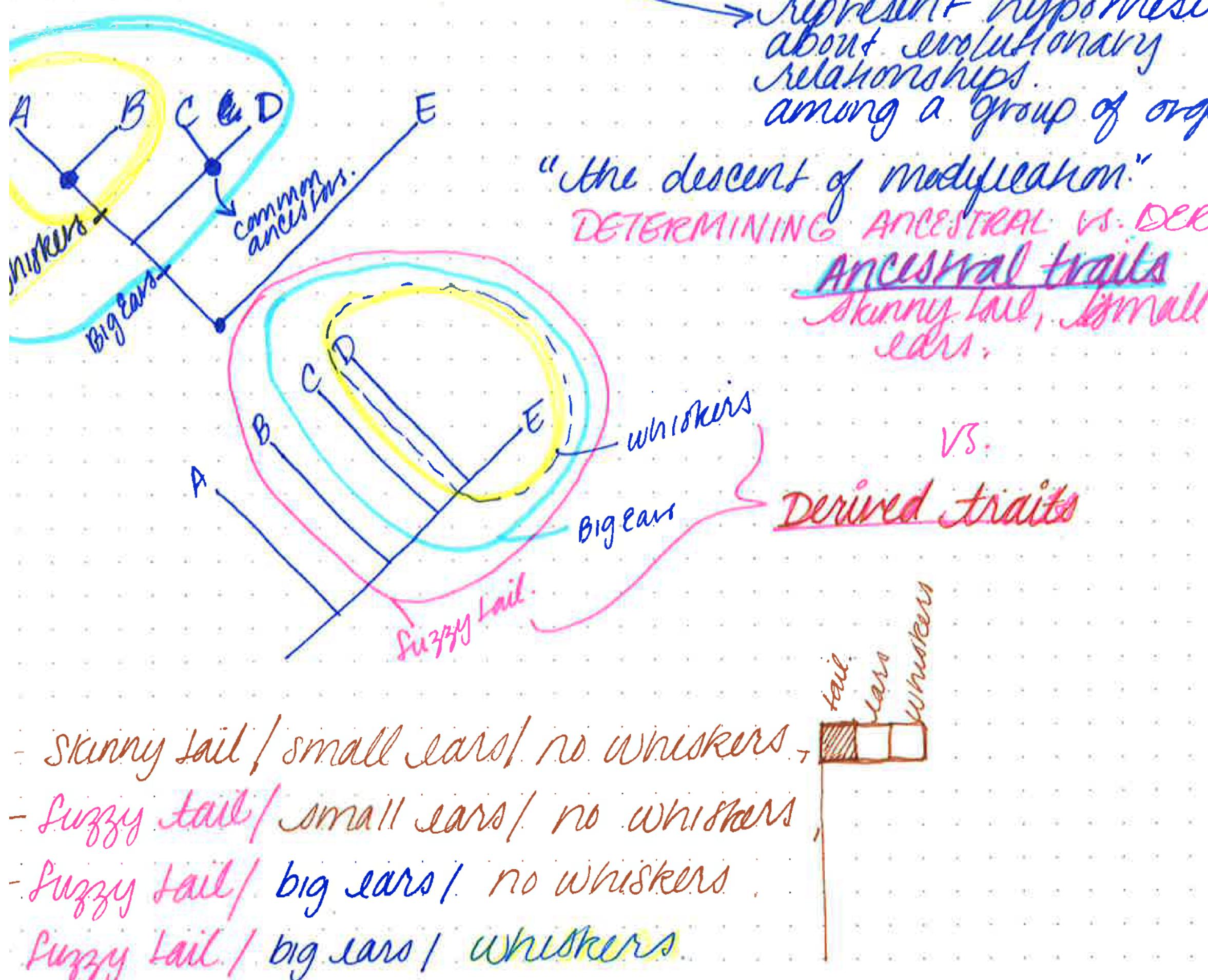
Main Points that came up in the workshop:

- Visualizing multiple traits at once / scaling for large trees
- Visualizing uncertainty in tree topology and estimated attributes
- Semantic zooming
- **Assessment of the status quo for charts and visualizations to communicate their research**

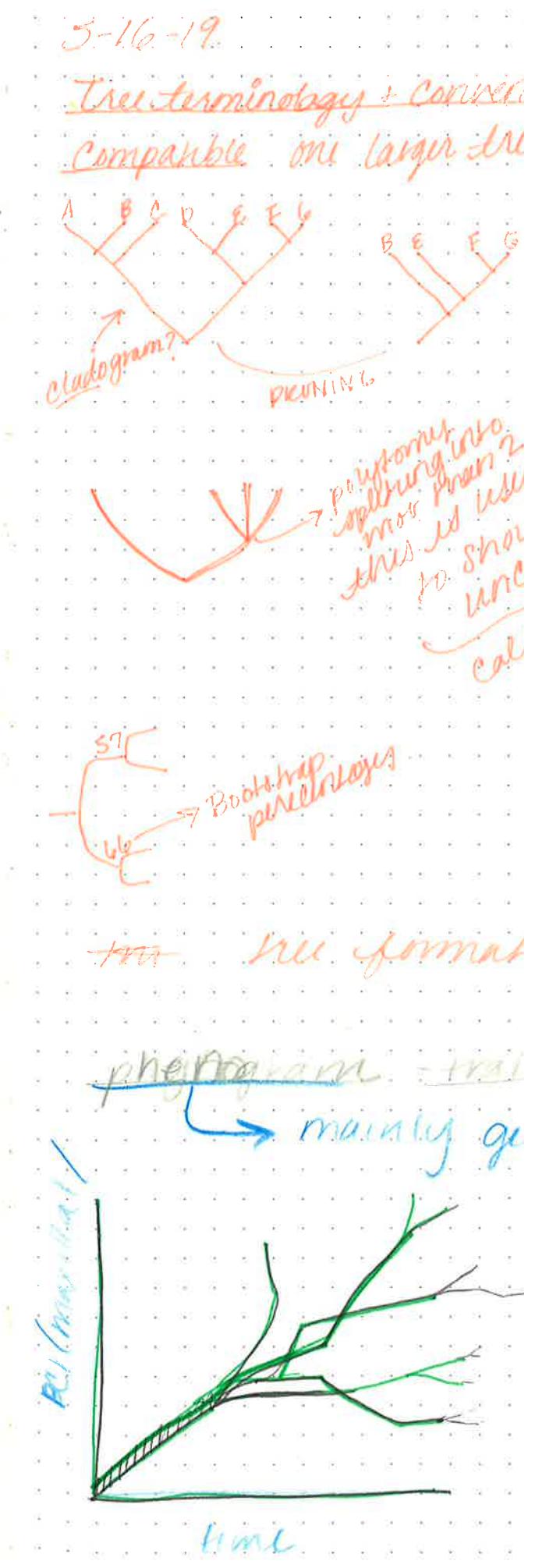
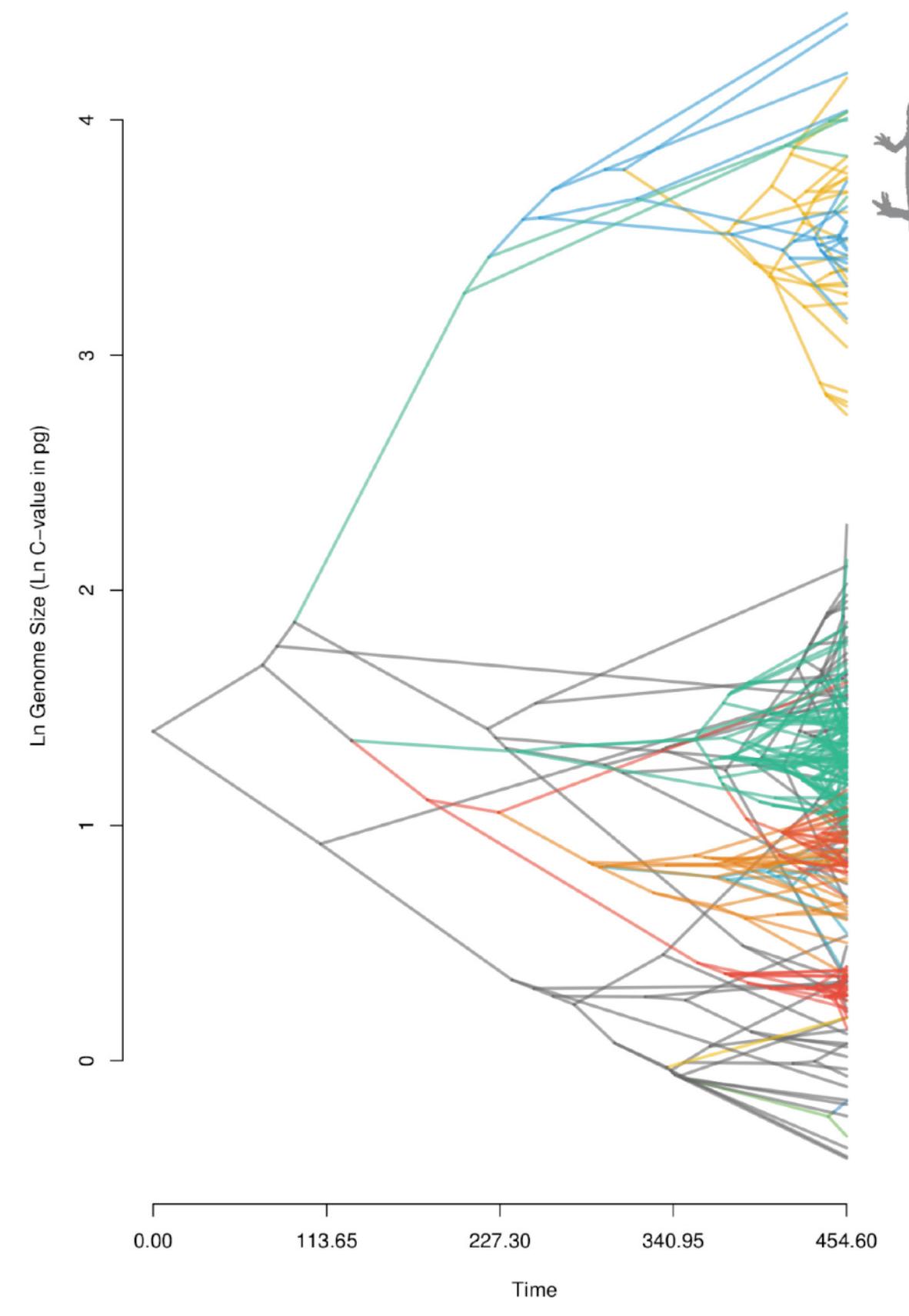
**“Figures first-
I think of it as a learning tool,
a way to organize your own
thoughts,”**

**“Yeah - and you can see what is
interesting in your data,”**

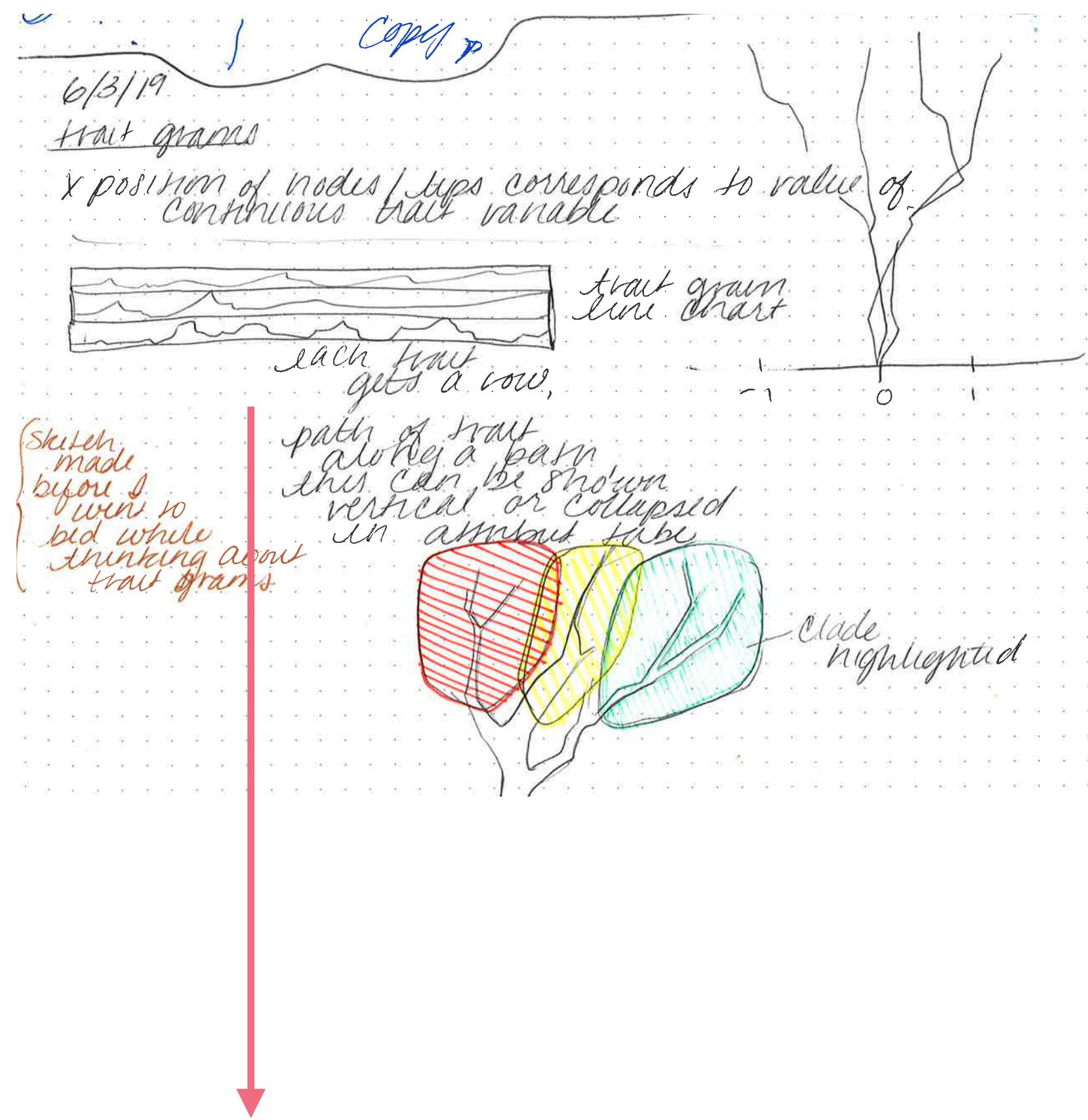
1-30-19. Building a Phylogenetic tree.



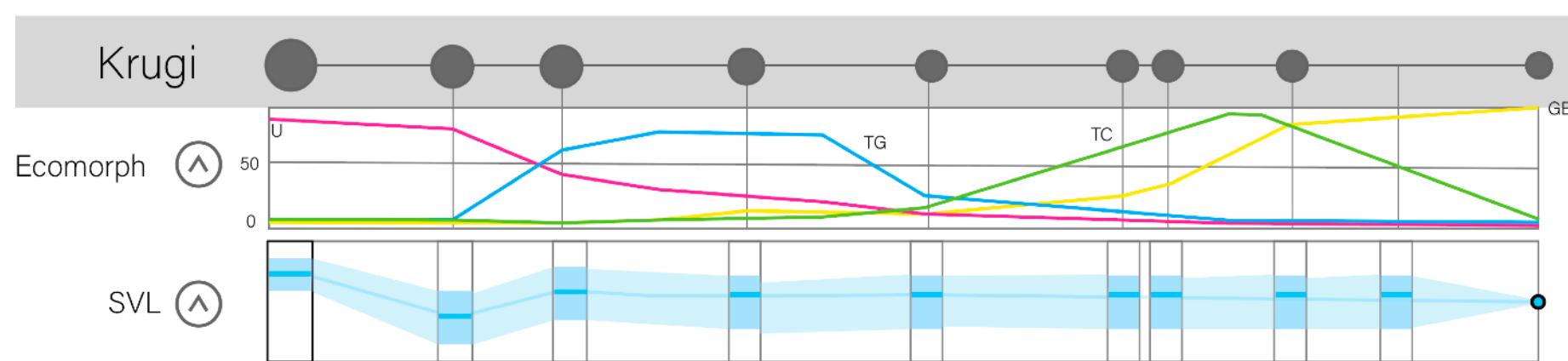
**Thoughts on the process:
Sketching from domain examples helped me understand their abstractions.**



Traitgrams express the change in value across time in a continuous path in the context of the topology.



realized that we could show the change in an attribute across the path from root to leaf in our linear layout.



Thoughts on the process:
Reflective memos help summarize what was important at that time, making it easier to build on. These thoughts would float to the top for the next meeting.

Thoughts on the process:

Second and third interviews helped build on existing ideas. This was where the concepts of convergence and state filters came up.

