



# Comparative studies reveal genetic divergence in pathogen response gene families amongst twenty forest tree species



UNIVERSITY OF  
NOTRE DAME

Mengmeng Lu

January 15, 2023

PAG 30, San Diego, CA

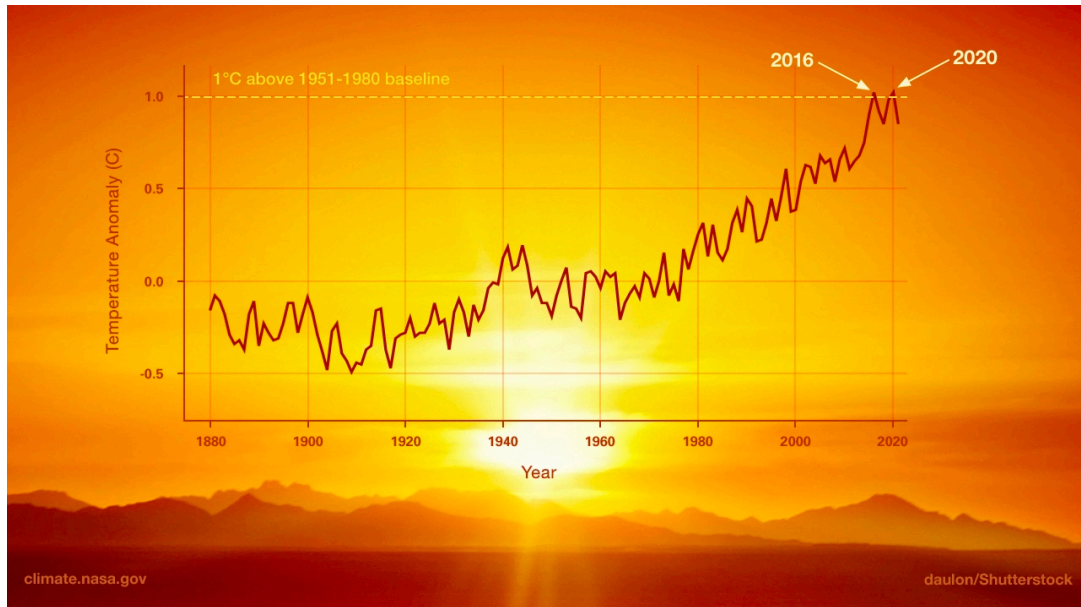


Forest trees provide key ecological benefits and services.



Pictures from <https://www.facebook.com/NDFieldEcology/photos>

# Global Warming & Climate Change



Macbeth: I will not be afraid of death and bane,  
till Birnam forest come to Dunsinane.



<http://theshakespeareblog.com/2017/12/preserving-macbeths-birnam-wood/>



<http://www.quickmeme.com/meme/3585b1>

“Conservation of forests is the need of the hour.”



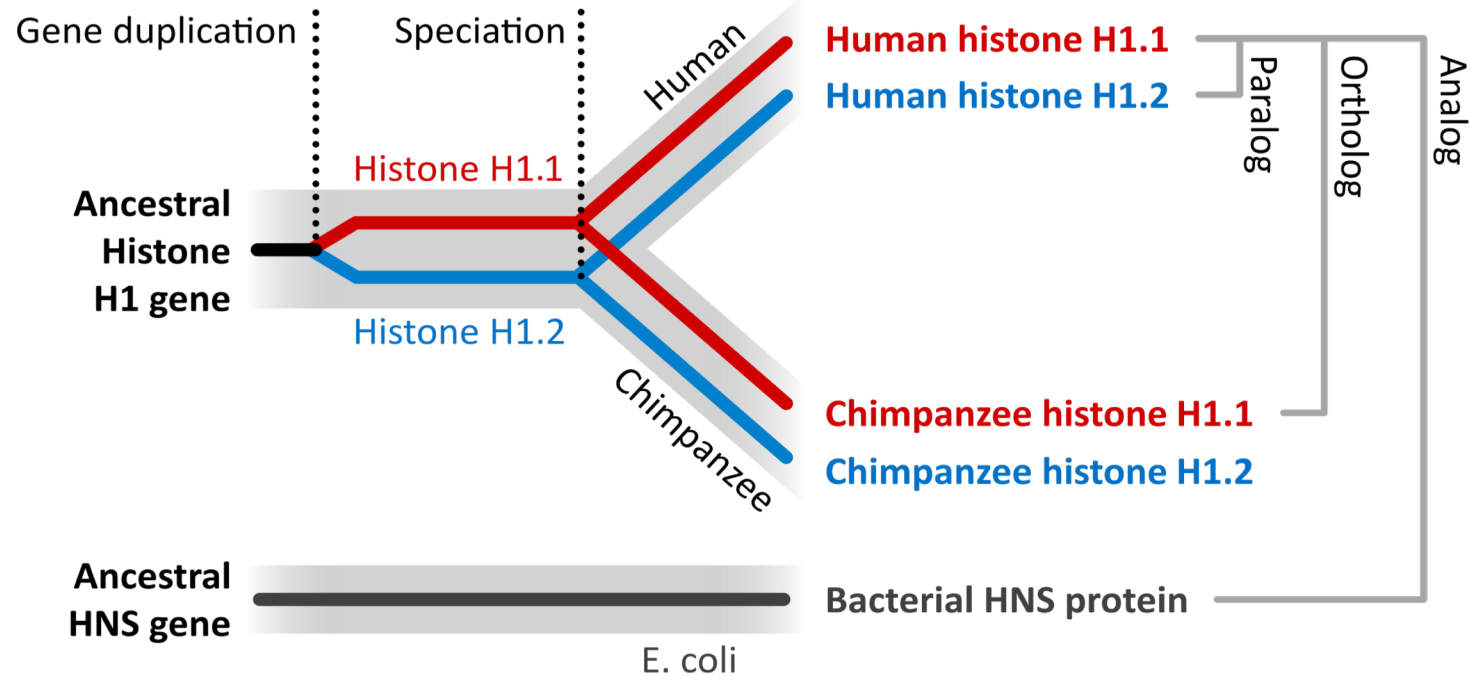
# Genetic diversity

- Heritable variation within species at DNA level
- Genetic differences within and among populations enable species to adapt to environmental changes
- Nucleotide substitutions per site
- Gene birth and death per gene (gene copy number changes)

# Aim

## ➤ Gene family variation across forest species

**Gene families** comprise a set of similar nucleotide or amino acid sequences. They arise as a result of gene or genome duplication events, sharing similar cellular and biochemical functions.



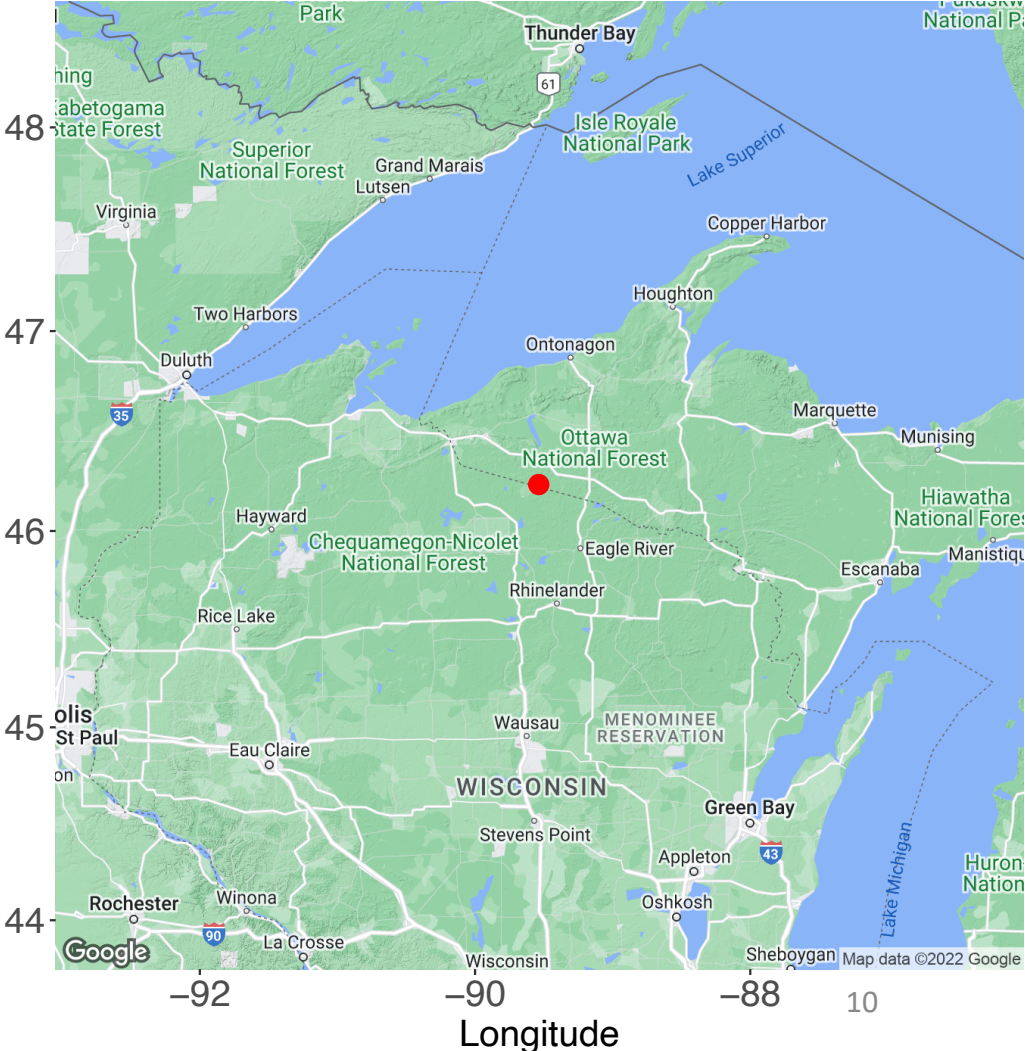


# Aims of this project

- Gene family variation across species
- What is the magnitude of gene duplication? What factors impact duplicate gene evolution? Are the gene gain and loss process random or controlled by natural selection?
- Overarching goal -- Dynamic relationships among three fundamental dimensions of biodiversity: genetic diversity, phylogenetic diversity, and functional diversity.

