

# Preliminary annotation and analysis of the big berry manzanita (*Arctostaphylos glauca*) genome

Yi Huang

University of California, Riverside

Coauthors: Glen Morrison; Sarah Wilson; Zheng Li; Michael Barker;  
Amy Litt

# Manzanitas, *Arctostaphylos* (Ericaceae)

- Evergreen shrubs and trees
  - Red and twisting branches
  - Urn-shaped flowers
- Most diverse woody genus in California Floristic Province (CFP)
  - Over 100 species and subspecies
  - 104 out of 105 species are found within CFP (center of manzanita biodiversity)



# Project goal: Species delimitation in *Arctostaphylos*

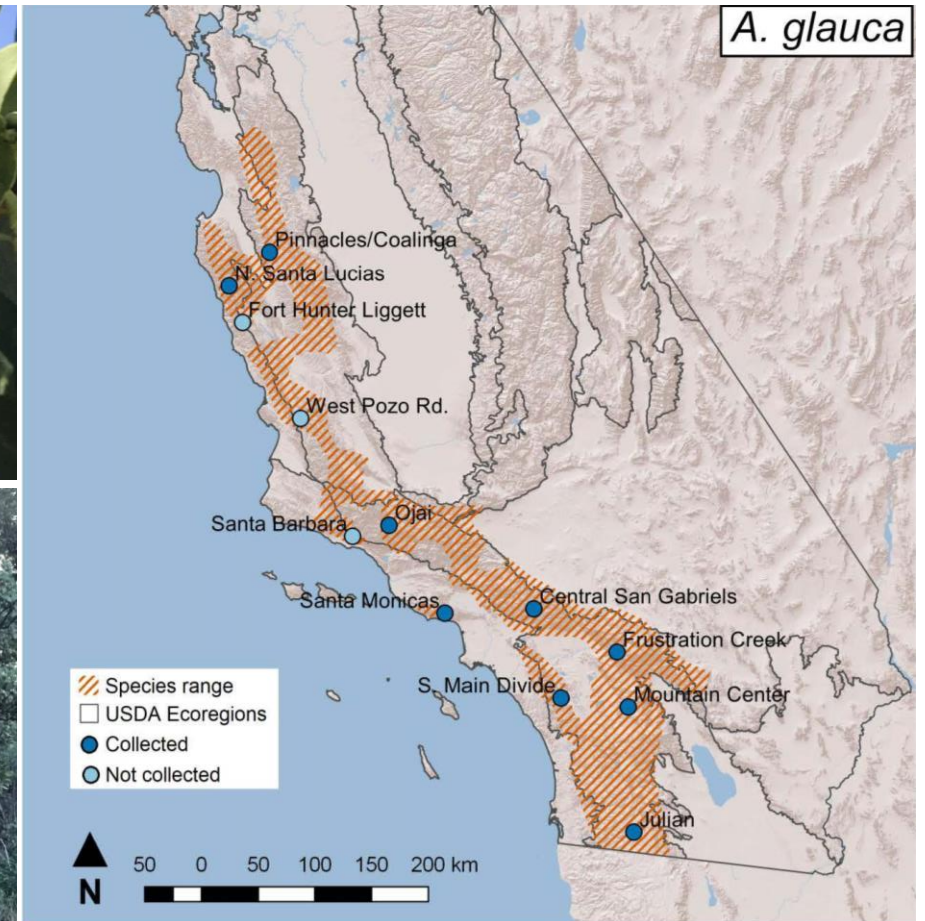
- Complicated patterns of interspecific and intraspecific variation
  - Hybridization
  - Morphological overlap
- Use reduced-representation genomic sequencing to test species delimitation
- Facilitate the detection of genetic variants across the genus
  - 10x more SNPs with reference genome





# *Arctostaphylos glauca*

- Diploid widespread manzanita species in CFP
  - $2n = 2x = 26$
- Ecological importance
  - Pollination
  - Restore vegetation in fire-disturbed area



# Outline of methods

- De novo genome assembly
  - Oxford nanopore sequencing and short-read Illumina sequence data by *Dovetail* company
  - BUSCO to assess completeness
- Genome annotation
  - Identify and mask repetitive element: RepeatMasker
  - Structural annotation: Funannotate pipeline (<https://funannotate.readthedocs.io/en/latest/index.html>)
- Preliminary intraspecific and interspecific genome comparison with other Ericales:
  - Synteny: MCScanX (<http://chibba.pgml.uga.edu/mcscan2/>)