Thoughts on the process:

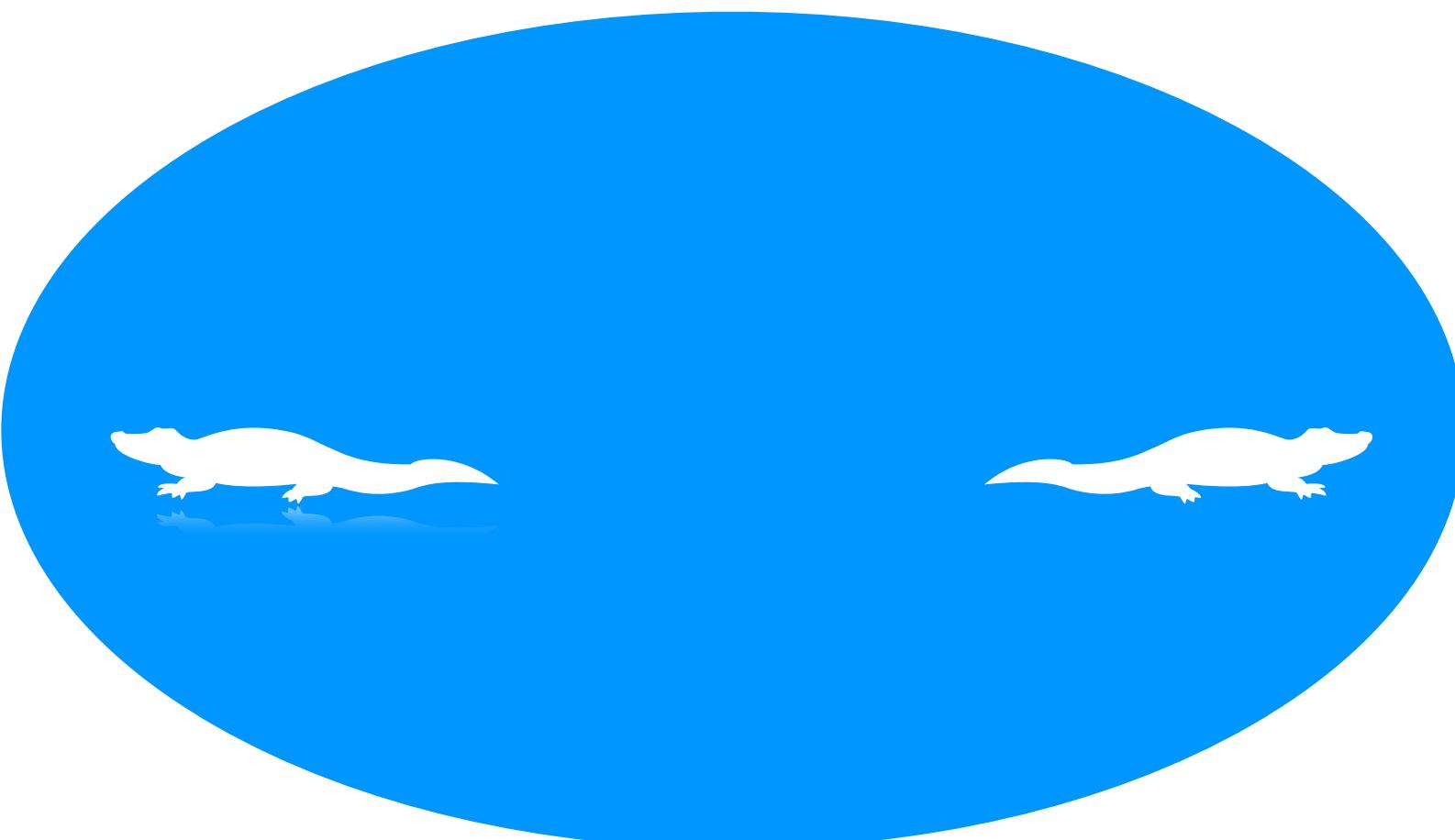
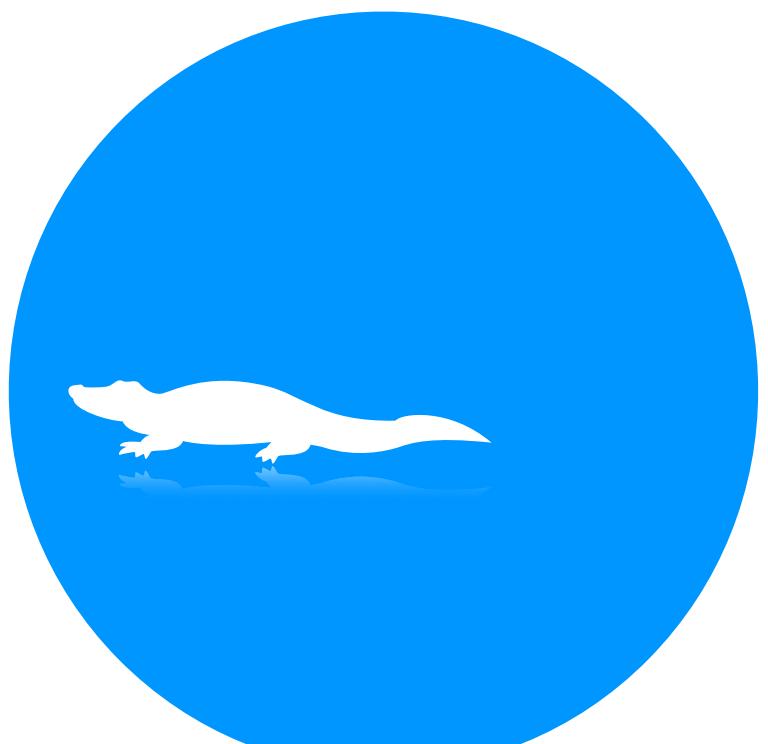
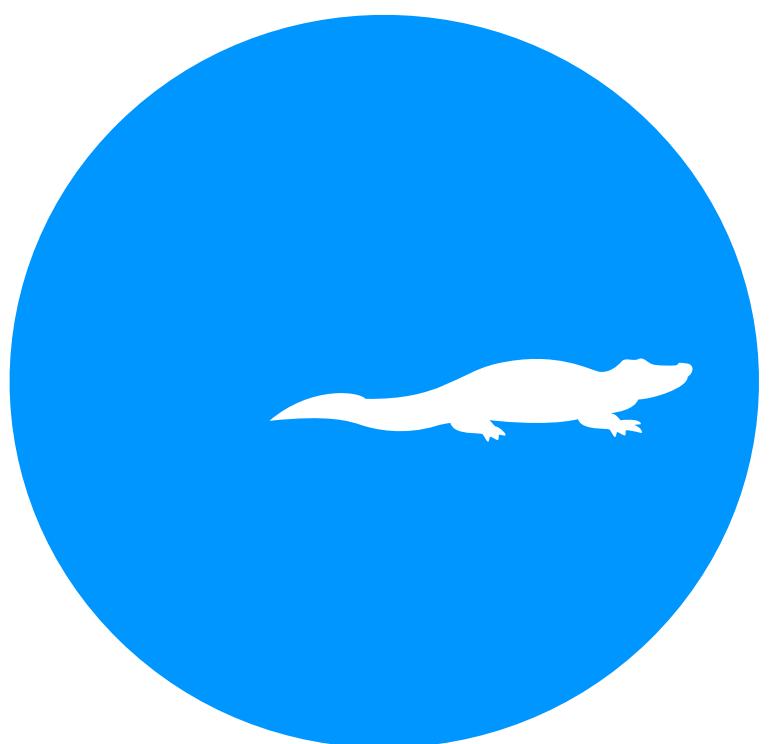
It was important to revisit notes to reflect on major themes in the context of new insight.

Example - character displacement.

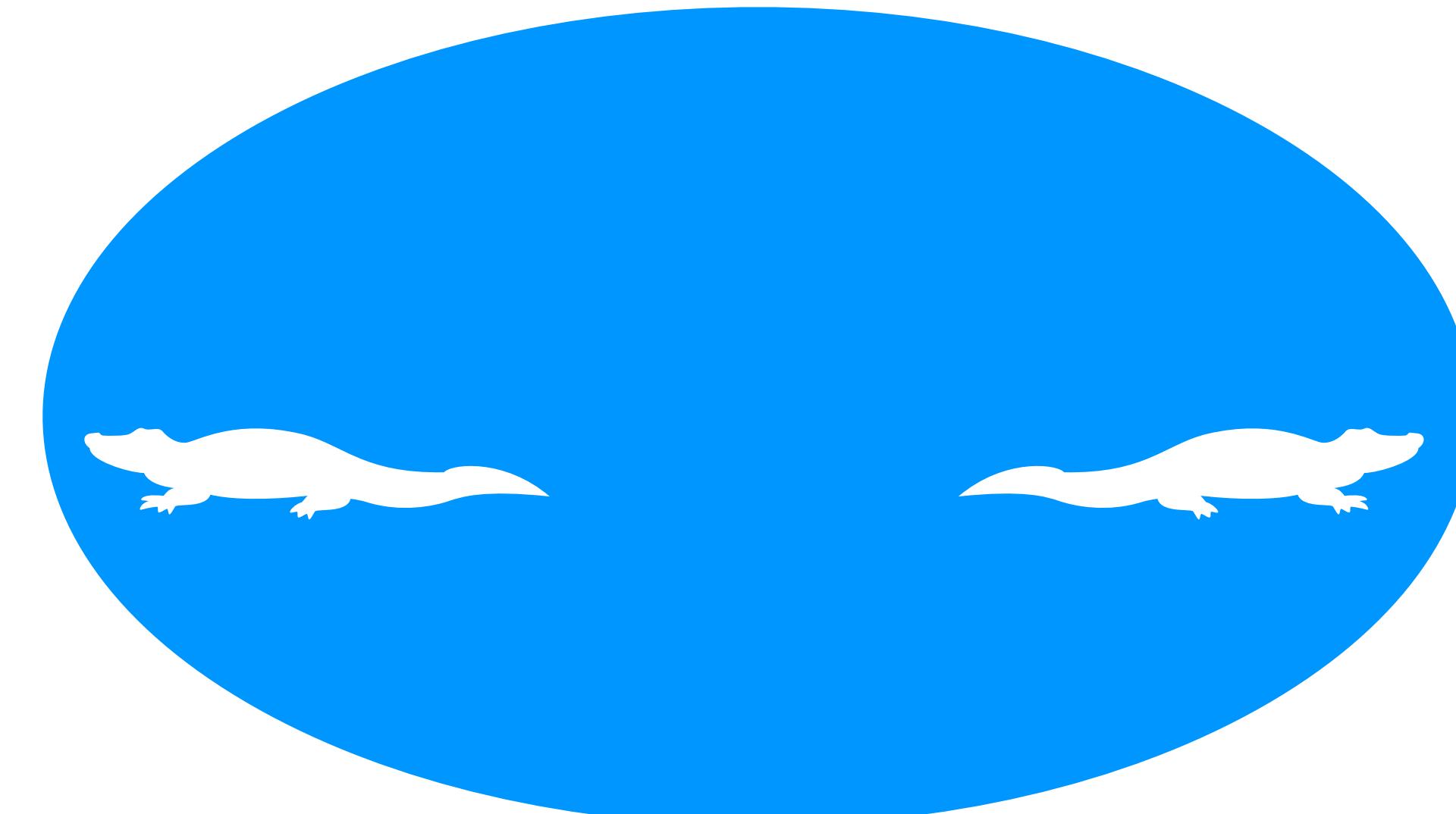
Domain goals.

*Central goal in their work:
Identify character transitions
and patterns that indicate
evolutionary adaption.*

*An example of this:
Character displacement.*

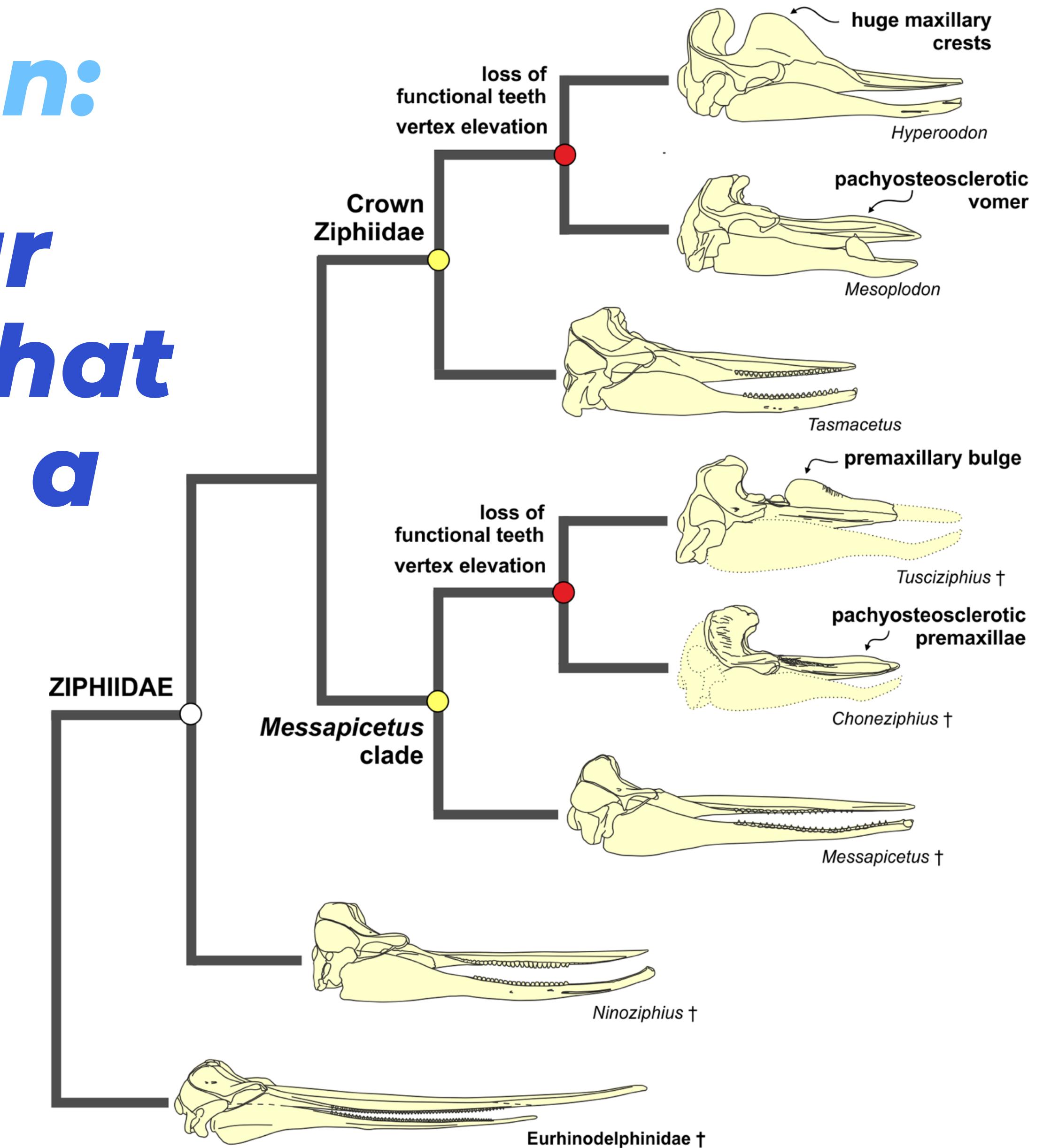


Sympatric Speciation:
**New species emerge from common
ancestor while inhabiting the same
geographic area.**



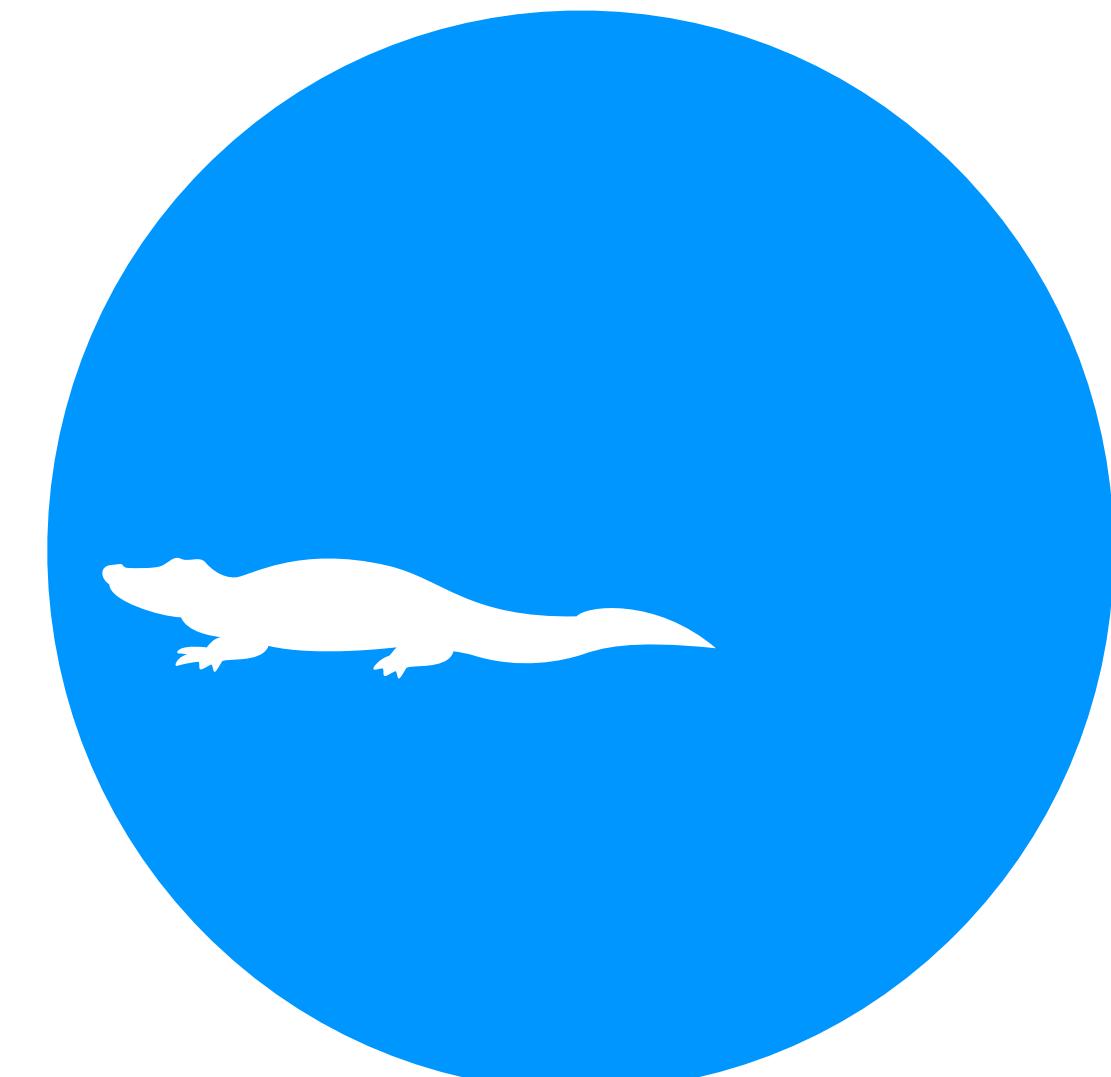
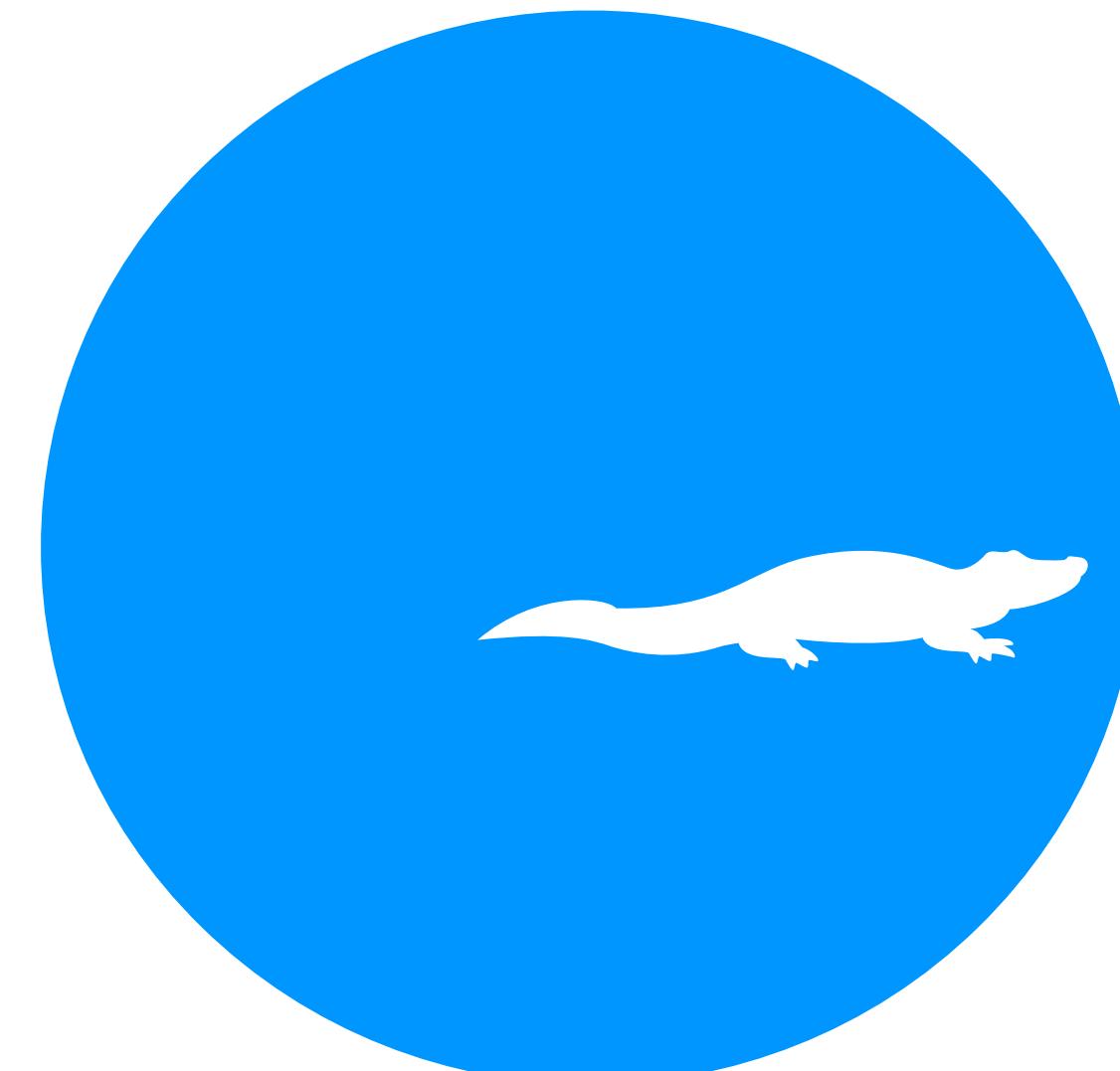
Convergent Evolution:

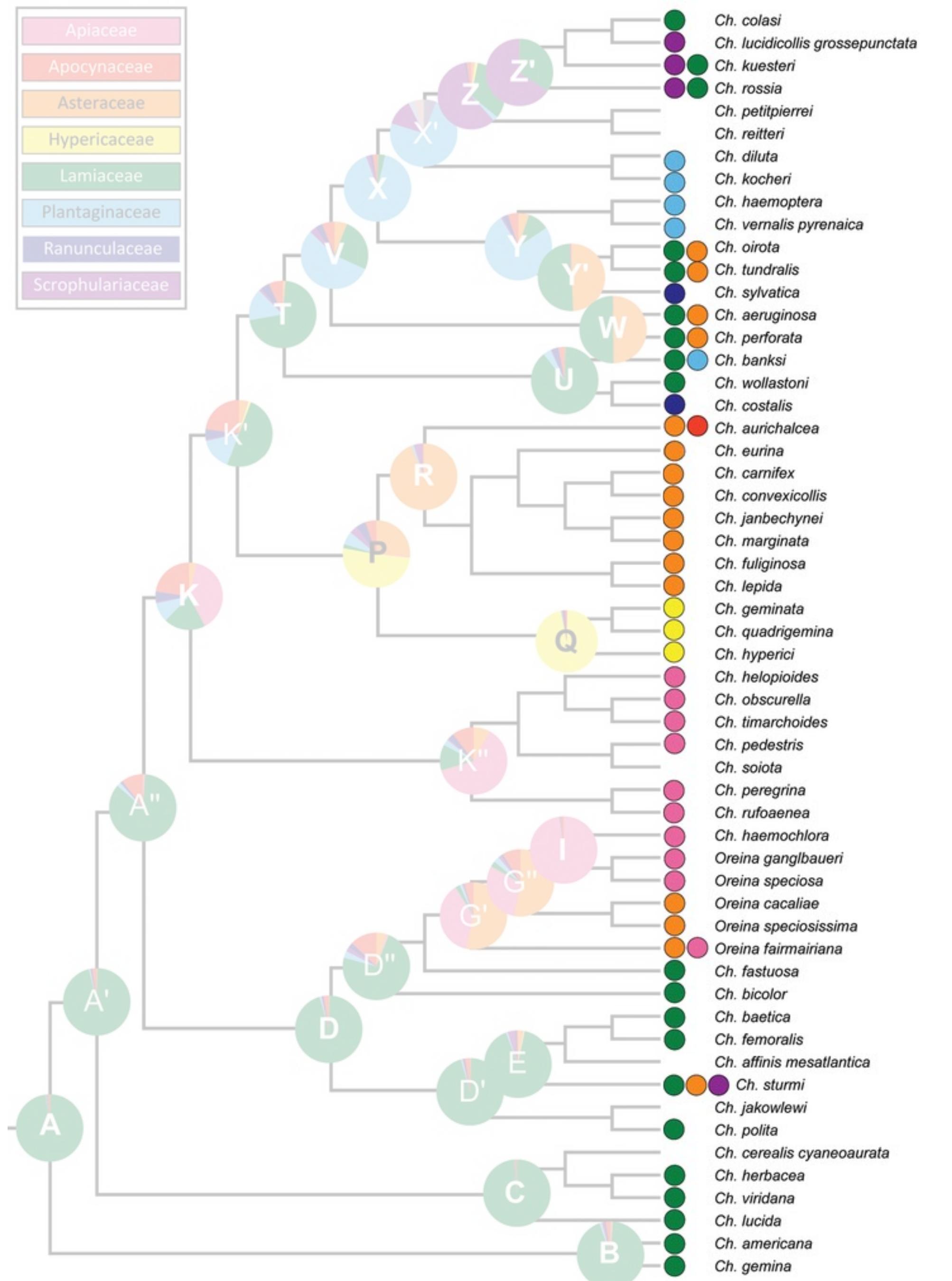
Emergence of similar features in species that have not come from a common ancestor



Convergent Evolution:

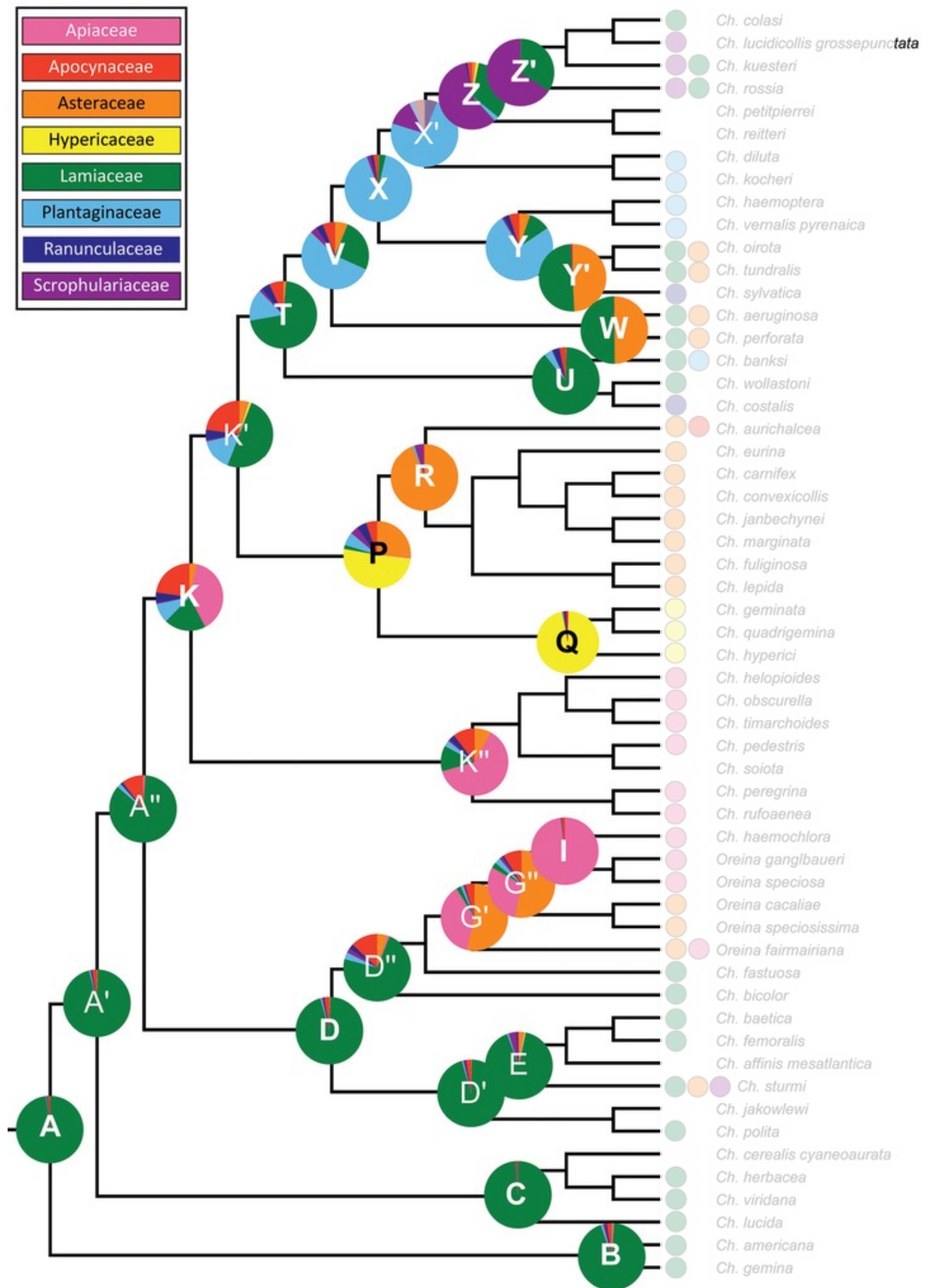
- **Trait or morphology emerges in separate areas of a tree.**
- **Find ecomorphs in separate geographic areas that have decreasing differences in trait values.**





Ancestral State Reconstruction

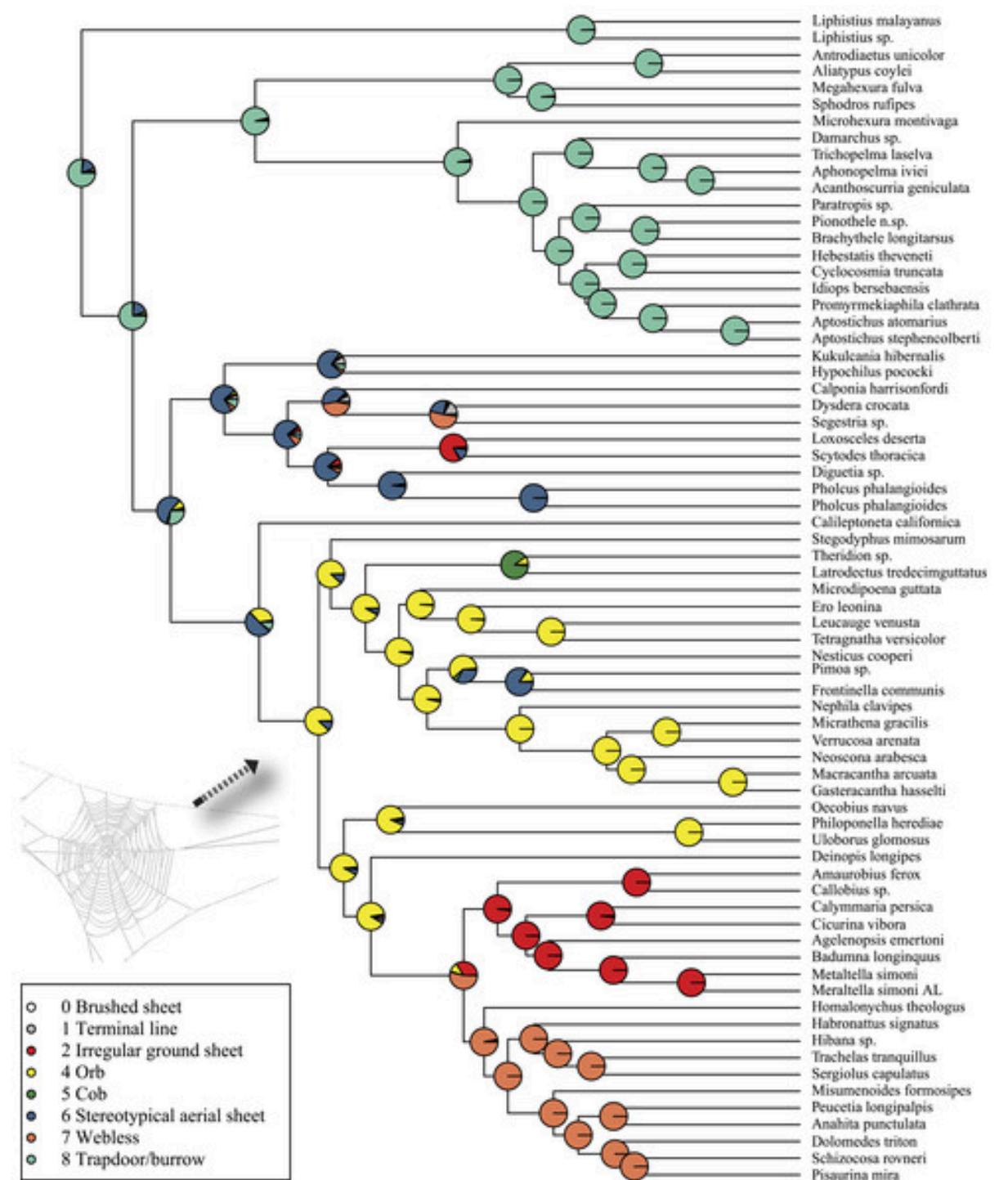
Use observed attributes from species samples



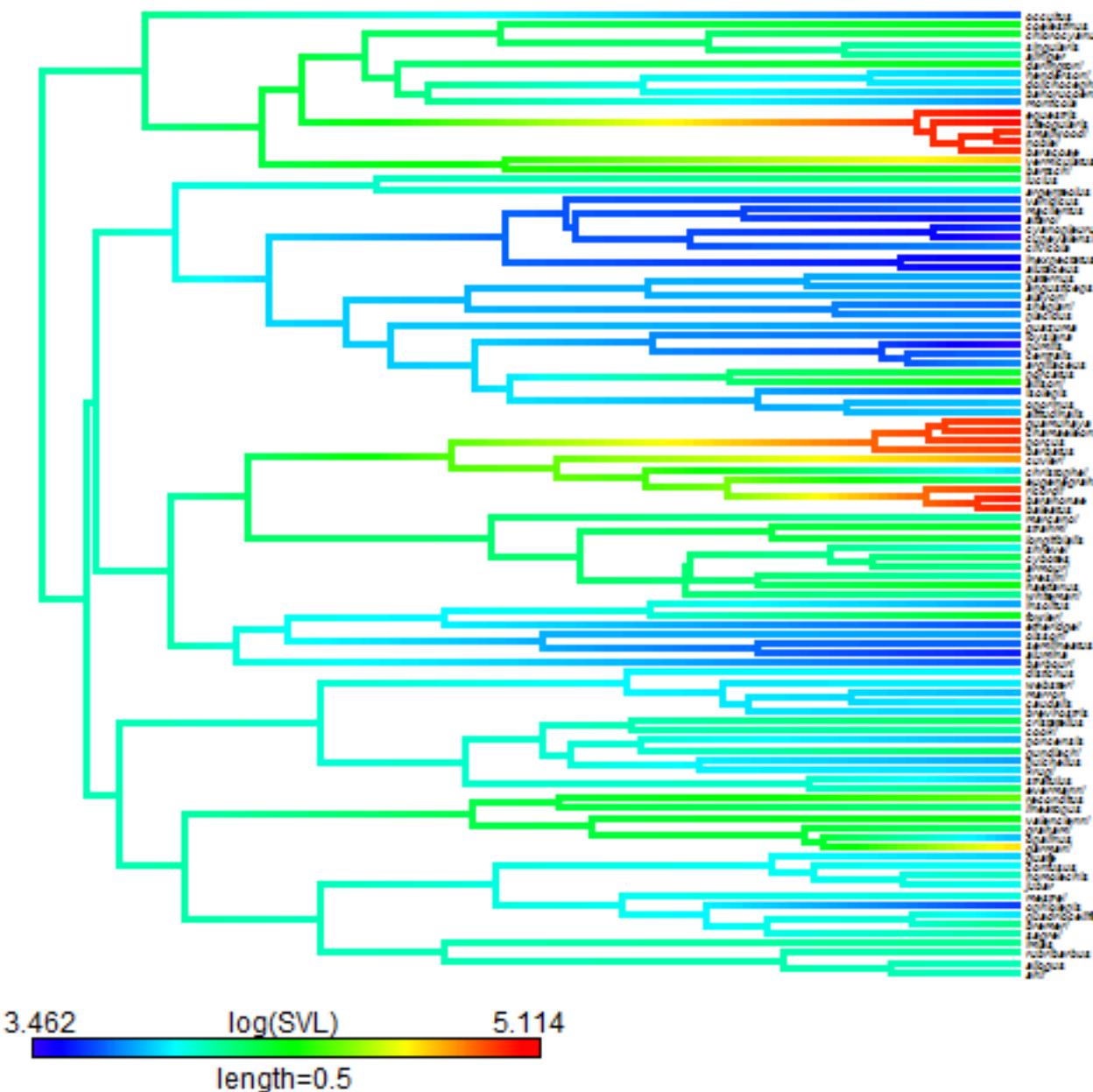
Ancestral State Reconstruction

Calculate state probabilities for the common ancestors up the tree.

