

# Open Green Genomes (OGG): expanding our framework for comparative plant genomics



**A plant science community initiative including > 100 contributors in addition to JGI and AGI collaborators**

# OGG Co-PIs/Contributors

Victor Albert  
Hamid Ashrafi  
Donovan Bailey  
William Baker  
Mike Barker  
Annette Becker  
David Bell  
Jeff Bennetzen  
Dominique Bergmann  
David Bertioli  
Soraya Bertioli  
Adam Bewick  
Crysten Blaby  
Siobhan Brady  
C Robin Buell  
Steven Cannon  
Andre Chanderbali  
Mark Chase  
Gloria Coruzzi  
John Cushman  
Roger Deal  
Claude dePamphilis  
Joshua Der  
Katrien Devos  
Liam Dolan  
Jeff Doyle

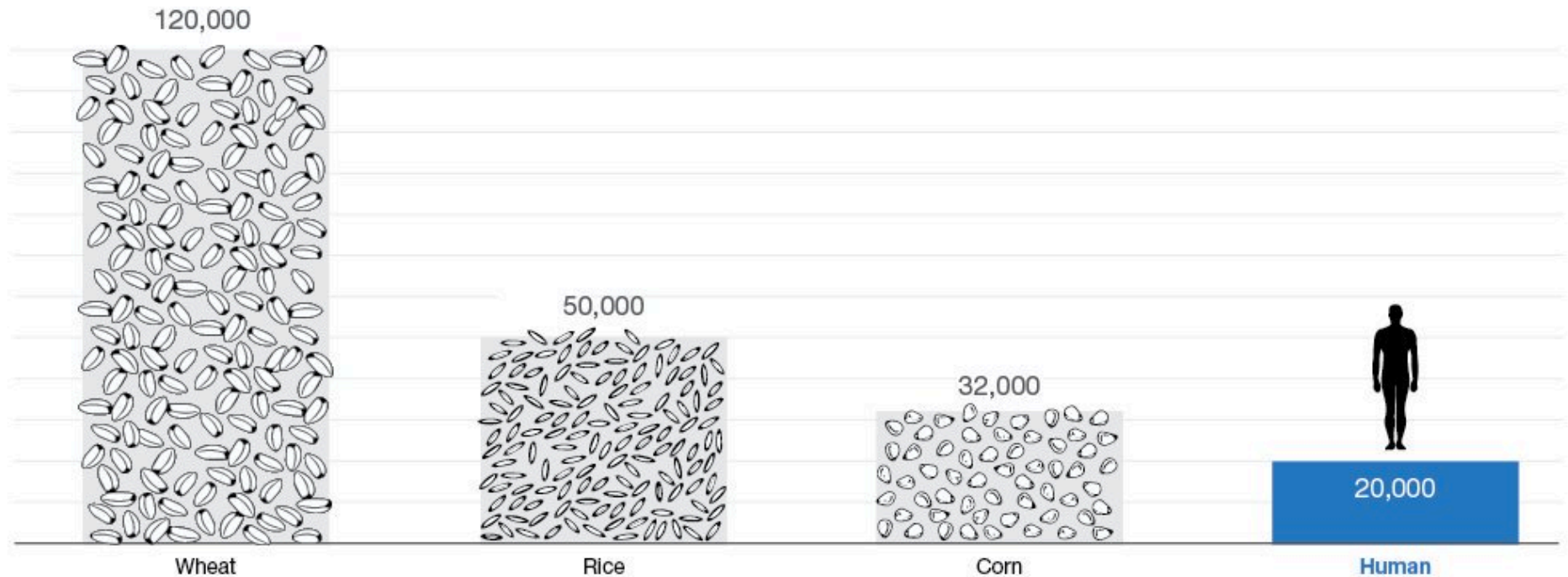
Patrick Edger  
Luis Eguiarte  
Paula Elomaa  
Zhangjun Fei  
Michael Freeling  
Mike Frohlich  
Vicki Funk  
Brandon Gaut  
Matthew Gitzendanner  
Thomas Givnish  
Olwen Grace  
Sean Graham  
Mike Grillo  
Alex Harkess  
James Hartwell  
Mitsuyasu Hasebe  
Samuel Hazen  
Katy Heath  
Karolina Heyduk  
Thomas Juenger  
Toby Kellogg  
Steven Kelly  
Hongzhi Kong  
Tianying Lan  
Jane Langdale  
Fay-Wei Li

Haiying Liang  
Charlotte Lindqvist  
Eric Lyons  
Hong Ma  
Jennifer Mandel  
Blaine Marchant  
Sarah Mathews  
Michael McKain  
Blake Meyers  
Ray Ming  
Siavash Mirarab  
Edwige Moyroud  
Lukas Mueller  
Chad Niederhuth  
Anne Osbourn  
François Parcy  
Gitte Petersen  
Chris Pires  
Lisa Pokorny  
Stefan Rensing  
Sue Rhee  
Carl Rothfels  
David Sankoff  
Scott Schlarbaum  
Bob Schmitz  
Eric Schranz