# University of Notre Dame Environmental Research Center (UNDERC)

## Field Research Site, Upper Peninsula of Michigan



## Principal Investigator



### **Nate Swenson**

Gillen Director of the University of Notre Dame Environmental Research Center

Professor of Biology

**CV**

**Google Scholar**



National  
Science  
Foundation

NOTRE DAME RESEARCH

UNIVERSITY OF NOTRE DAME

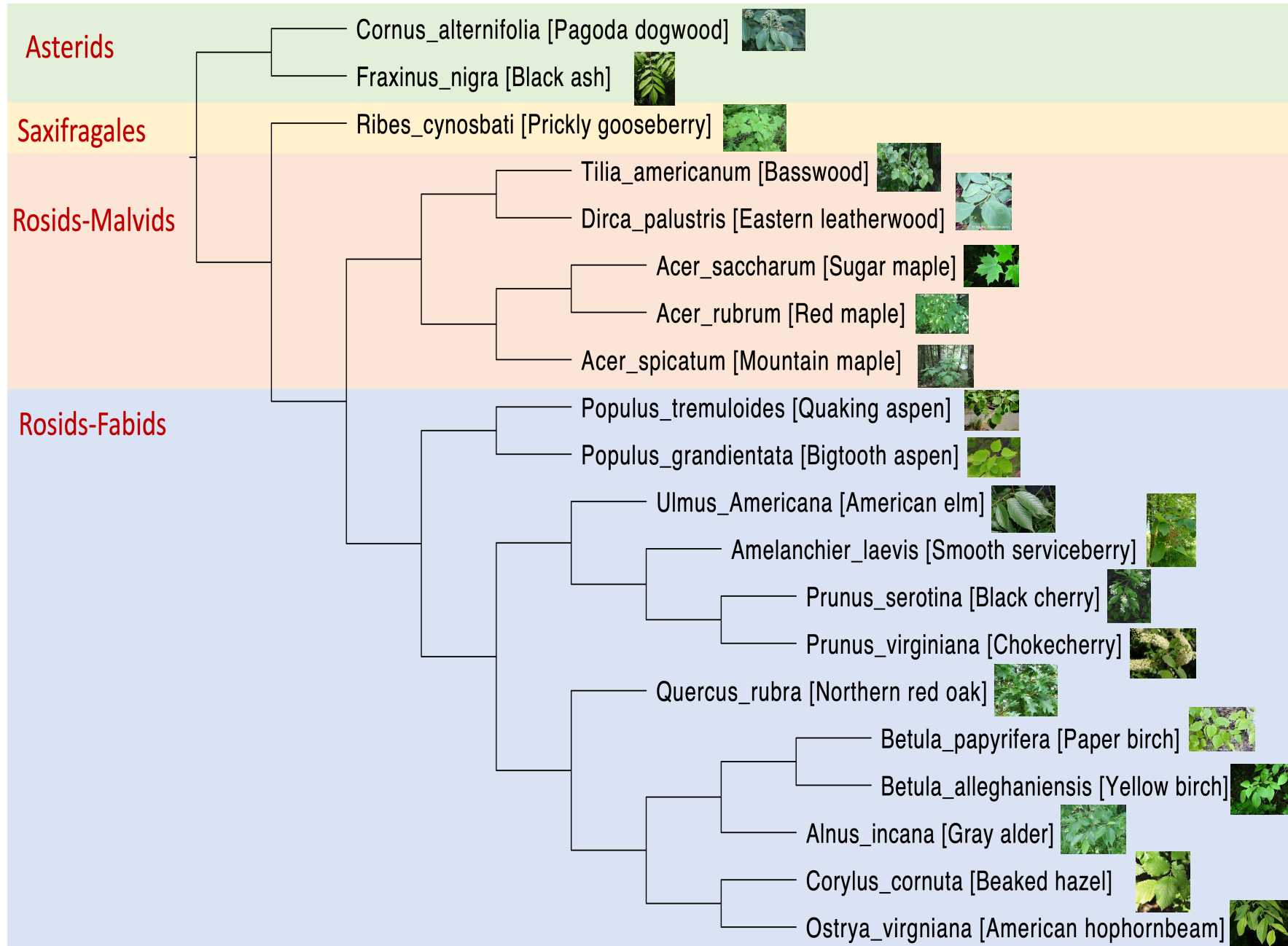
University of Notre Dame  
Environmental Research Center

Home About Research Education News and Events Contact

**Pristine  
Environments.  
World-Changing  
Research.**

Promoting environmental  
research and education

# Study species --- angiosperms, eudicots



# Methods

- Perform transcriptome assembly and annotation
- Investigate gene duplication and gene family size
- Assess gene family expansion and contraction

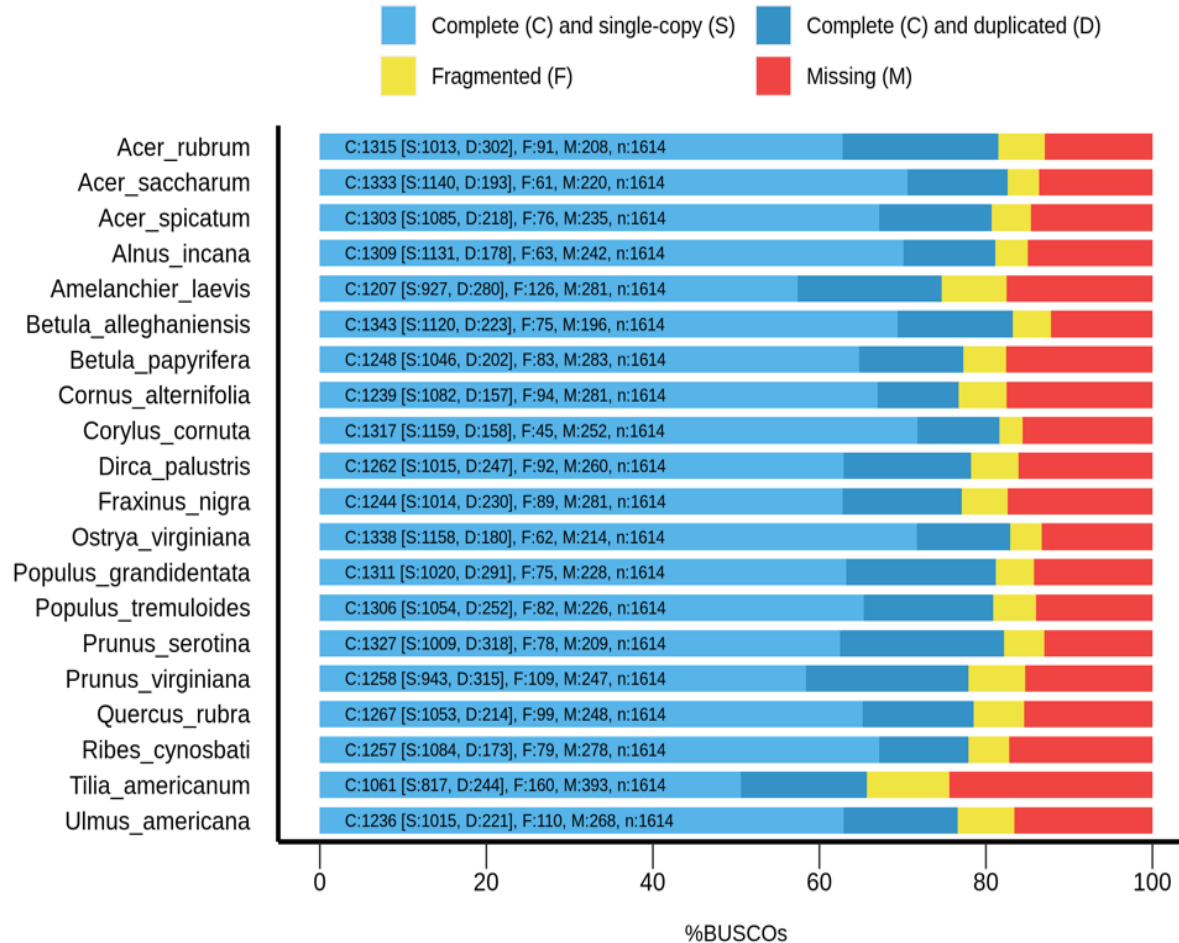
# Outline

- Perform transcriptome assembly
- Investigate gene duplication and gene family size
- Assess gene family expansion and contraction