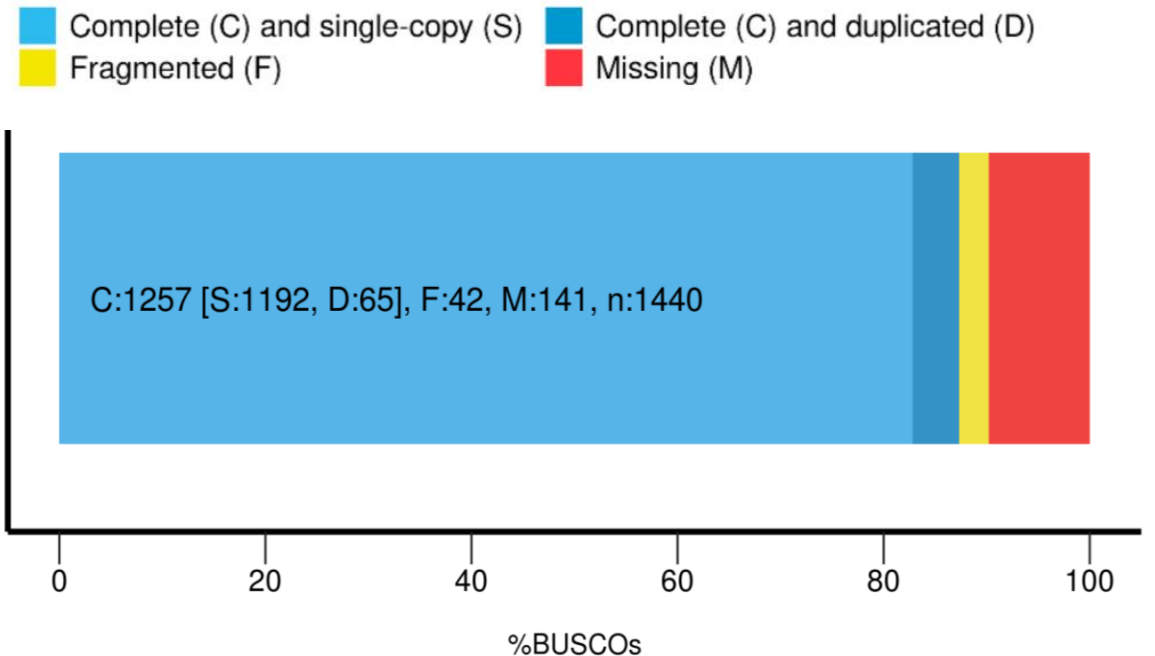
# De novo genomic assembly of *A. glauca*

- Good contiguity after assembly
- 87.3% completeness using embryophota\_odb9 lineages in BUSCO

## Statistics of Assembled Genome

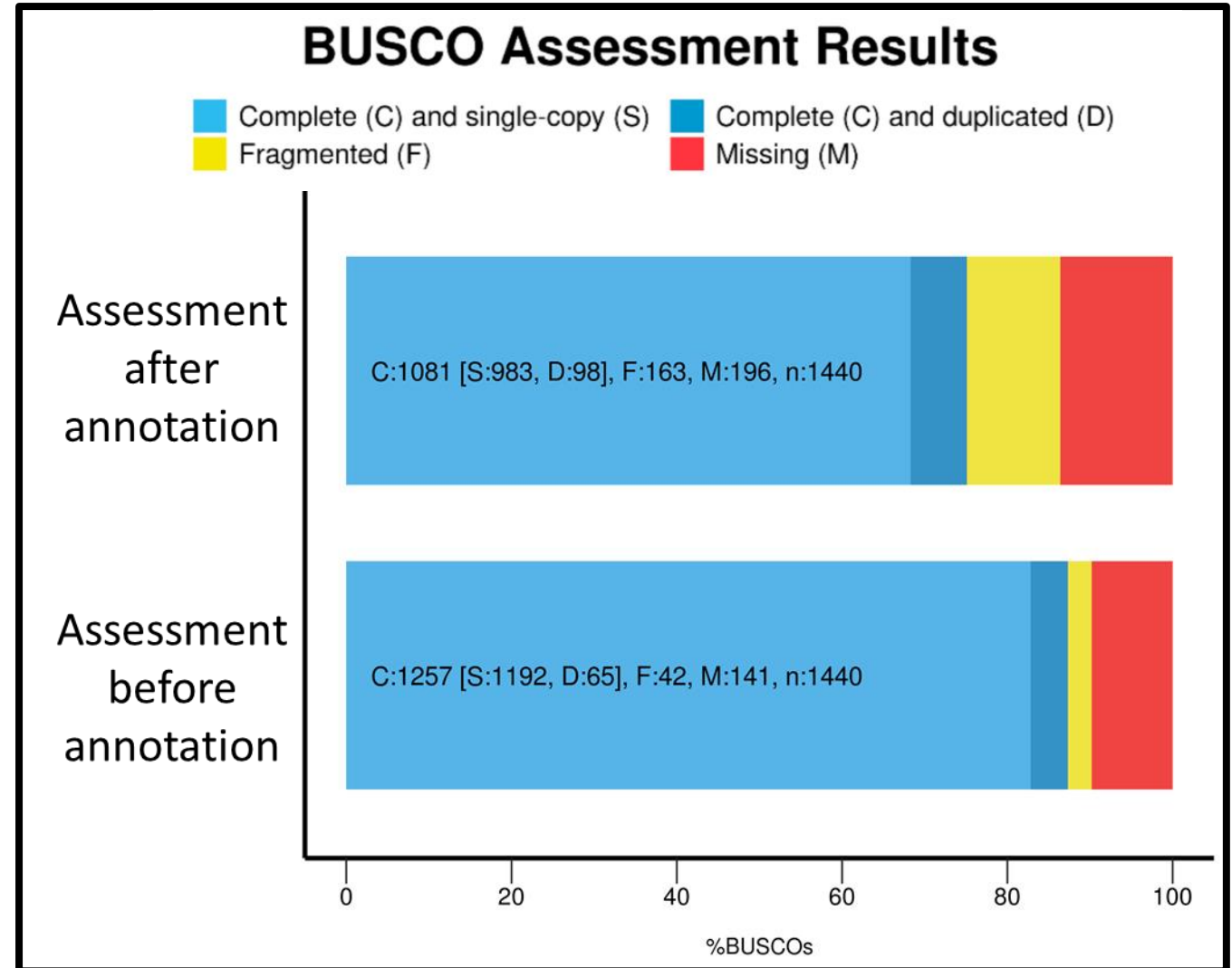
Genome size	567.62Mb
N50	38.767Mb
N90	33.601Mb
Chromosome-level assembly	13 scaffolds > 1 Mb
Percent of genome in gaps	0.09%

