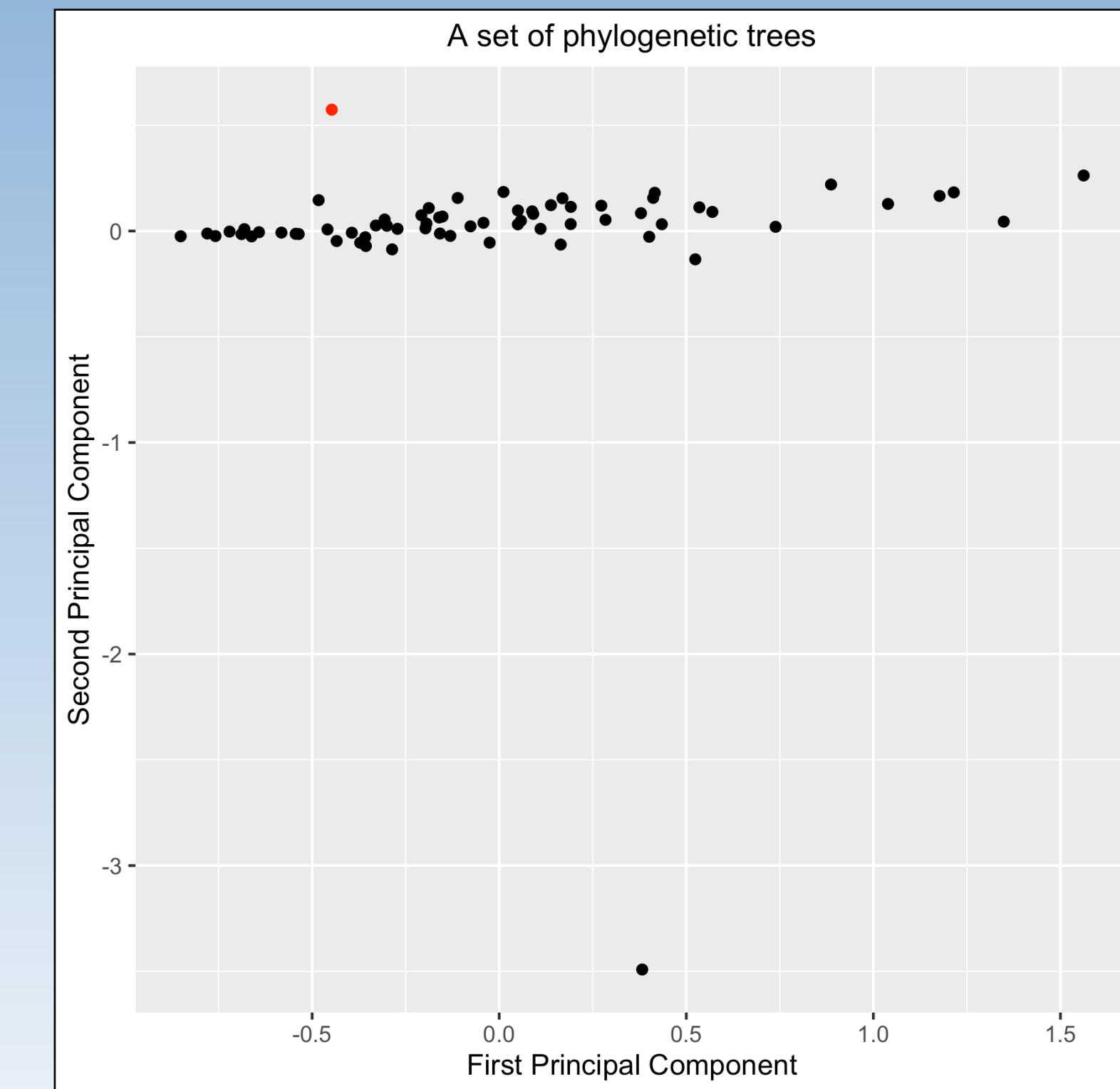
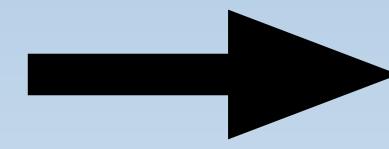
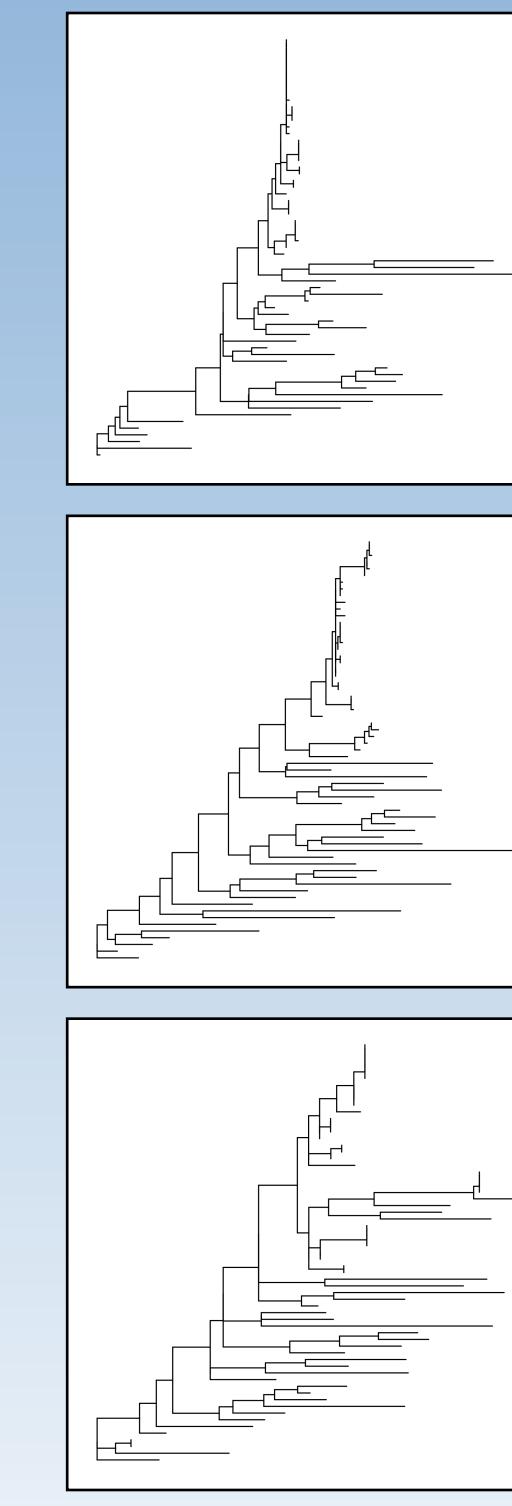
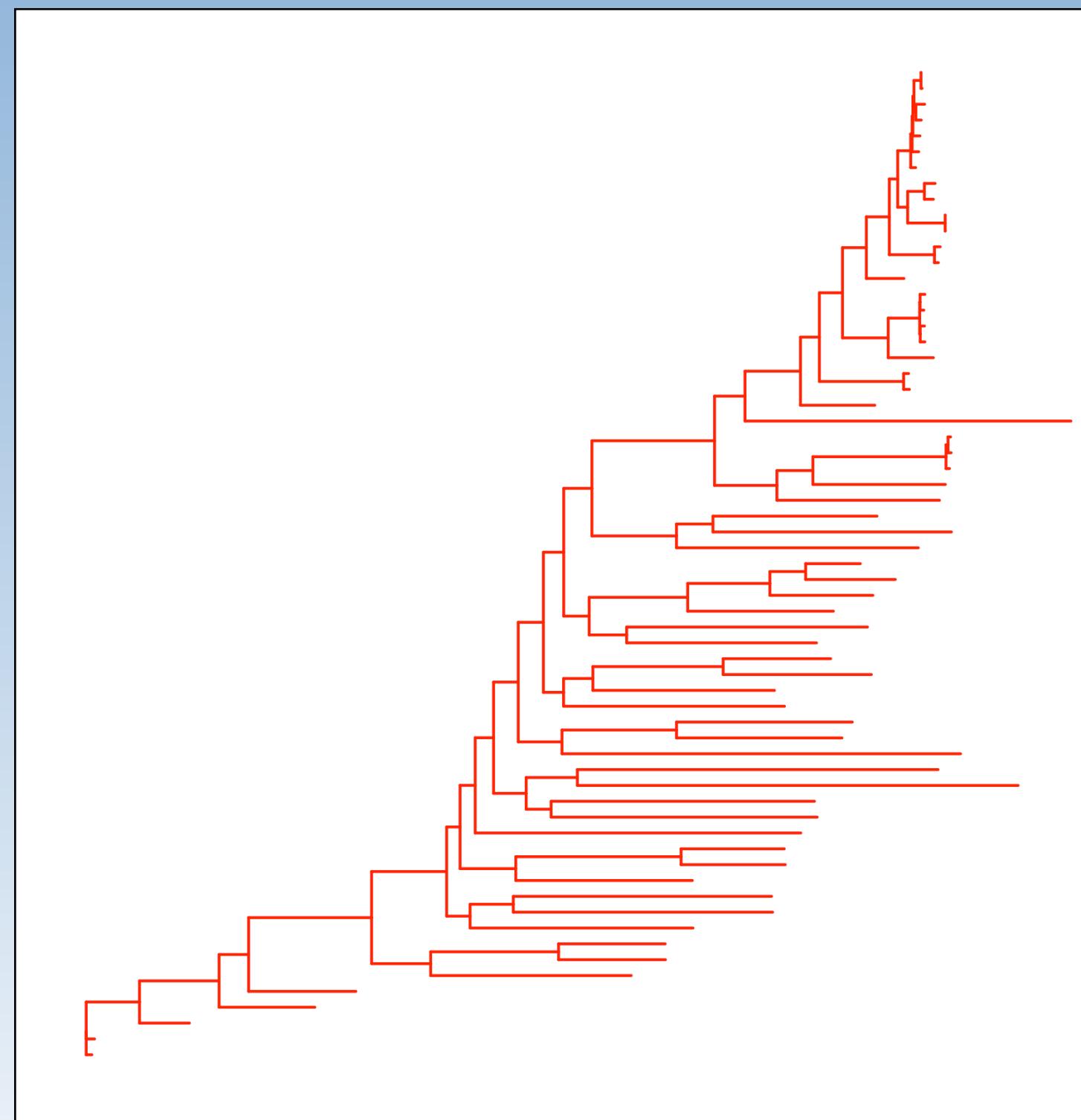


