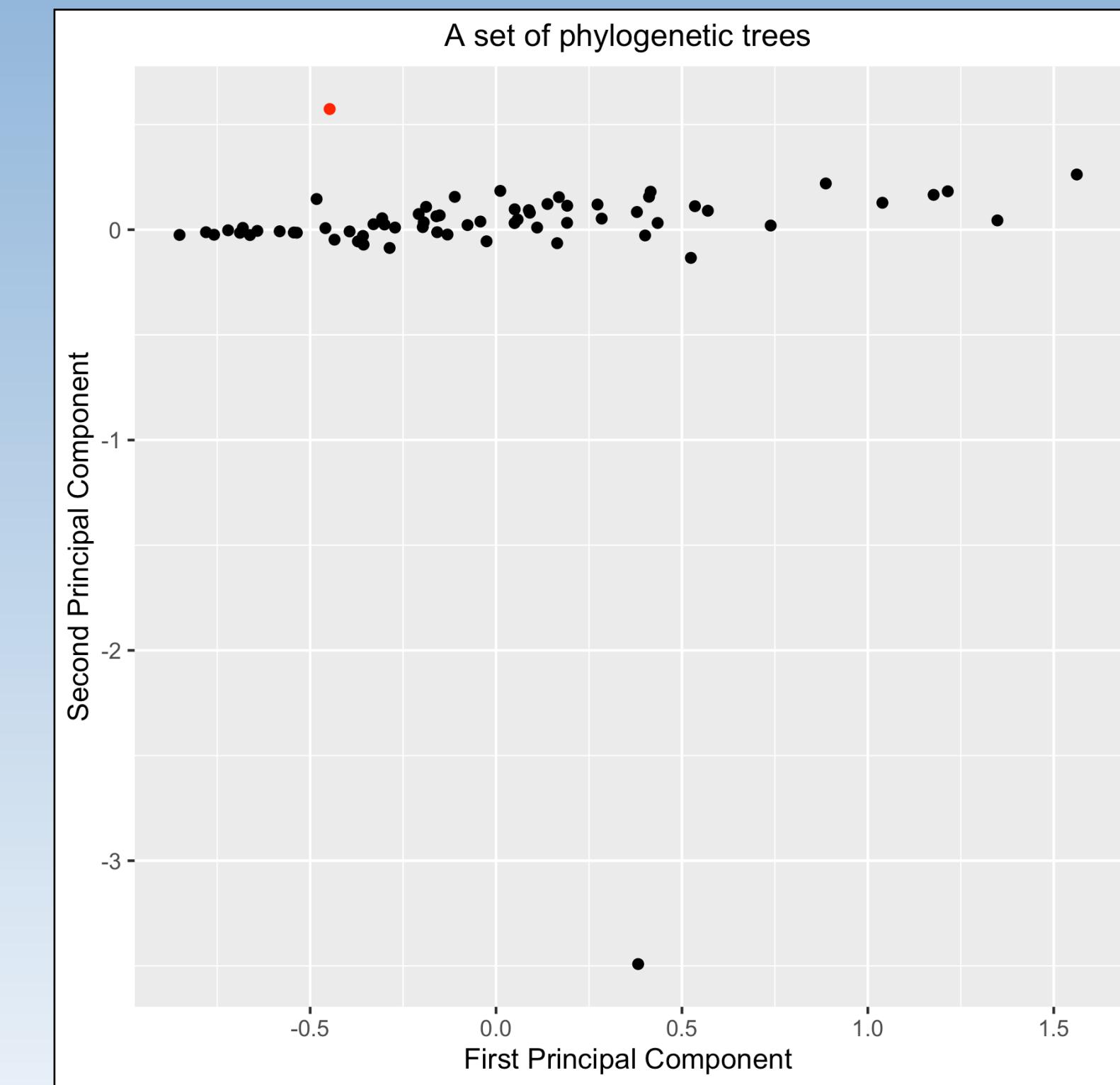
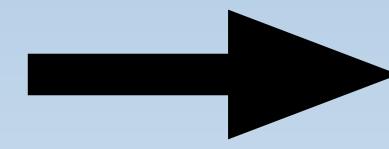
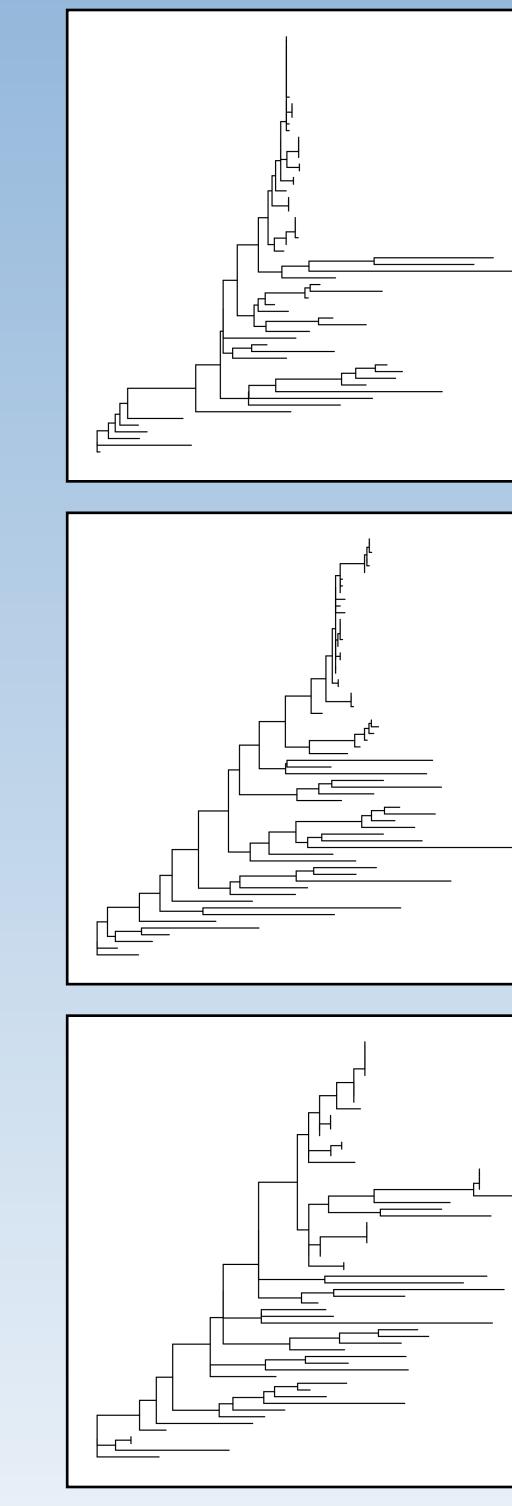
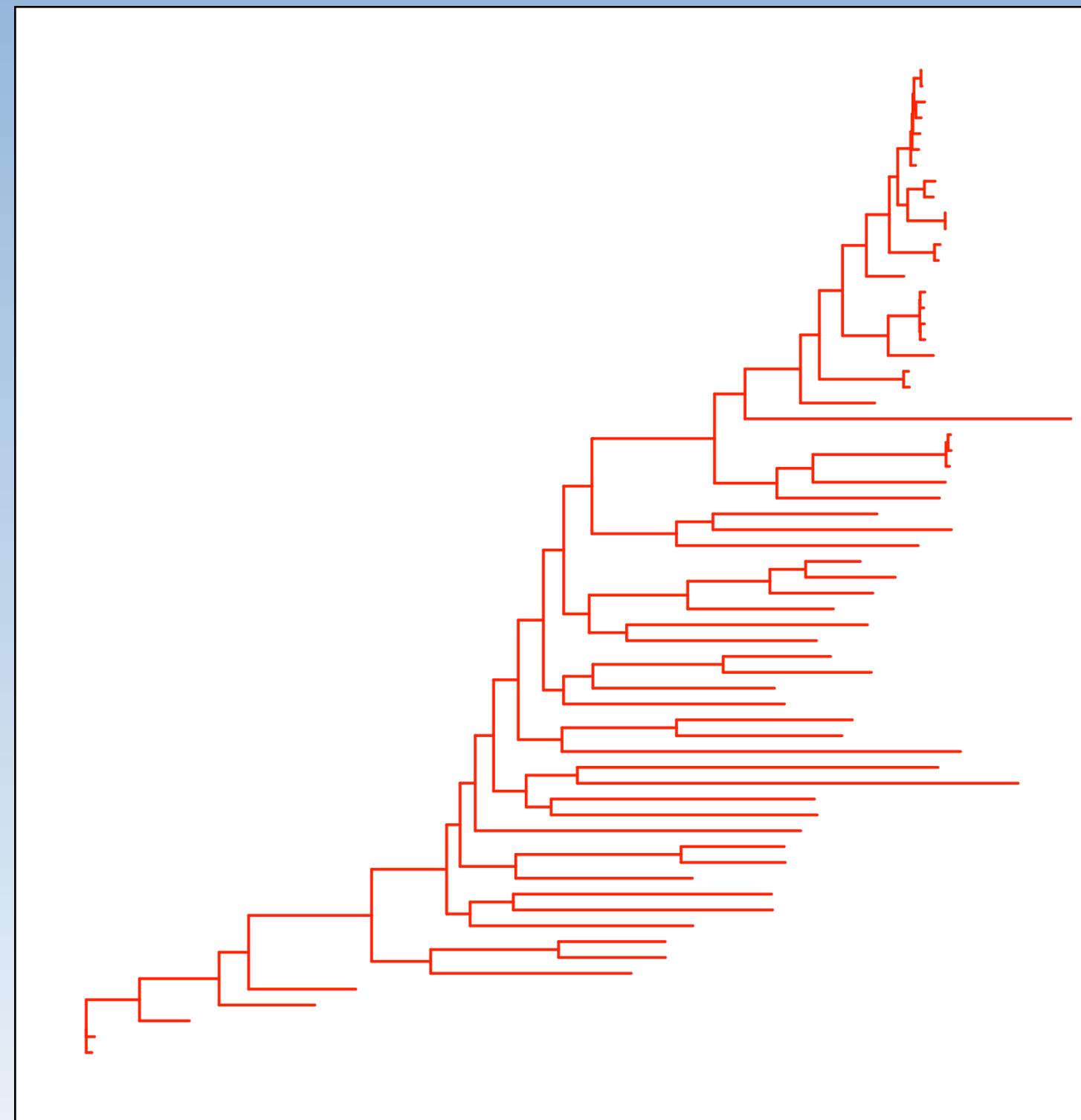


# Phylogenomics: Gene Tree Exploration

A visualization tool to investigate gene and genome level evolutionary histories



Sarah Teichman

[teichs@uw.edu](mailto:teichs@uw.edu)