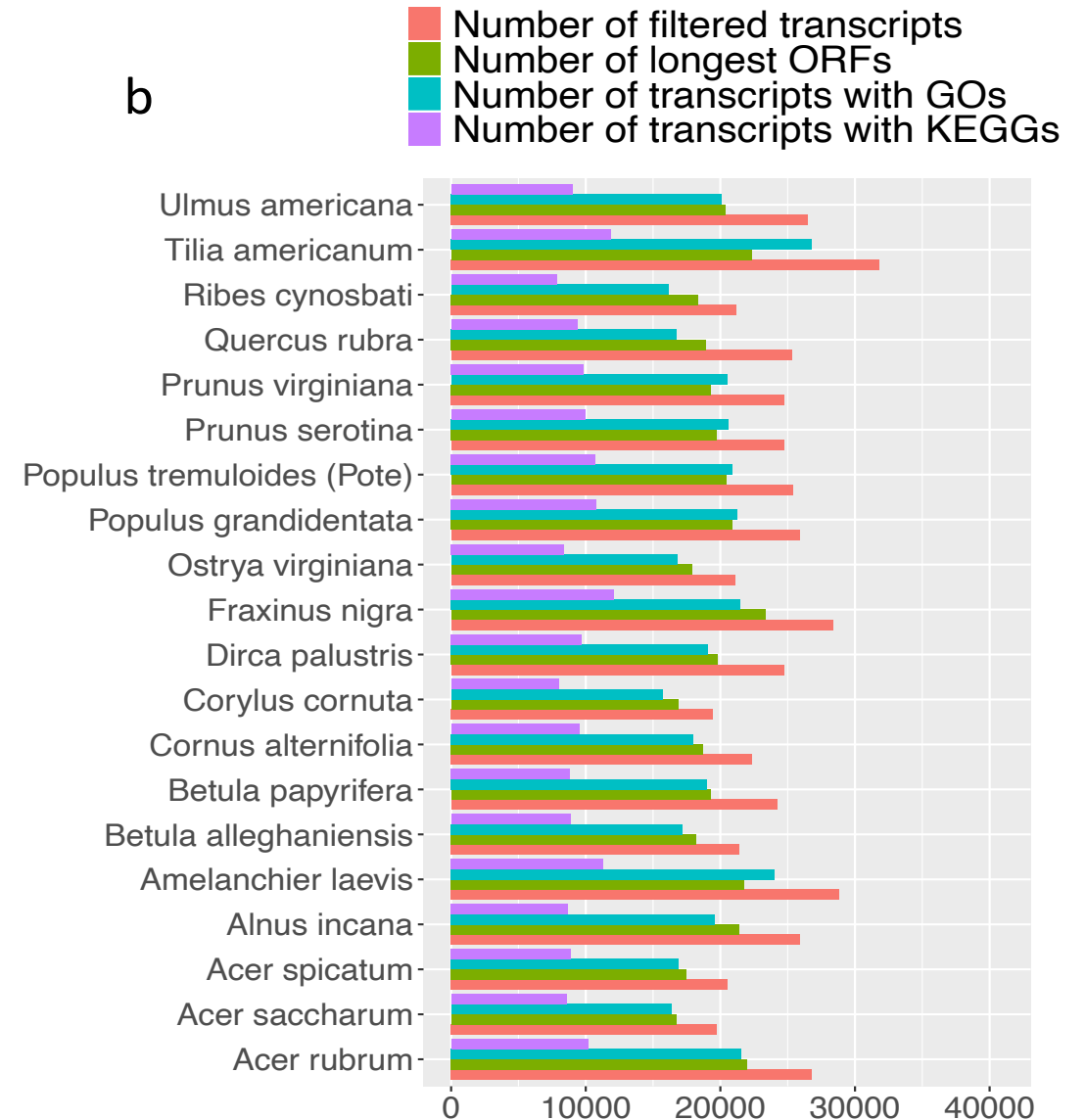
# Transcriptome assembly and annotation

a

## BUSCO Assessment Results



b



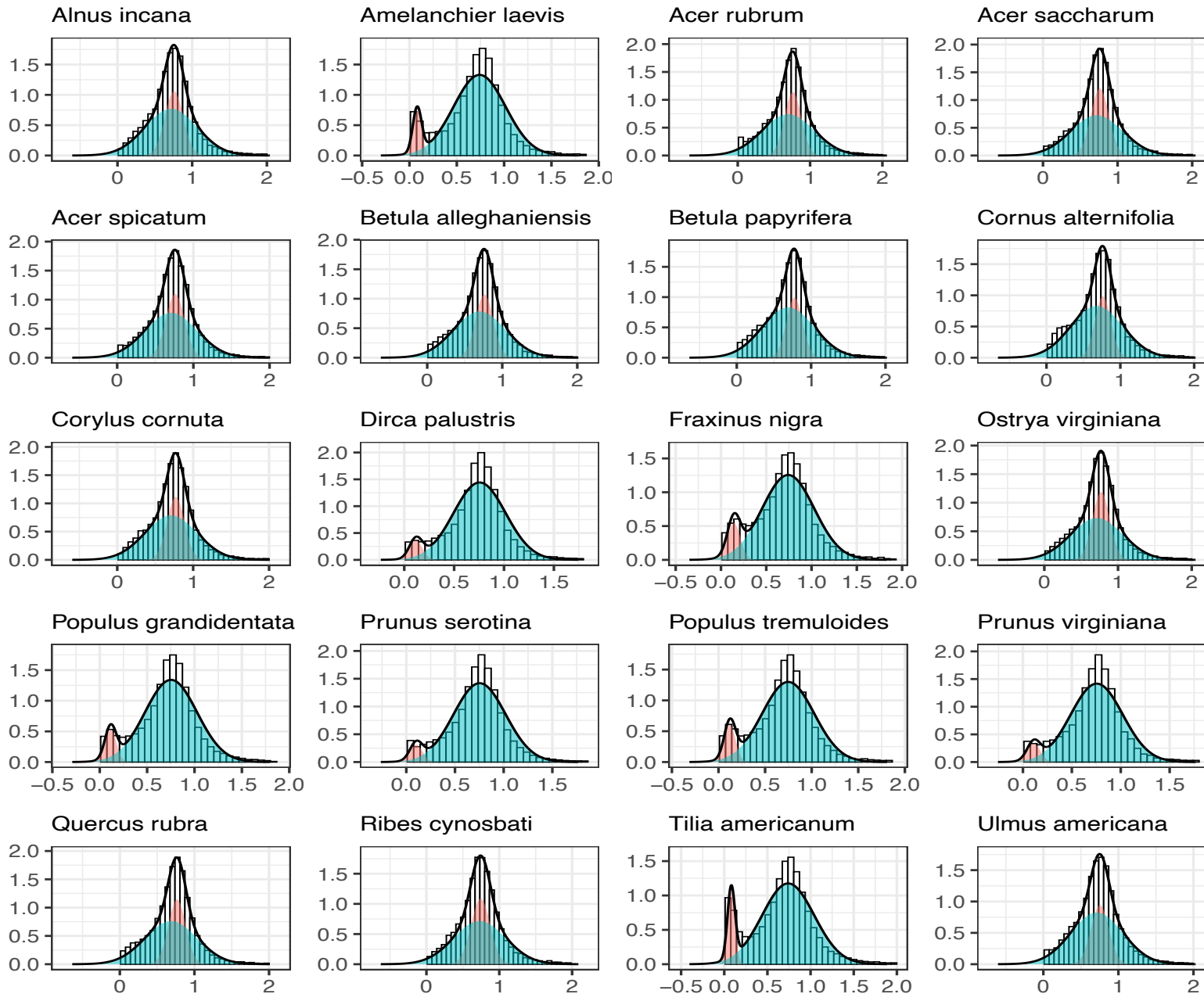
# Outline

- Perform transcriptome assembly
- Investigate gene duplication and gene family size
- Assess gene family expansion and contraction