Phylogenomics: Gene Tree Exploration

A visualization tool to investigate gene and genome level evolutionary histories



Sarah Teichman

teichs@uw.edu

@sarah_teichman

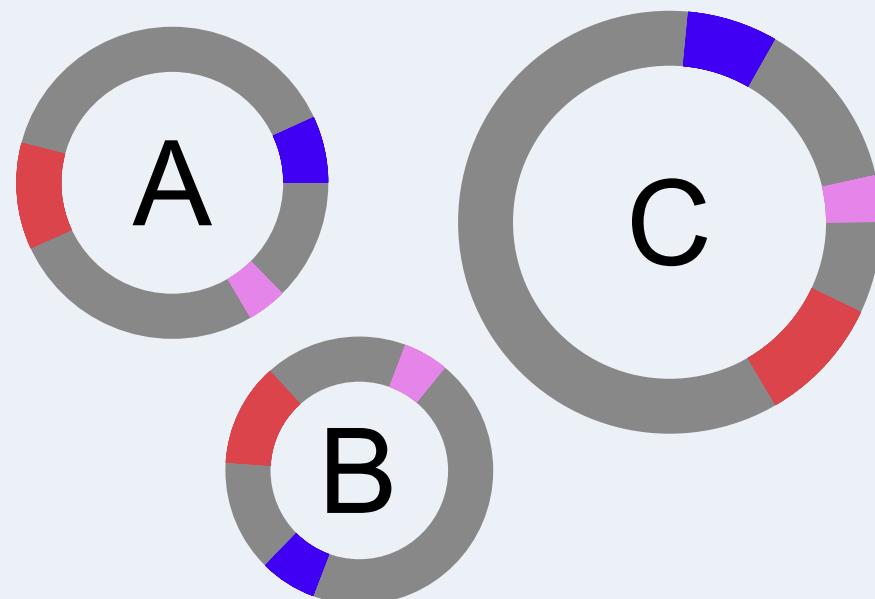
Phylogenetics

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

Phylogenetic trees

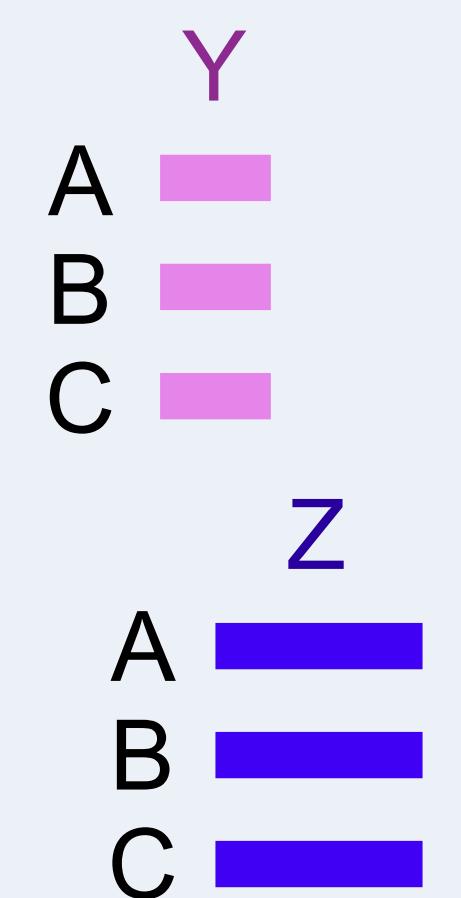
Visual representations of hypotheses about evolutionary relationships

1. Identify target gene-types in genomes of interest

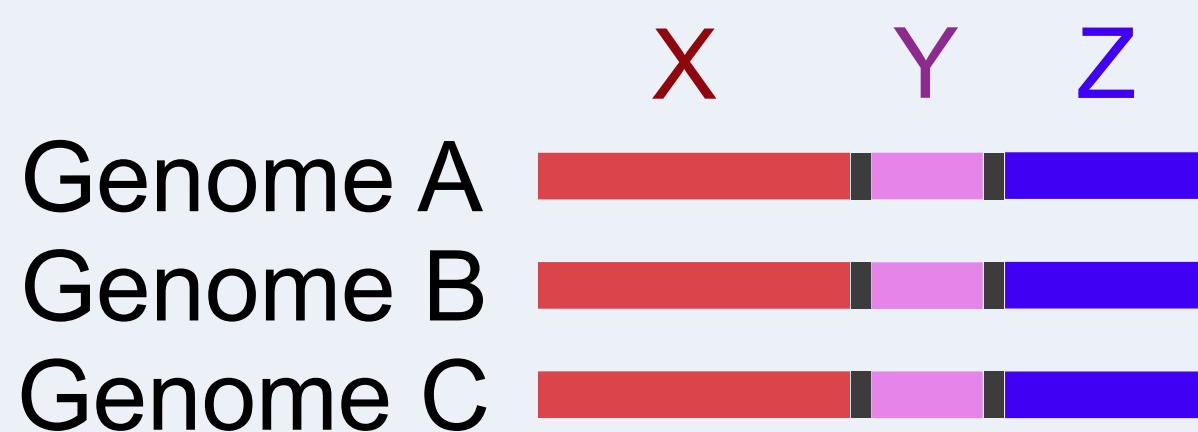


2. Align individual target gene-sets

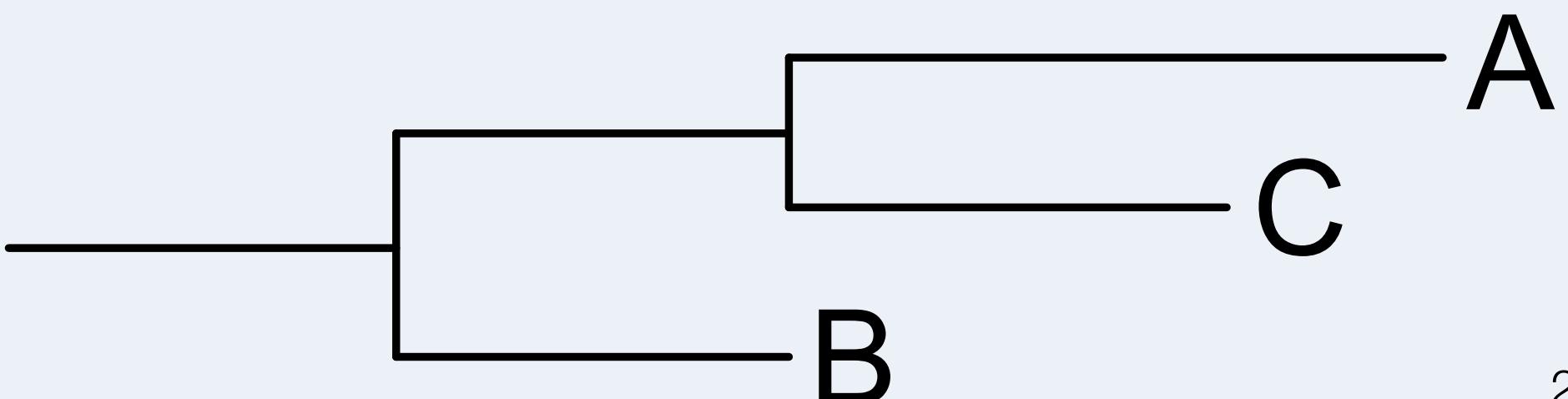
	Gene X
Genome A	—
Genome B	—
Genome C	—



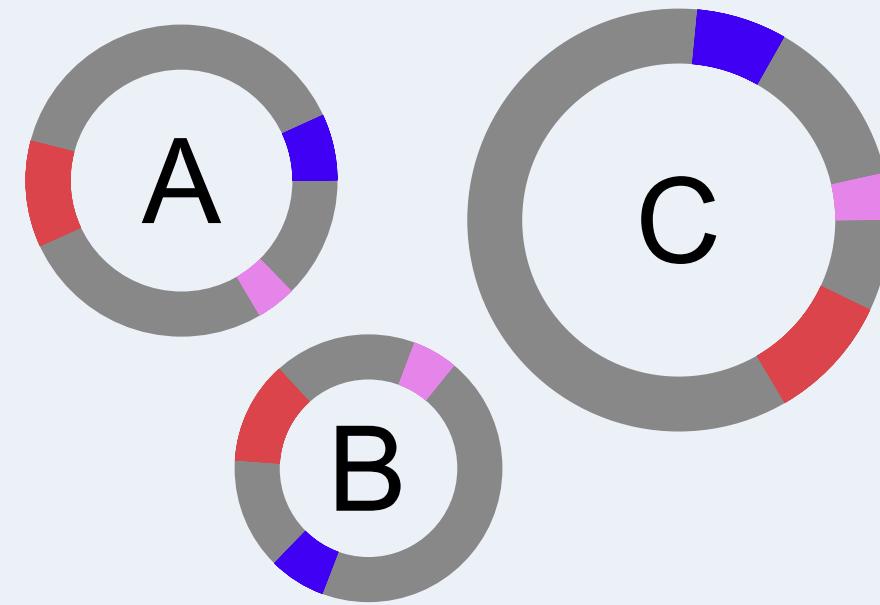
3. Stick alignments together



4. Infer evolutionary relationships



**1. Identify target gene-types
in genomes of interest**



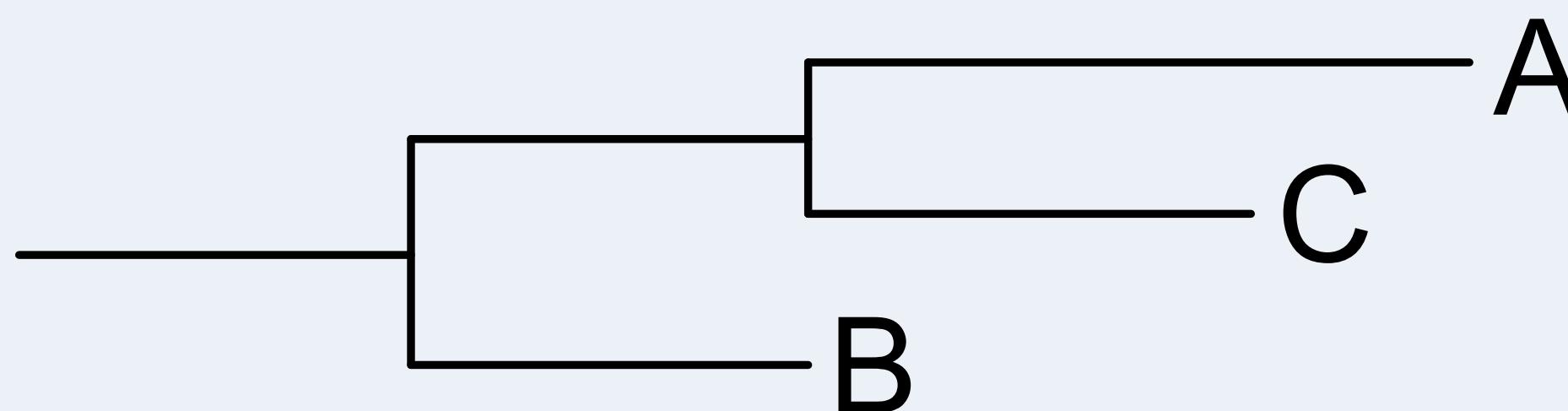
**2. Align individual
target gene-sets**

Gene X

Genome A	[red bar]
Genome B	[red bar]
Genome C	[red bar]

Y	A [pink bar]
B [pink bar]	C [pink bar]
Z	
A [blue bar]	B [blue bar]
C [blue bar]	

4. Infer evolutionary relationships

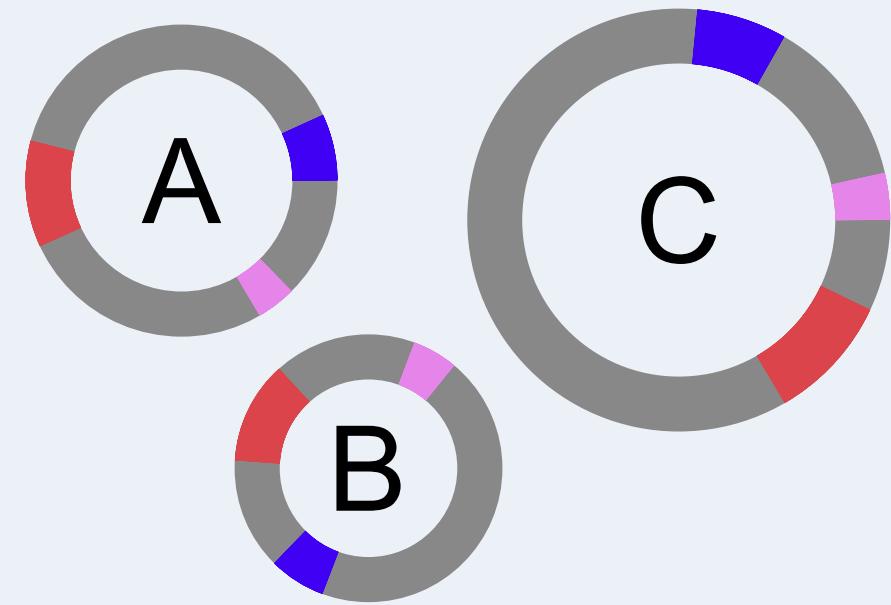


3. Stick alignments together

X	Y	Z	
Genome A	[red bar]	[pink bar]	[blue bar]
Genome B	[red bar]	[pink bar]	[blue bar]
Genome C	[red bar]	[pink bar]	[blue bar]