@sarah\_teichman

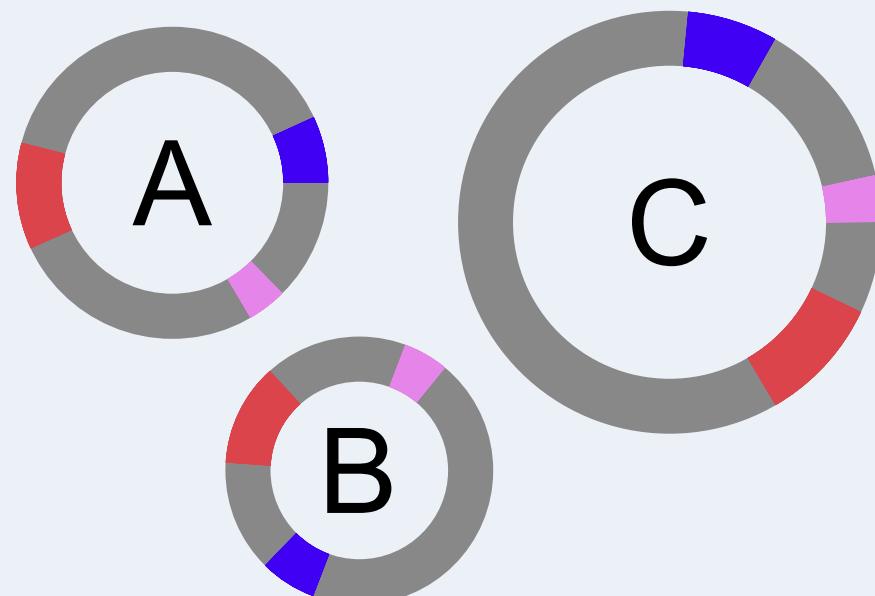
## Phylogenetics

*The practice of trying to infer evolutionary relationships between organisms based on heritable traits or characteristics*

### Phylogenetic trees

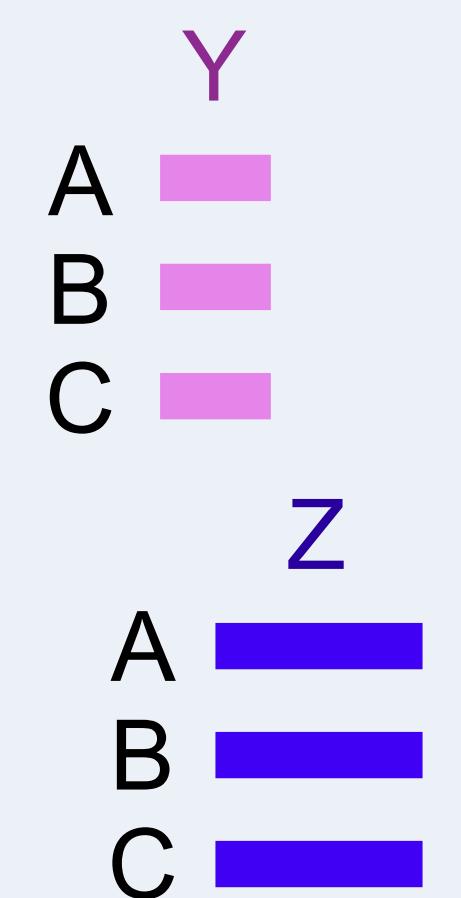
*Visual representations of hypotheses about evolutionary relationships*

1. Identify target gene-types in genomes of interest

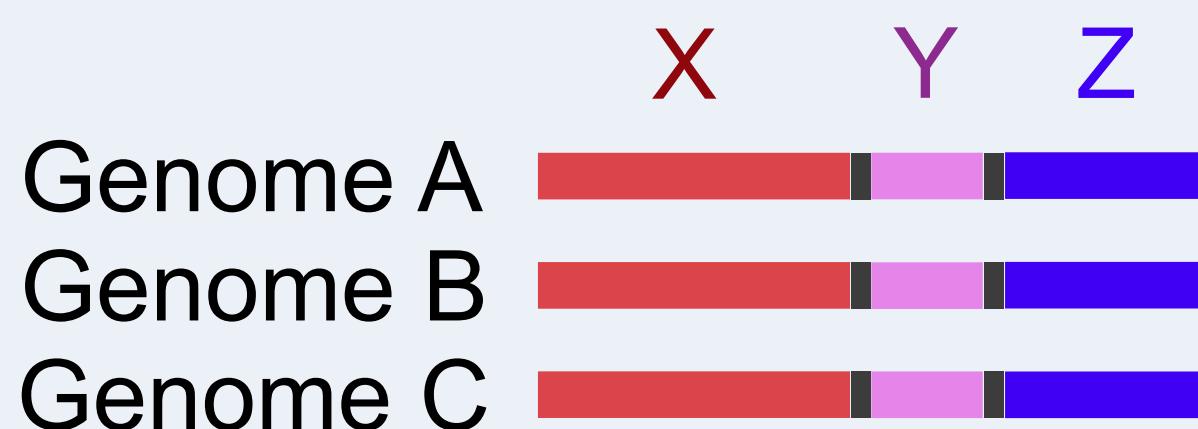


2. Align individual target gene-sets

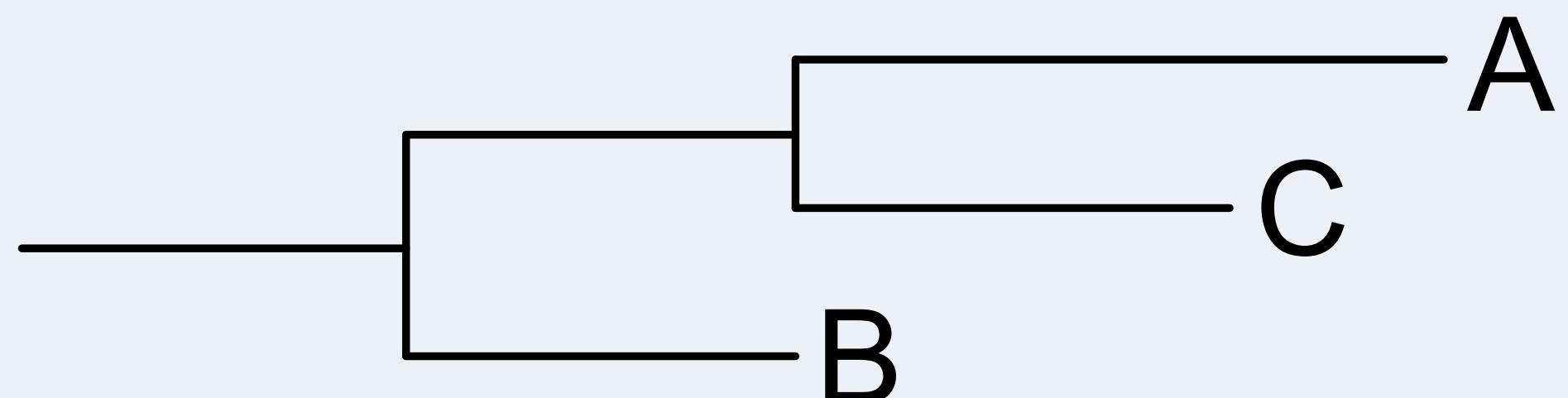
	Gene X
Genome A	—
Genome B	—
Genome C	—