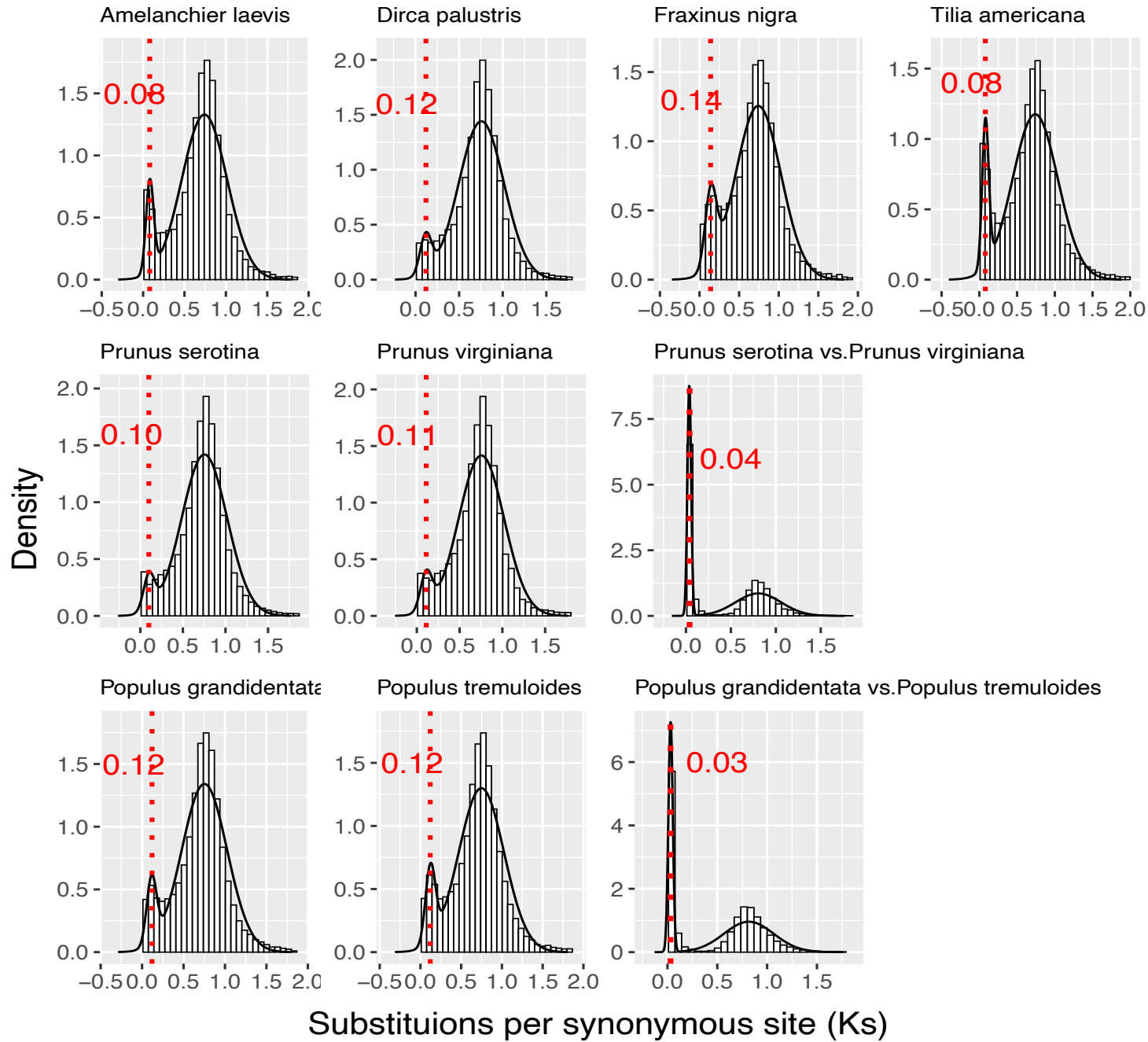
- Plot synonymous substitutions distribution ( $K_s$ ) to estimate the age of divergence.
- Estimate gene duplication events based on orthology.

# Distribution of Ks (the level of synonymous substitutions)

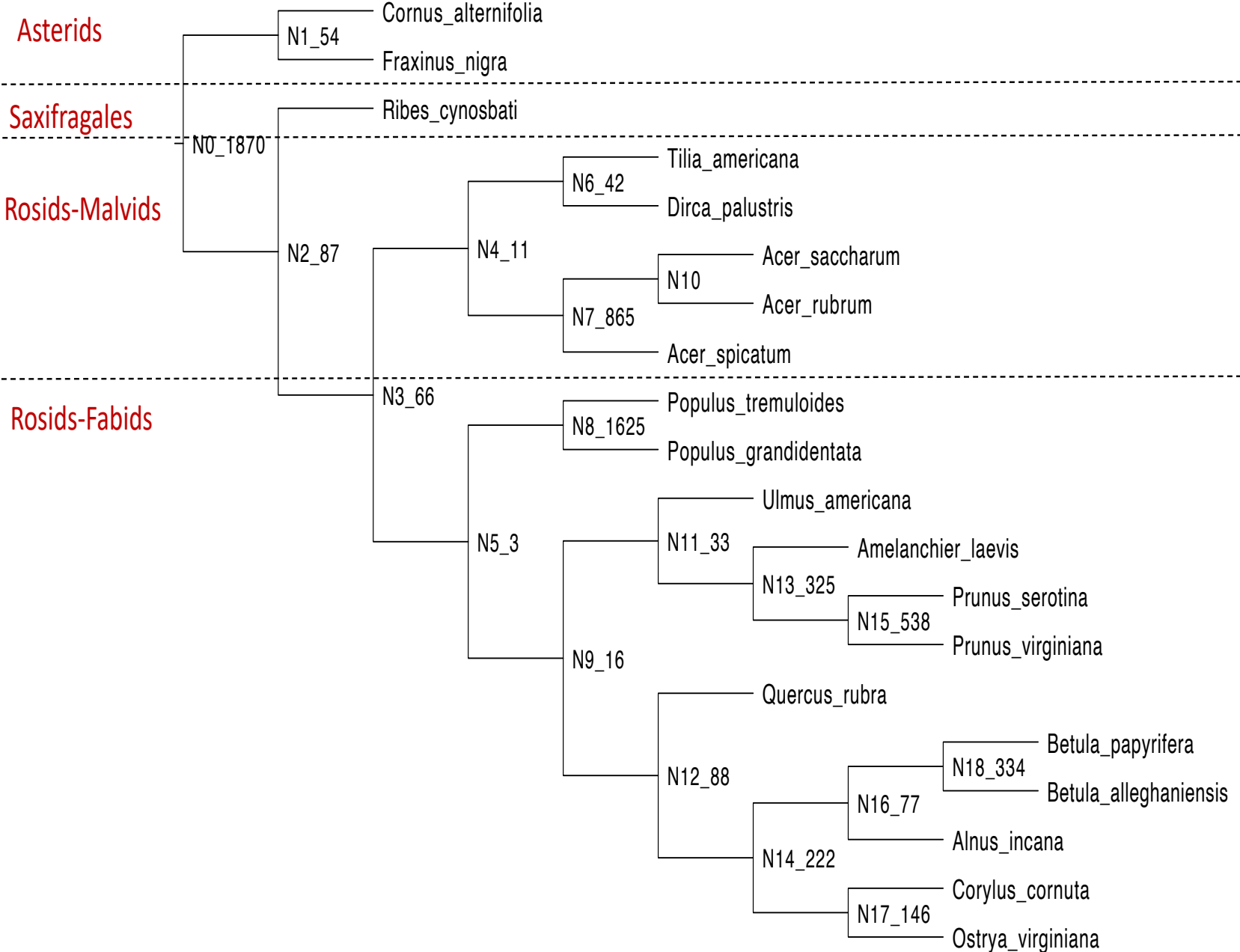


From the left to right on X-axis, larger peaks represent ancient gene duplication events; small peaks represent recent gene duplication

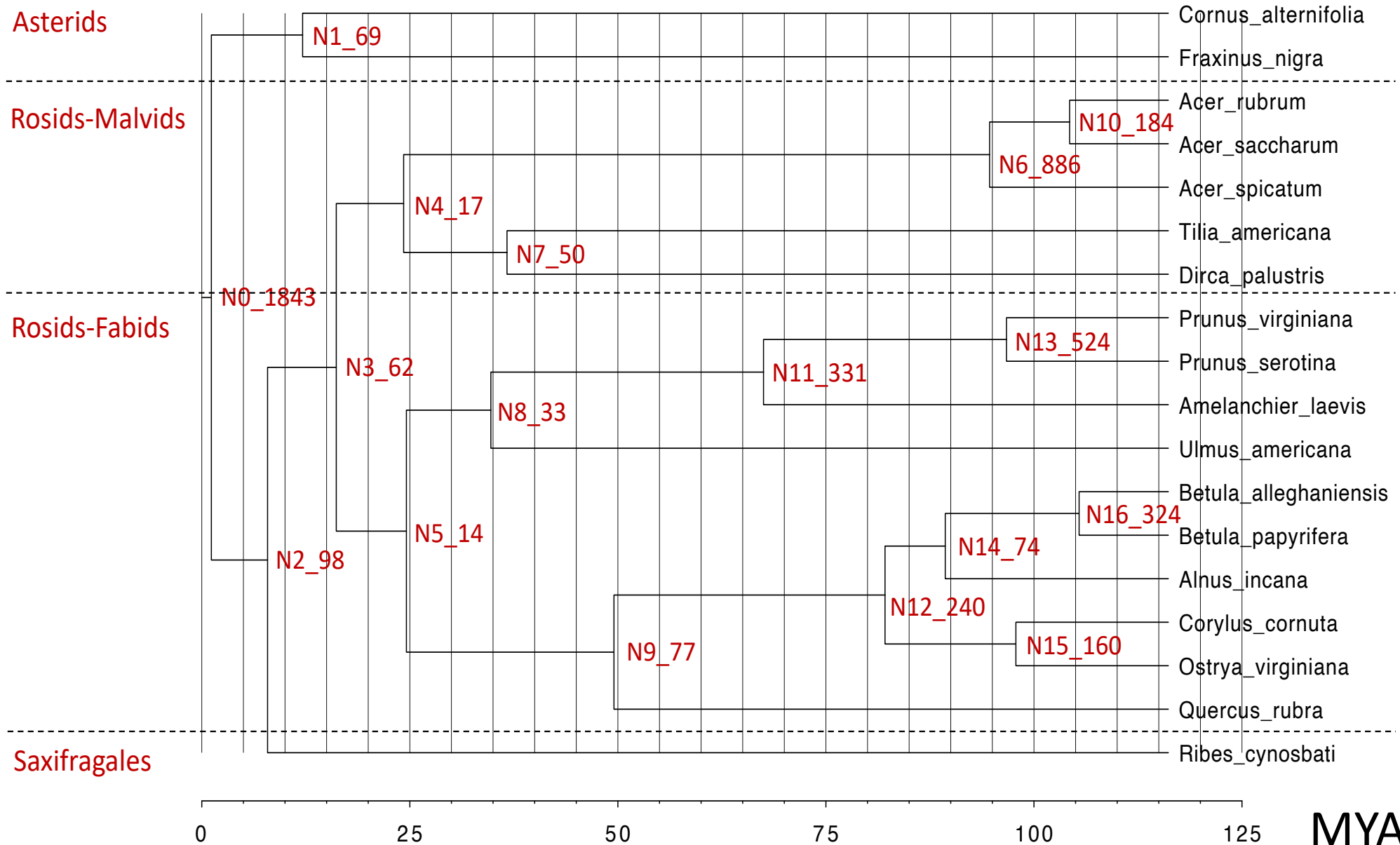
# Distribution of Ks (the level of synonymous substitutions)