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

**1. Identify target gene-types
in genomes of interest**



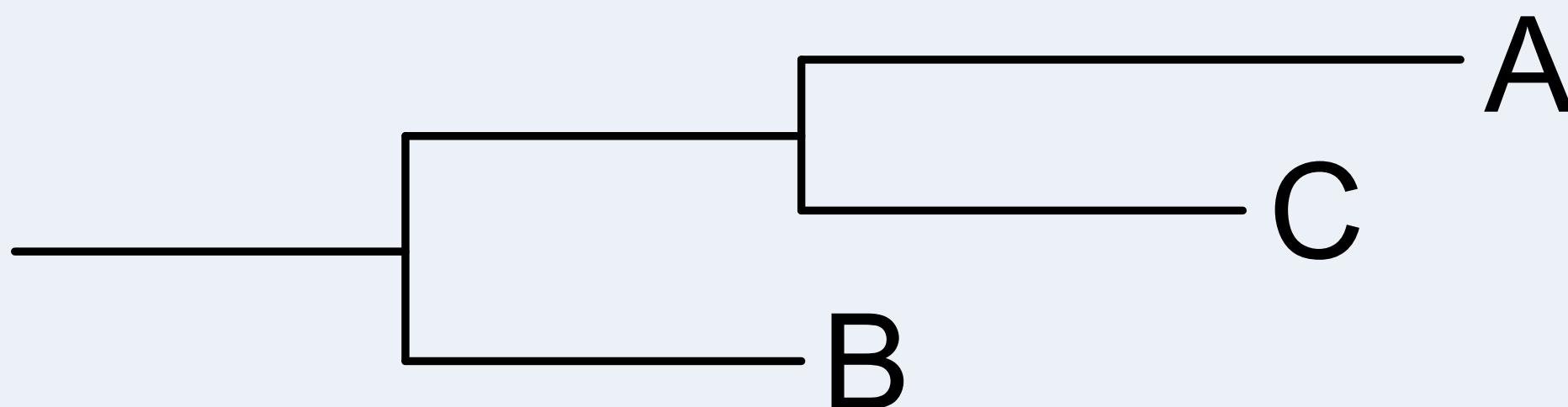
**2. Align individual
target gene-sets**

Gene X

Genome A	[red bar]
Genome B	[red bar]
Genome C	[red bar]

Y	A [pink bar]
B [pink bar]	C [pink bar]
Z	
A [blue bar]	B [blue bar]
C [blue bar]	

4. Infer evolutionary relationships



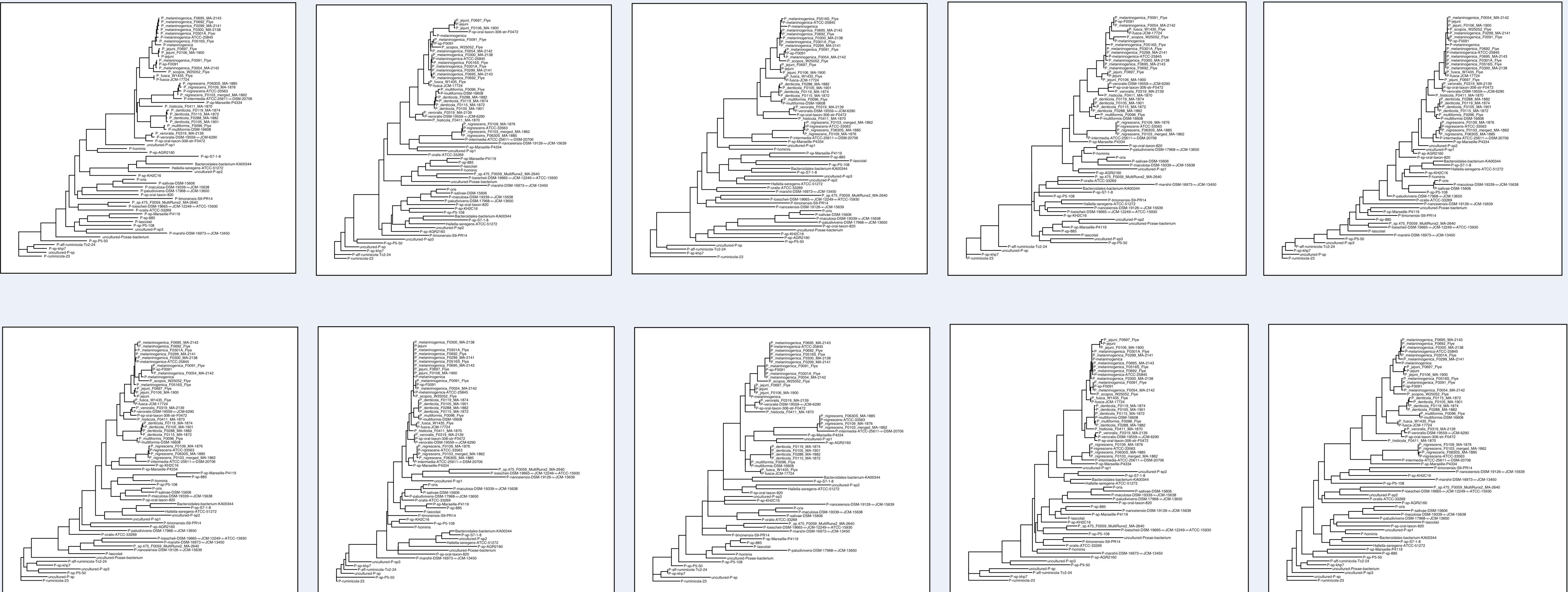
3. Stick alignments together

X	Y	Z	
Genome A	[red bar]	[pink bar]	[blue bar]
Genome B	[red bar]	[pink bar]	[blue bar]
Genome C	[red bar]	[pink bar]	[blue bar]

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

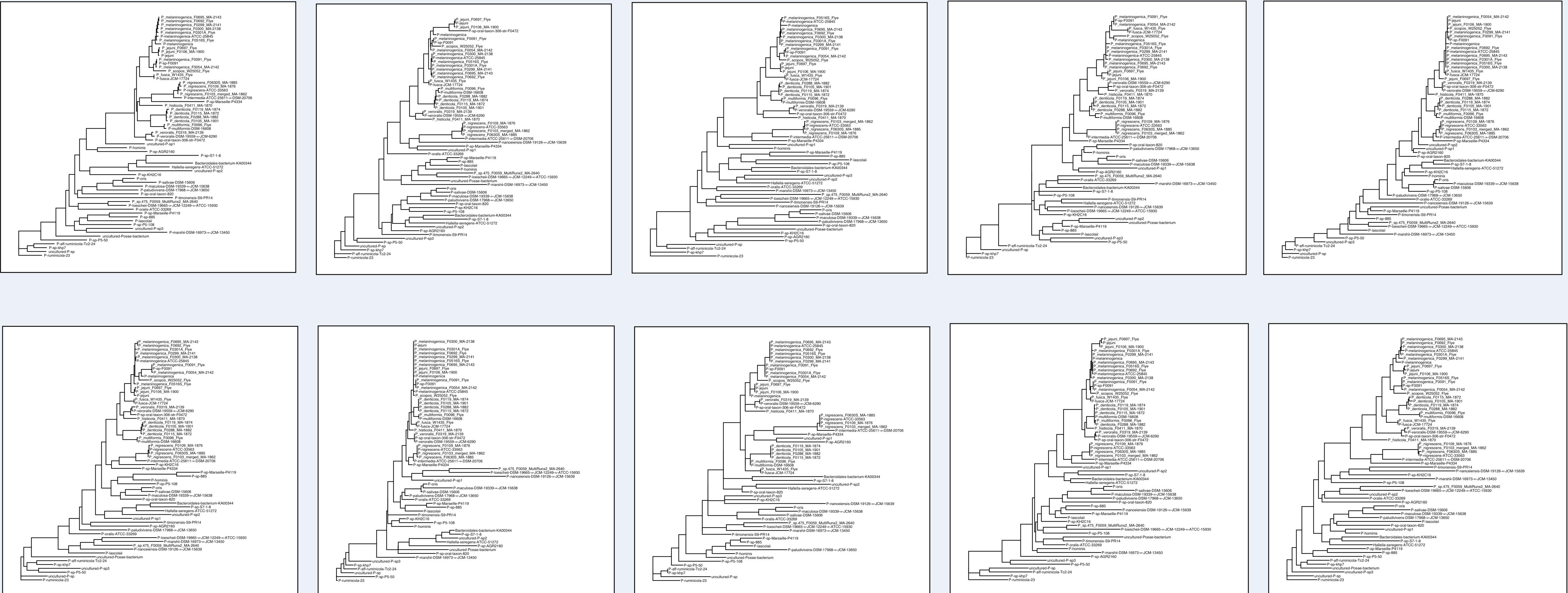
**Why?
Because we might
learn something new
from our data**

Phylogenomics: Gene Tree Exploration



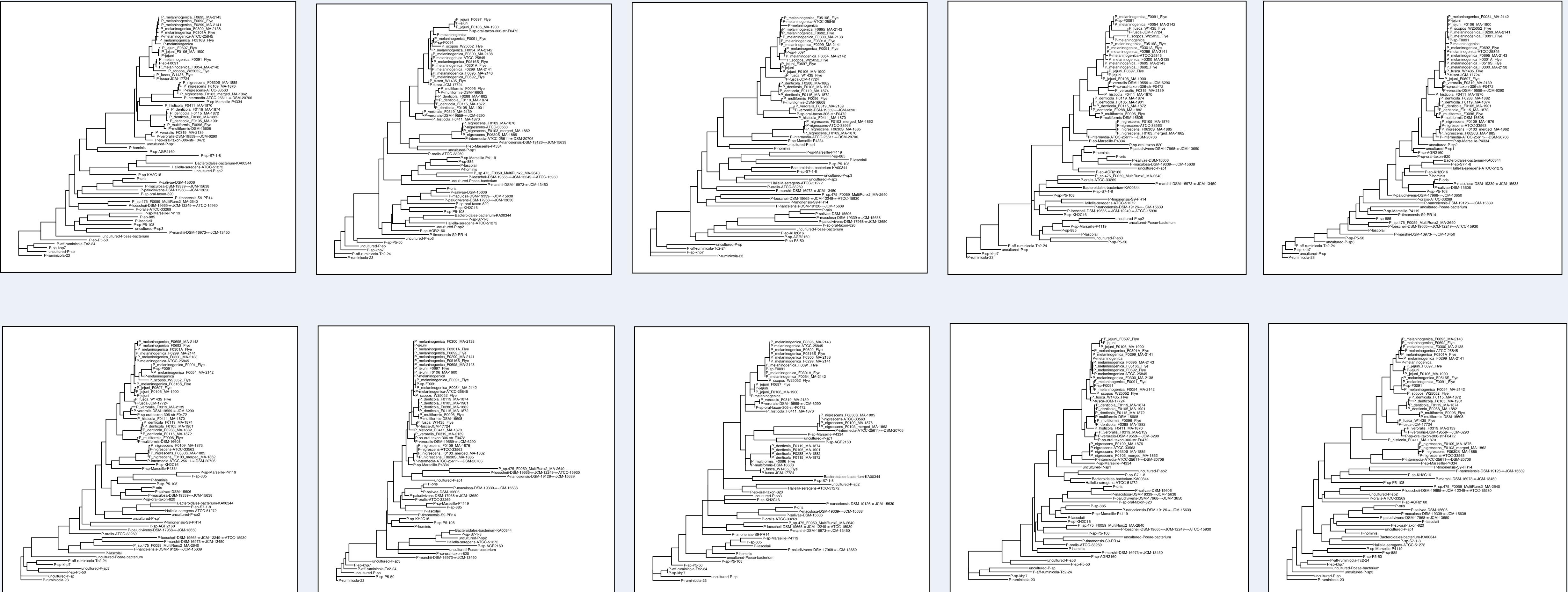
Phylogenomics: Gene Tree Exploration

What can we learn from these plots?