## BUSCO Assessment Results



# Genome annotation summary

- Repetitive elements: 45.62% of the genome
- Number of protein coding genes: 40,069
- Average length of protein coding genes: 3024.83bp
- BUSCO assessment: 75.1% completeness



# Pattern of collinearity within *A. glauca* genome

---

## Statistics of Synteny Analysis within *A. glauca* genome

---

Number of Collinear Genes	8387
Number of All Genes	40,069
Percentage of collinear genes	20.93

---

*MCSanX*: a toolkit for detection and evolutionary analysis of gene synteny and collinearity

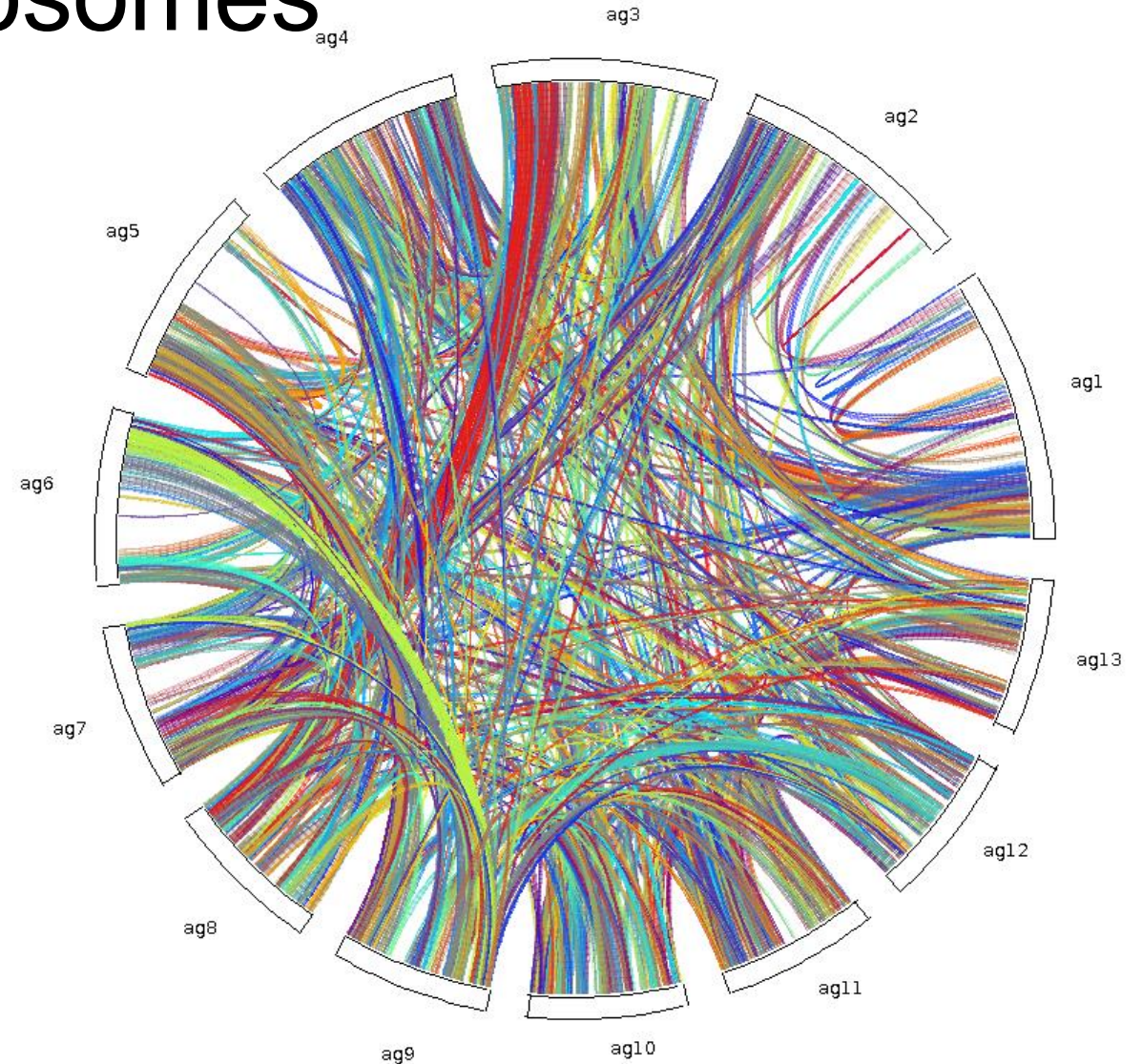
[Yupeng Wang](#),<sup>1,2</sup> [Haibao Tang](#),<sup>1,3,4</sup> [Jeremy D. DeBarry](#),<sup>5</sup> [Xu Tan](#),<sup>1,3</sup> [Jingping Li](#),<sup>1,2</sup> [Xiyin Wang](#),<sup>1,6</sup> [Tae-ho Lee](#),<sup>1</sup>  
[Huizhe Jin](#),<sup>1,2</sup> [Barry Marler](#),<sup>1</sup> [Hui Guo](#),<sup>1,3</sup> [Jessica C. Kissinger](#),<sup>2,5,7</sup> and [Andrew H. Paterson](#)<sup>1,2,3,7,8,\*</sup>