3. Stick alignments together

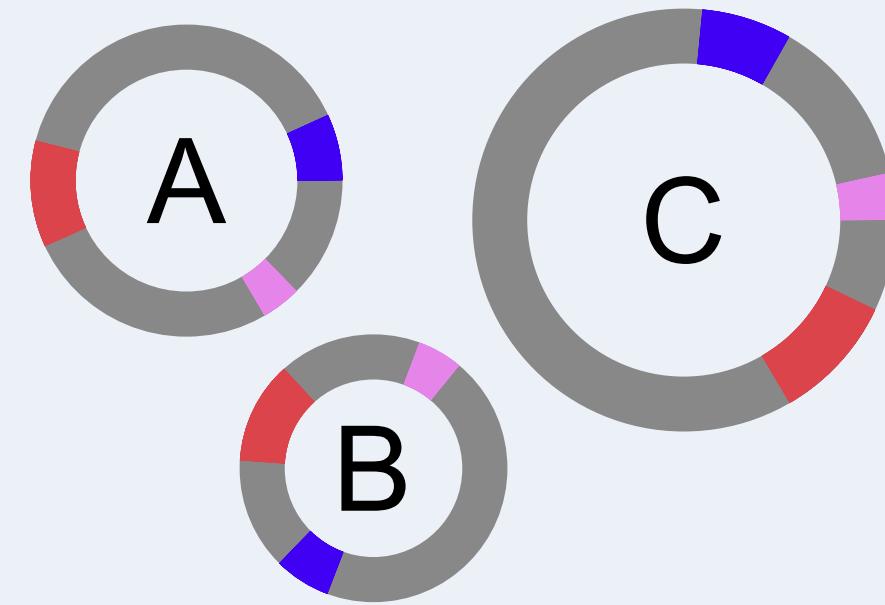


4. Infer evolutionary relationships



# Phylogenomics: Gene Tree Exploration

1. Identify target gene-types  
in genomes of interest



2. Align individual  
target gene-sets

Gene X

Genome A	—
Genome B	—
Genome C	—

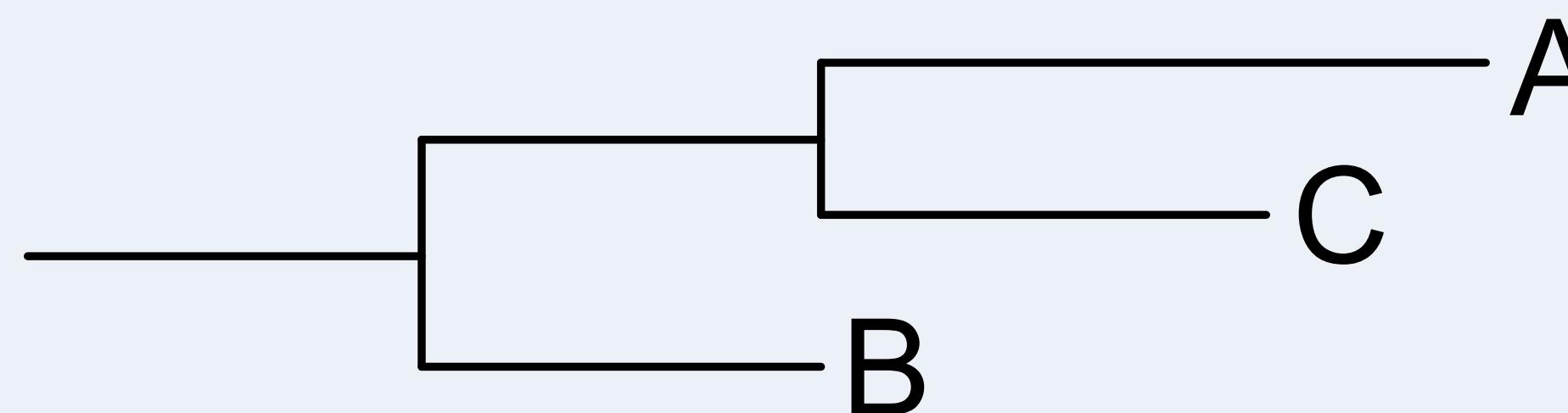
Y

A	—
B	—
C	—

Z

A	—
B	—
C	—

4. Infer evolutionary relationships



3. Stick alignments together

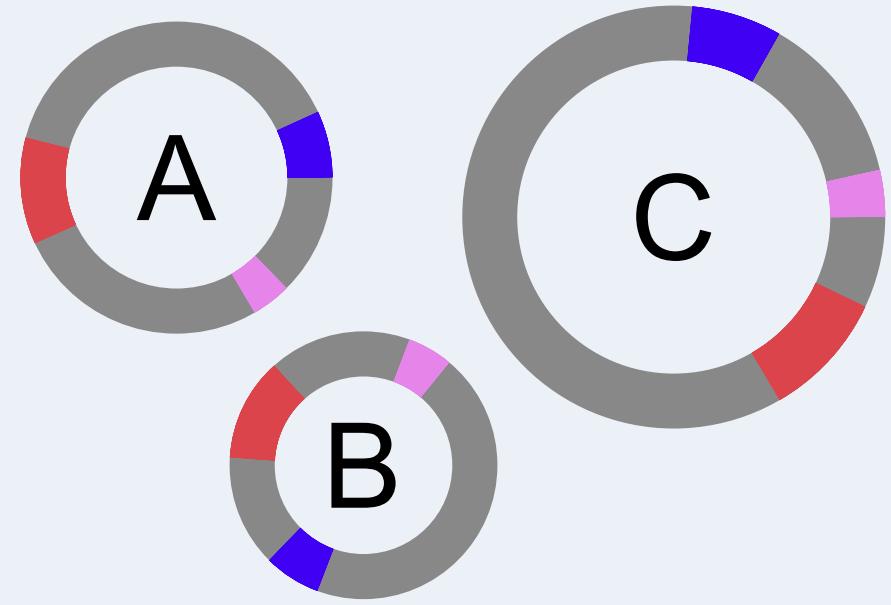
X Y Z

Genome A	—	X	—	Y	—	Z
Genome B	—	X	—	Y	—	Z
Genome C	—	X	—	Y	—	Z

What if along with  
concatenating,  
we estimate all  
of our gene-level  
phylogenies?

# Phylogenomics: Gene Tree Exploration

1. Identify target gene-types  
in genomes of interest



2. Align individual  
target gene-sets

Gene X

Genome A	—
Genome B	—
Genome C	—

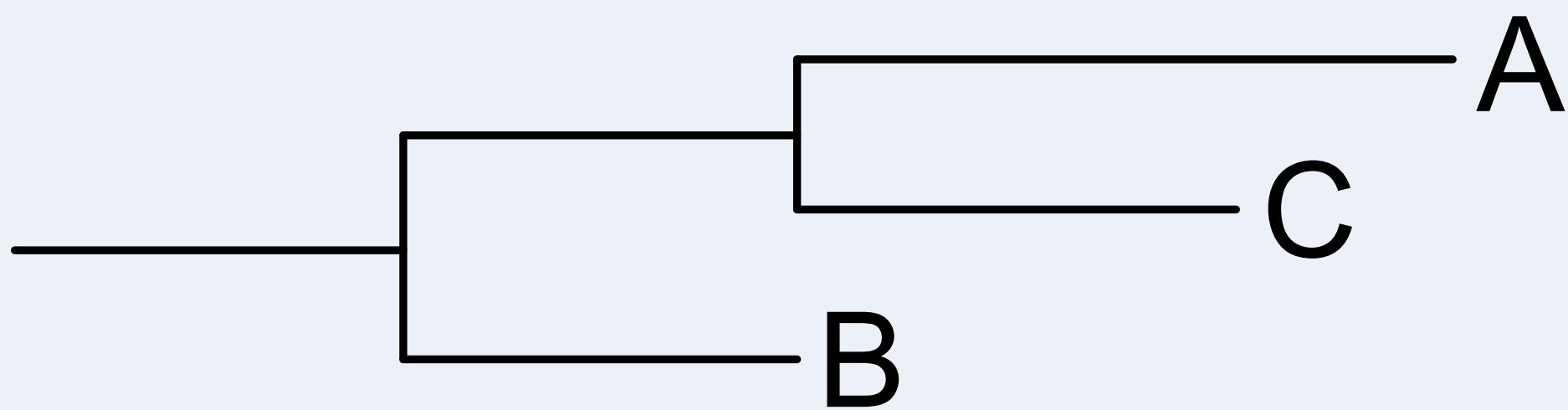
Y

A	—
B	—
C	—

Z

A	—
B	—
C	—

4. Infer evolutionary relationships



3. Stick alignments together

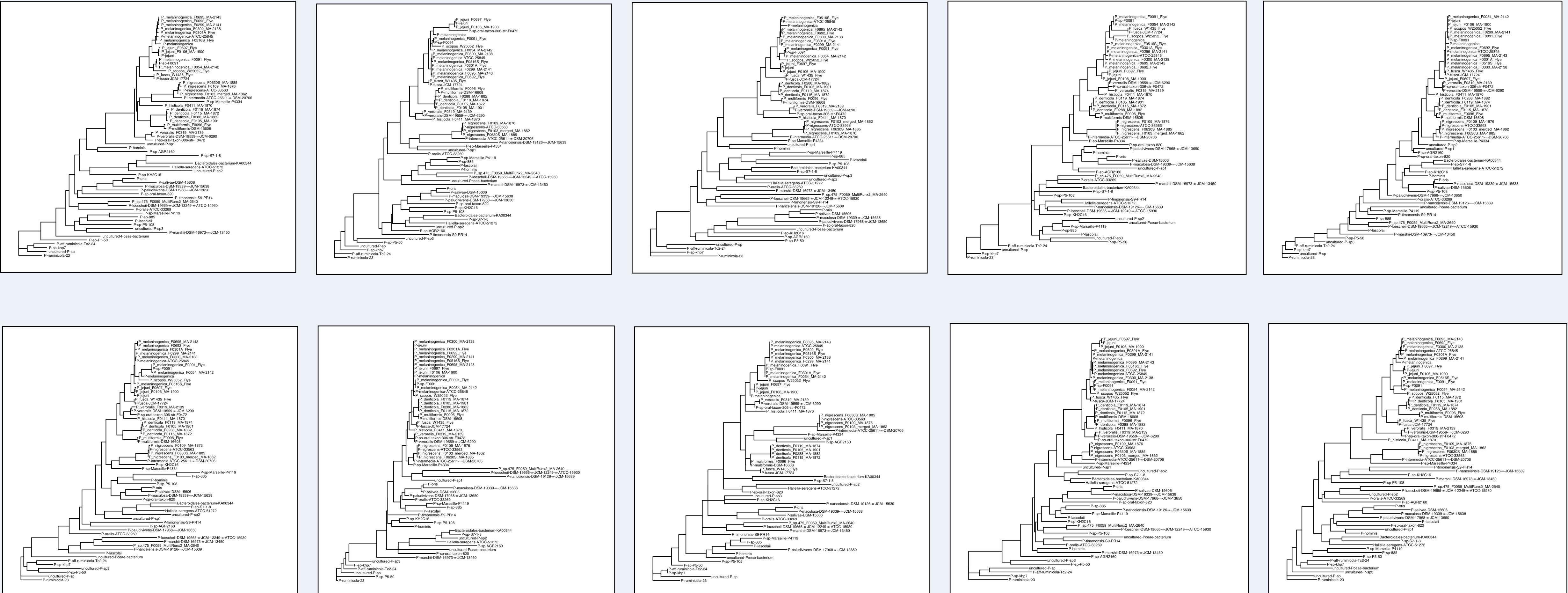
X Y Z

Genome A	—	—	—
Genome B	—	—	—
Genome C	—	—	—

What if along with  
concatenating,  
we estimate all  
of our gene-level  
phylogenies?

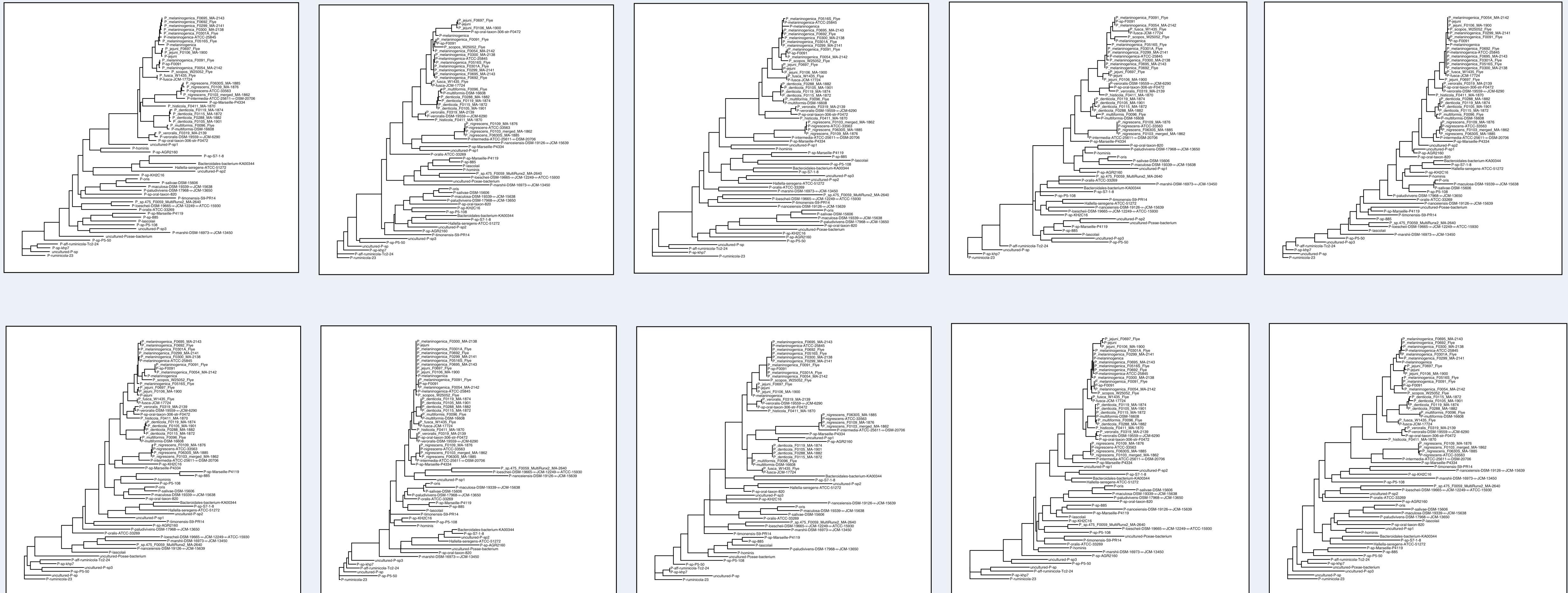
Why?  
Because we might  
learn something new  
from our data