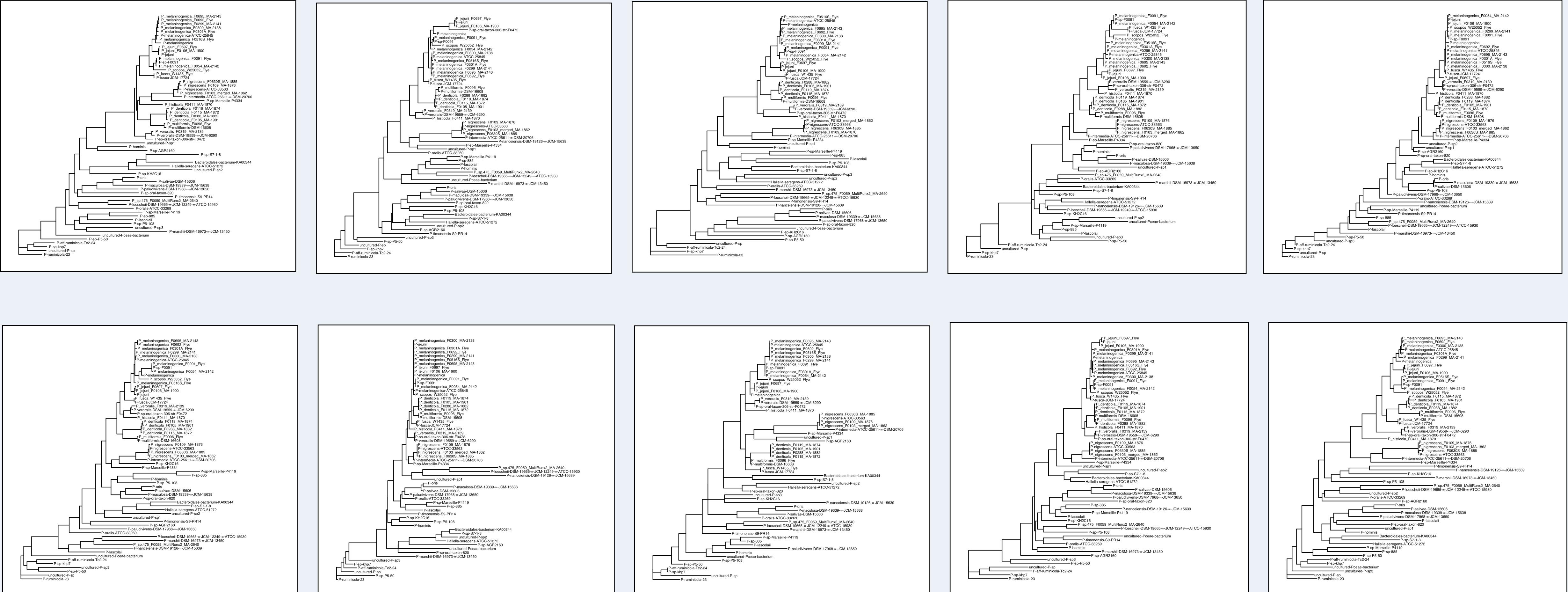
Phylogenomics: Gene Tree Exploration

What can't we learn from these plots?



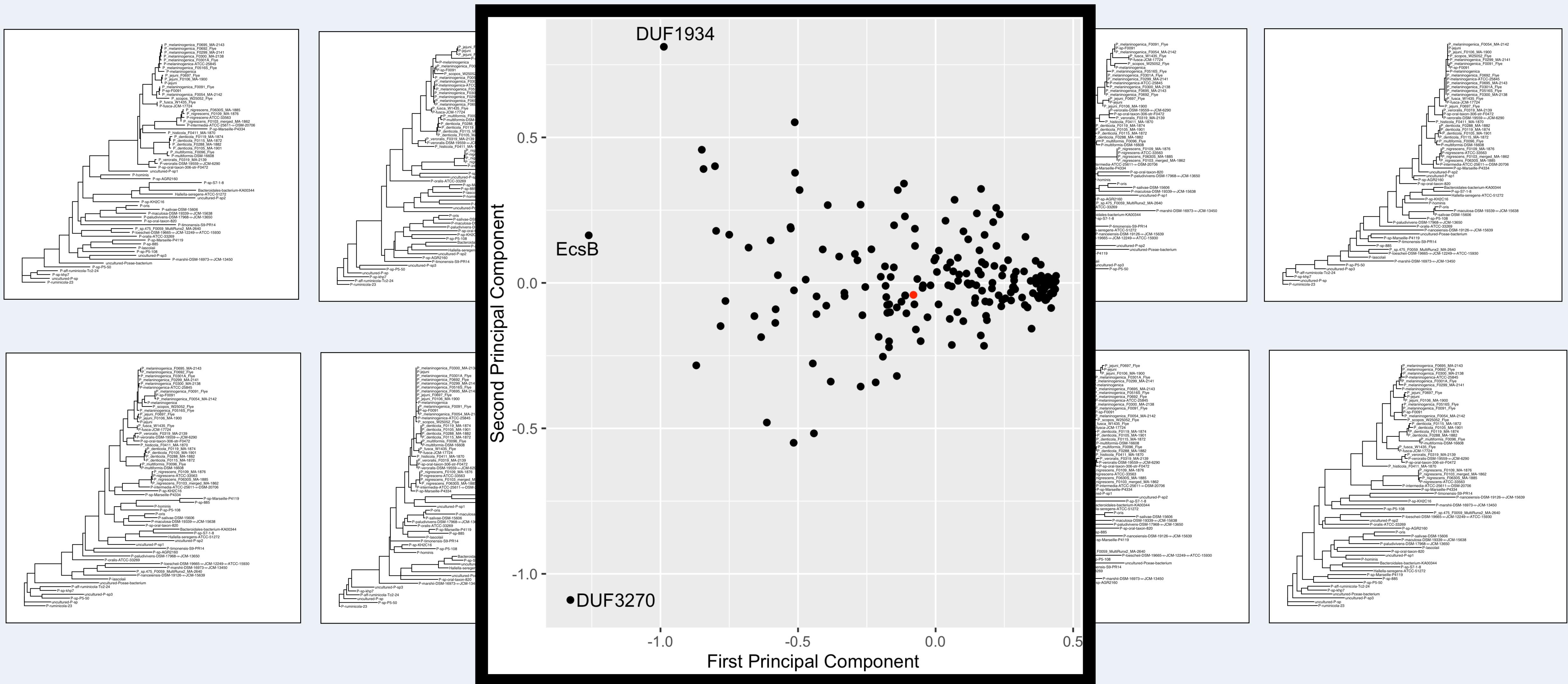
Phylogenomics: Gene Tree Exploration

How can we reduce this data to compare gene trees?