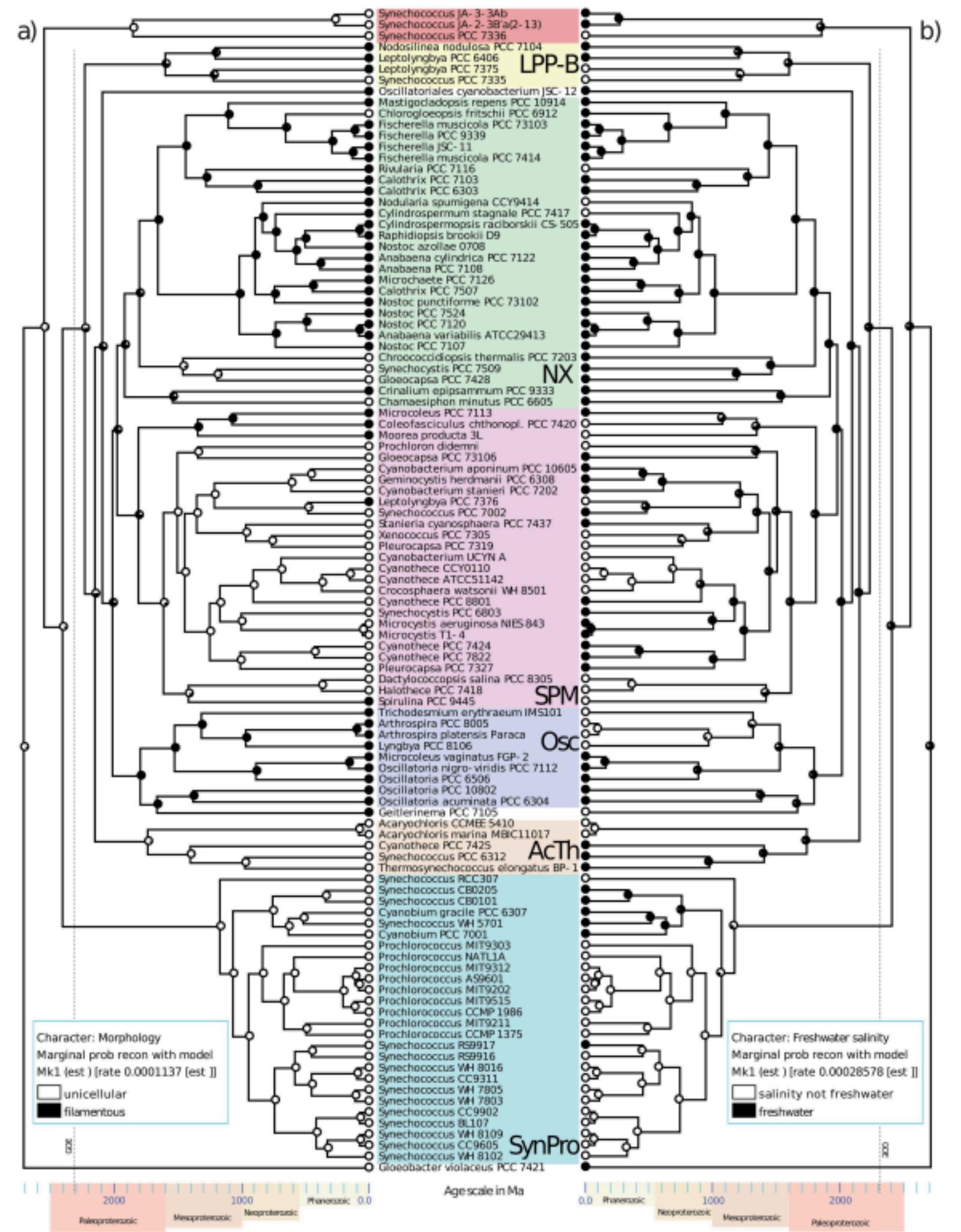
Discrete



Continuous



“Selective pressures involved in the origin of some traits may only be inferred with a comparative approach,” (Rezende et. al. 2012)



Hard to make comparisons using current representations.

Currently compare states of different attributes on separate trees.

This falls apart after more than 2 traits.

“If you have one continuous trait you can do things. If you two - ok. If you have three or four or five, there is nothing really sufficient” (from workshop)

**Lack the tools to view
multiple discrete and
continuous characters
together.**

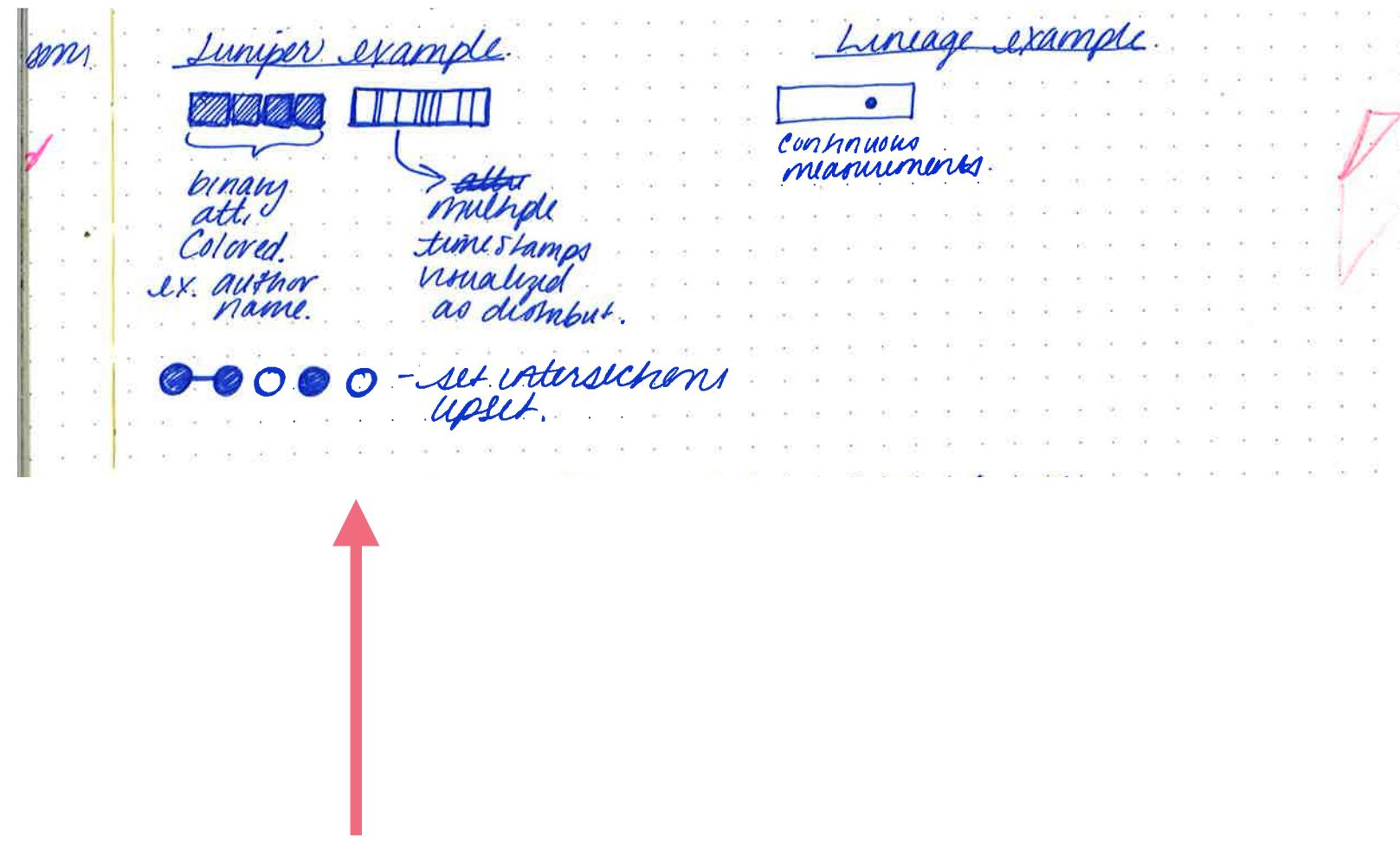
We want to build for this:

***Identify character transitions
and patterns that indicate
evolutionary adaption.***

Need the ability to:

- ***See how traits change through time***
- ***Group and filter based on ancestral and species traits***
- ***Compare values of different groups*****
- ***Retain context of tree topology***

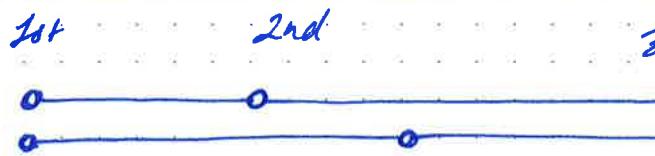
Initial focus was on view of multiple traits:



Started by sketching data representations used in existing tools.

Viewing multiple attributes at once seemed like a use case for Lineage/Juniper style visualization.

sort/
align by branch points.
by common ancestors?



- ④ ability to align by attributes
- ability to align by branches
- ④ filter on attributes.

current data sample

species

gymnodactylus

Chamaeleo chamaeleon

lippiae

can sort by a branch or attributes.

attribute

chamaeleon.

ambulus.

chamun.

nodondinae

align by branching

align by

linearize to compare the

branching

implicit tree view:

implied tree

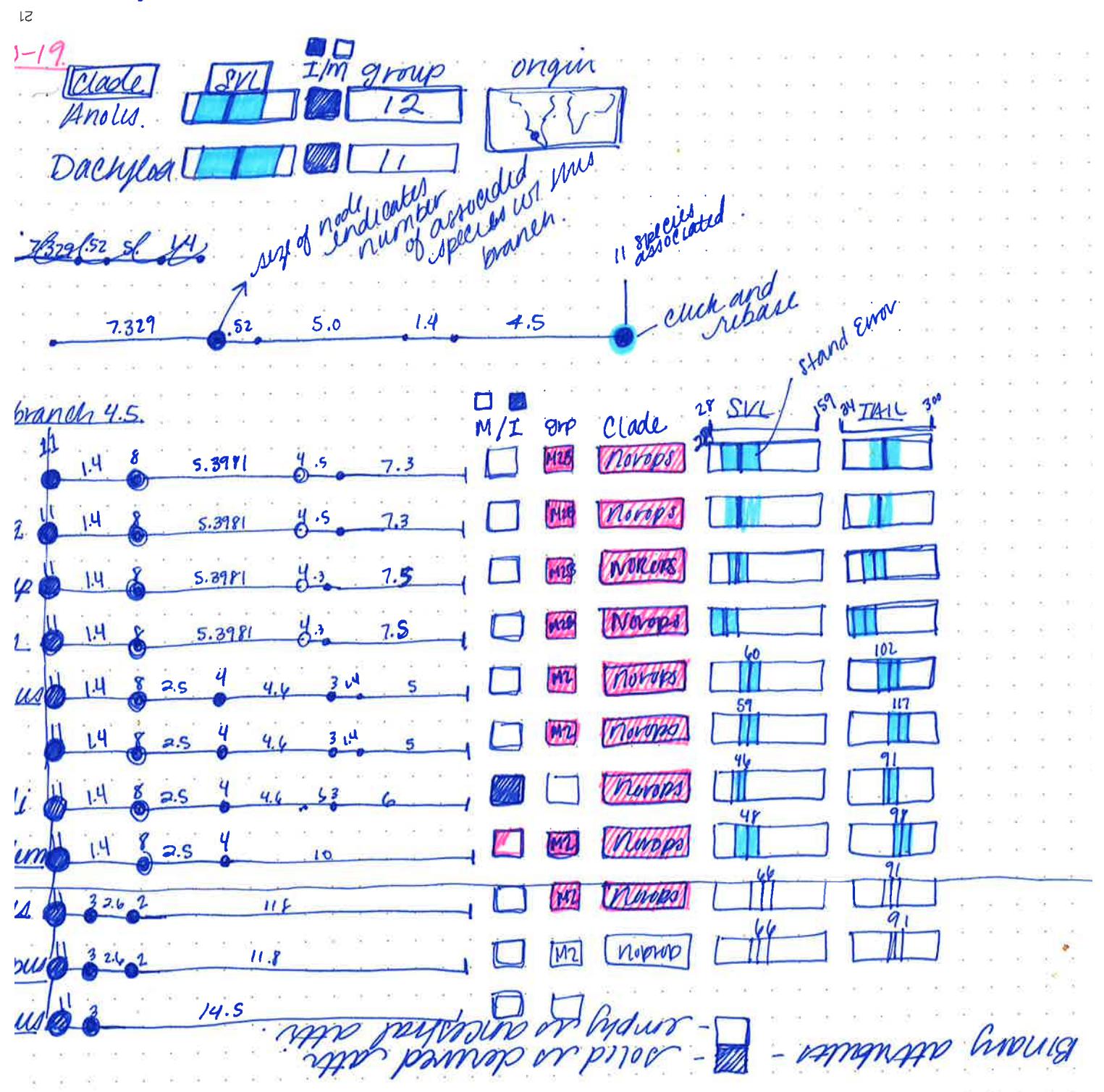
view:

branching

attribute

branching

↳ how do you determine ancestral vs derived traits.



**Separate linear
lines for each path
to allow sorting by
traits.**

Species as a reference point seems not as relevant.



- **See how traits change through time**
- **Compare values of different groups?**

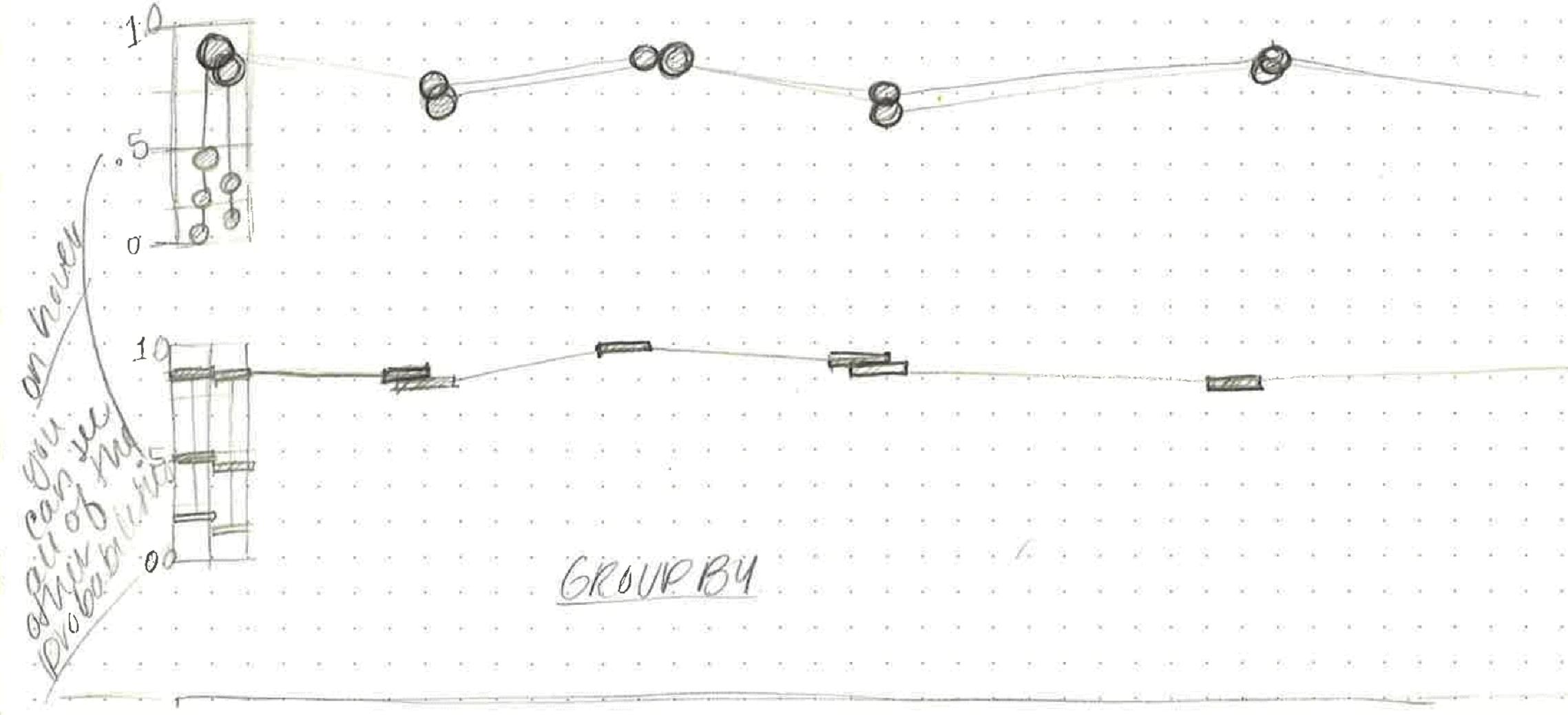
Filter/group by traits and trait shifts.

Filter based on state transition.

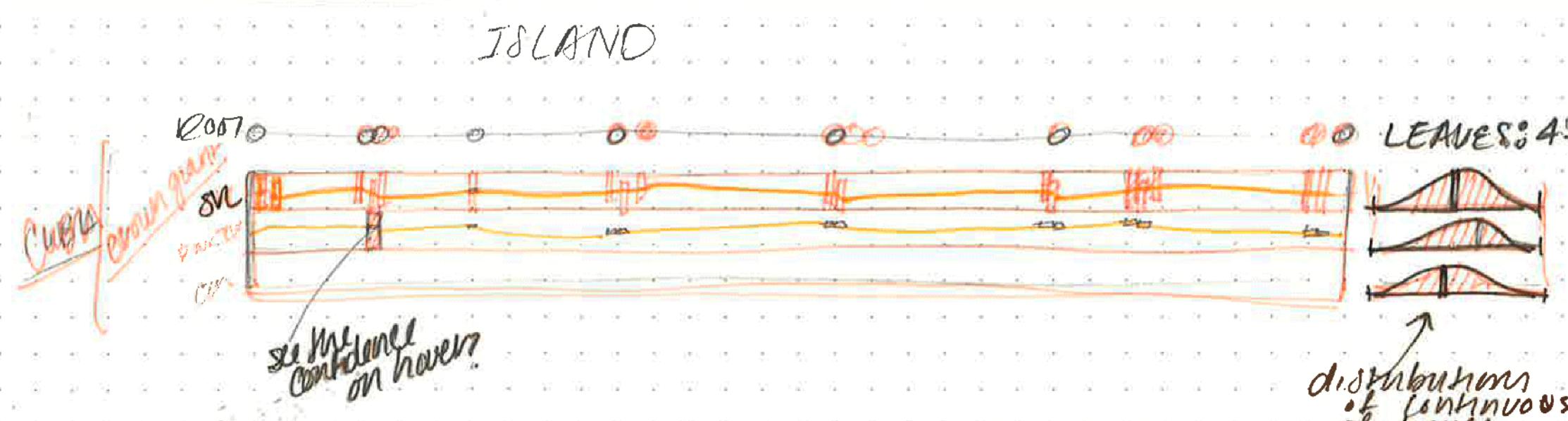


View the context of the tree.