► [Author information](#) ► [Article notes](#) ► [Copyright and License information](#) [Disclaimer](#)



# Visualize collinear regions between *A. glauca* chromosomes

- Circle plot from MCScanX suggests substantial WGD events
  - 13 white bars on the edge of plot represent 13 chromosomes of *A. glauca*
  - Each colored bundle represents a collinear block of a syntenic gene pair



# Complicated pattern of collinearity may result from the large number of total genes

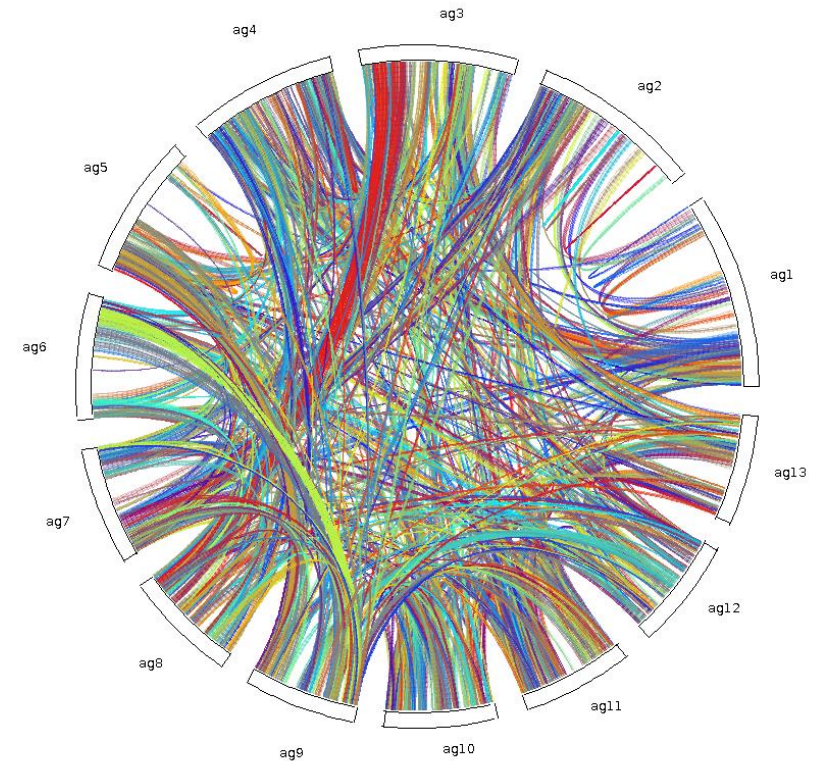
---

## Statistics of Synteny Analysis within *A. glauca* genome

---

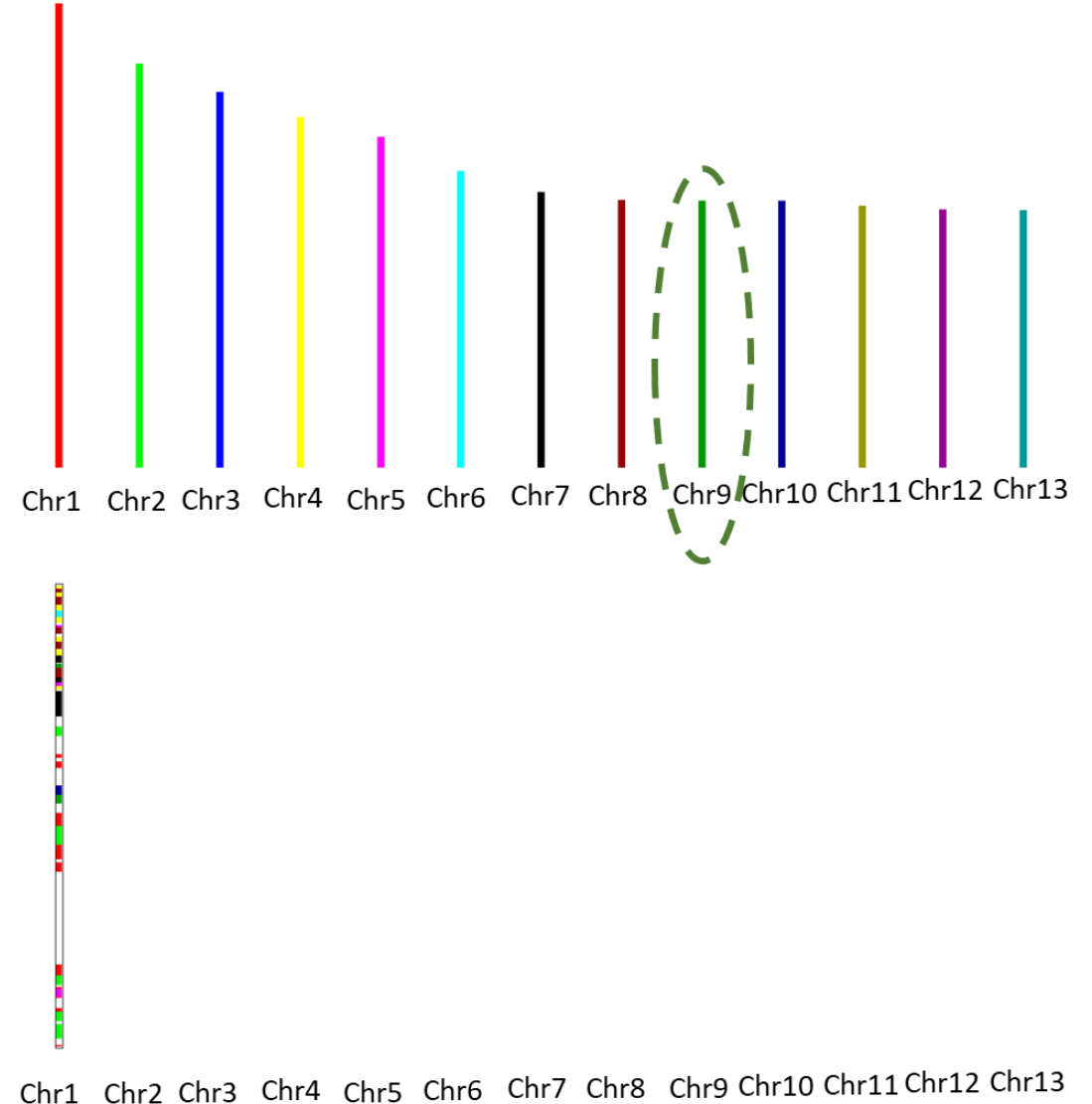
Number of Collinear Genes	8387
Number of All Genes	40,069
Percentage of collinear genes	20.93