# Phylogenomics: Gene Tree Exploration



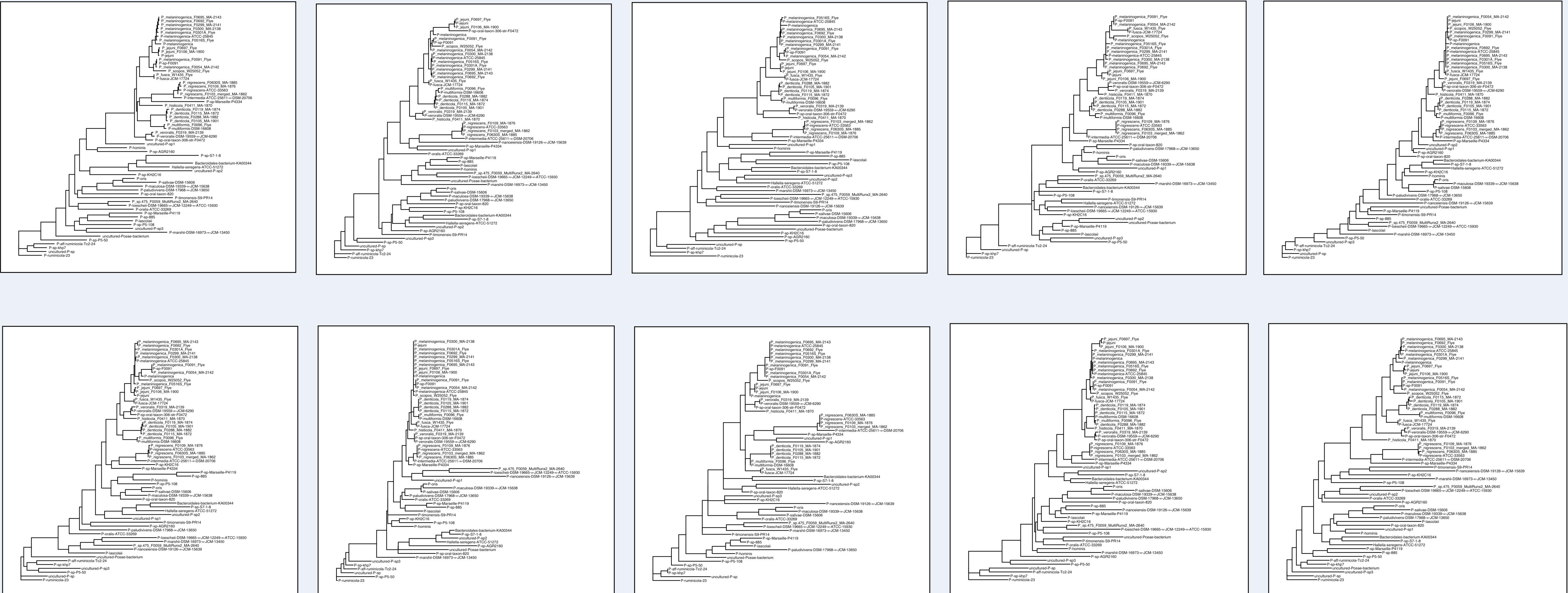
# Phylogenomics: Gene Tree Exploration

# What can we learn from these plots?



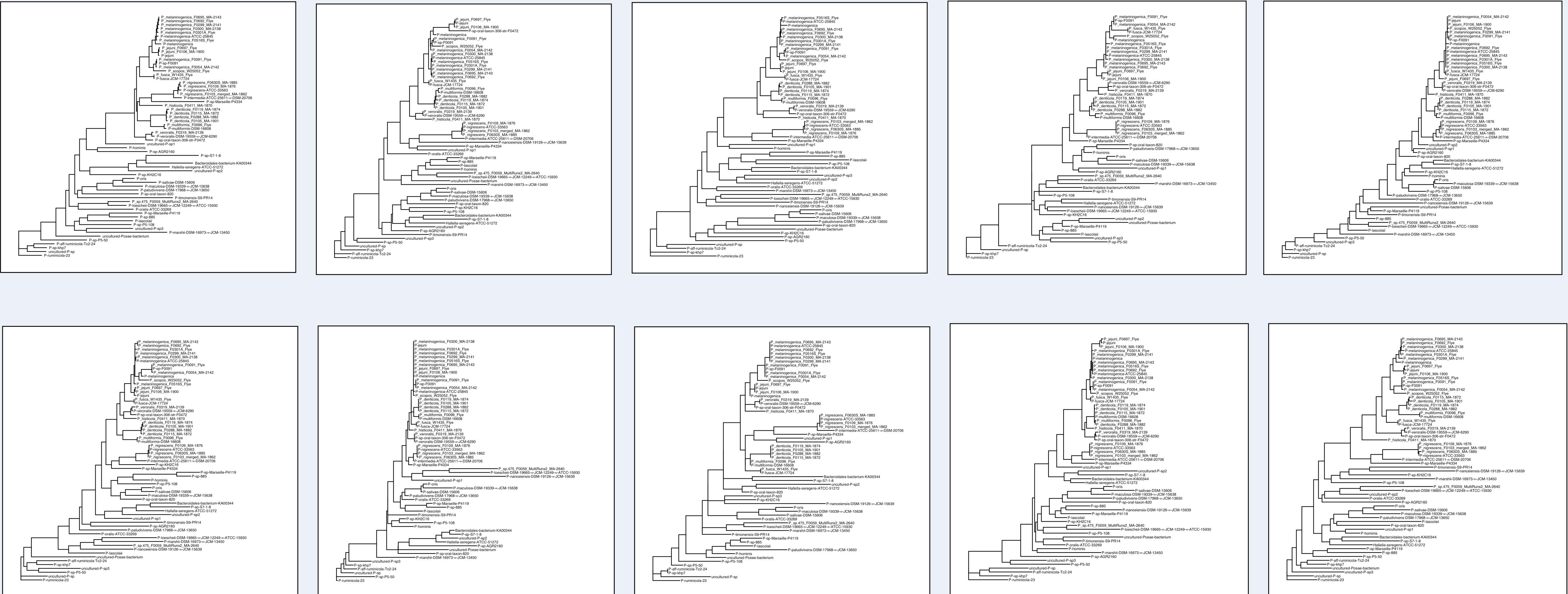
# Phylogenomics: Gene Tree Exploration

What can't we learn from these plots?



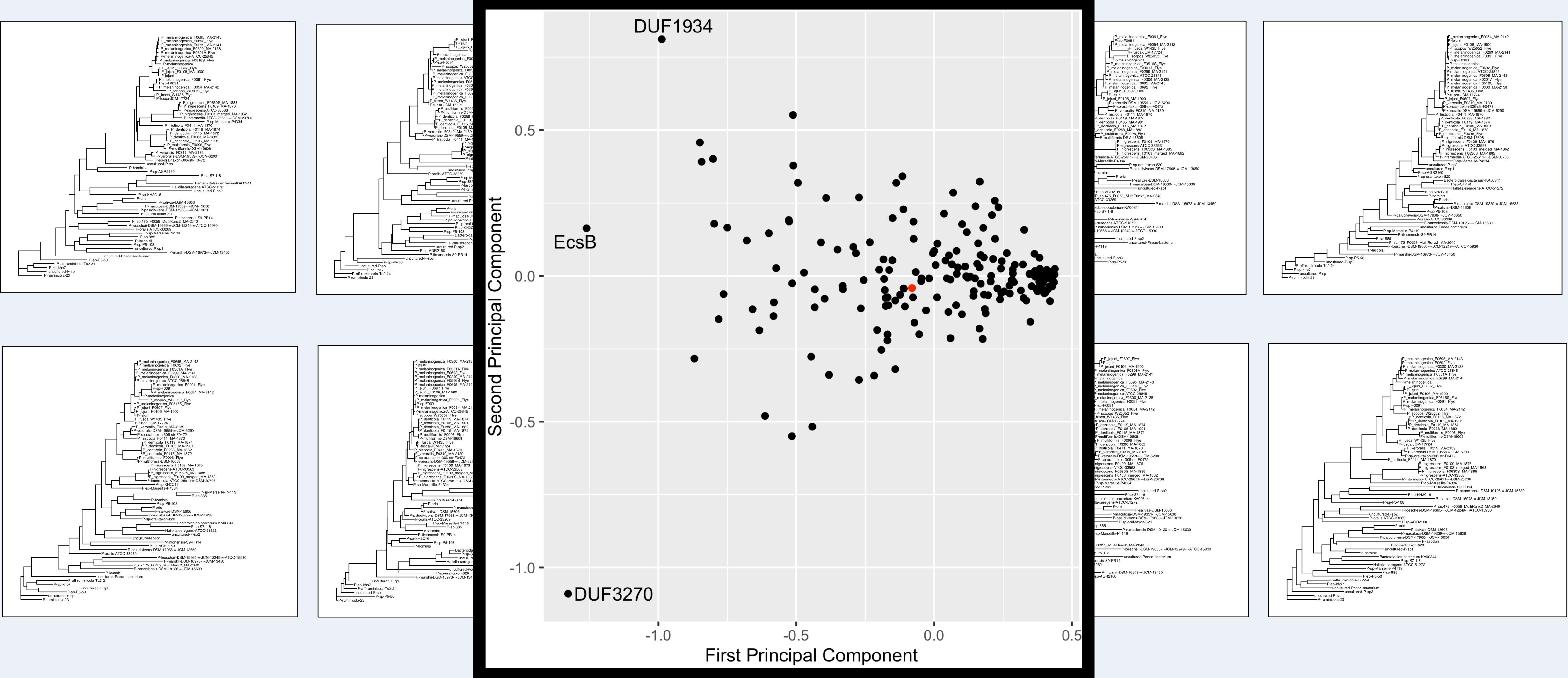
# Phylogenomics: Gene Tree Exploration

How can we reduce this data to compare gene trees?