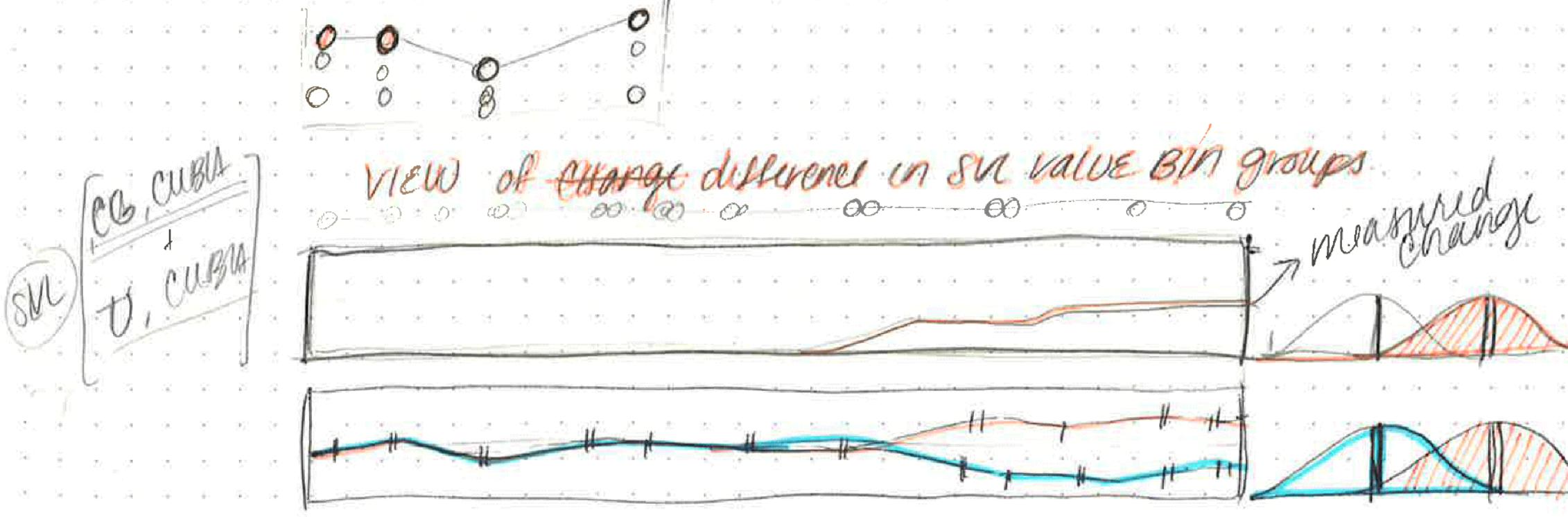
reconsidering the encoding of nodes



GROUP BY

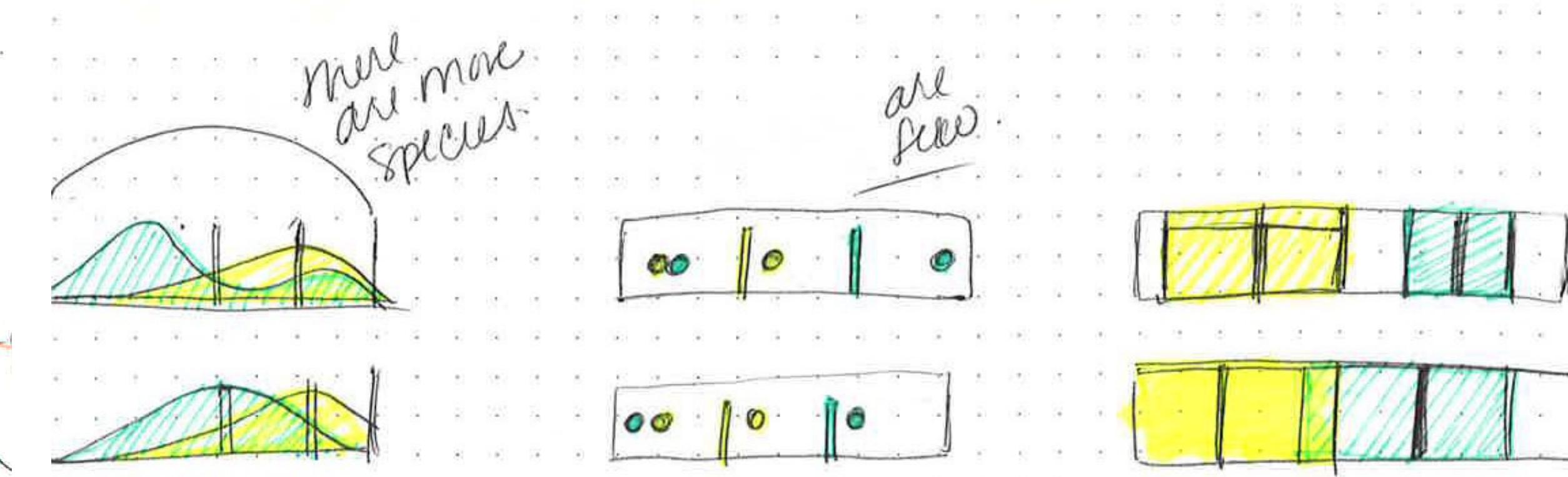


④ gray out everything but the transition shifts

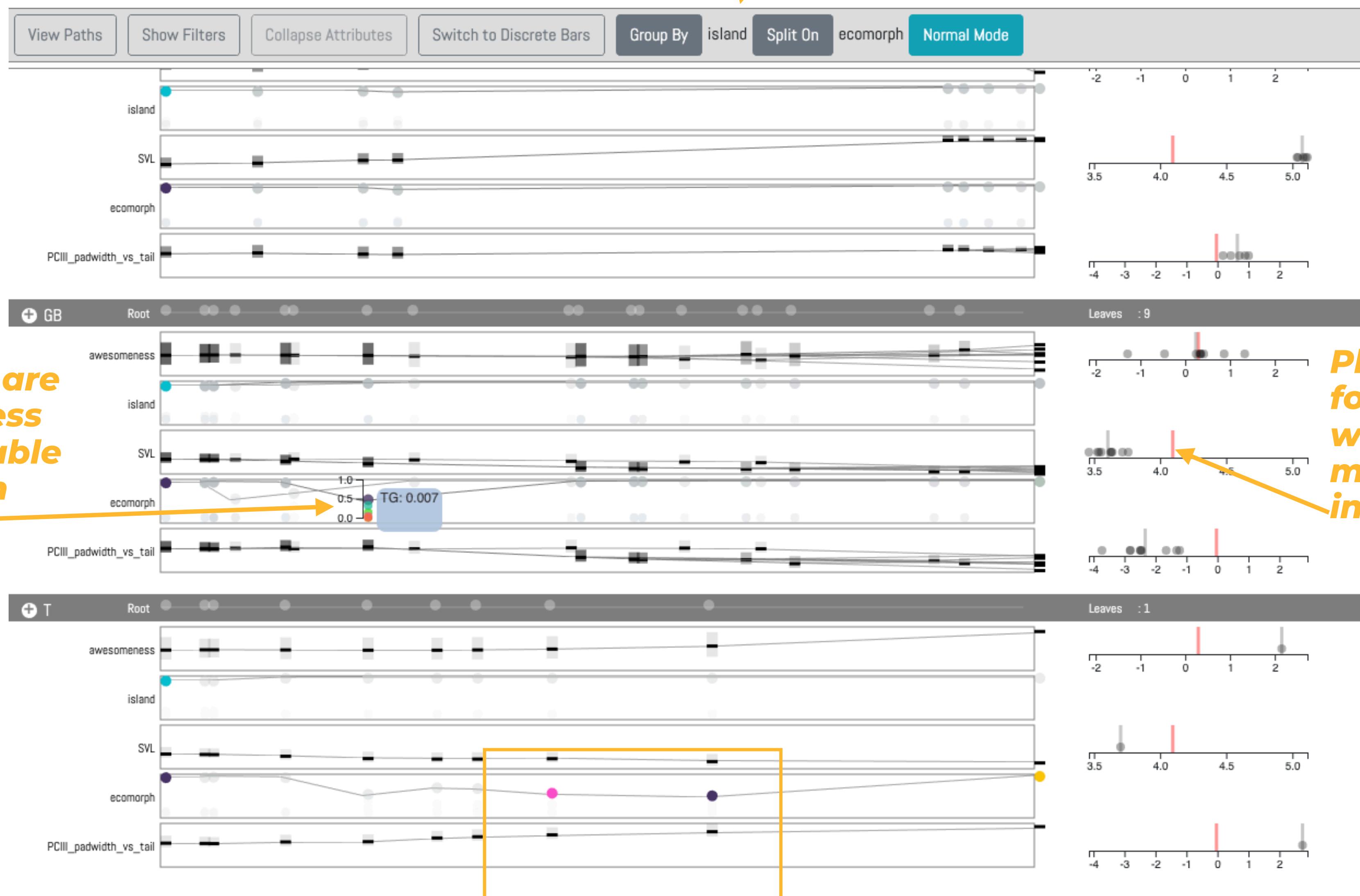


Compare trait-defined groups.

- Emphasis on comparison of continuous.
- Group by one discrete trait and split on another.
- From that, select groups to compare.
- Ability to see the measured difference in a continuous value across time.

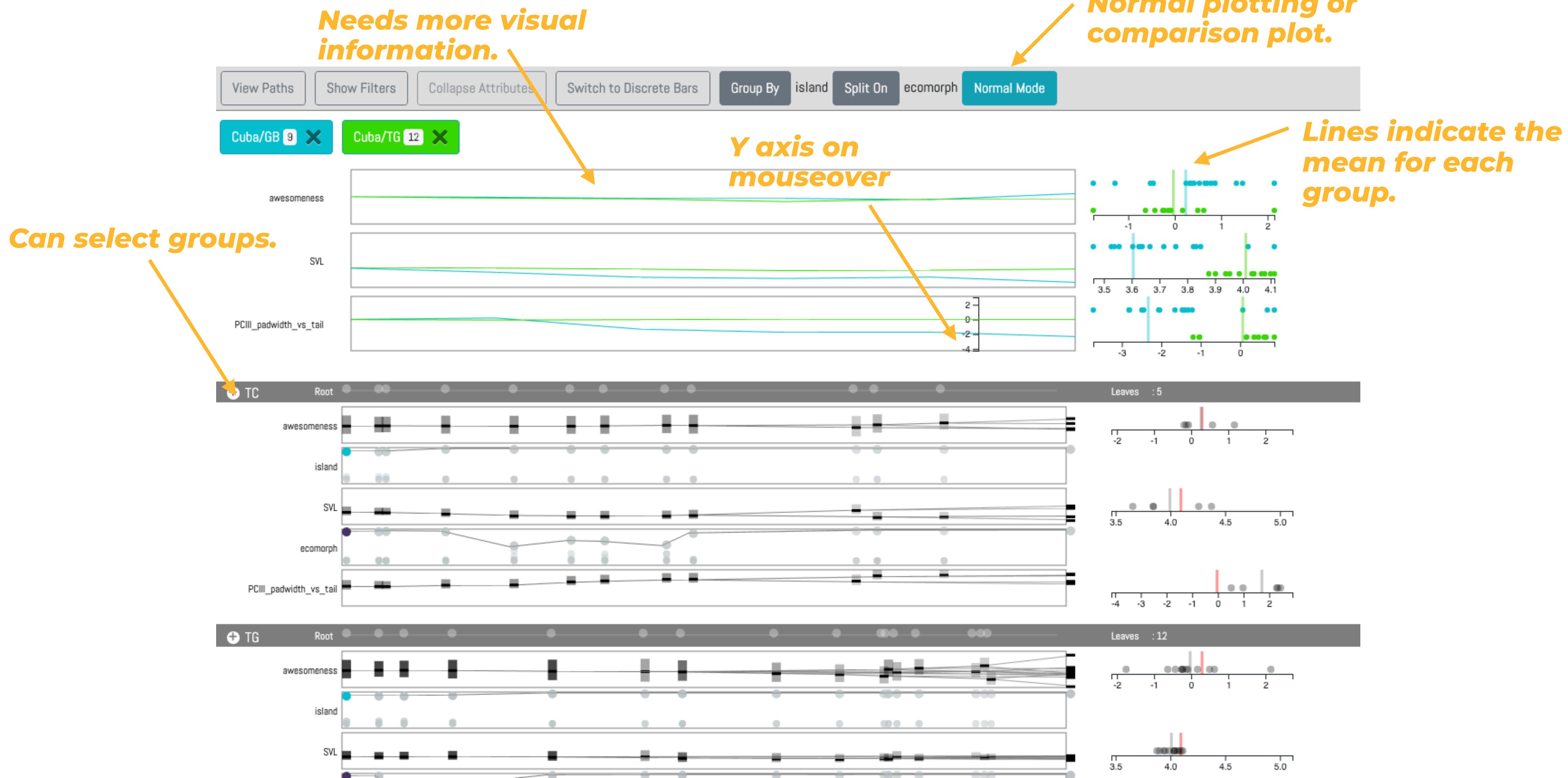


Grouping on discrete traits for comparison.



Circles that are colored indicate a probable state shift.

Select Groups to see differences in values.



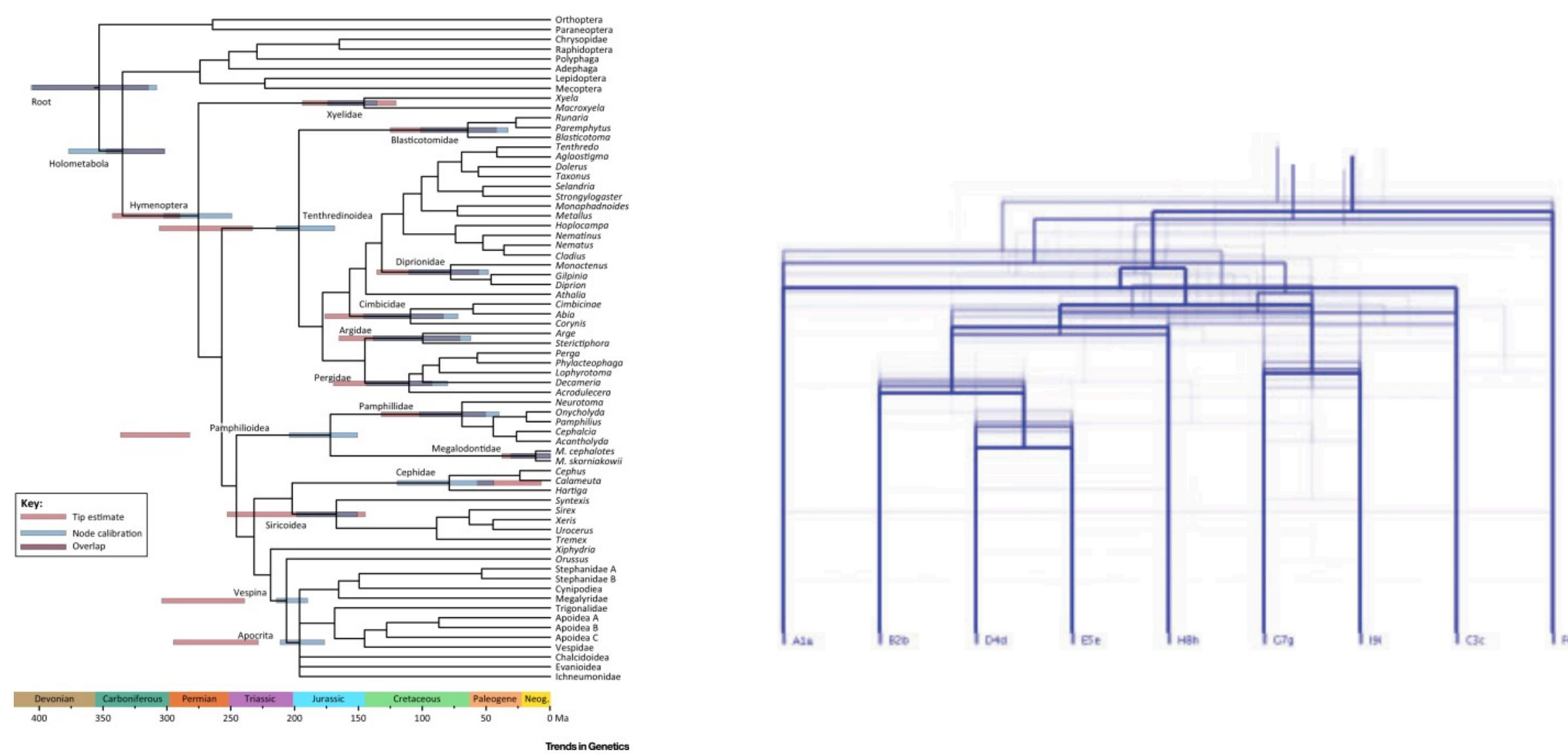
Select Groups to see differences in values.



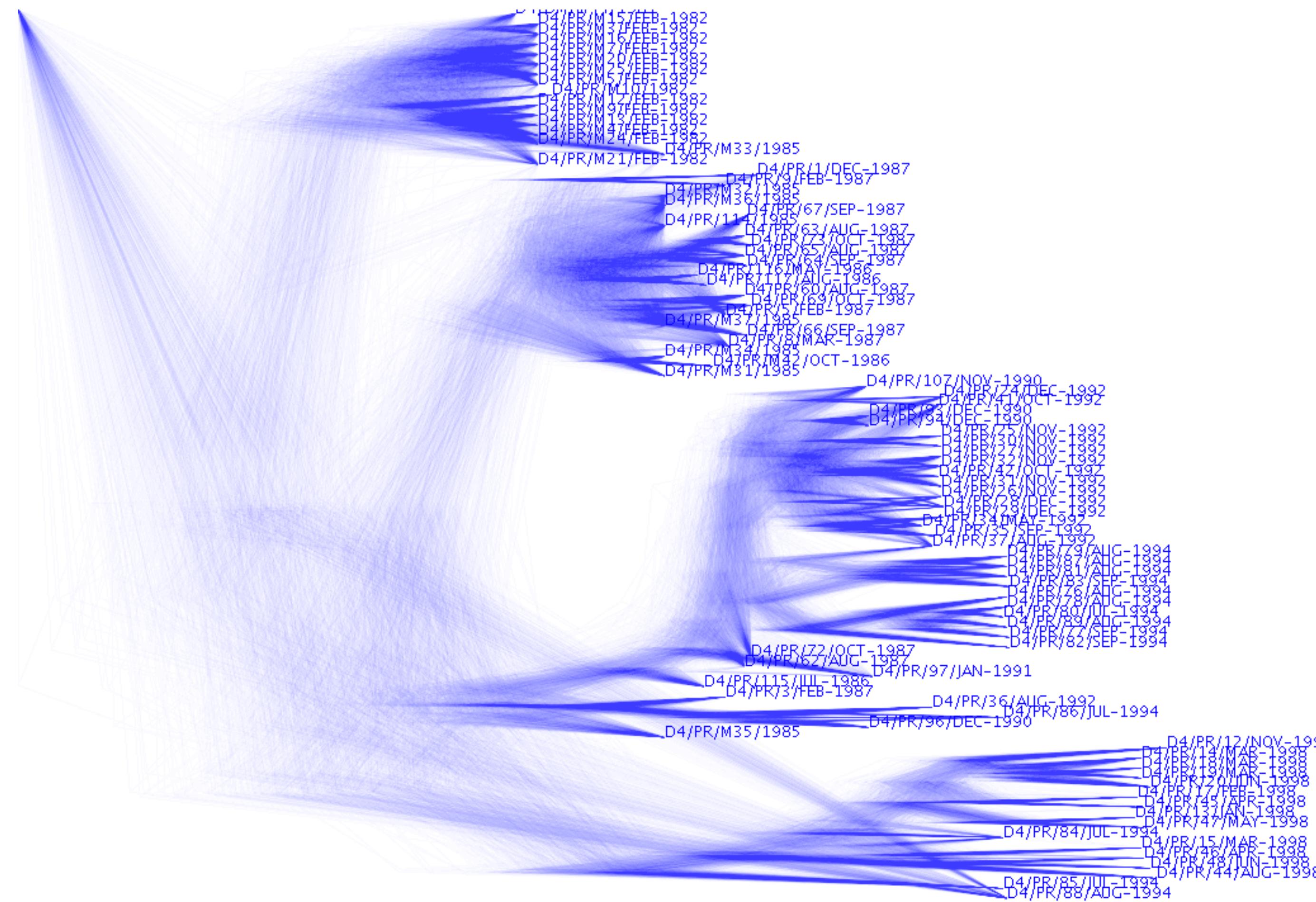
Support primary tasks:

- ***Visualizing uncertainty***
- ***Making the tool scalable to large trees***

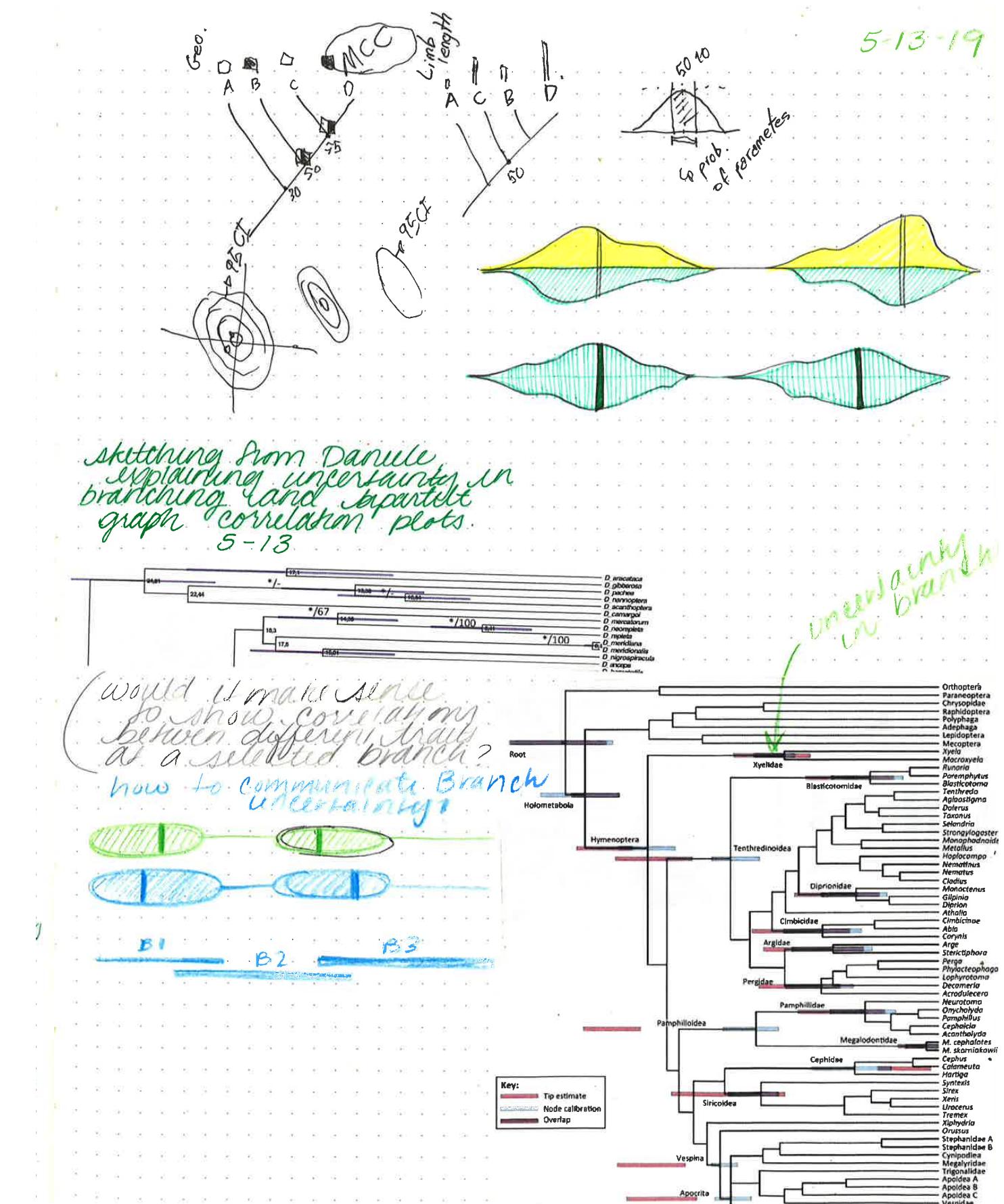
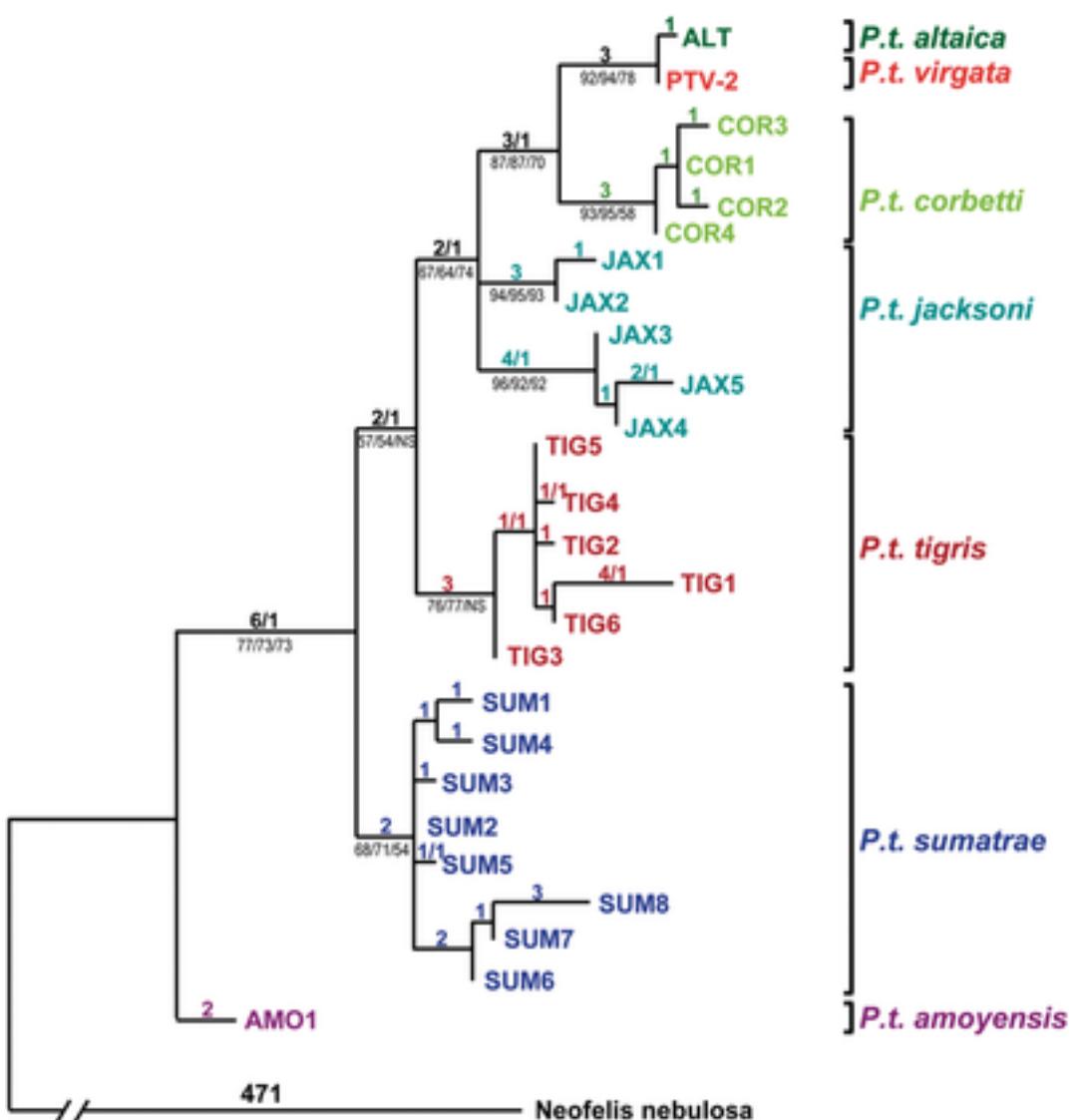
The Uncertainty in the reconstructed traits and the topology of the tree.



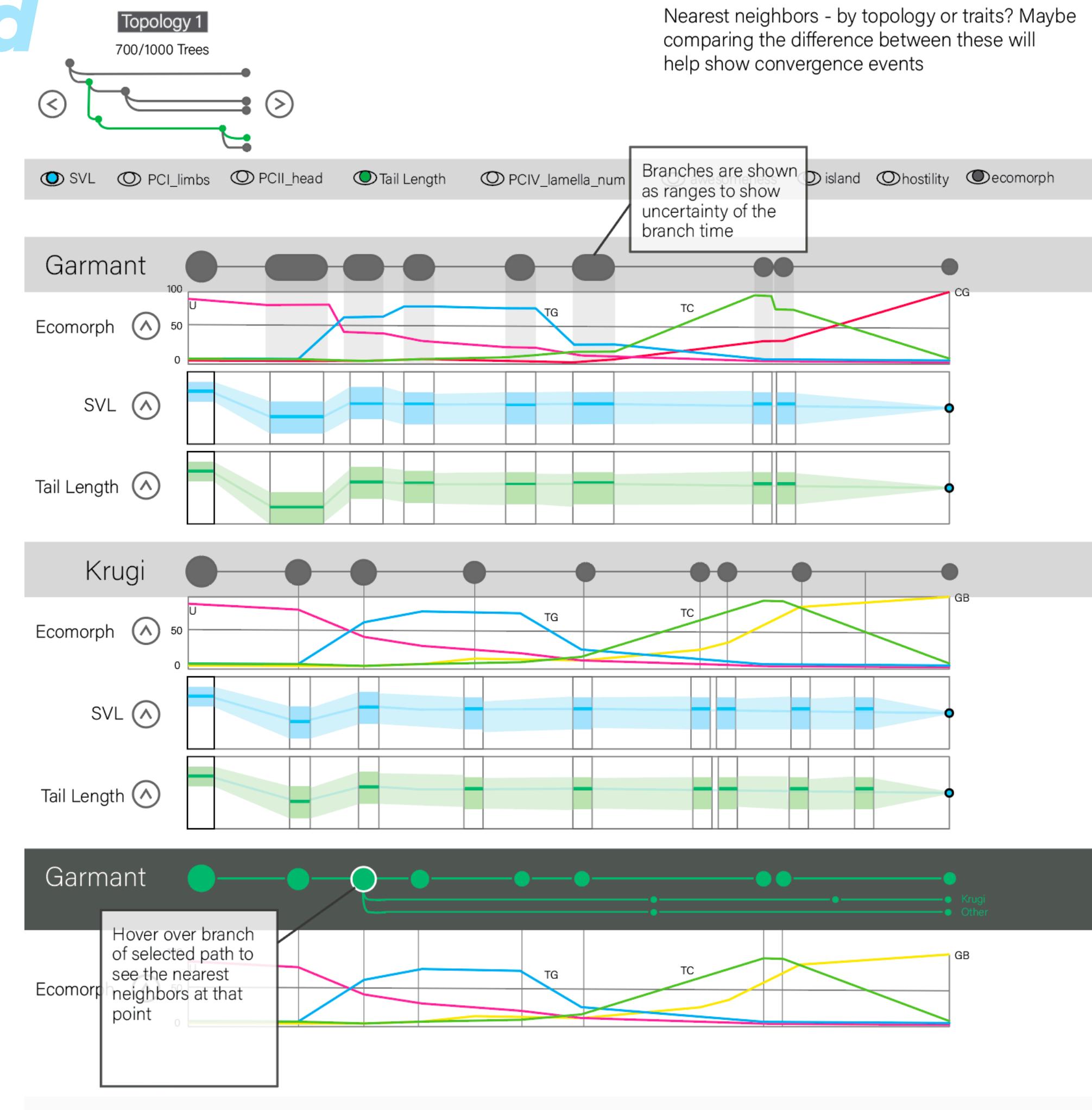
Uncertainty in topology

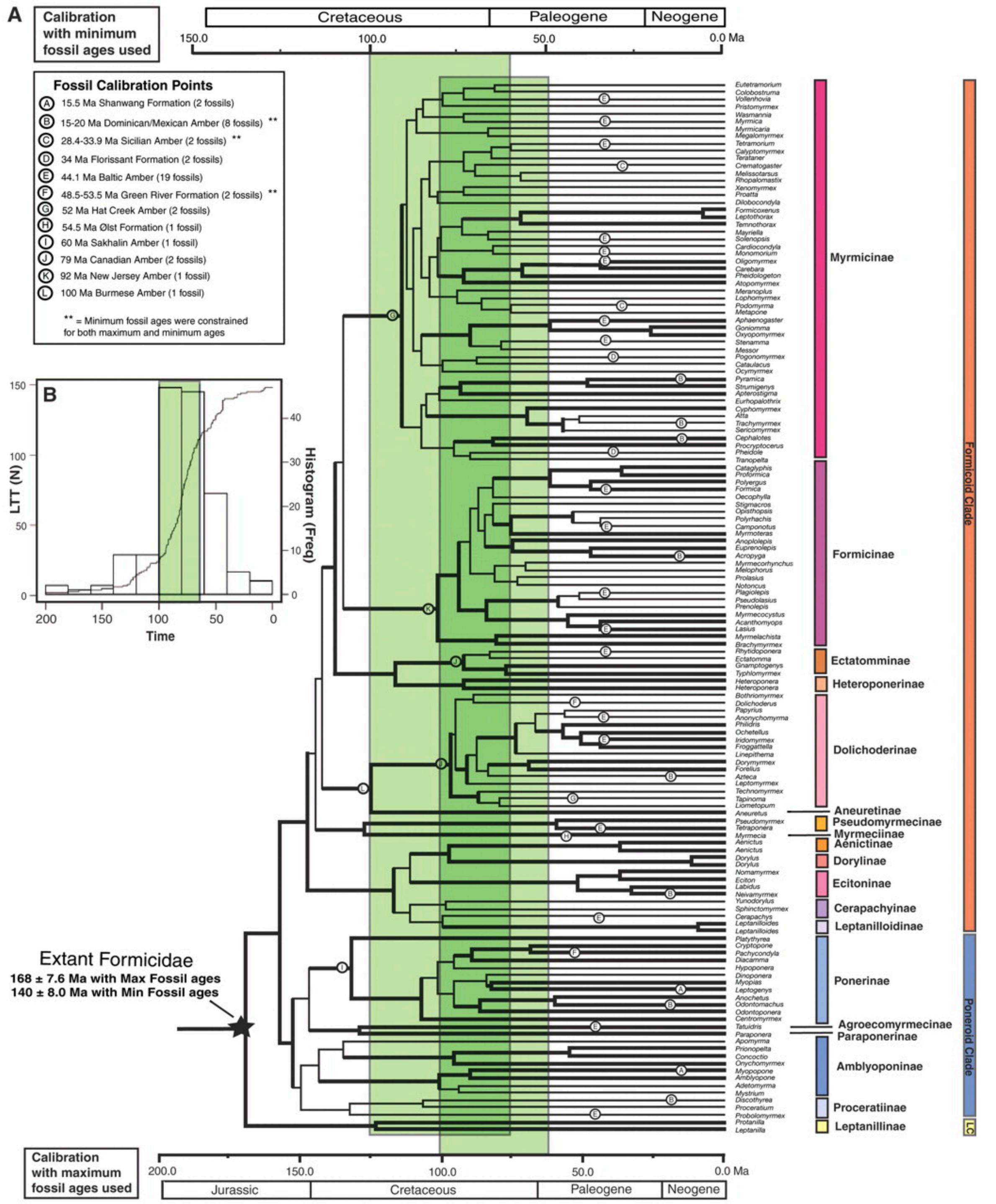


Uncertainty in branch points



Top tree topologies can be navigated through.





Yet another consideration:

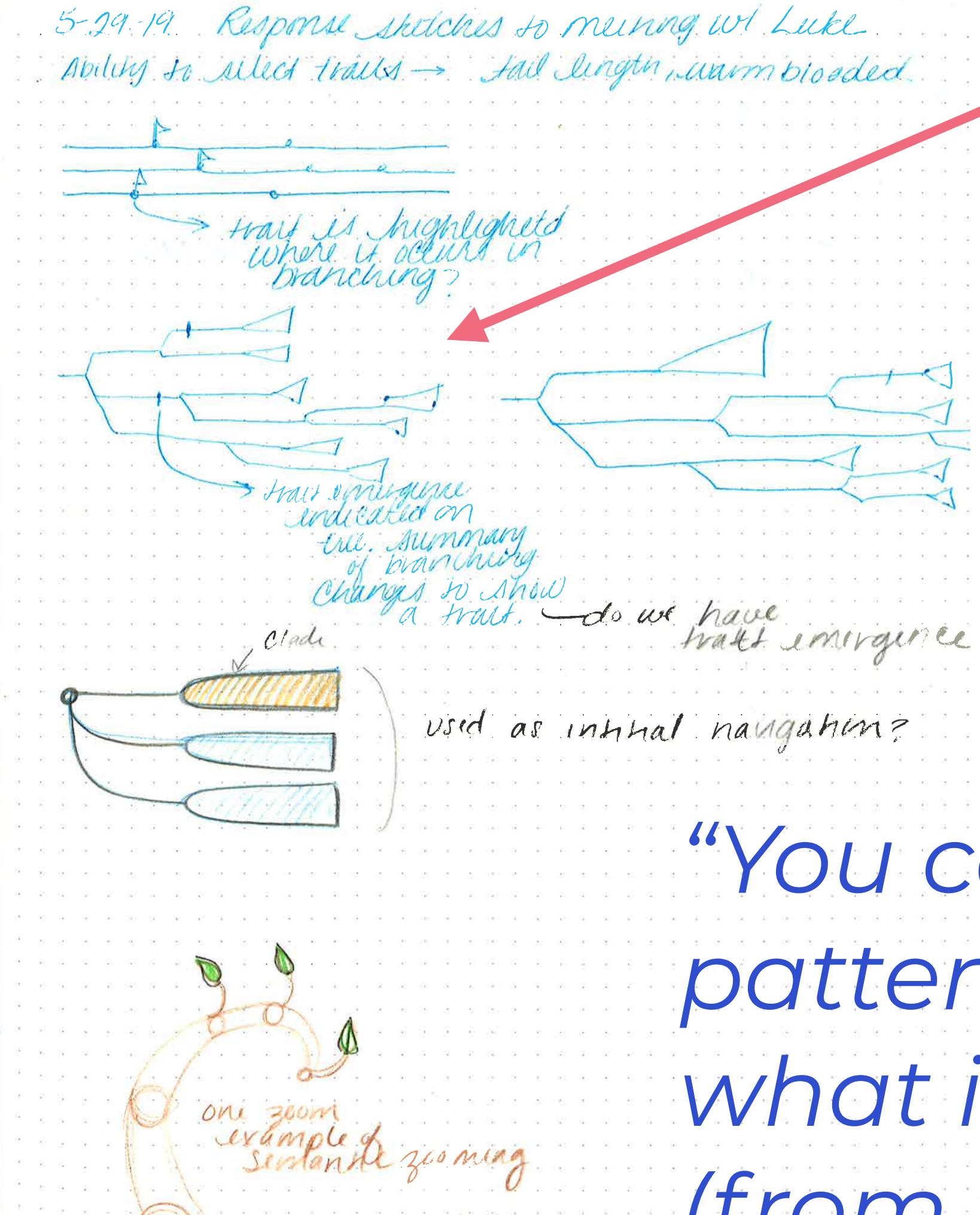
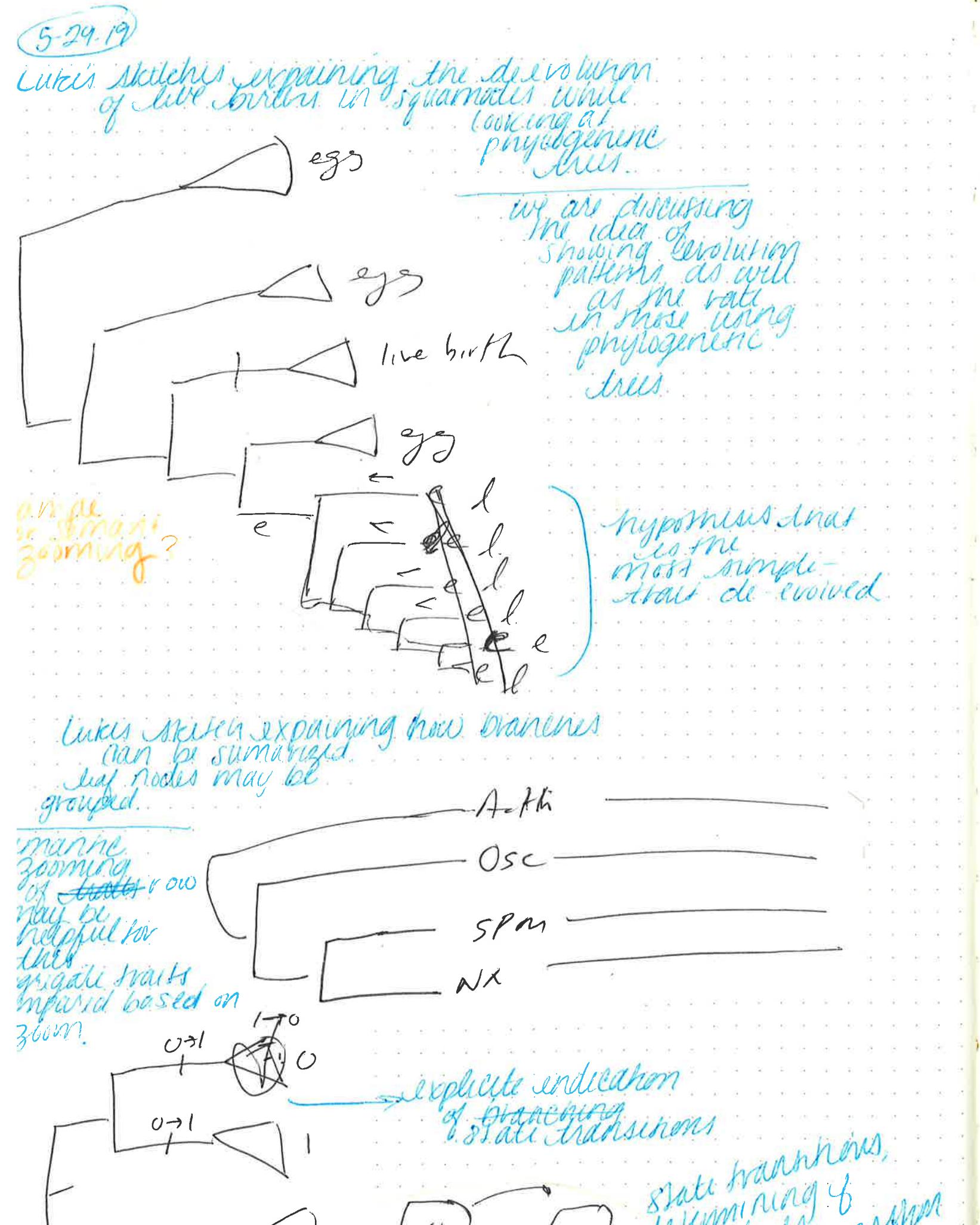
This needs to be scalable to view large trees (>300) with multiple attributes.

Making it Scalable:

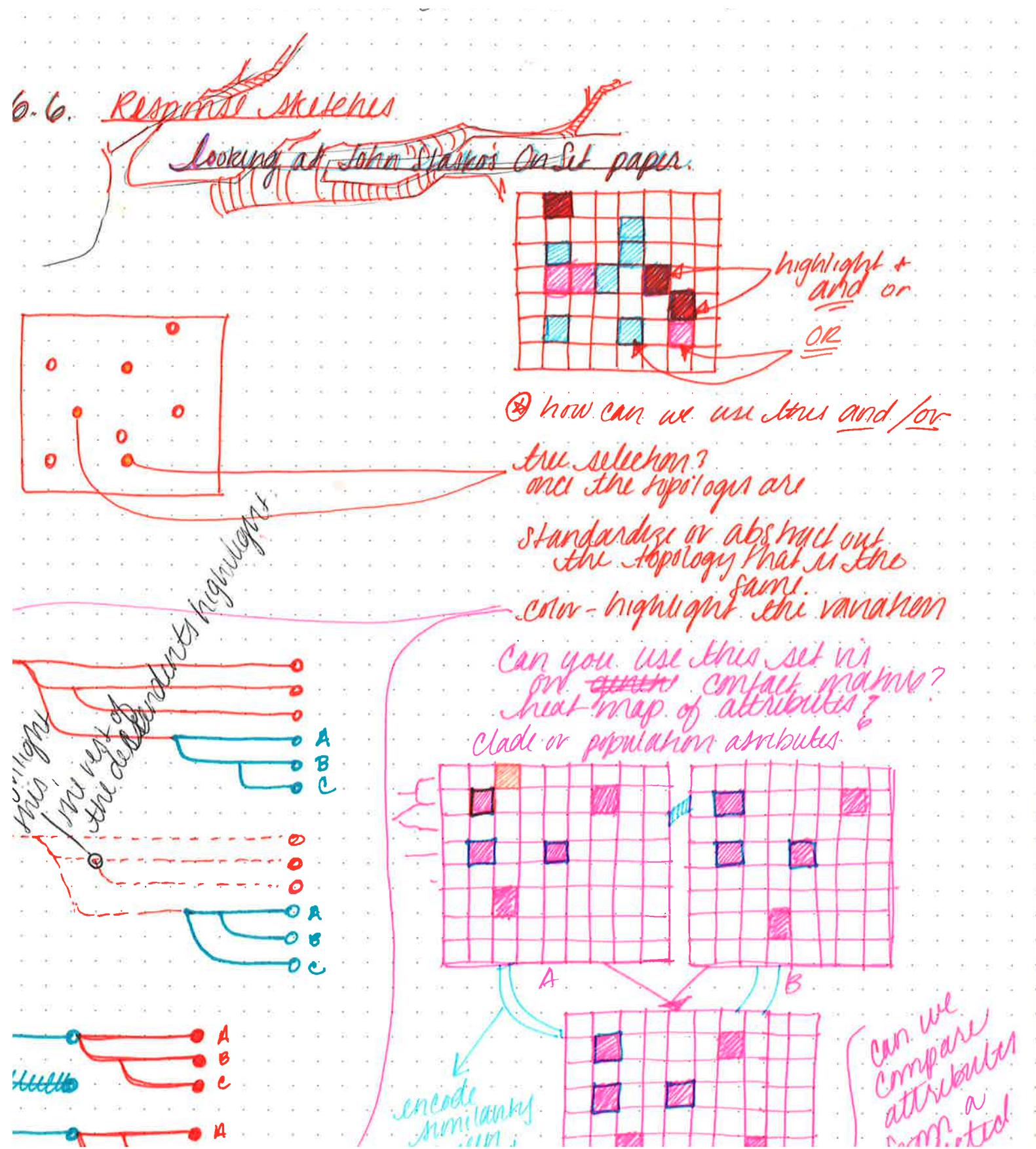
- **Semantic zooming.**
- **Collapse what is not important for the task at hand.**
- **Aggregation.**

View responsive
to the level
shown in the
tree.

“You could visualize large
patterns by collapsing
what is not important”
(from the workshop)



“Visualization [would depend] on the context”



If you start with 100 species, the individual path view is less helpful.

Austin's sketches.

7-16-19

how to

this is similar to path finder
sydney nelson

ABUS - explorer

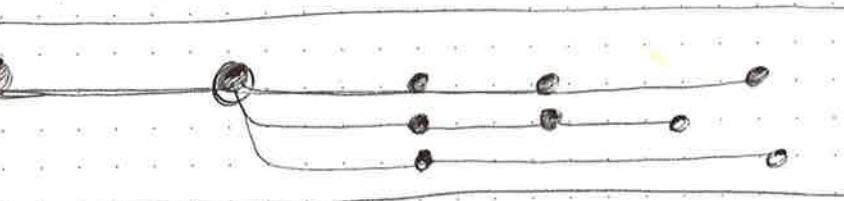
fix overlap accompanying attri

position nodes precisely

space attri

① calculate distance metric

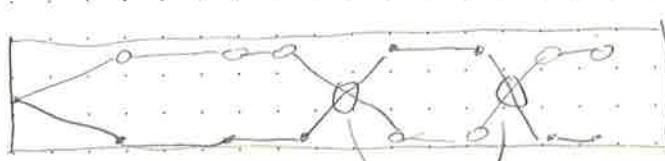
Spoke w/ DMO, we talked about



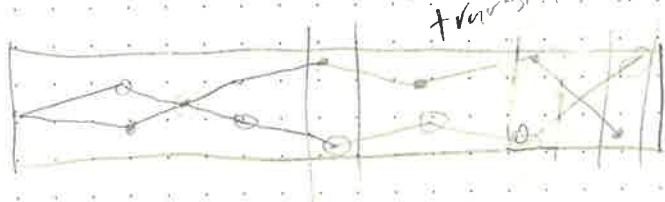
7-17-19

Filter by list of species? (interactive?)

Filter by discrete traits (at present/observed)



② transitions? (when?)



I want transitions first
filled (green) filled
Sign flips (blue) too
Sensitive
2 trans
75%, 25%

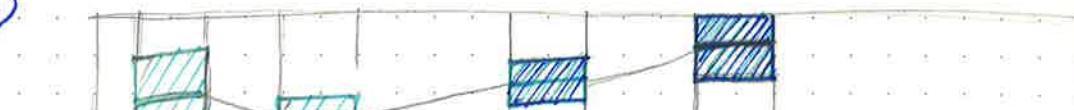
7-17-19

subtracting
the probabilities
to see stuff
happen in year would be import

time in year when me grnts happen

"Certain" & greater than prob

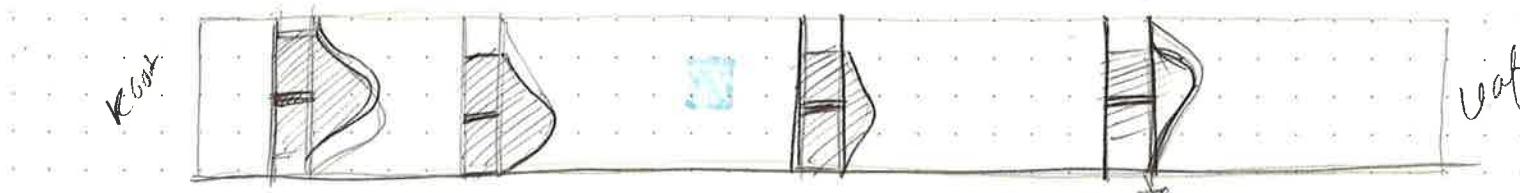
how do we handle the
continuous variables?



secondary transition
encodes higher values dark
it can be collapsed
down to coarse
bins

lines stacked
up or down to increase
increase

③ filter system to focus on state switches
should the distribution now be
the filter off



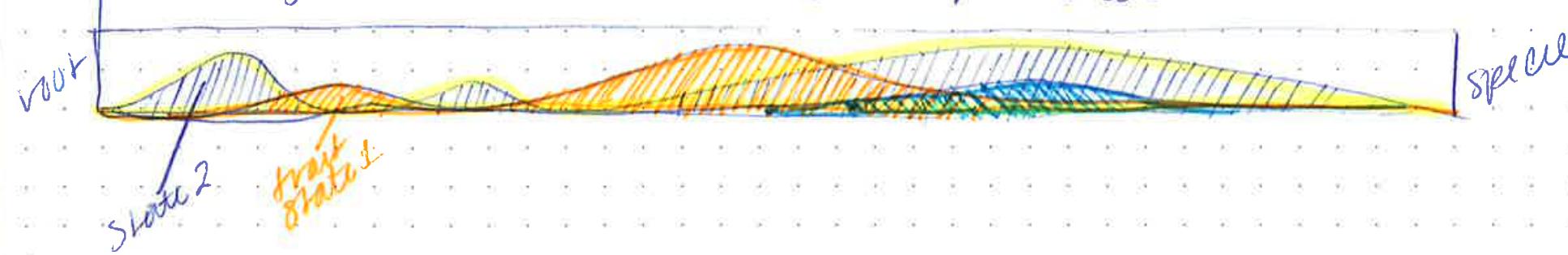
SEARCH BASE TO FILTER BY LIST OF NAMES

not yet
partitions not yet
done

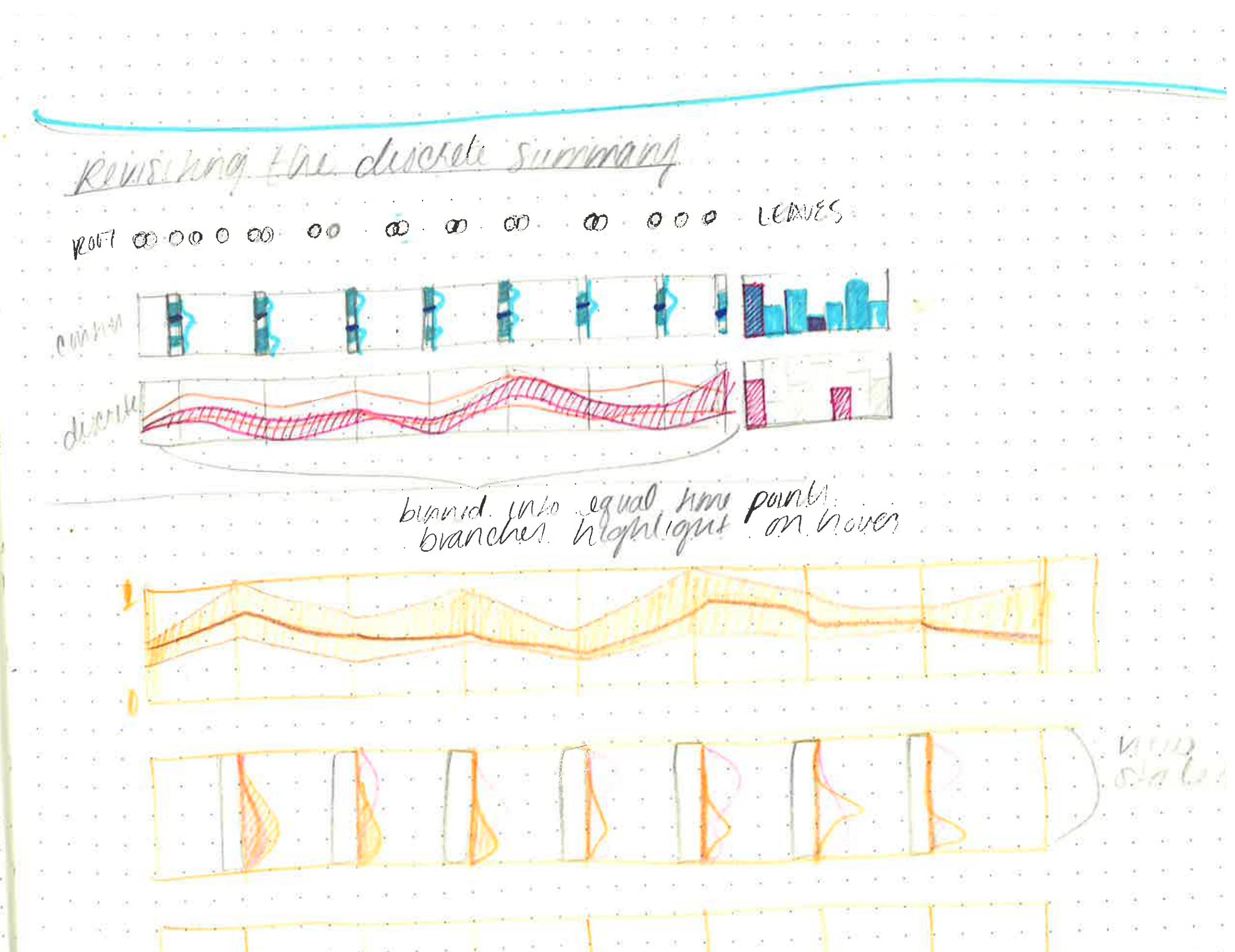
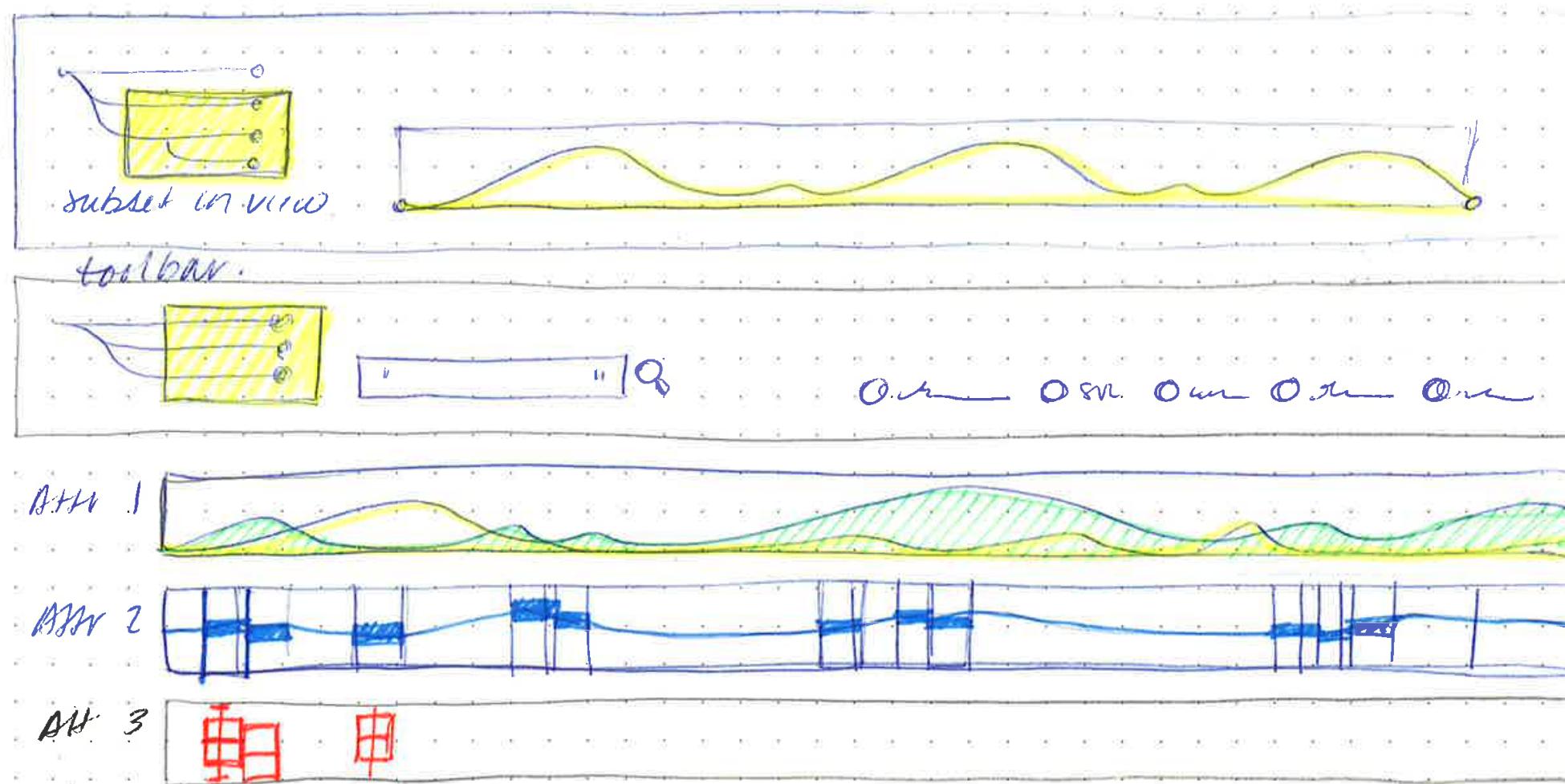
7-5-19

meeting w/ Daniel

doodle to show me summary view we talked about
you have distributions of trait probabilities for ~100 species