---



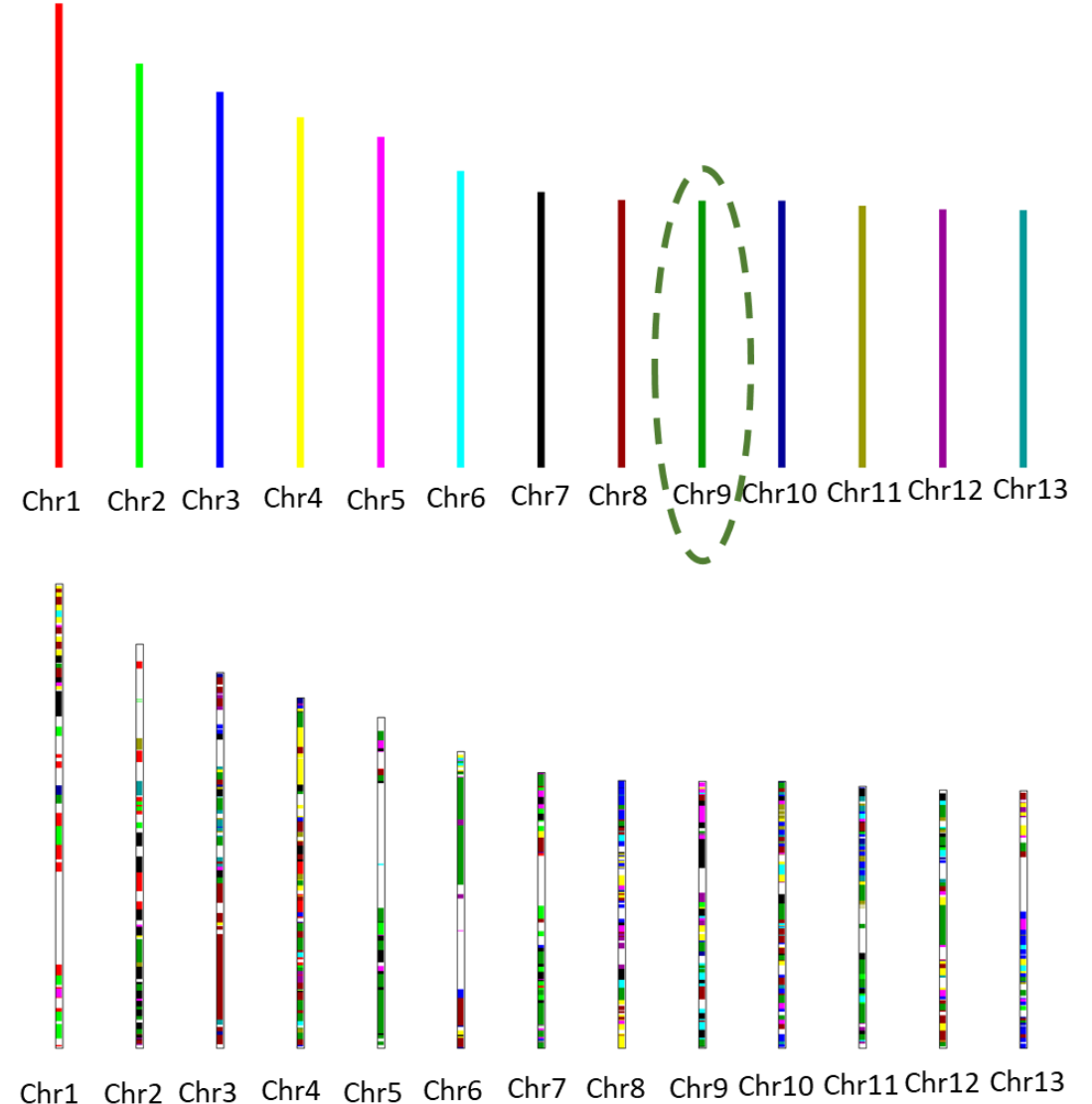
# The estimated proportion of segmental duplications on each chromosome

- Bar plot from MCScanX also suggests substantial duplication events within *A. glauca*
  - The corresponding origin for each chromosome is indicated in color
  - Segmental regions of chromosome 9 have been observed across other chromosomes



# The estimated proportion of segmental duplications on each chromosome

- Bar plot from MCScanX also suggests substantial duplication events within *A. glauca*
  - The corresponding origin for each chromosome is indicated in color
  - Segmental regions of chromosome 9 have been observed across other chromosomes