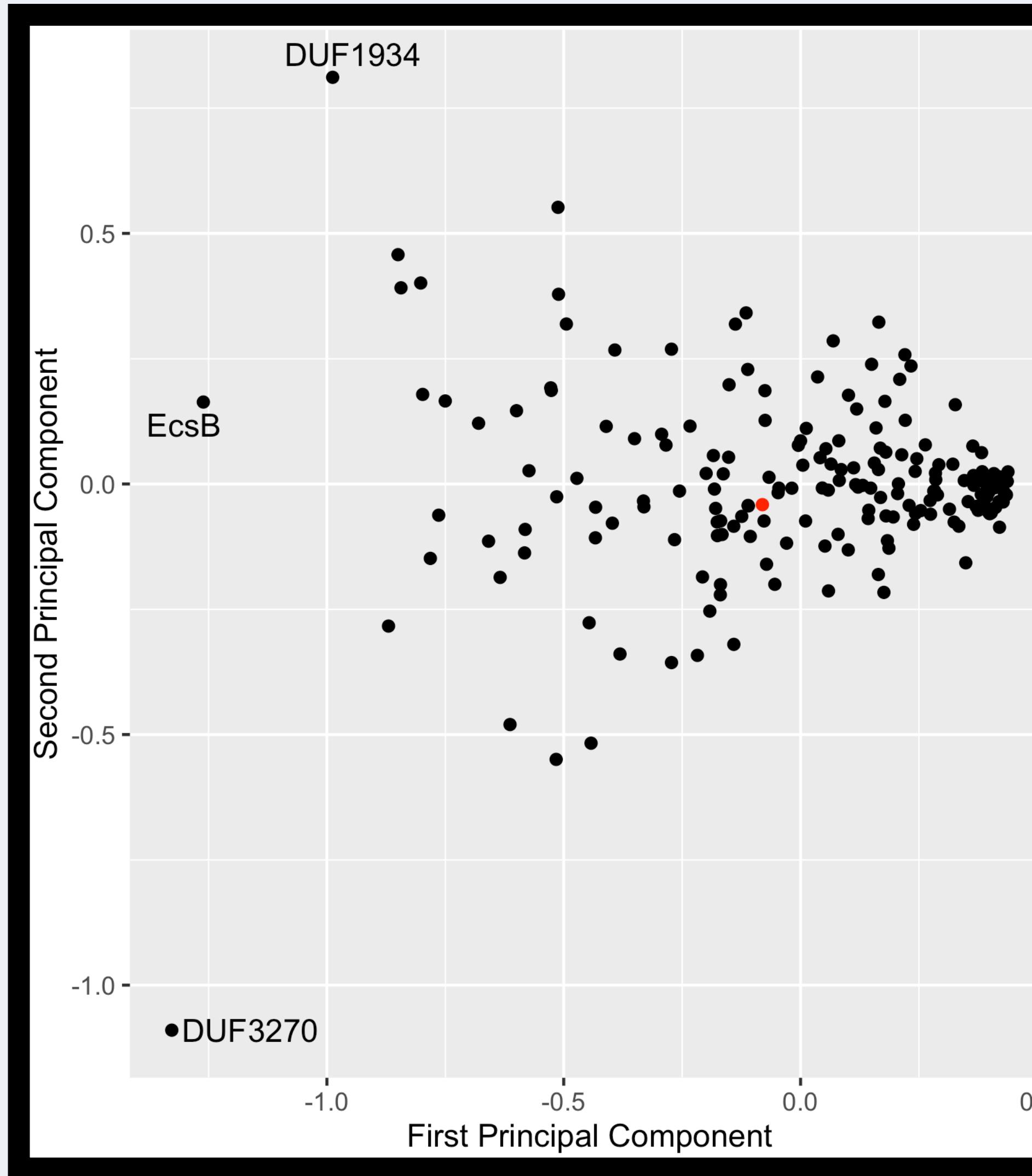
# Phylogenomics: Gene Tree Exploration

How can we reduce this data to compare gene trees?

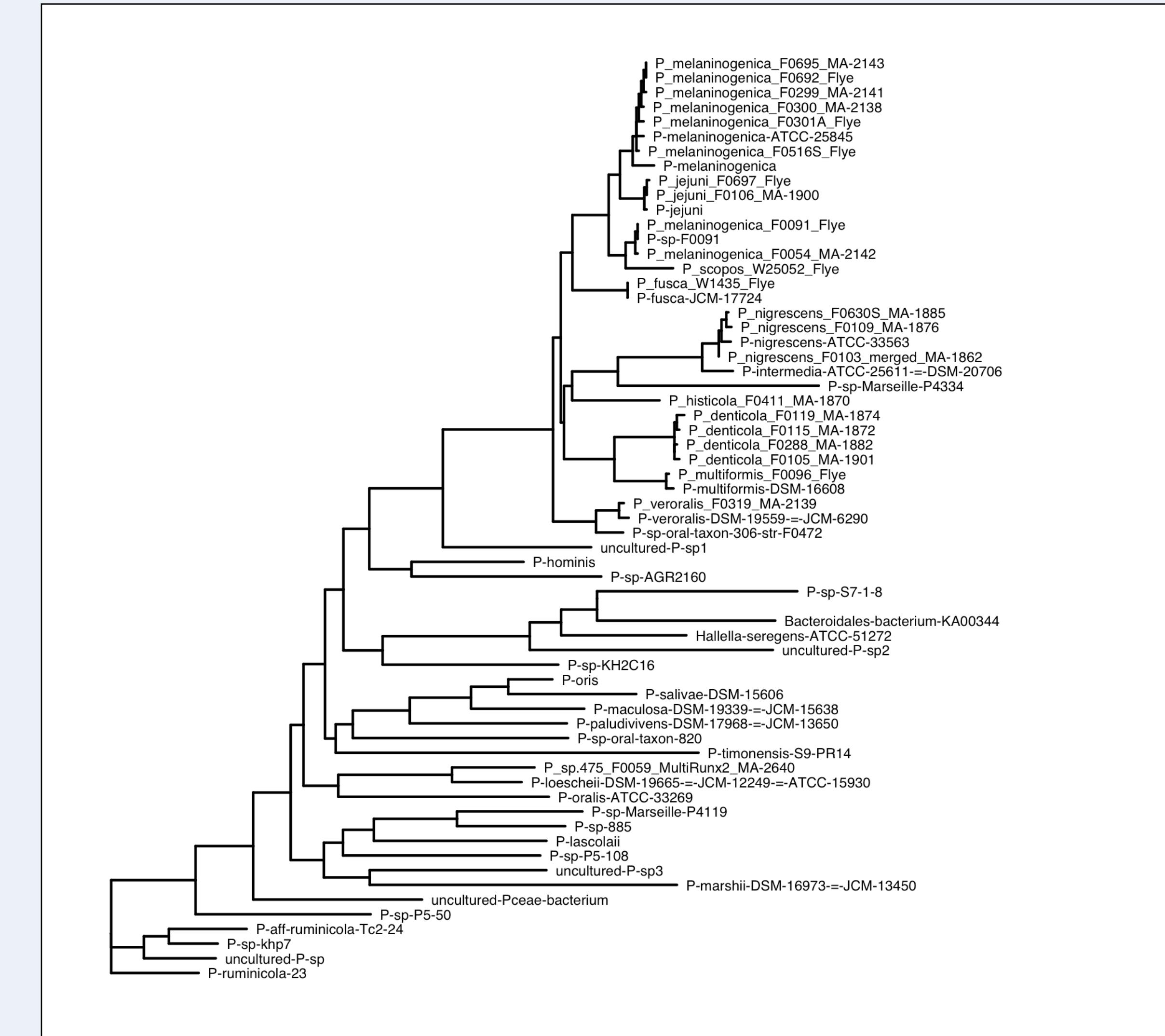


# Phylogenomics: Gene Tree Exploration

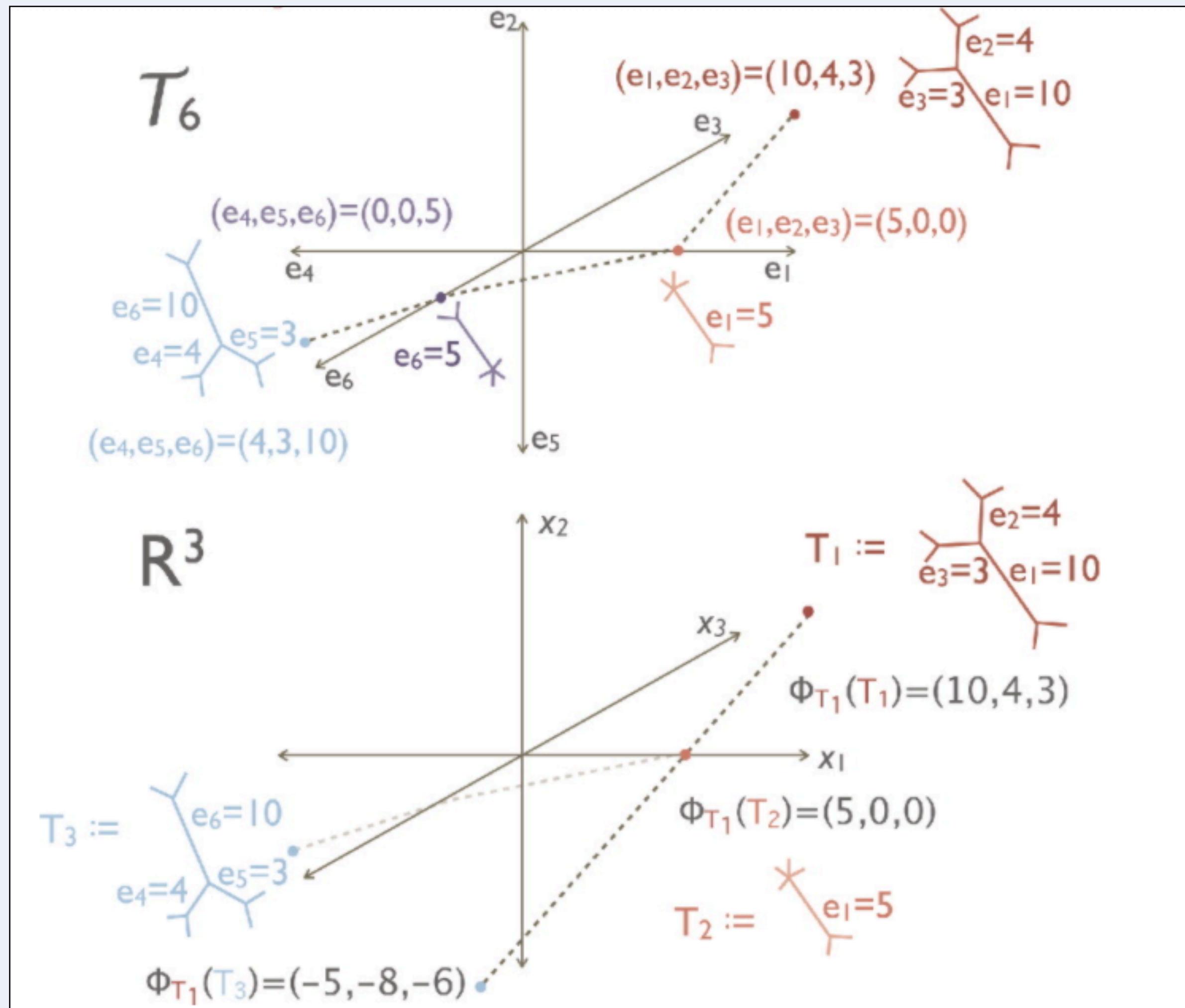
Why is this hard?



A tree is more complex than a vector!