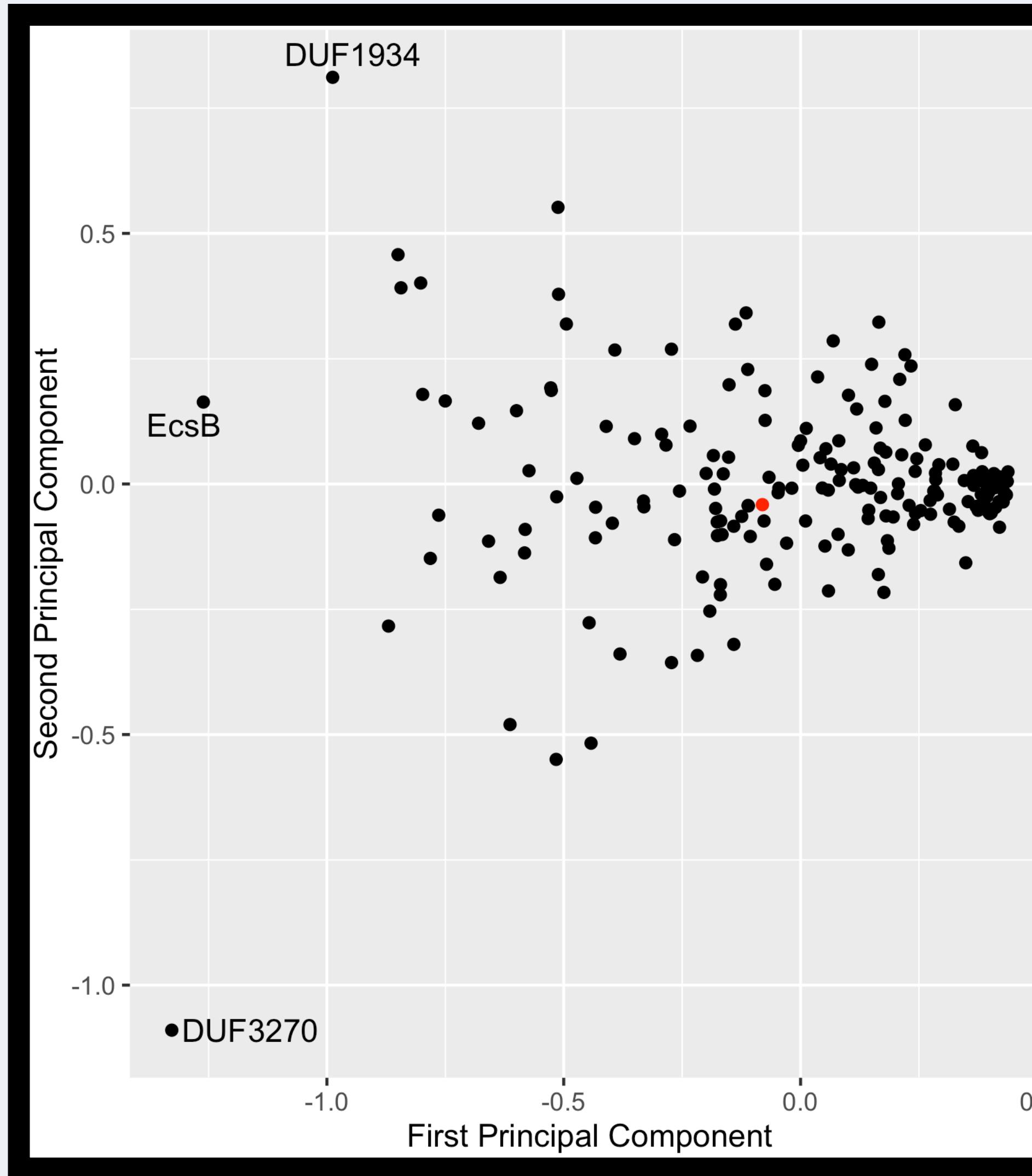


Phylogenomics: Gene Tree Exploration

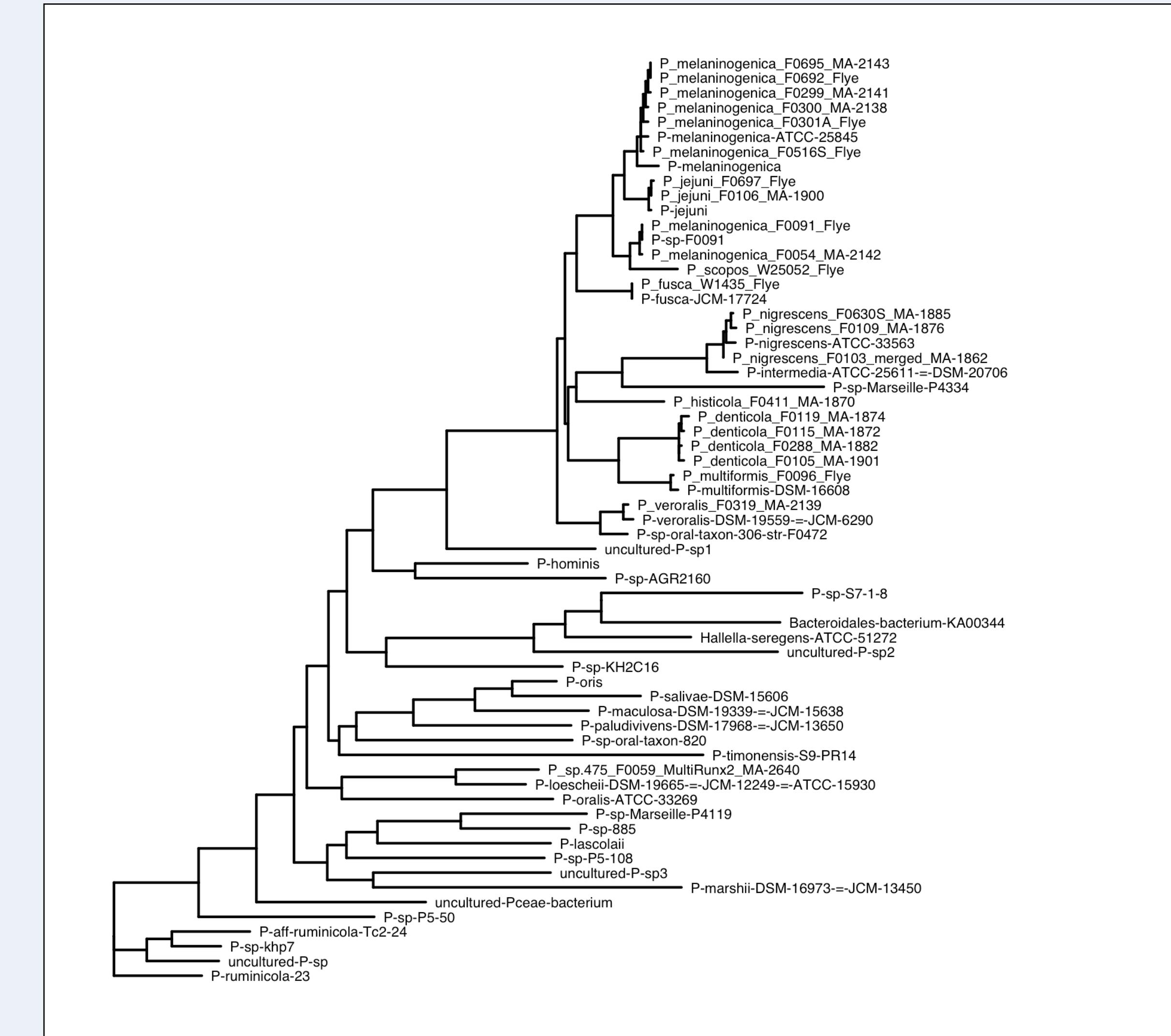
How can we reduce this data to compare gene trees?



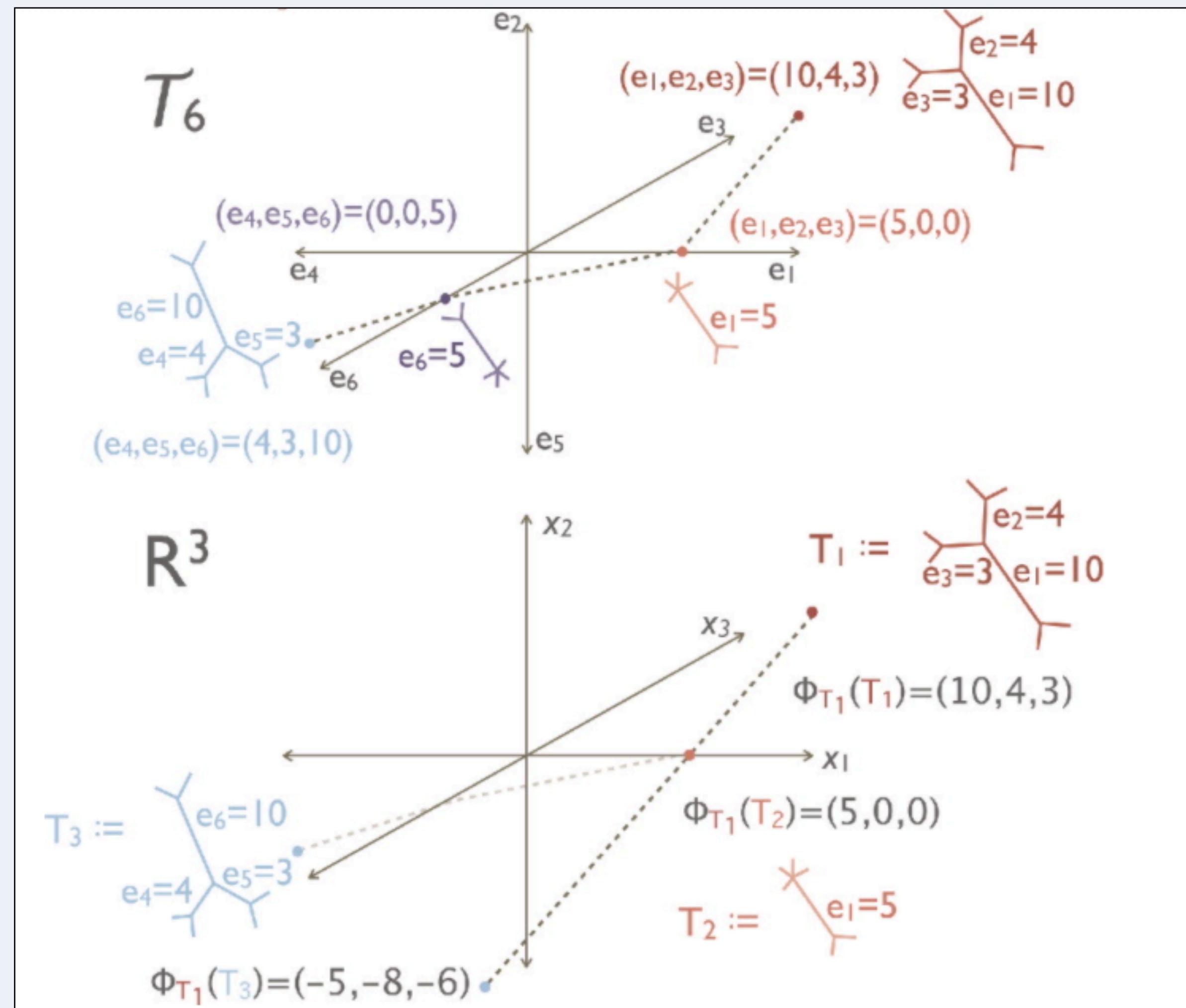
Why is this hard?



A tree is more complex than a vector!



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



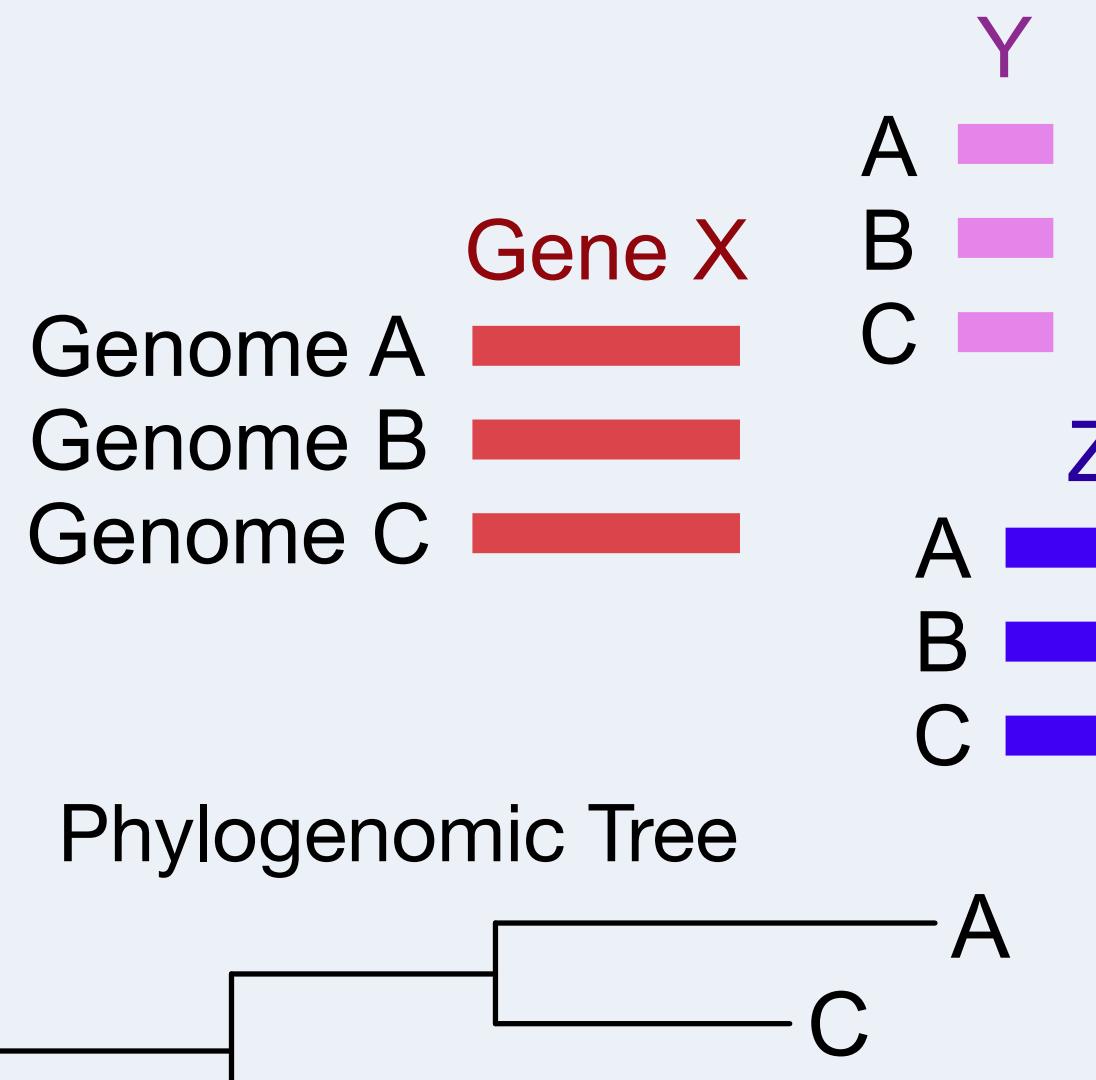
BHV space:
Constructed in
Billera et al. (2001)

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

Figure: Willis and Bell (2018)