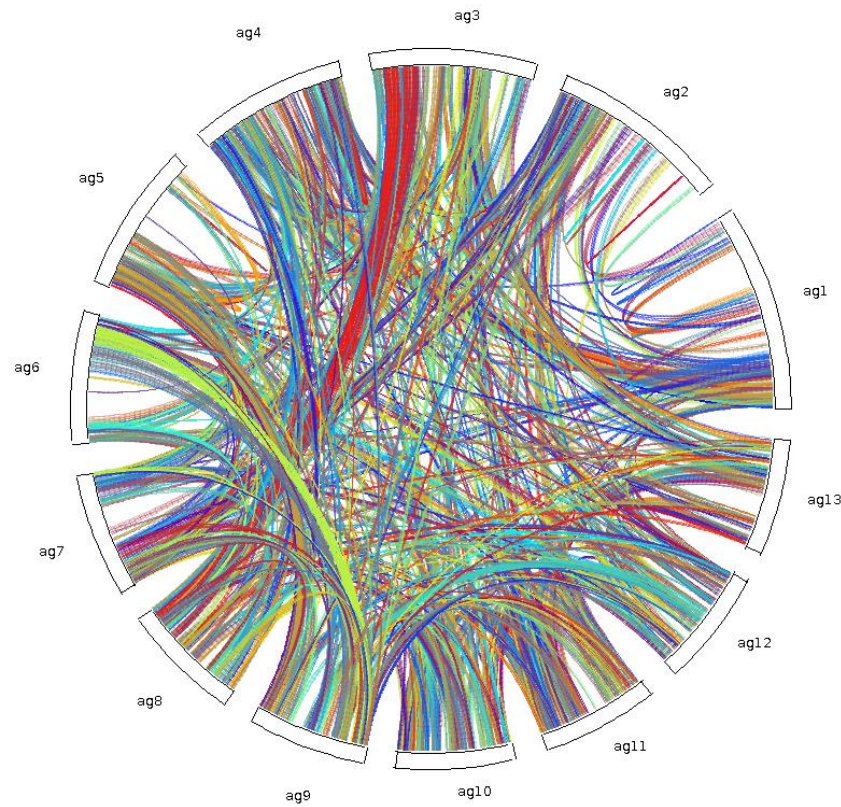


# Differences in duplicate gene origins among Ericales

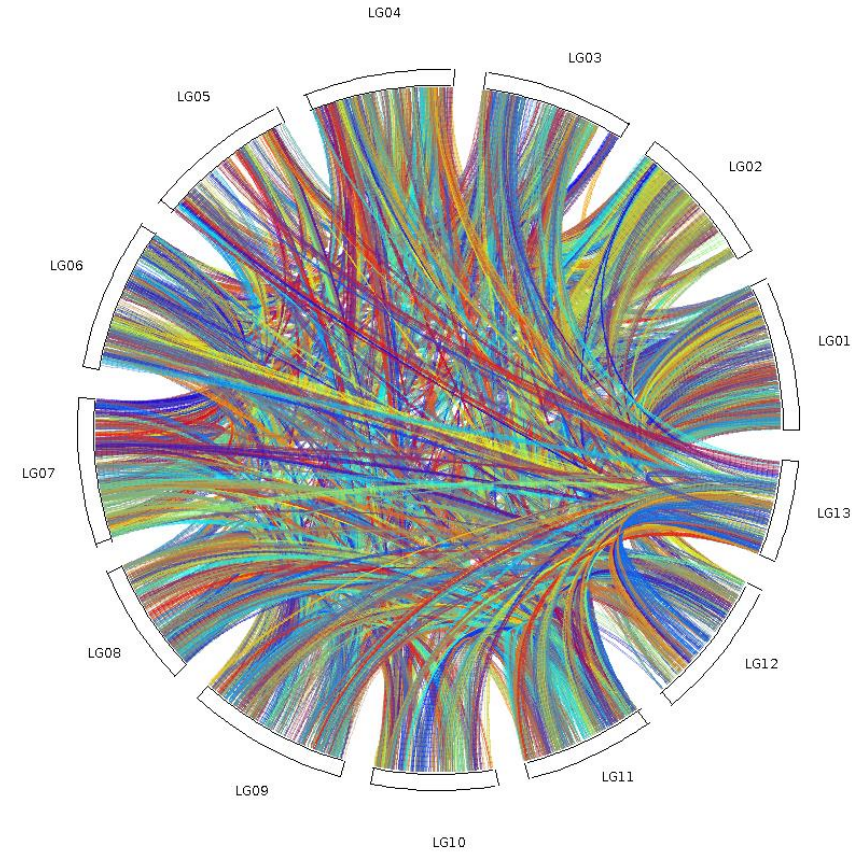
Species	No. of duplicate genes	No. of genes from different Origins (percentage)				
		Singleton	Dispersed	Proximal	Tandem	WGD
<i>Arctostaphylos glauca</i>	39,807	7,306 (17.7)	13,544 (34.0)	5285 (13.3)	5285 (13.3)	8387 (21.1)
<i>Rhododendron williamsianum</i>	21,375	3,037 (14.2)	6,608 (30.9)	651 (30.5)	2,558 (12.0)	8,521 (39.9)
<i>Actinidia chinensis</i> var. <i>chinensis</i>	33,116	1,600 (4.8)	2,295 (6.9)	274 (0.8)	1,655 (3.5)	27,292 (82.41)
<i>Actinidia eriantha</i>	42,943	3,802 (8.9)	8,381 (19.5)	2,072 (4.8)	2,078 (4.8)	26,655 (62.1)
<i>Primula veris</i>	18,285	2,871 (15.7)	13,319 (72.8)	385 (2.1)	1,126 (6.17)	584 (3.0)