## Log map transformation: map object from BHV tree space to Euclidean space



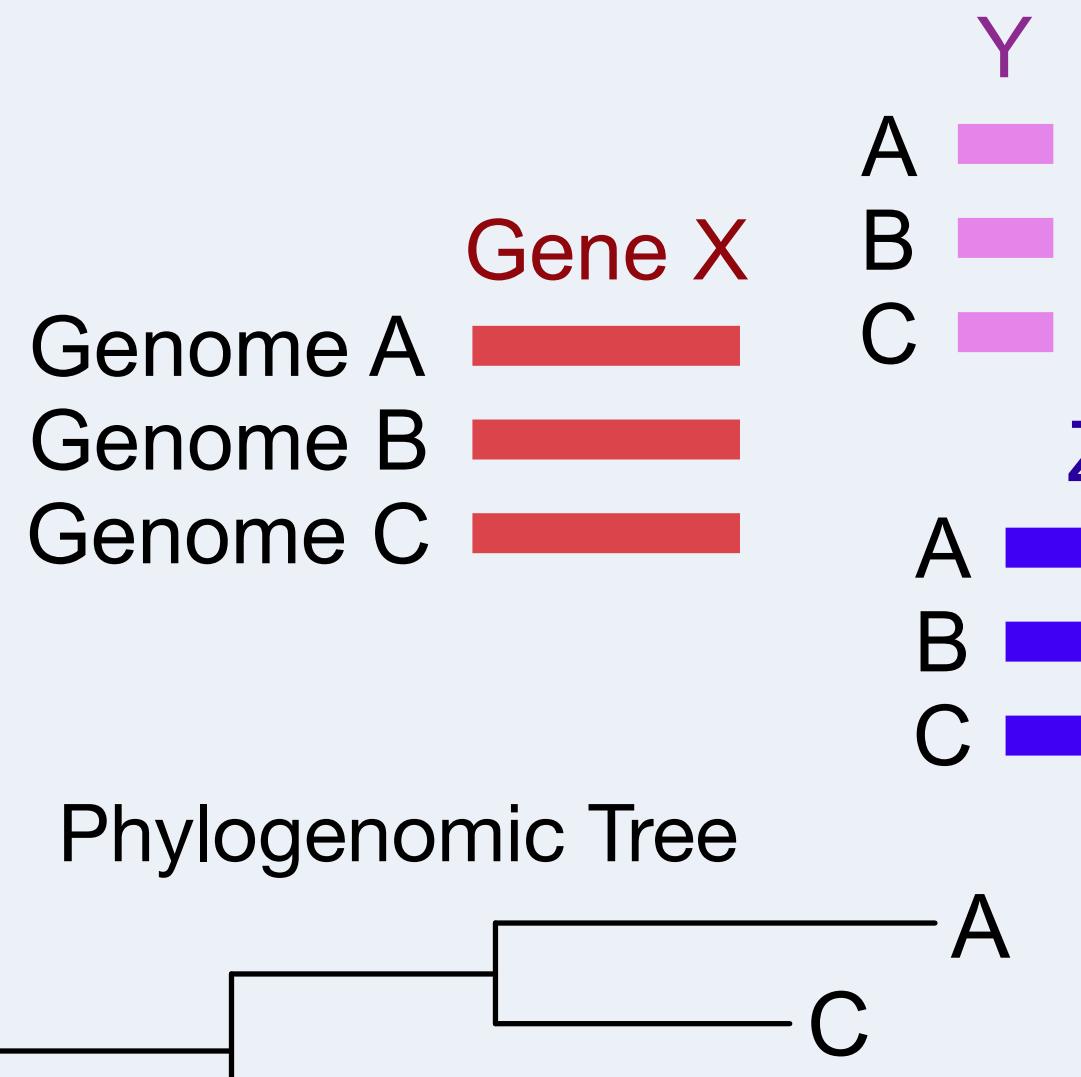
**BHV space:**  
Constructed in  
Billera et al. (2001)

**Log map:**  
Introduced in  
Barden et al. (2018)

Figure: Willis and Bell (2018)

## GROVE: Gene tRee Ordination Visualization for Exploration

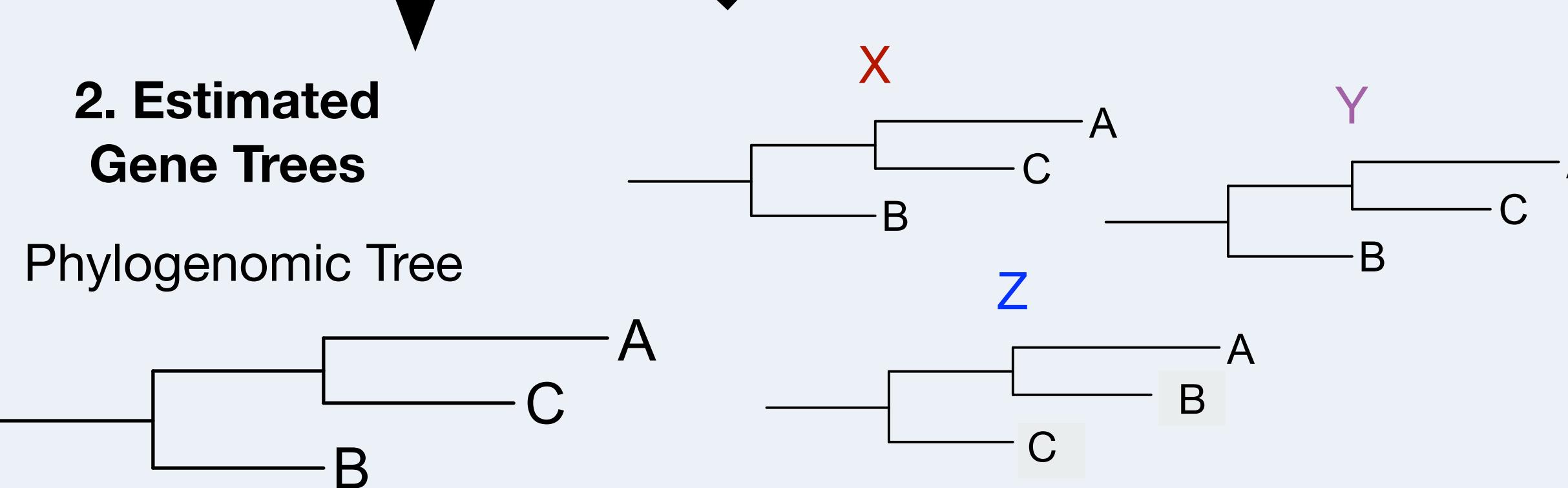
### 1. GToTree Output



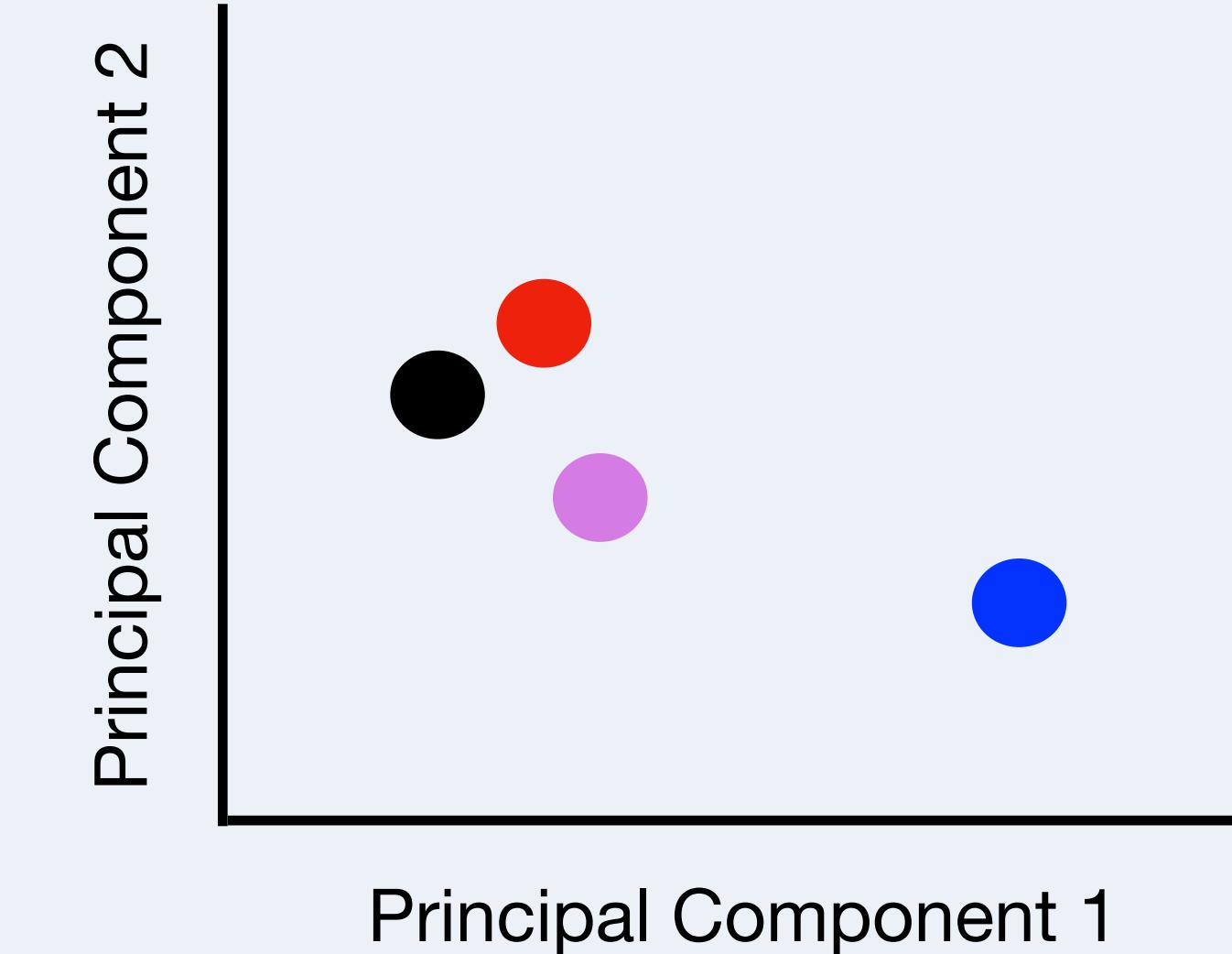
### 3. Vectors in $\mathbb{R}^5$ Using the Log Map