# Gene duplication events shown on a given species tree



# Number of duplication events at node and divergence time

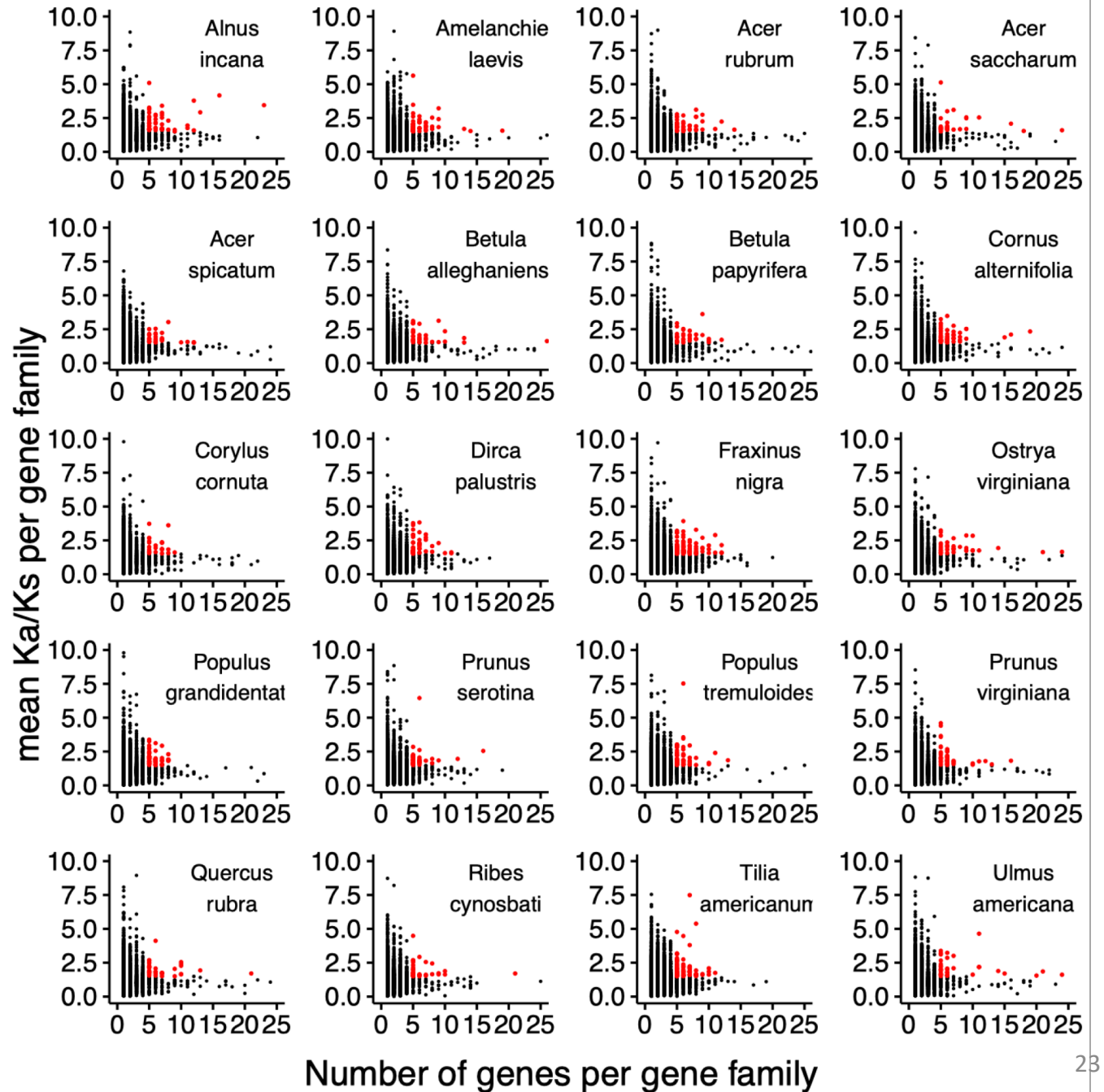


# Why do we care about gene duplication?

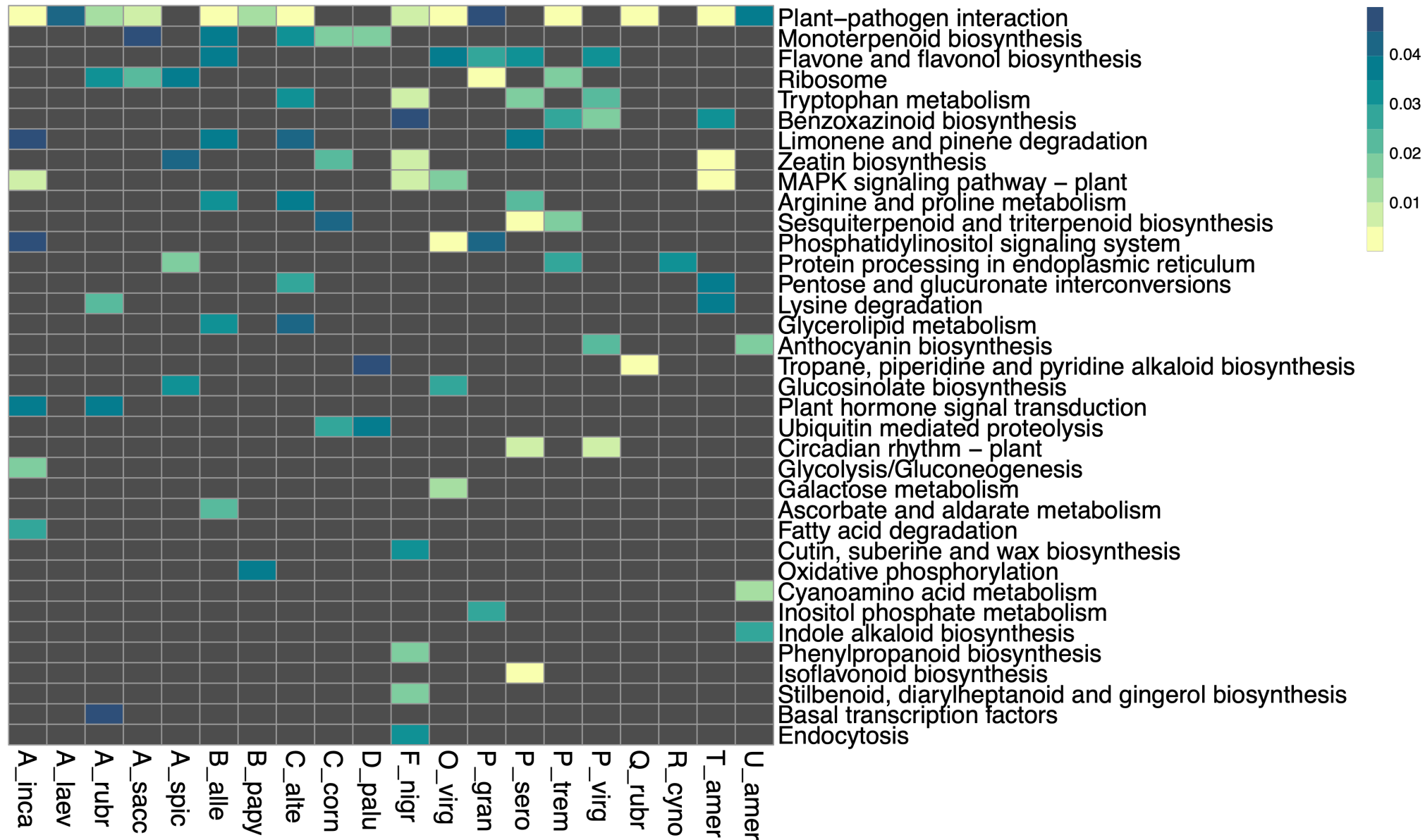
- Gene duplication are the origins of genetic diversity, providing materials for speciation and lineage-specific innovation.
- Most duplicate genes ended up being "dead", that's said, they were lost or pseudonized, yet a few could survive.