GROVE: Gene tRee Ordination Visualization for Exploration

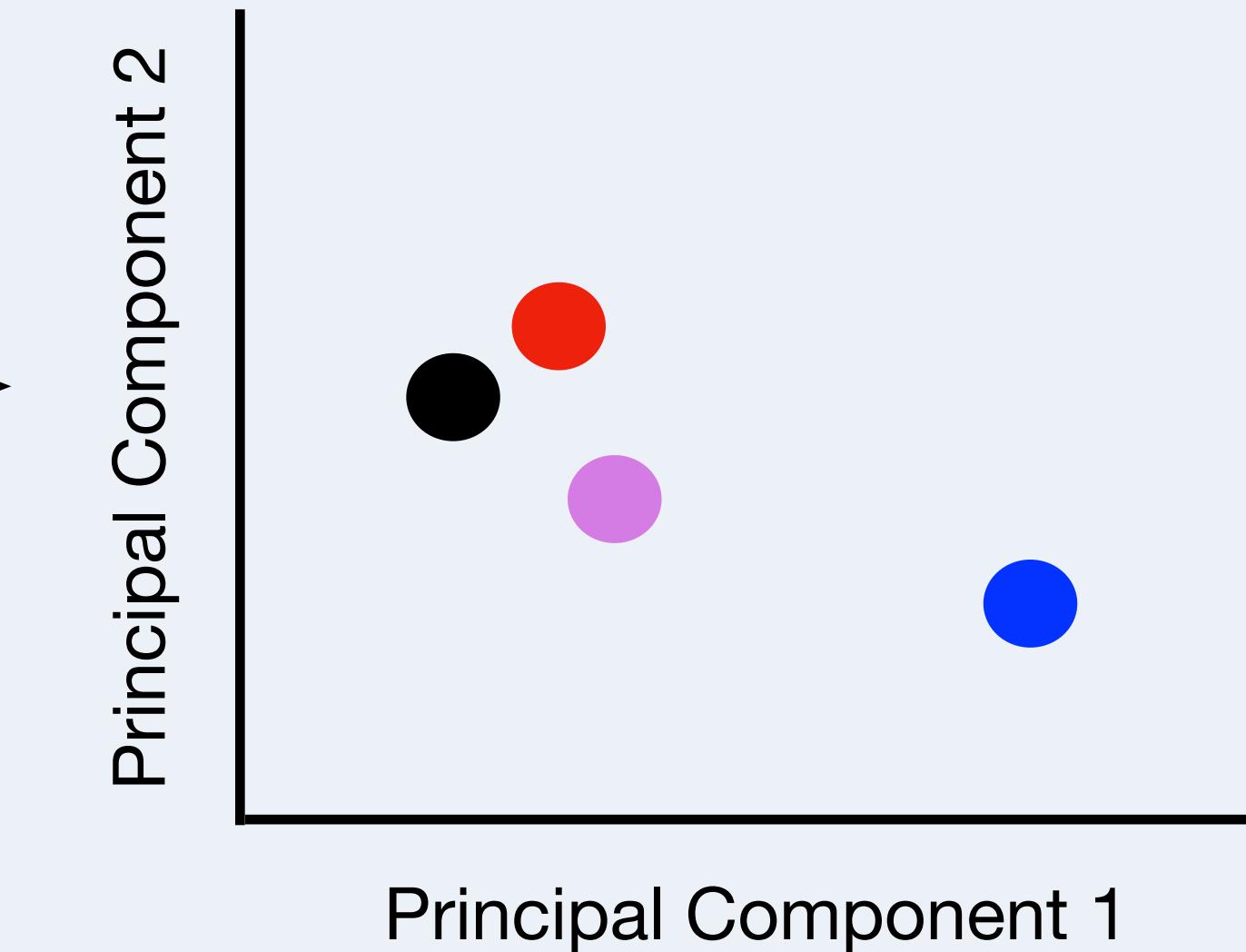
1. GToTree Output



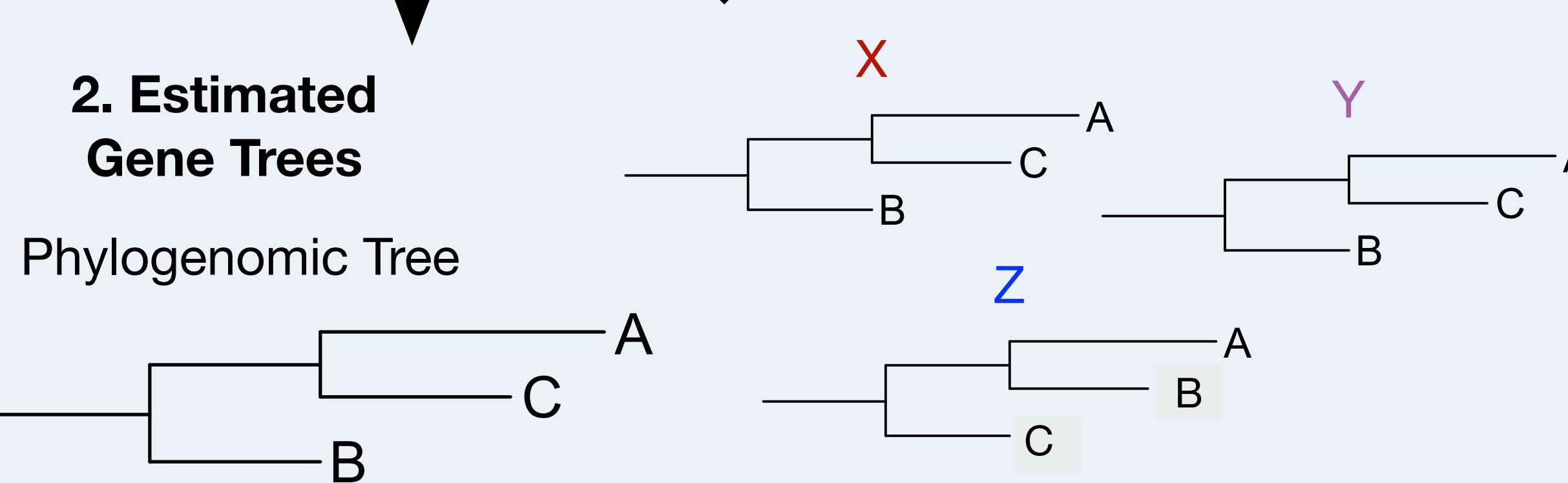
3. Vectors in \mathbb{R}^5
Using the Log Map

	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

4. Ordination Plot



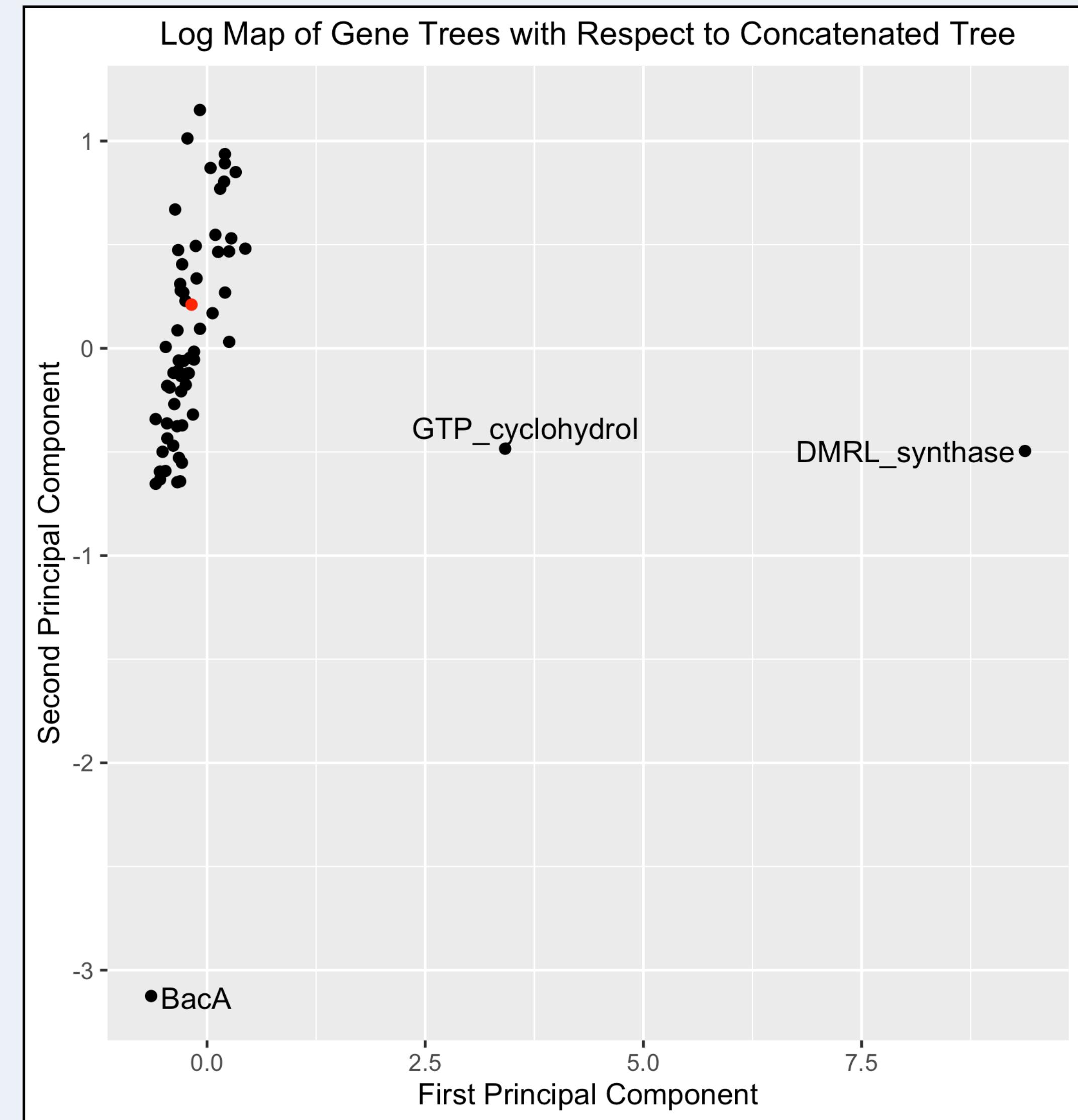
2. Estimated Gene Trees



R package
in development

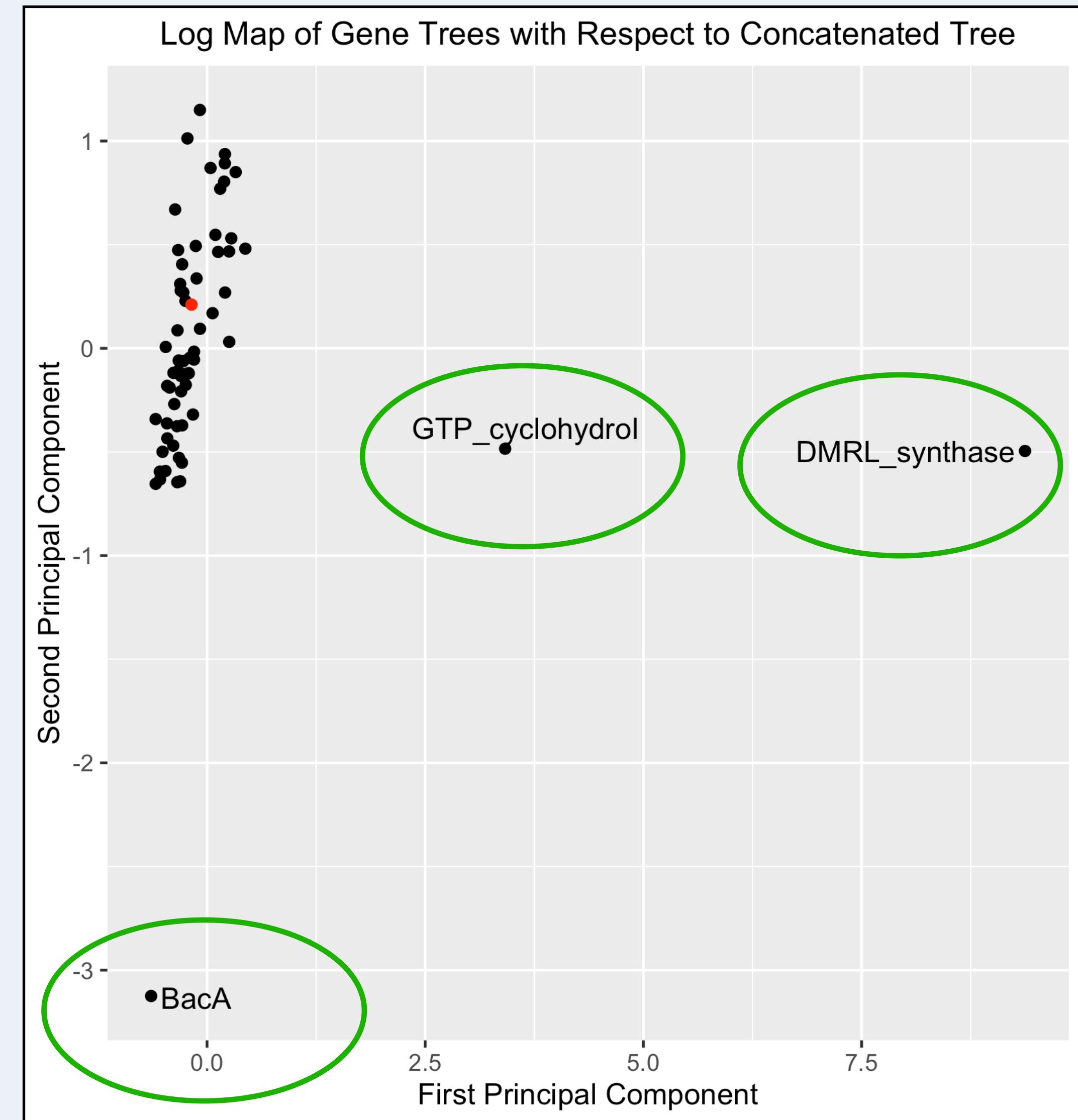
An Analysis of Prevotella Genomes

What might you find?