Begin w/ summary view of distributions



[View Paths](#)

[Show Filters](#)

[Collapse Attributes](#)

[Switch to Discrete Bars](#)

[Group By](#)

Shown: 100 /100

Root

awesomeness

island

SVL

ecomorph

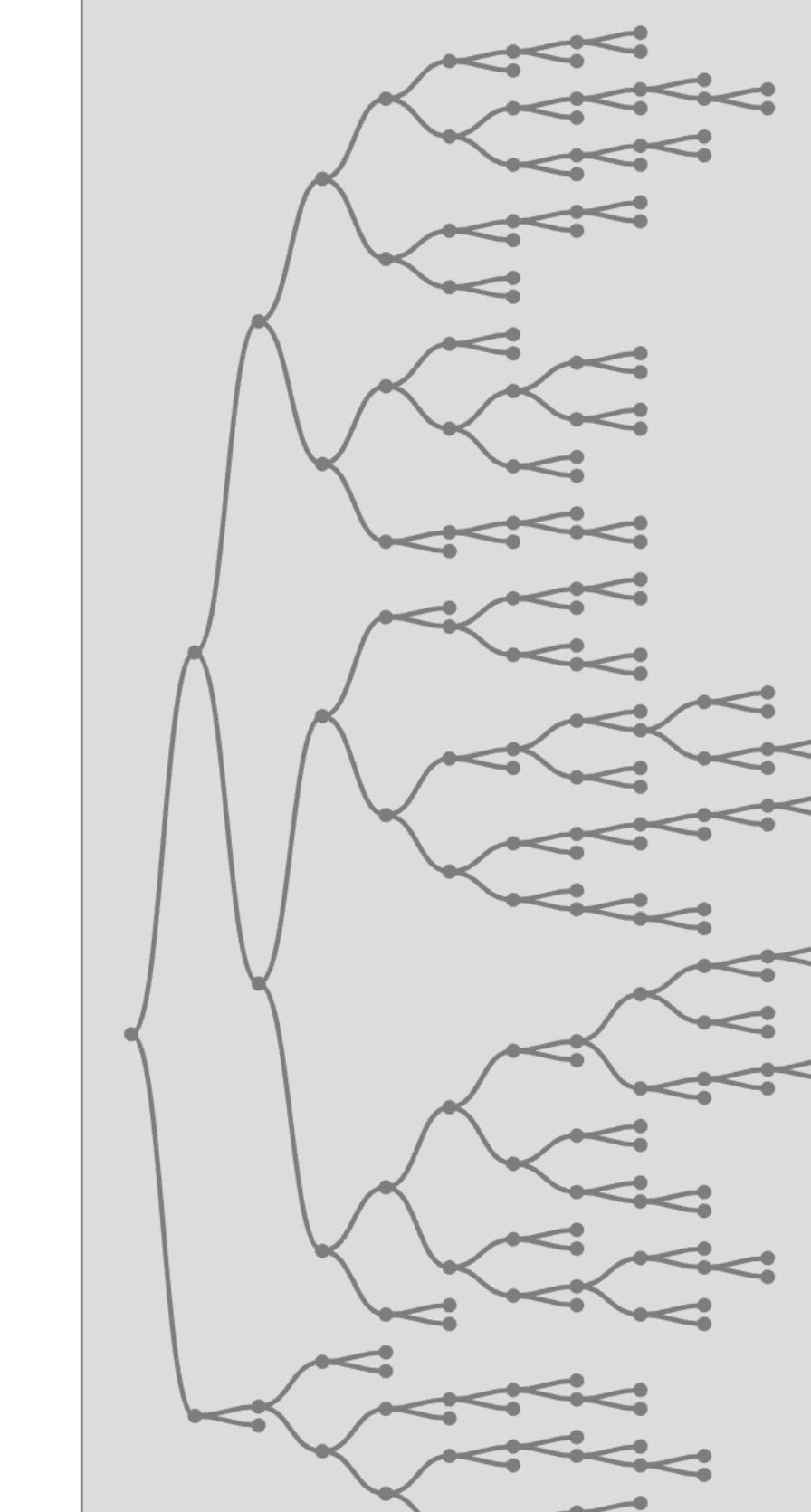
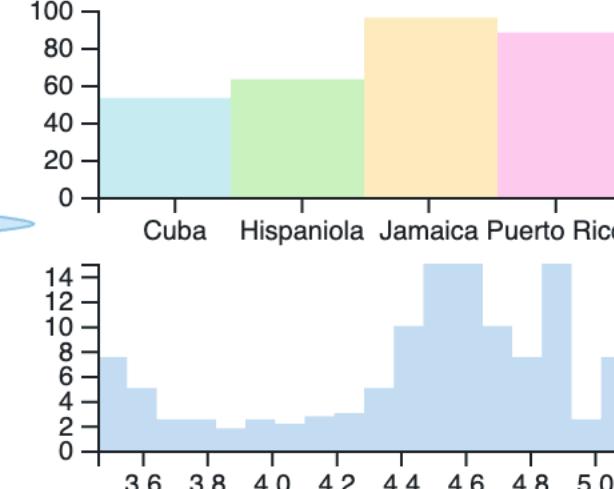
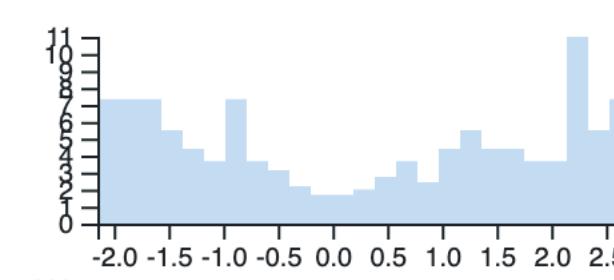
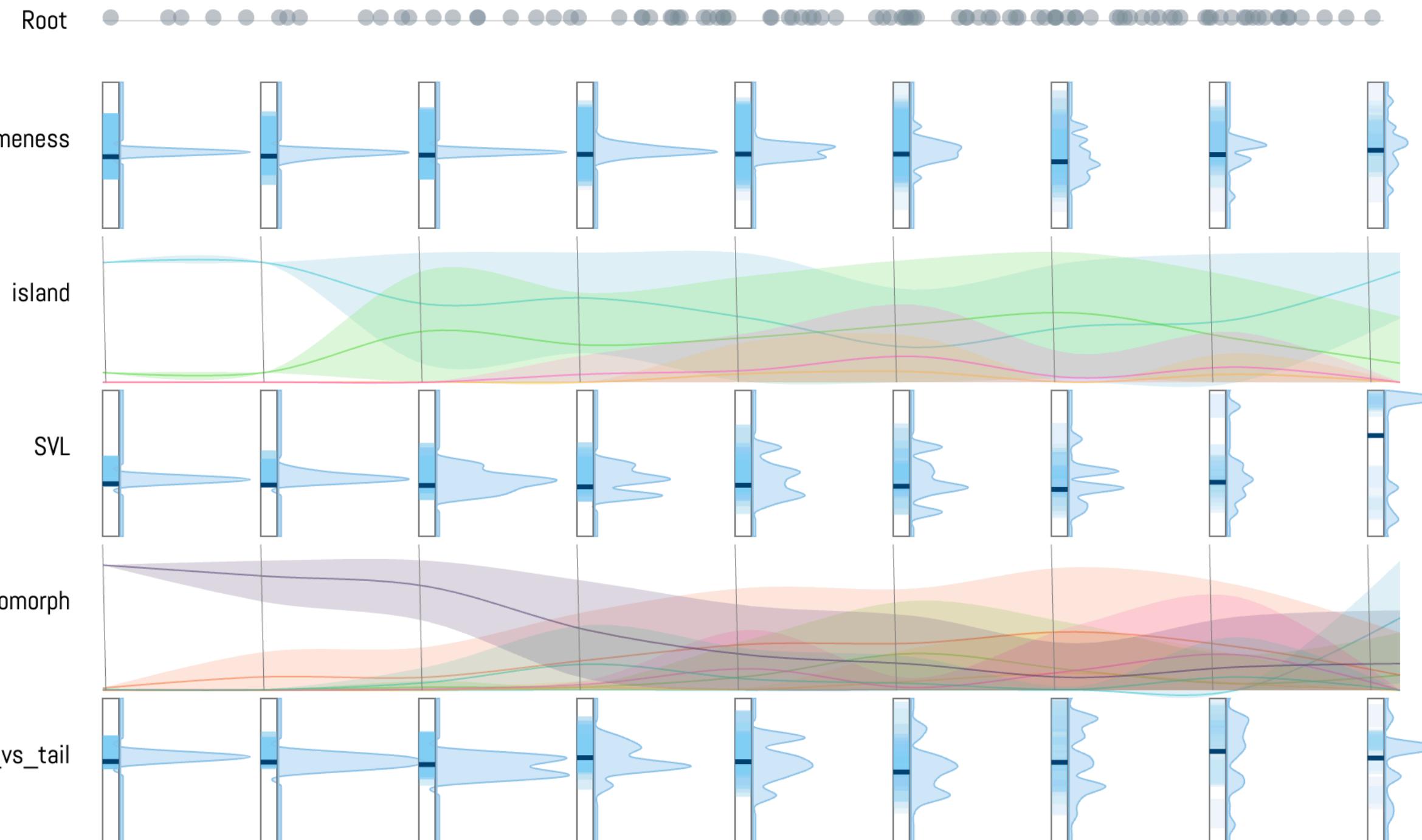
lwidth_vs_tail

Leaves

[Filter by Tree](#)

[Show Lengths](#)

[See Values](#)

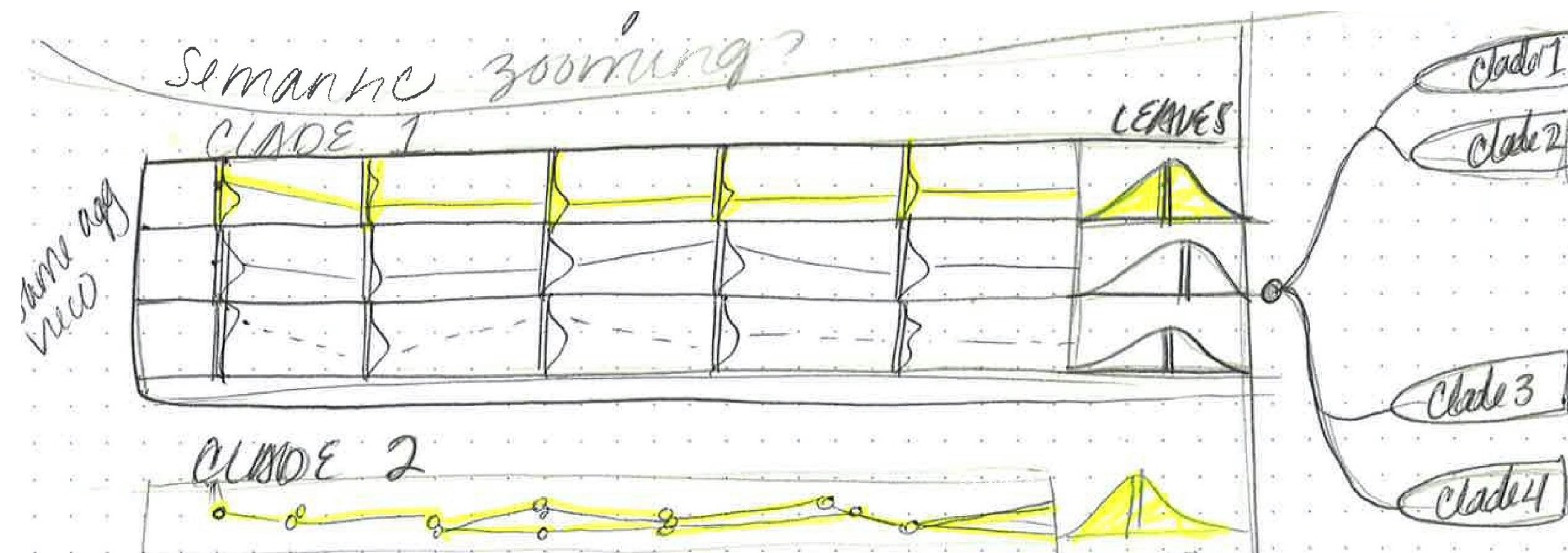




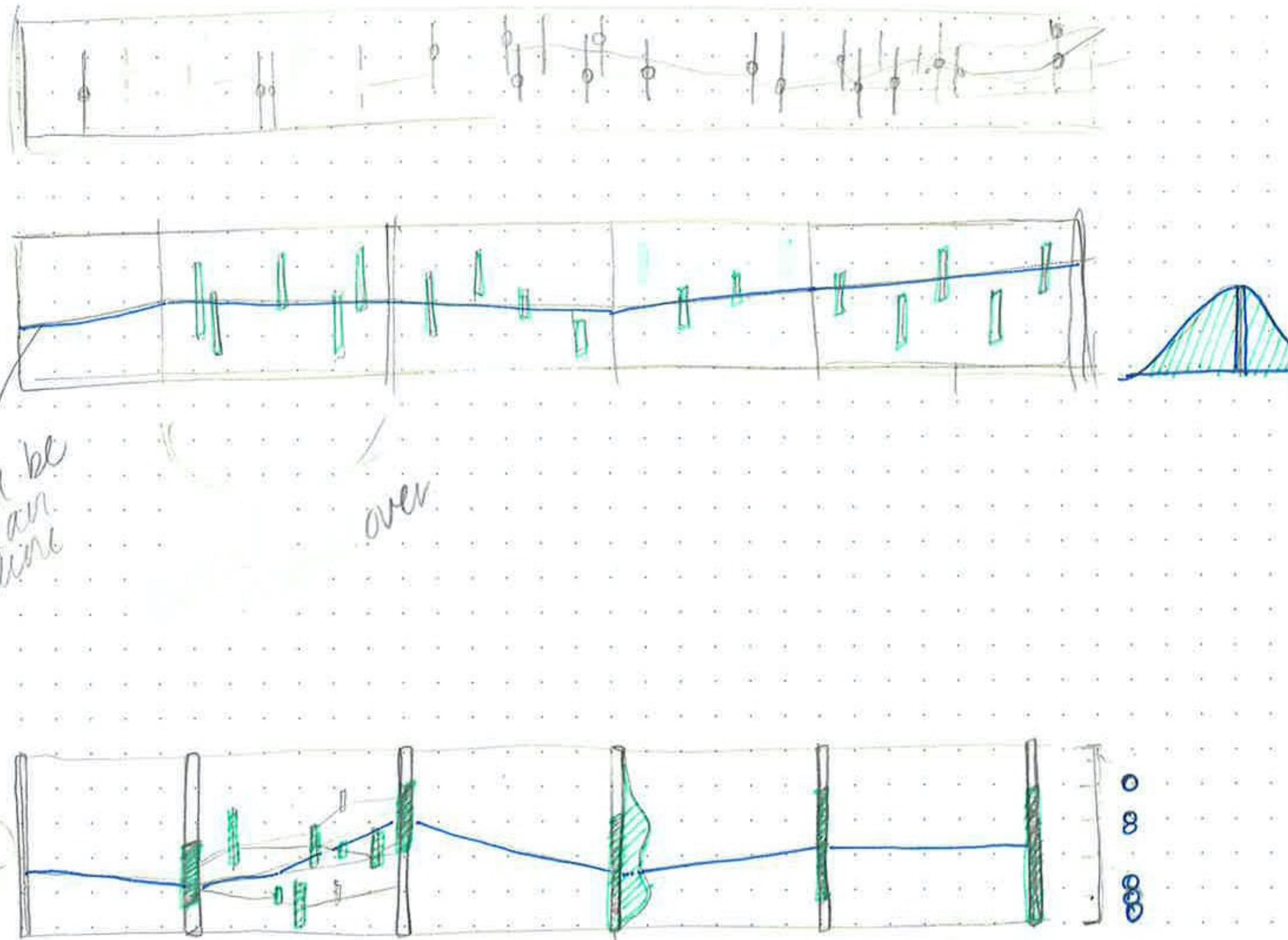
- ***What is important in the overview of the tree-
What are they looking for at that level?***
- ***How can we adjust the view at varying levels
of group size?***

Thinking about semantic zoom...

*Instead of a threshold number for the aggregated view,
The path view could be responsive to the tree topology.
If the tree is collapsed to clade, the view would show
aggregated data grouped by clade.*



What would a visualization for a mid-range number of species.



never over
bin bar
to get
distribution

Things to do next.

Incorporate larger dataset with the clades

Restructure the distribution view.

Add state shifts to the grouped view.

Design the group visualization for the initial grouping.

***I've been looking though some of the
papers that use the anoles dataset.
Can we test the tool with questions
posed in these papers ?***