# Scatter plot between gene size and Ka/Ks per gene family

Highlighted gene families with a size  $\geq 5$  and Ka/Ks  $\geq 1.5$

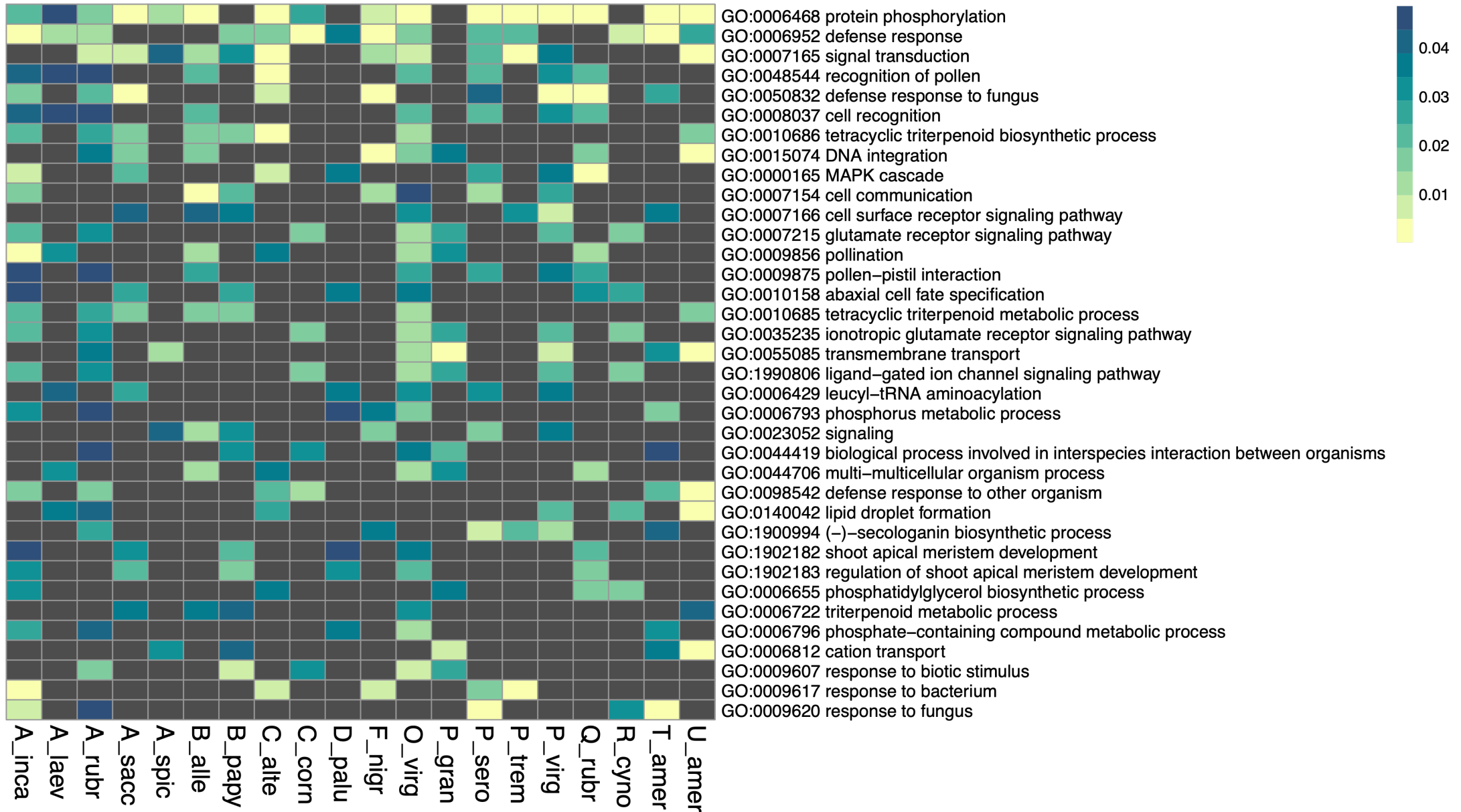


# KEGG pathway enrichment





# GO-biological process enrichment



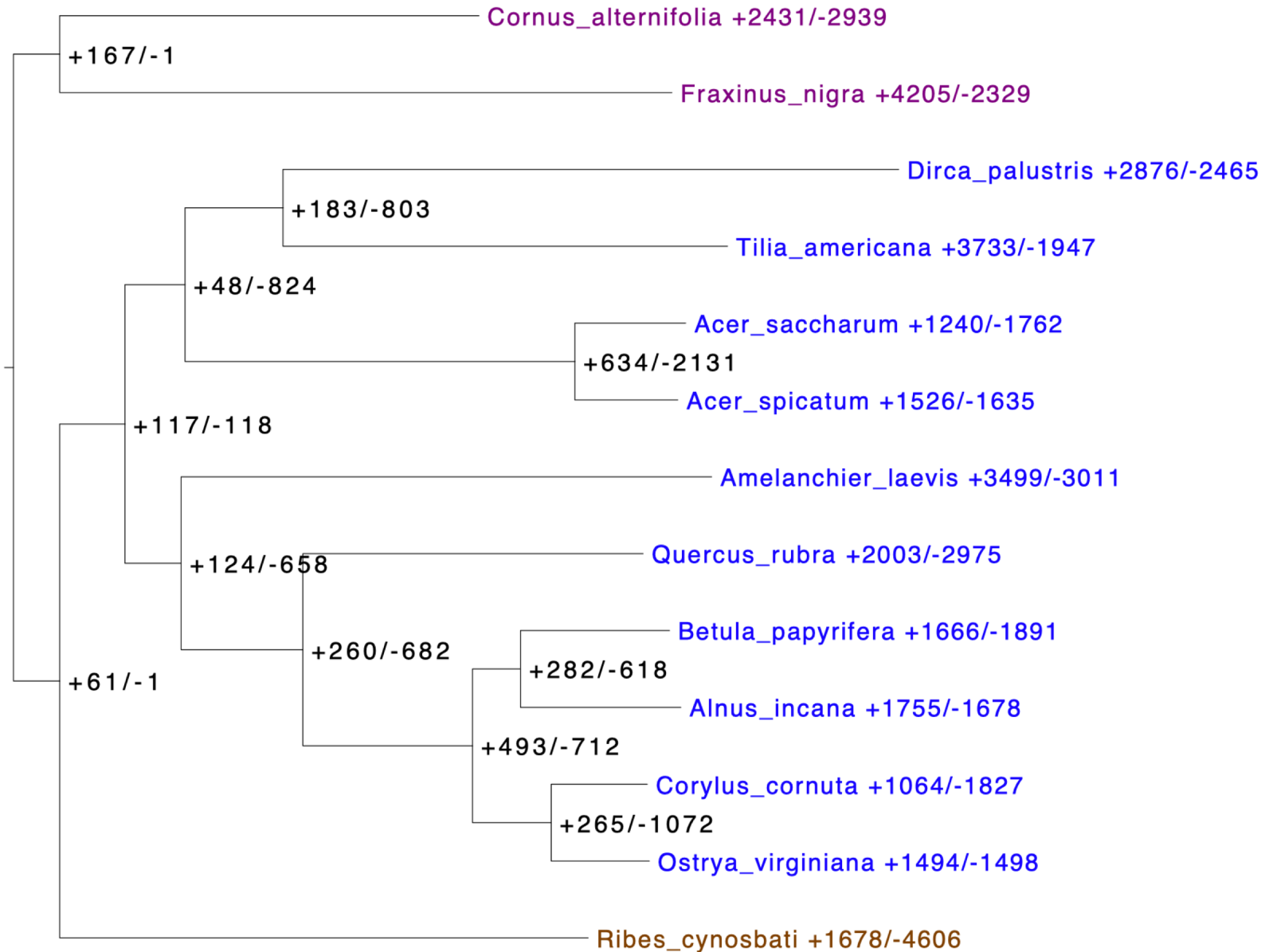
It seems gene families involved in plant-pathogen interaction and metabolite biosynthesis actively remain extensive variation.

# Outline

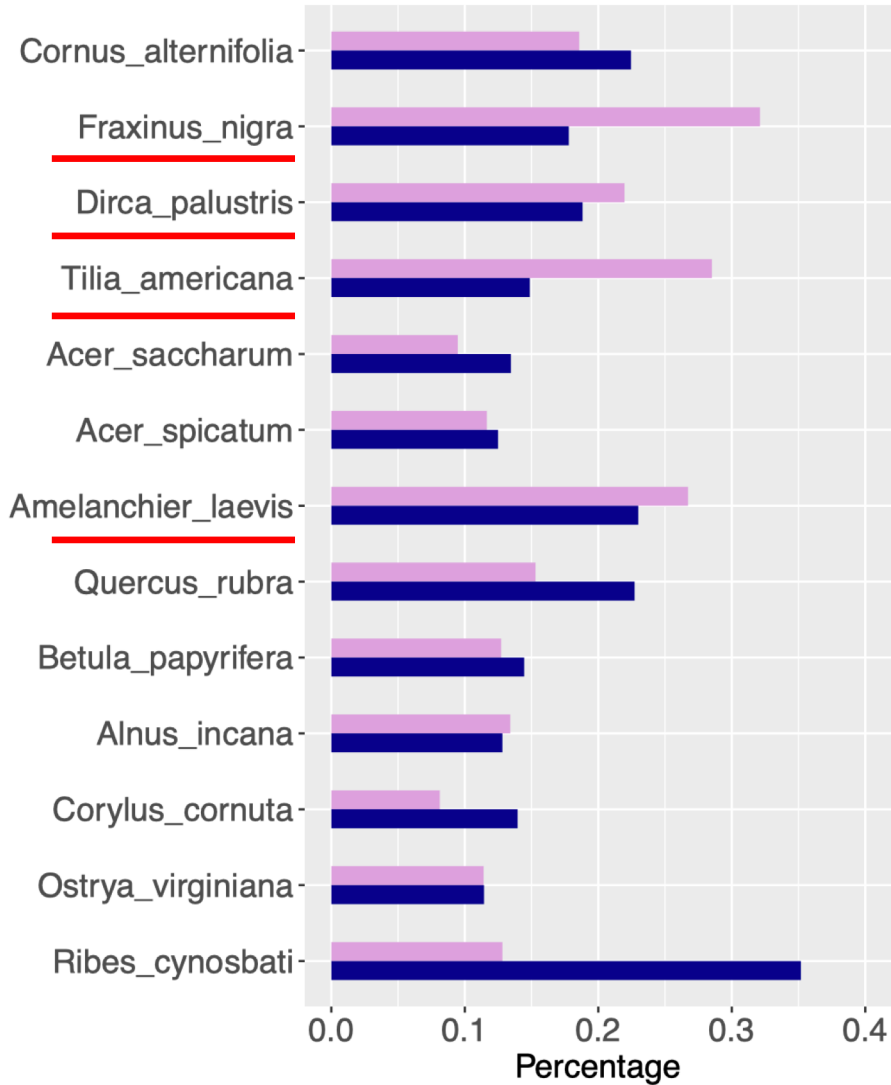
- Perform transcriptome assembly
- Investigate gene duplication and gene family size
- **Assess gene family expansion and contraction**