An Analysis of Prevotella Genomes

What might you find?

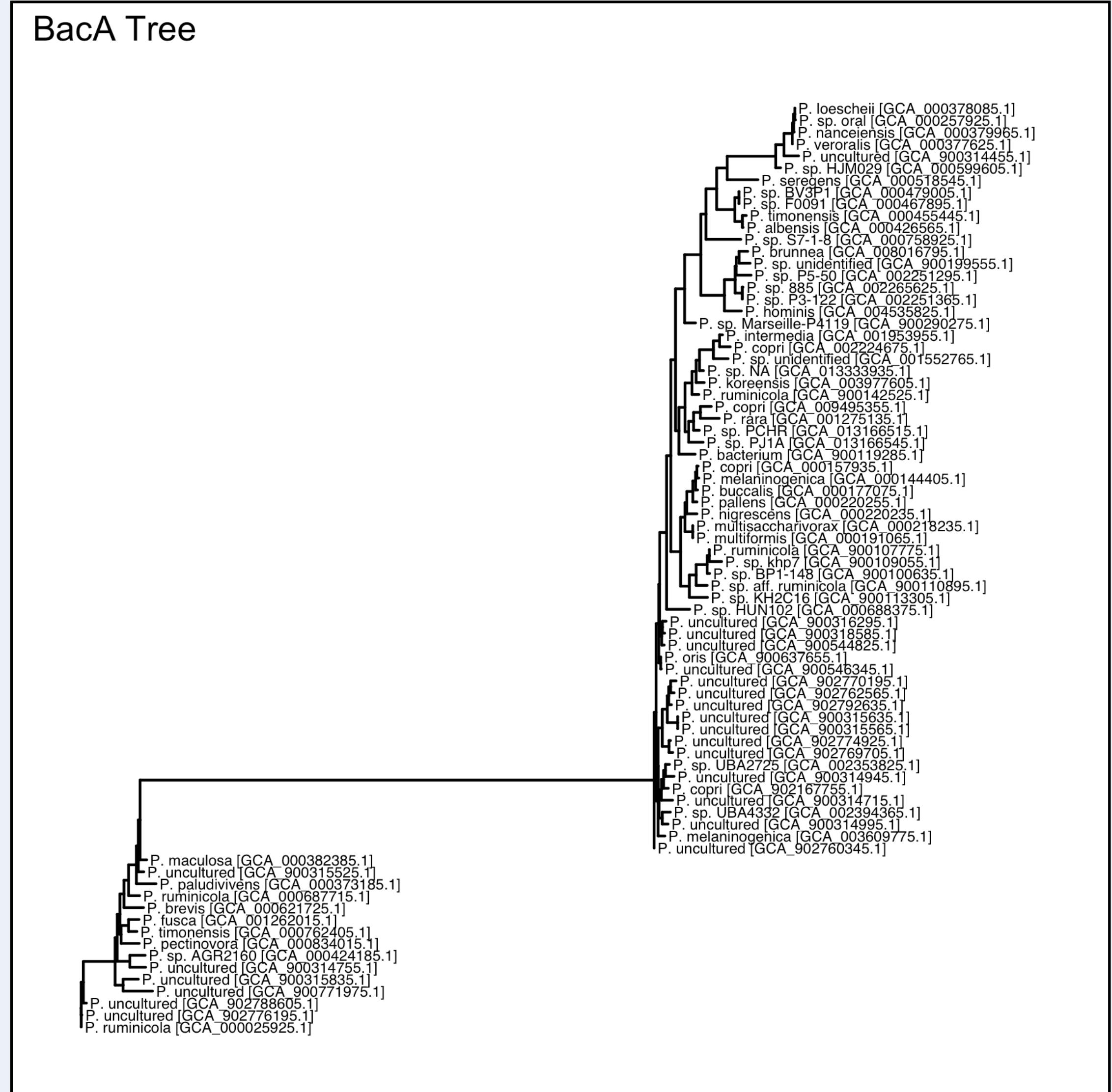
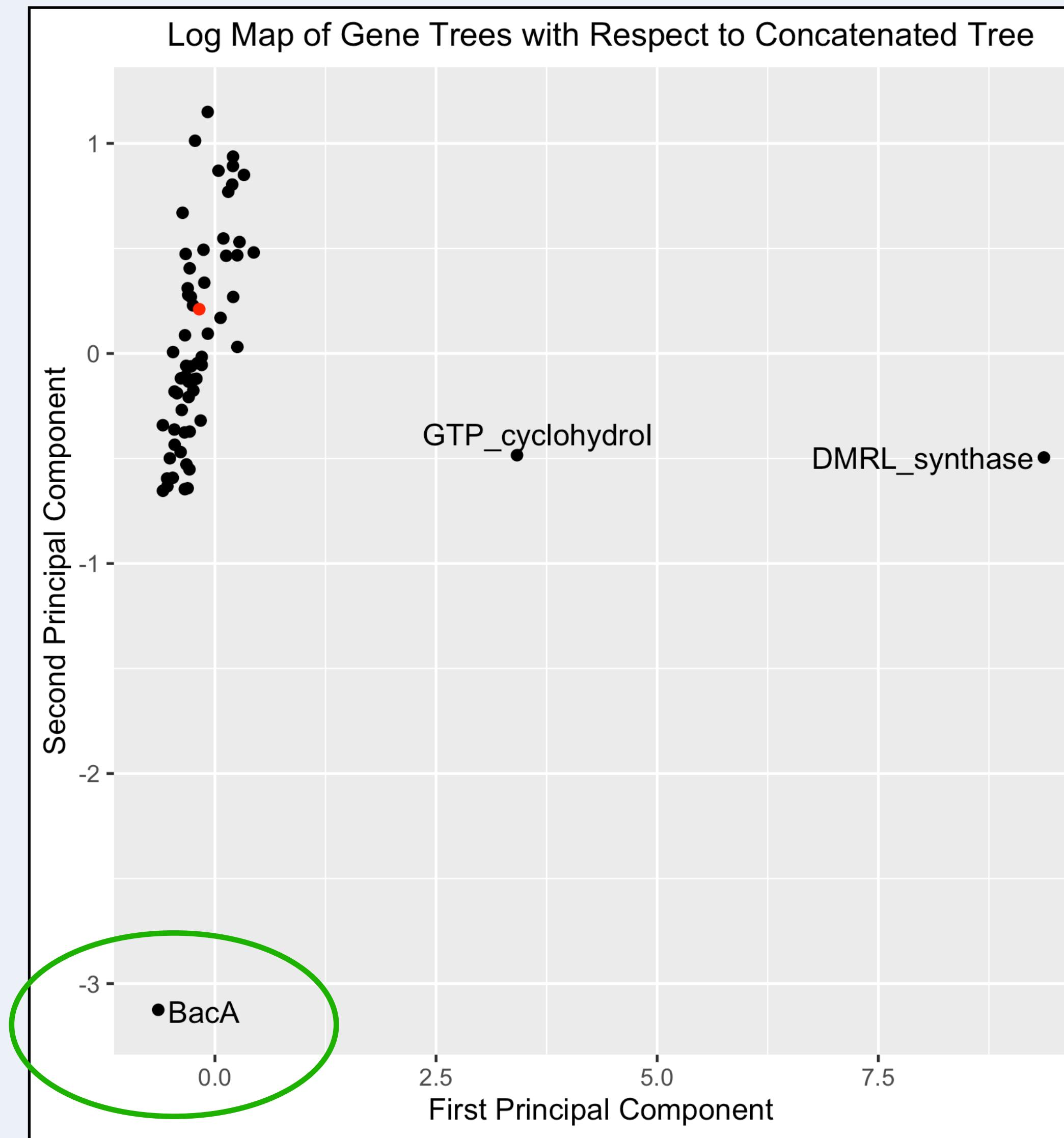


We have some
outlying gene trees!

Phylogenomics: Gene Tree Exploration

An Analysis of Prevotella Genomes

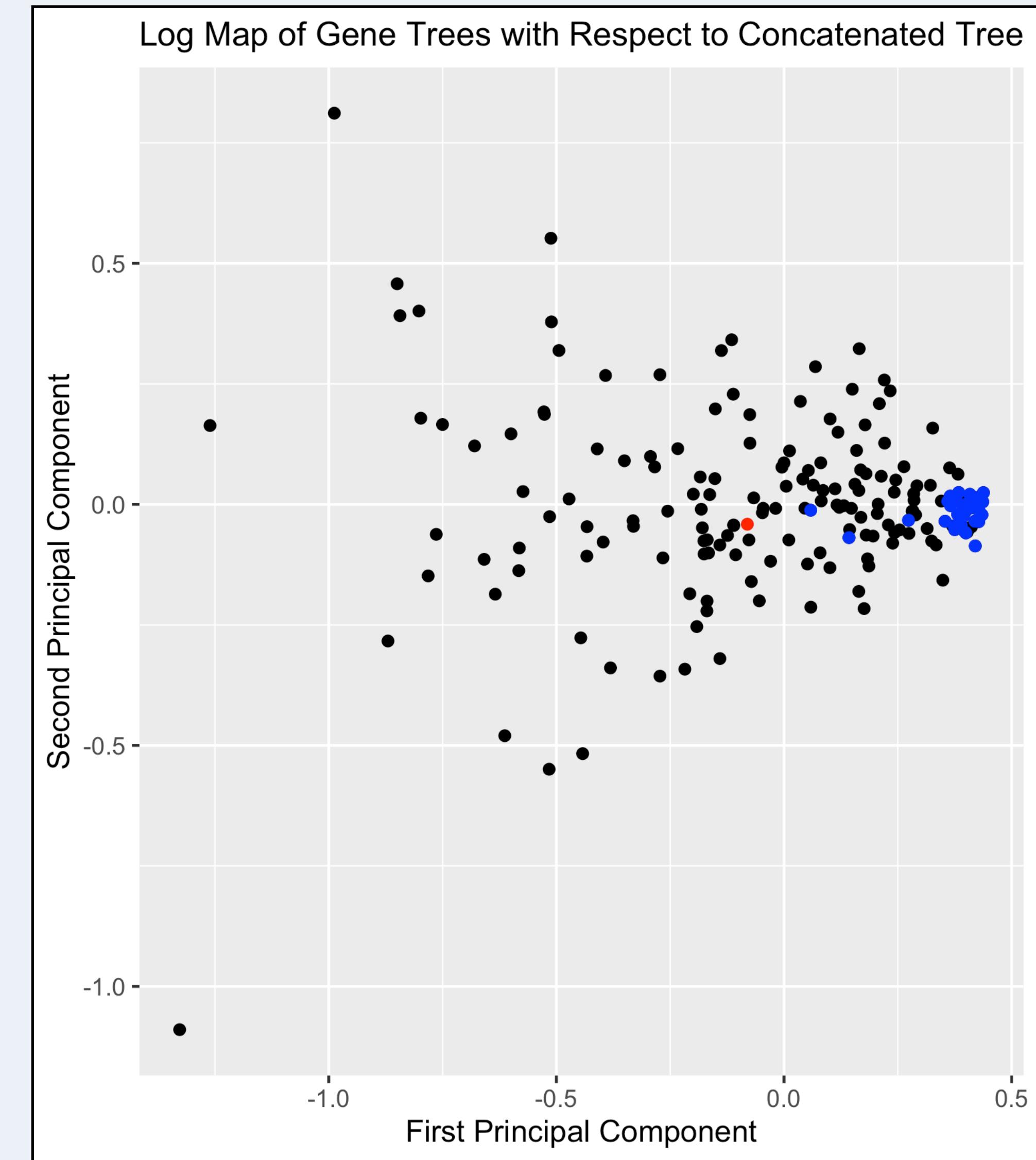
What might you find?