# Visualization of WGD in *A. glauca* and *R. williamsianum*

*A. glauca*

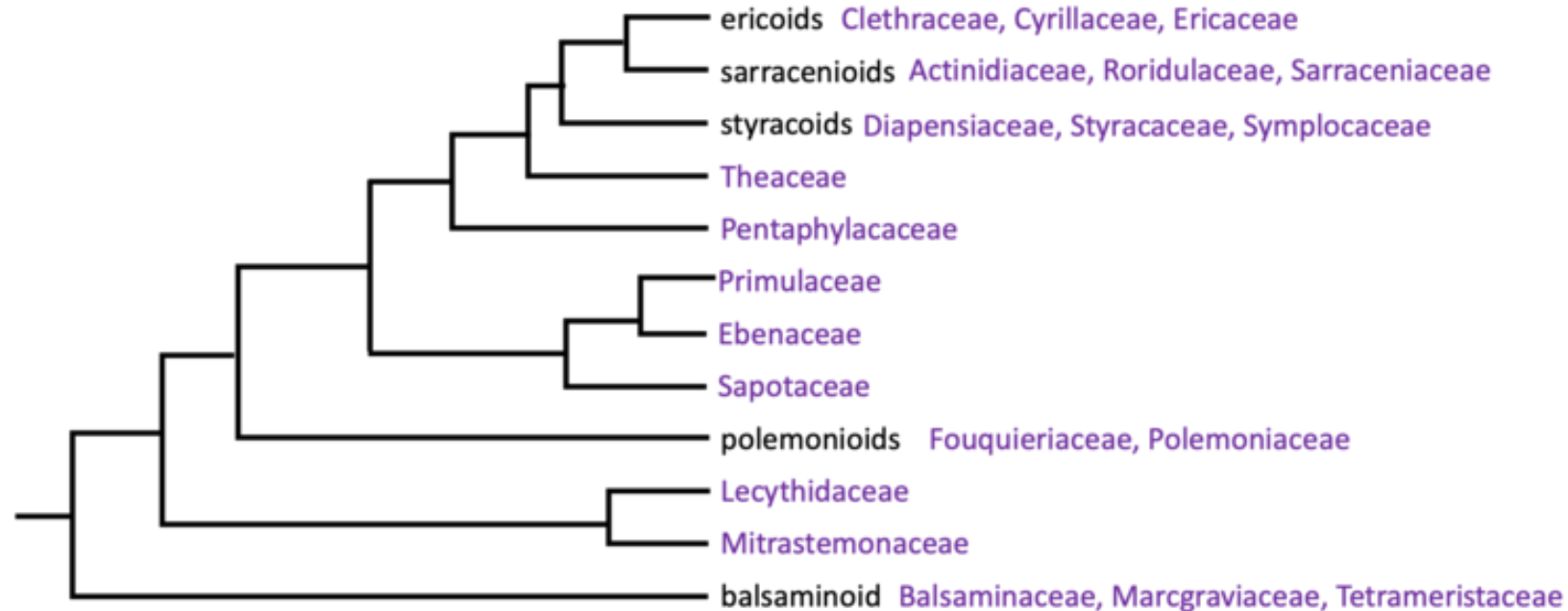


*R. williamsianum*



# Limitation and future work

- Too many genes
  - RNA-seq data to facilitate structural and functional annotation
  - Additional genome sequencing to reduce the gap
- Include more genomes of Ericales



# Acknowledgement

- PI: Amy Litt
- Glen Morrison
- Sarah Wilson
- Alex Rajewski
- Zheng Li
- Michael Barker



## California Conservation Genomics Project

Leading the nation's first statewide conservation genomics initiative