	V1	V2	V3	V4	V5
X		10	4	3	5
Y		2	3	1	2
Z		-4	-8	2	3
Phylogenomic	5	8	2	1	4

### 2. Estimated Gene Trees



### 4. Ordination Plot

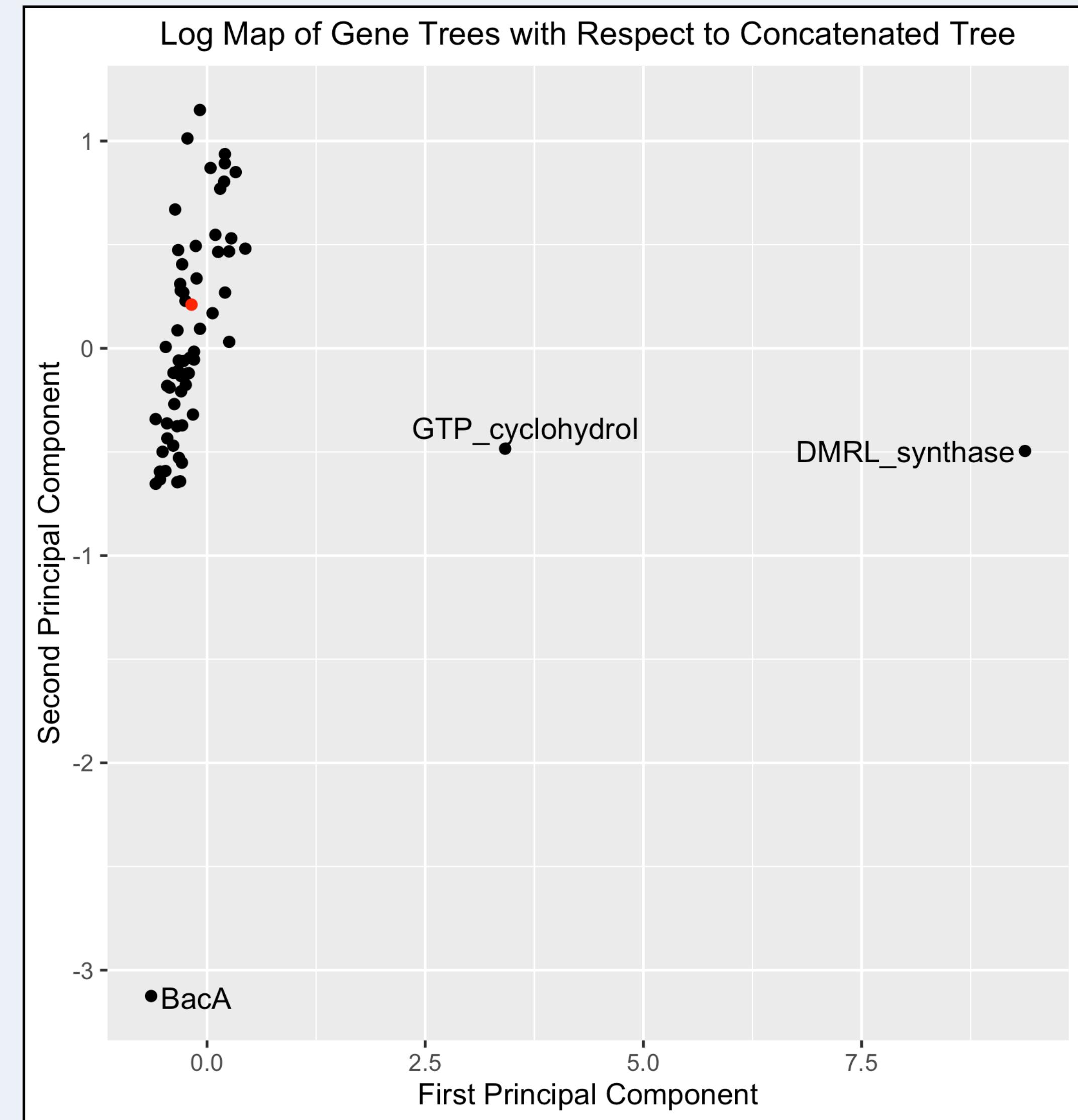


R package  
in development

# Phylogenomics: Gene Tree Exploration

An Analysis of  
Prevotella Genomes

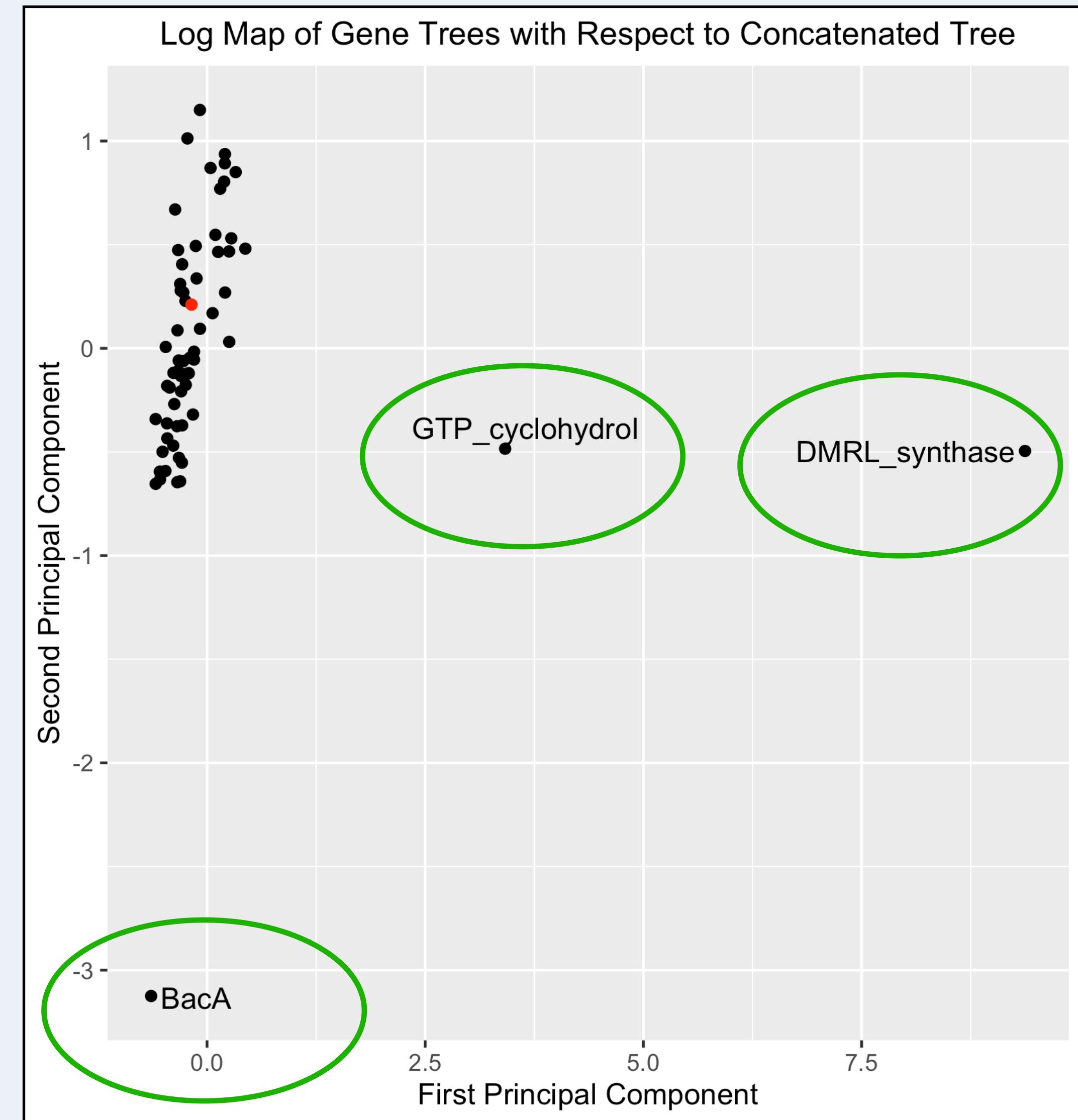
What might you find?



# Phylogenomics: Gene Tree Exploration

An Analysis of  
Prevotella Genomes

What might you find?

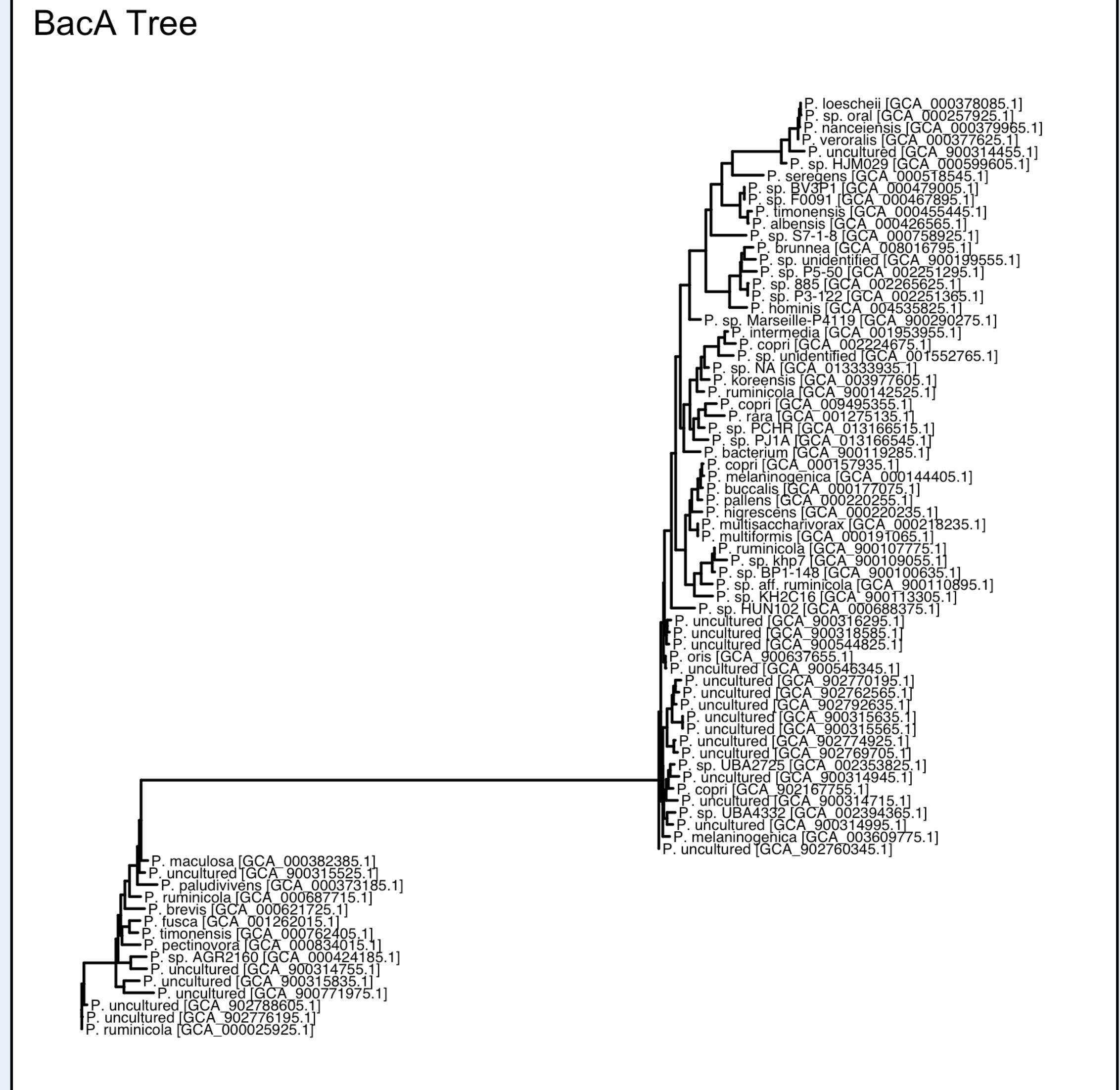
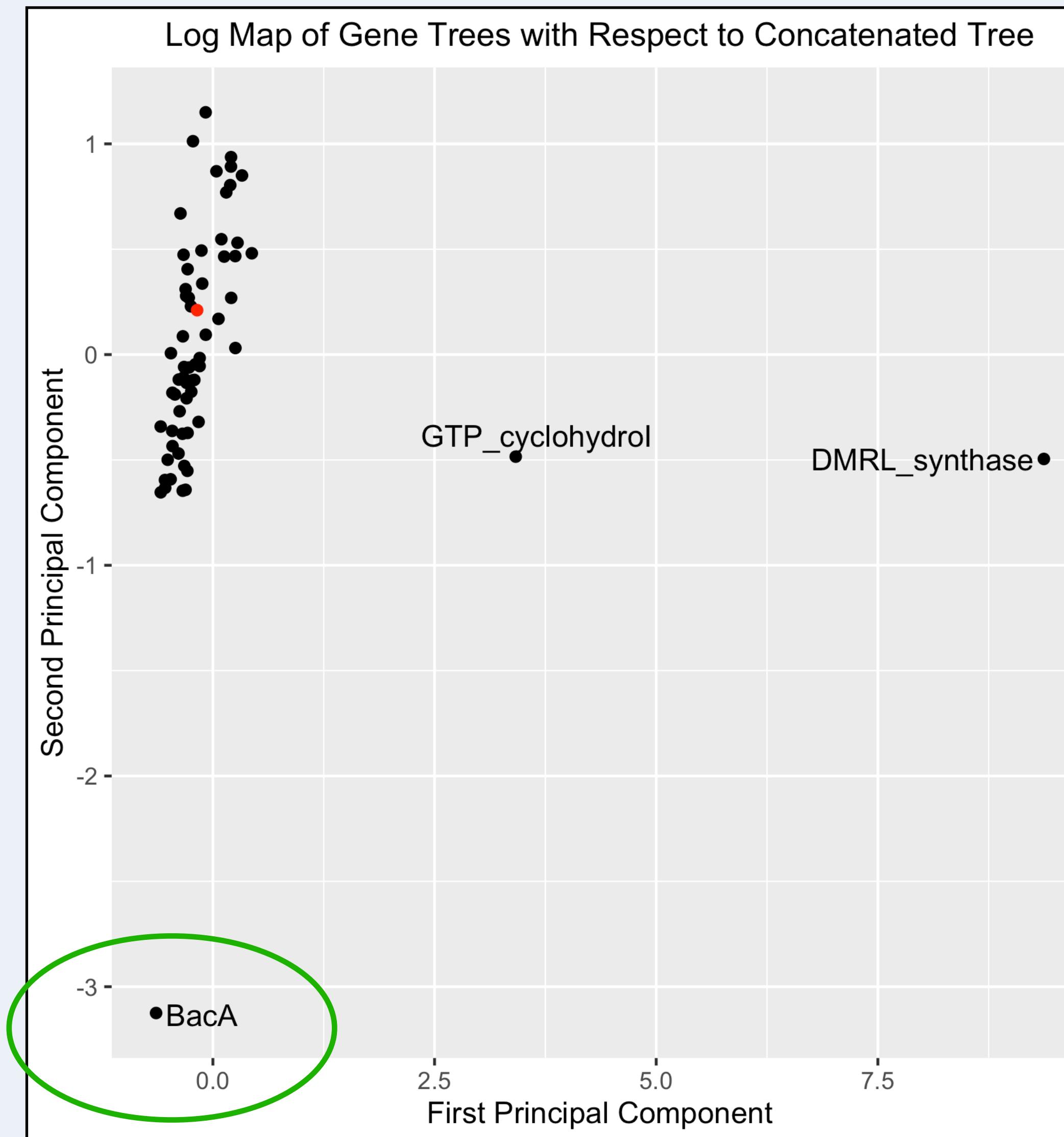