An Analysis of Streptococcus Genomes

Ribosomal genes
shown in blue

What might you find?



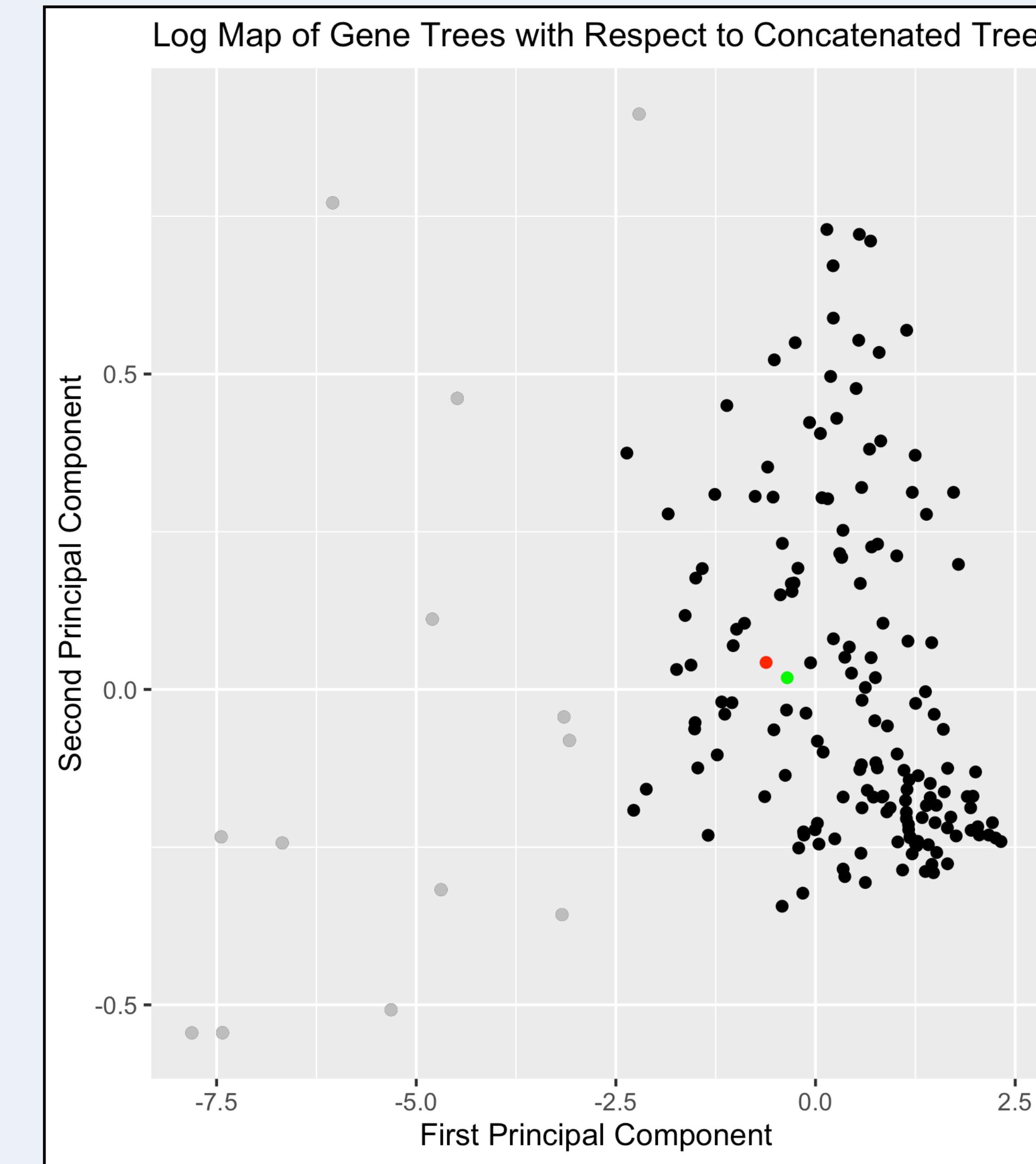
**The ribosomal genes
mostly cluster together!
(But not all of them)**

An Analysis of Cyanobacteria Genomes

What might you find?

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

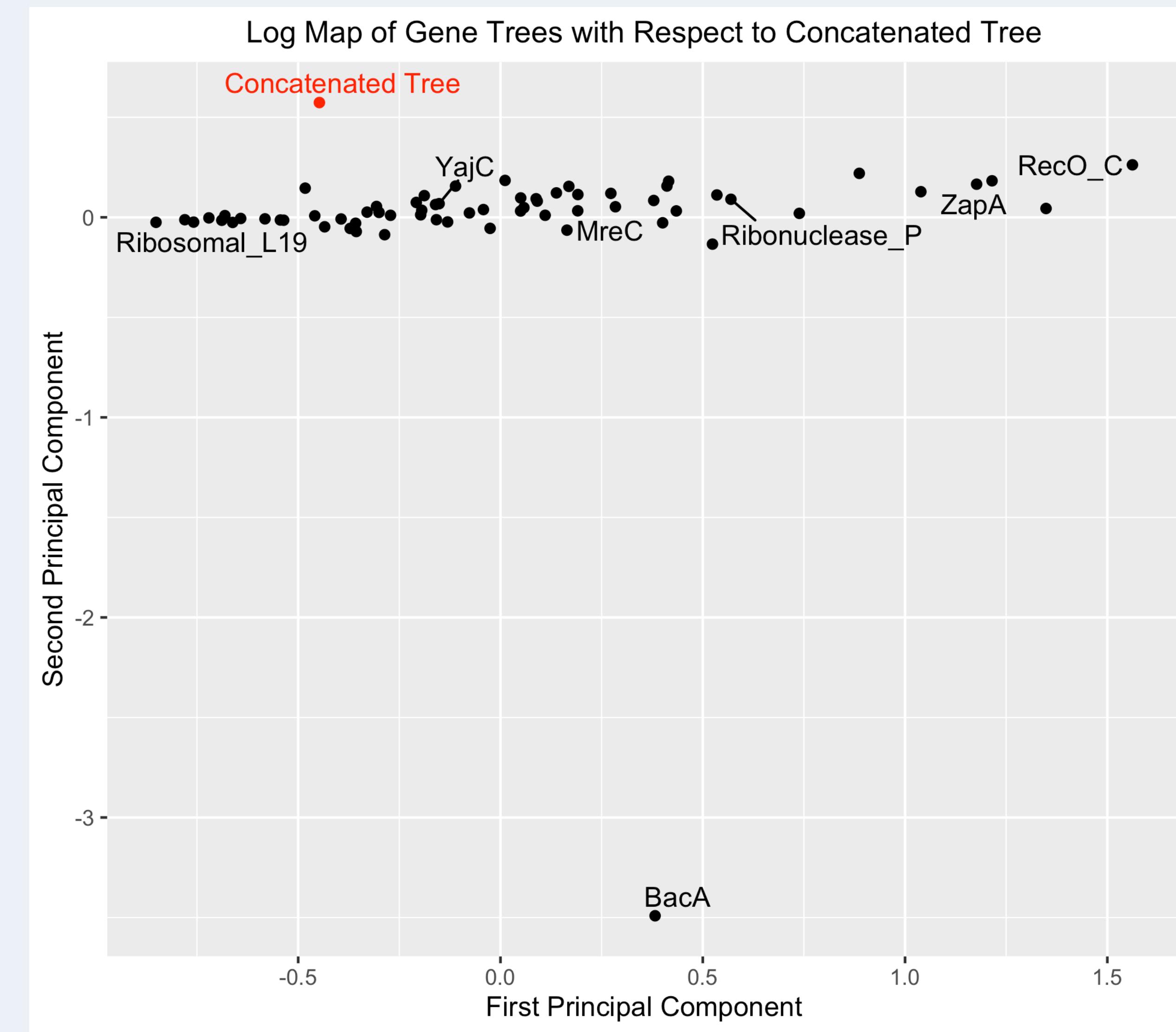
**Green point is phylogenomic
tree from reduced SCG set**



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

An Different Analysis of Prevotella Genomes

What might you find?



The concatenated tree looks pretty different from gene trees! (Turns out that IQTree got stuck at a local optima)

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

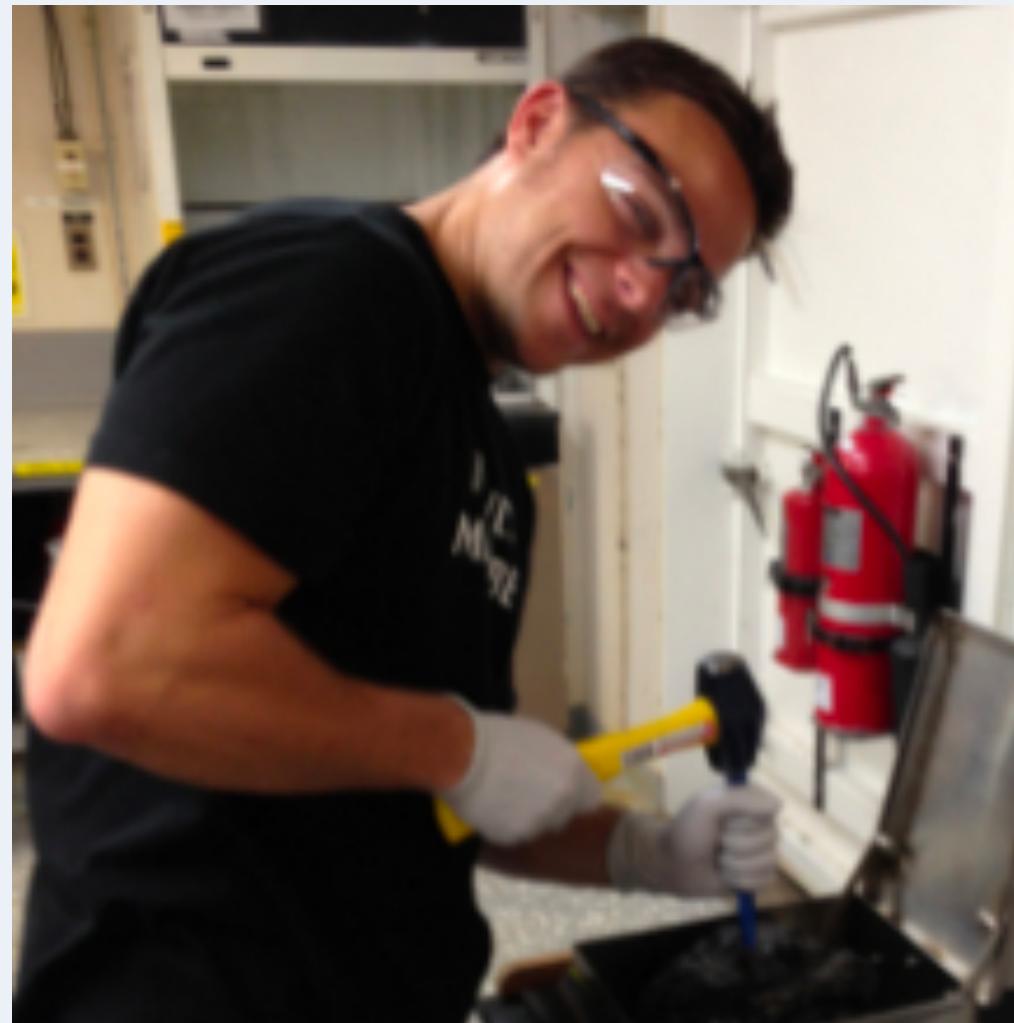
Summary

- Examining gene trees complements building a phylogenomic tree
- For a large number of genes and genomes, a data reduction technique can help search for patterns
- This tool can help you:
 - Find evolutionary signal in individual genes to investigate further
 - Compare and contrast SCG genes used for phylogenomic tree construction
 - Identify sequences to examine closer for errors

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. I'd love to chat!