- Construct gene families based on orthologues using the program OrthoFinder
- Count gene numbers across species for each gene family
- Estimate the birth-death parameters (gene gain or gene loss) with the species tree and gene family counts using the program CAFE5

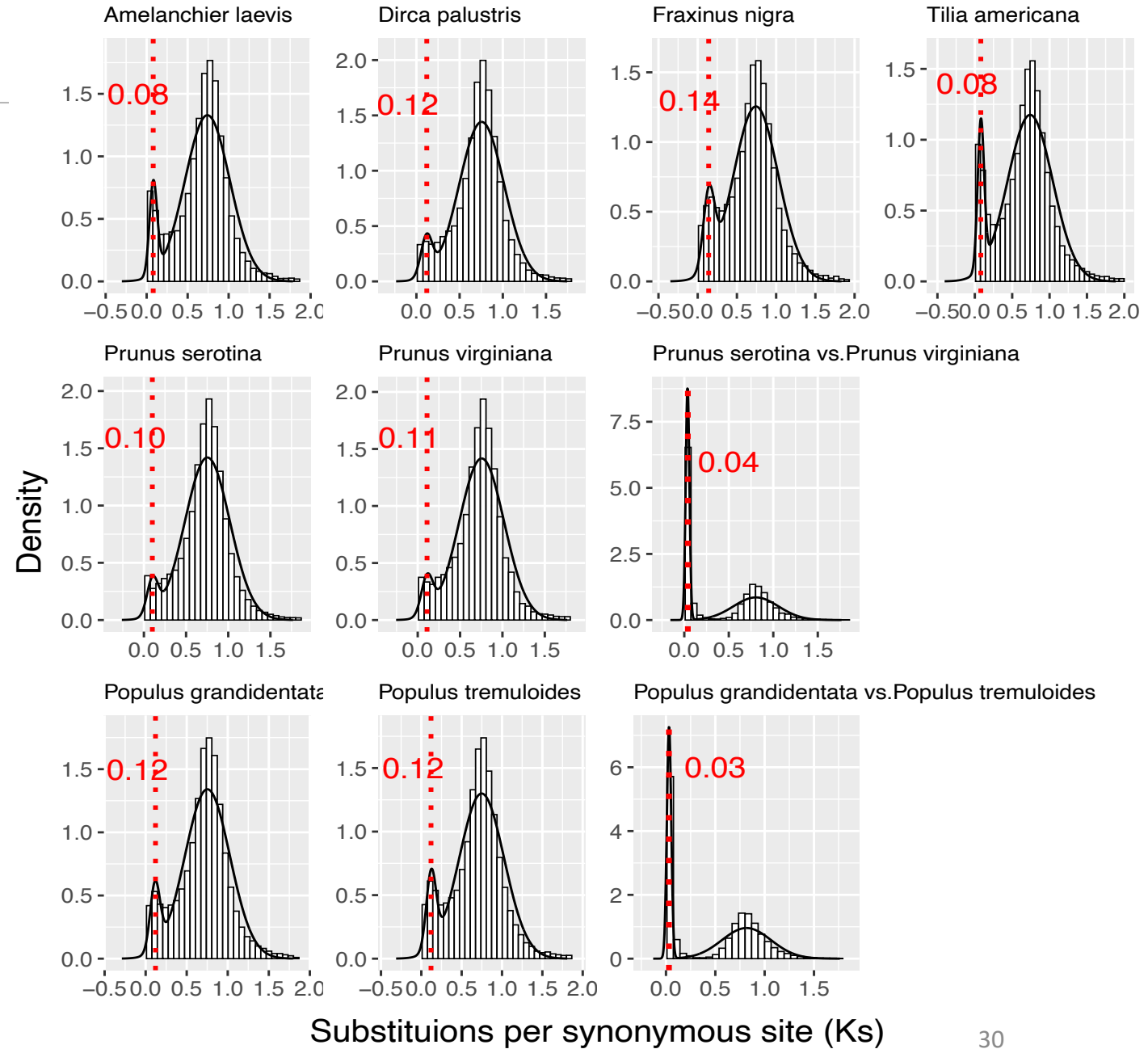
# Gene family expansion and contraction amongst diploid species



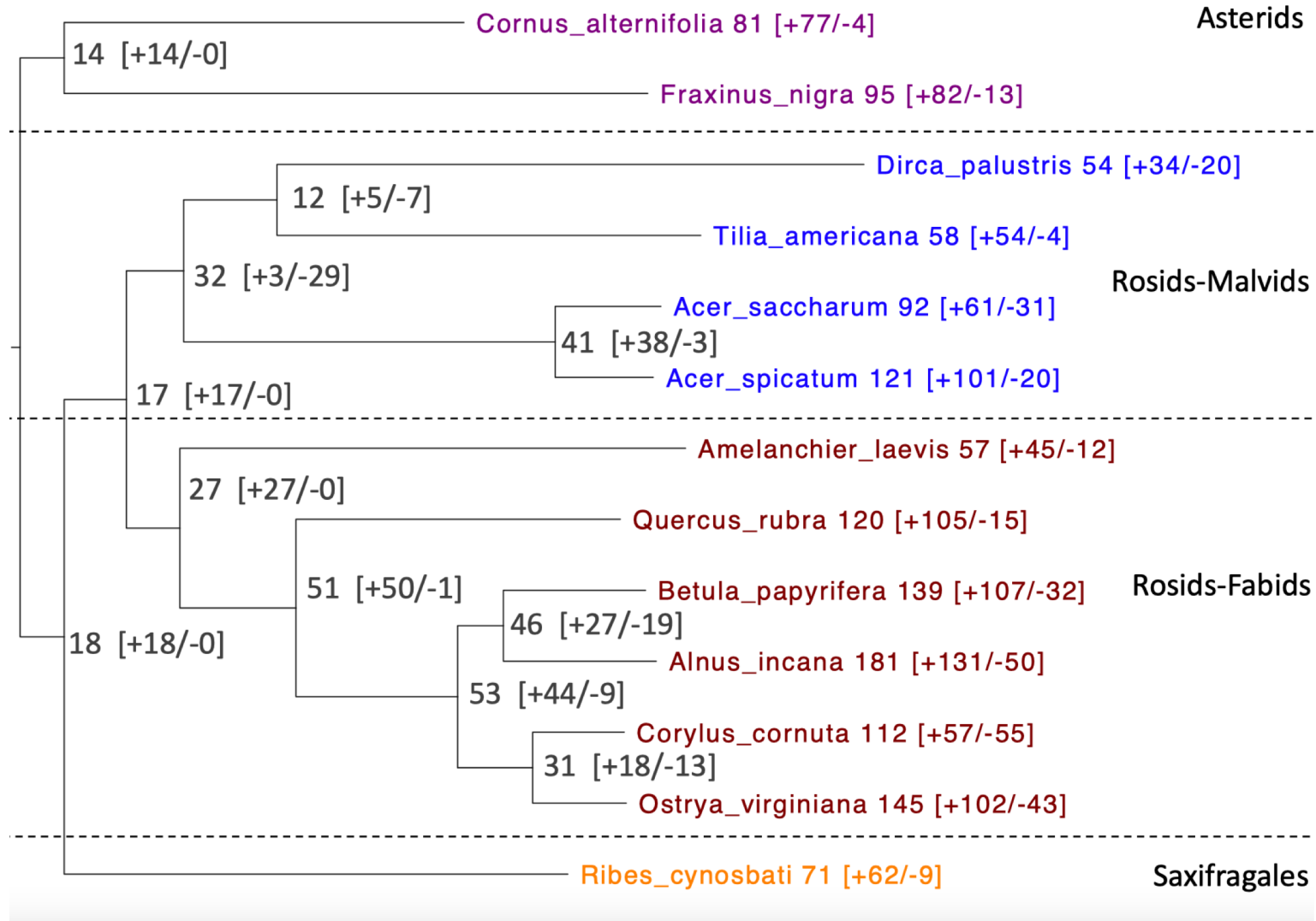
# Gene family expansion and contraction rate