We have some  
outlying gene trees!

# Phylogenomics: Gene Tree Exploration

## An Analysis of Prevotella Genomes

What might you find?

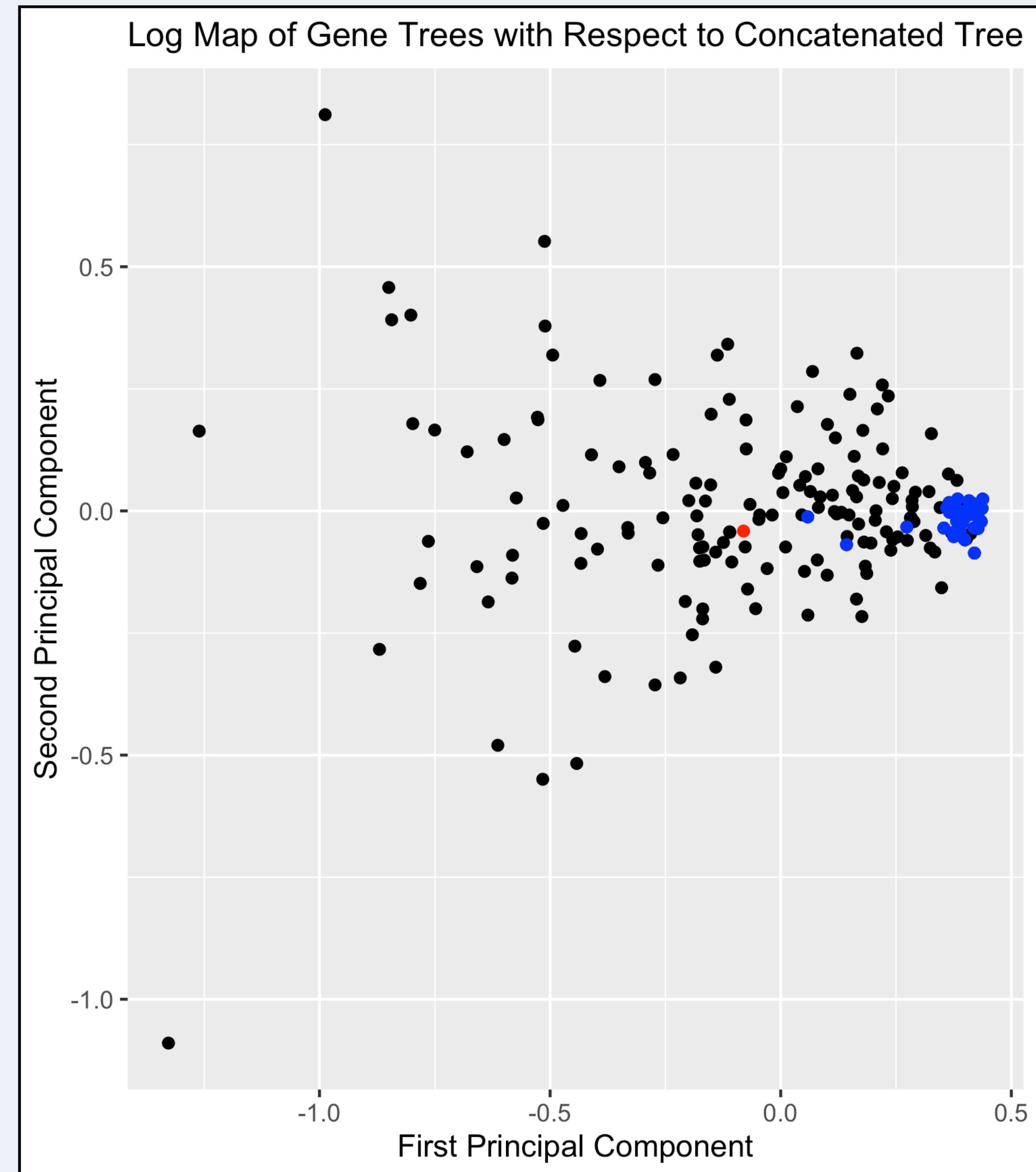


# Phylogenomics: Gene Tree Exploration

An Analysis of  
Streptococcus Genomes

What might you find?

Ribosomal genes  
shown in blue



The ribosomal genes  
cluster together!

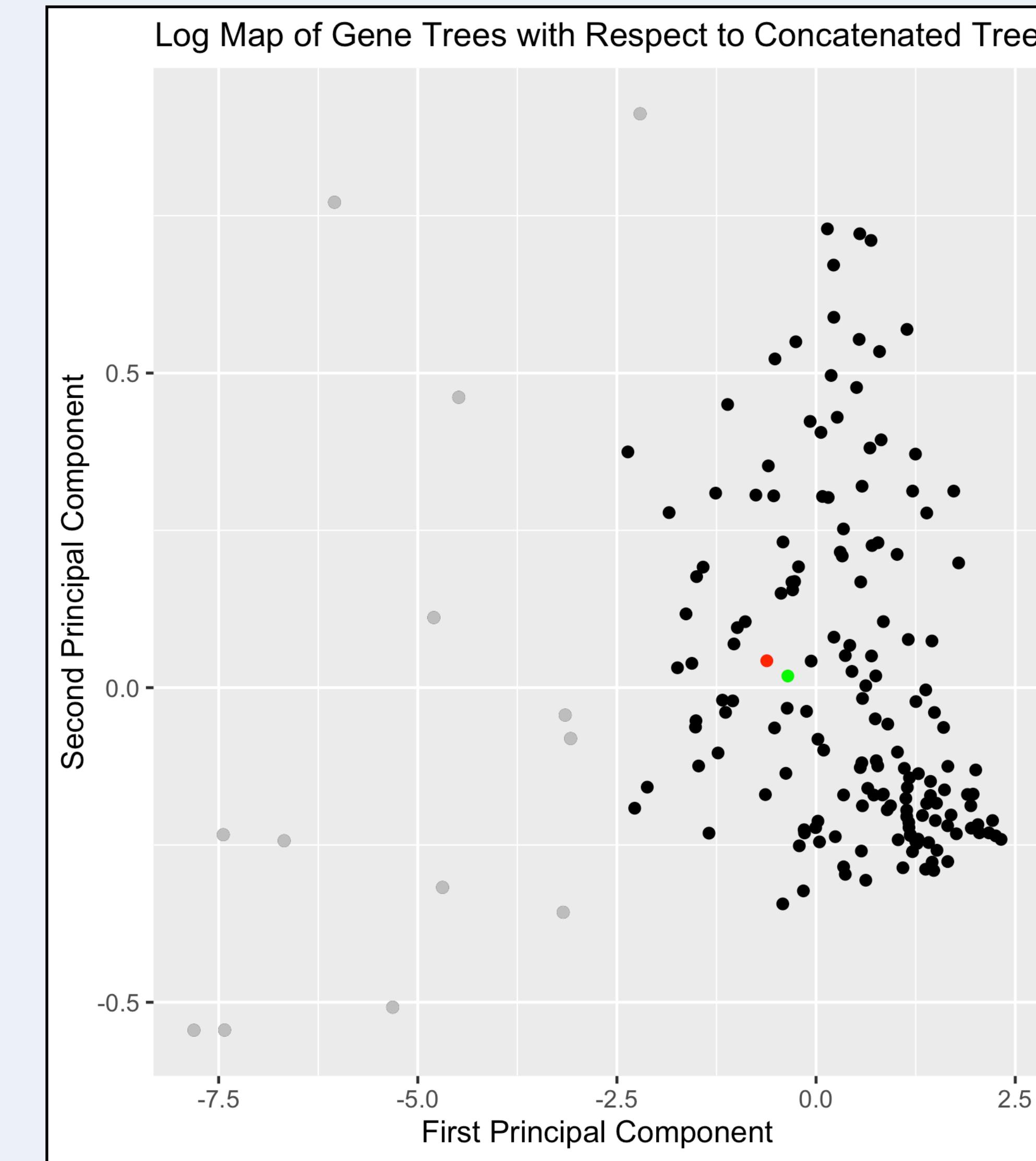
# Phylogenomics: Gene Tree Exploration

## An Analysis of Cyanobacteria Genomes

What might you find?

**Red point is phylogenomic  
tree from full SCG set**

**Green point is phylogenomic  
tree from reduced SCG set**



**The phylogenomic tree seems  
to be robust to gene choice!**

## Limitations

- This can generate hypotheses. Not confirm hypotheses!
- Can only consider genes that are present in alignments for all genomes
  - As taxonomic breadth goes up, number of shared genes goes down

## Thank you to:



Amy Willis



Mike Lee

## Do you:

- Have questions?
- Have data you think could be interesting to investigate in this way?
- Want to know when this R package will be out of development?

Send me an email at  
[teichs@uw.edu](mailto:teichs@uw.edu). I'd love to chat!