[ABOUT US](#) [LEADERSHIP TEAM](#) [PROJECTS](#) [SPECIES](#) [PRESS](#) [DOWNLOADS](#) [CONTACT](#)







Thank You!

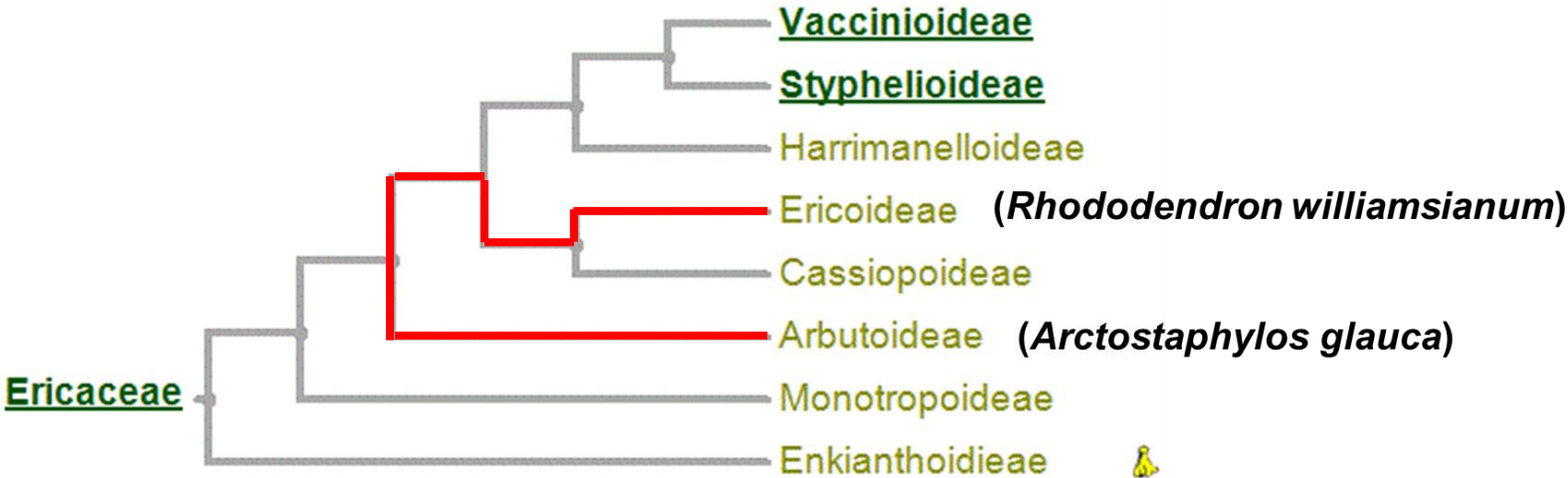


# Self-genome comparison and cross-genome of *A. glauca* & *R. williamsianum*

Reference genome	Target genome	Maximum duplication depths	Duplication depth with maximum number of genes
A. glauca	R. williamsianum	37	2
R. williamsianum	A. glauca	26	3
R. williamsianum	R. williamsianum	46	4
A. glauca	A. glauca	22	0

Difference: 9

small scale  
chromosomal  
duplications



At least 9  
WGD events  
within lineage