■ Contraction  
■ Expansion



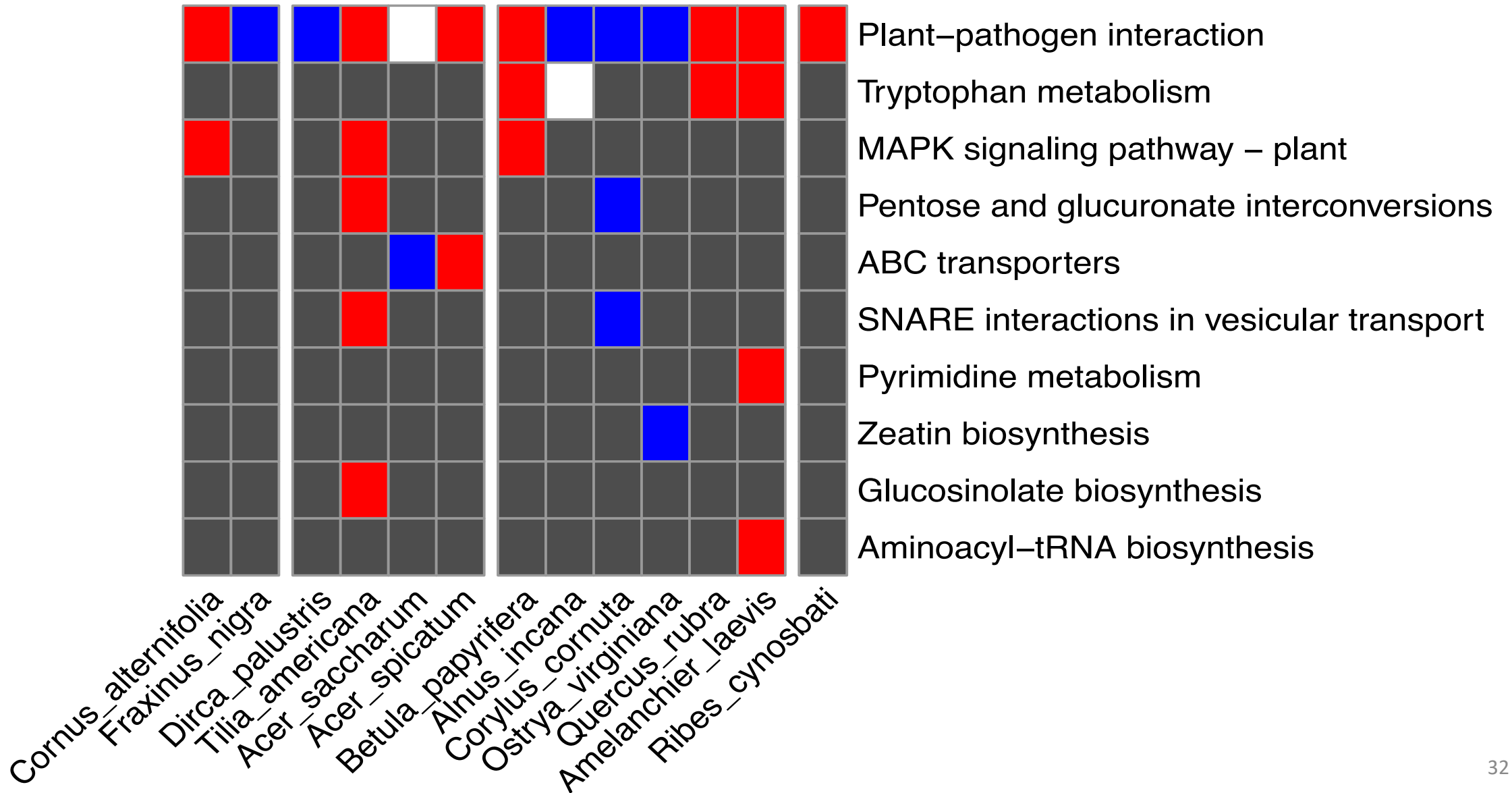
# Significant (Rapid) Gene family expansion and contraction



What are these fast evolving genes?

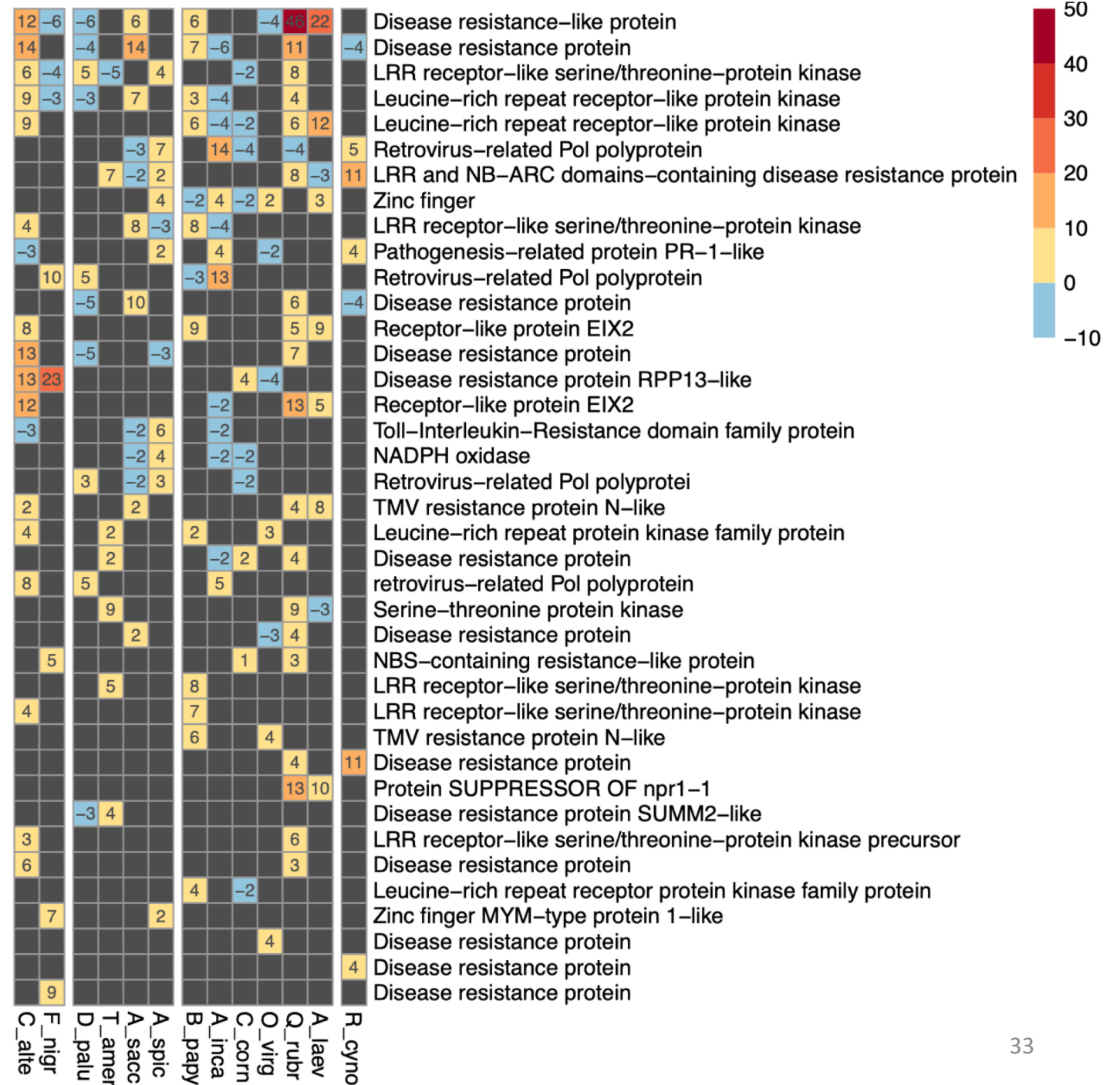
# Gene pathways enriched in expanded and contracted gene families

Red -- expanded; blue -- contracted; white – expanded and contracted





# Gene families belong to the fast evolving plant-pathogen interaction



# Summary

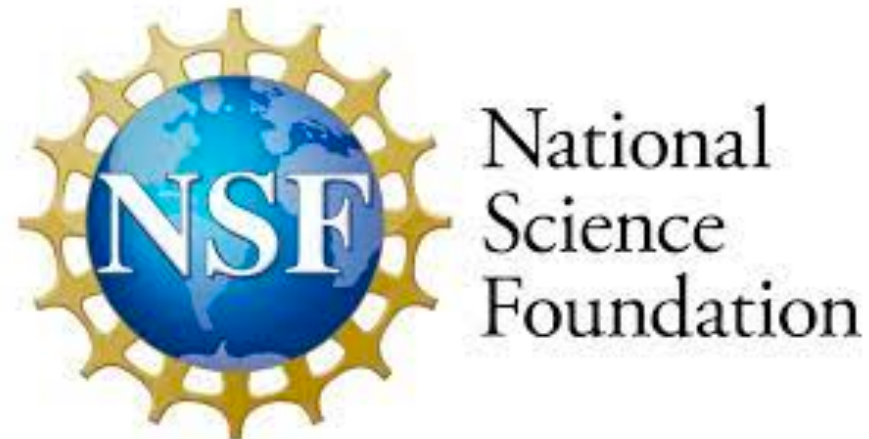
- The common WGDs were shared by all the study species, some species have recent duplication events, which contribute to the gene family expansions.
- Gene families involved in plant-pathogen interaction and metabolite biosynthesis actively remain extensive variation.
- Fast evolution gene families, namely expansion or contractions, have enrichment of plant-pathogen interaction genes.

Call attention to ...

Forest genetic diversity for adaptive capacity

- Nucleotide substitutions per site
- Gene birth and death per gene (gene copy number changes)

# Acknowledgement





# Thank you and questions?



UNIVERSITY OF  
NOTRE DAME

Mengmeng Lu  
January 15, 2023  
PAG 30, San Diego, CA





<http://www.quickmeme.com/meme/3585b1>