Charlie Scutt  
Ole Seberg  
Emily Sessa  
Erin Sigel  
Neelima Sinha  
Stephen Smith  
Doug Soltis  
Pamela Soltis  
Dennis Stevenson  
Peter Szovenyi  
Teemu Teeri  
Chung-Jui Tsai  
Jerry Tuskan  
Yves Van de Peer  
Esther van der Knaap  
Juan Carlos Villarreal  
Tandy Warnow  
Norm Wickett  
Alex Widmer  
Rod Wing  
Stephen Wright  
Todd Michael  
Shu-Miaw Chaw  
....Your Name Here.....

# Three things with more genes than you

Total number of genes per organism



Sources: AAAS, *The Gene* by Siddhartha Mukherjee



**Bill Gates** @BillGates · 4 Dec 2016

This surprised me...wheat, rice, and corn all have more complex genomes than humans: [b-gat.es/2h7PZSI](https://b-gat.es/2h7PZSI)

490

5.3K

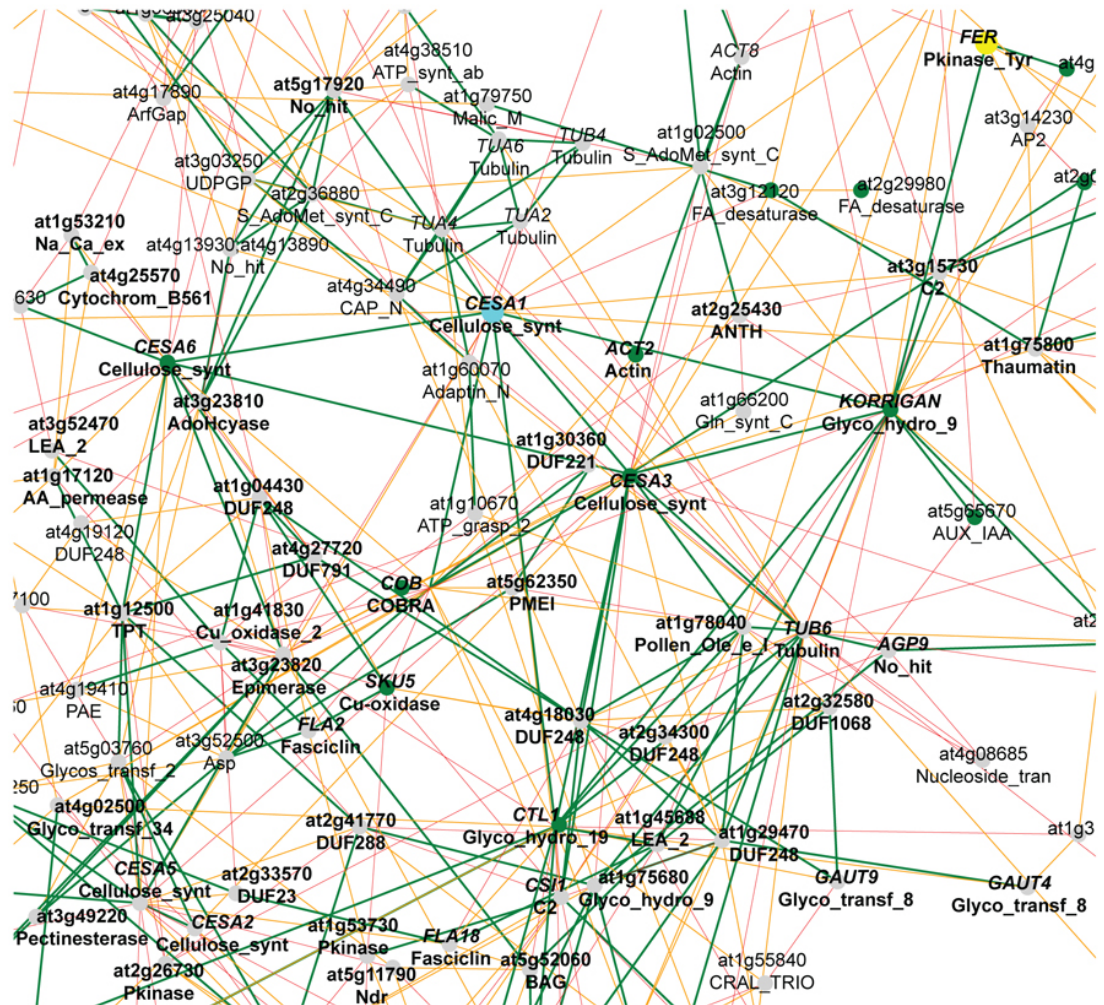
9.1K



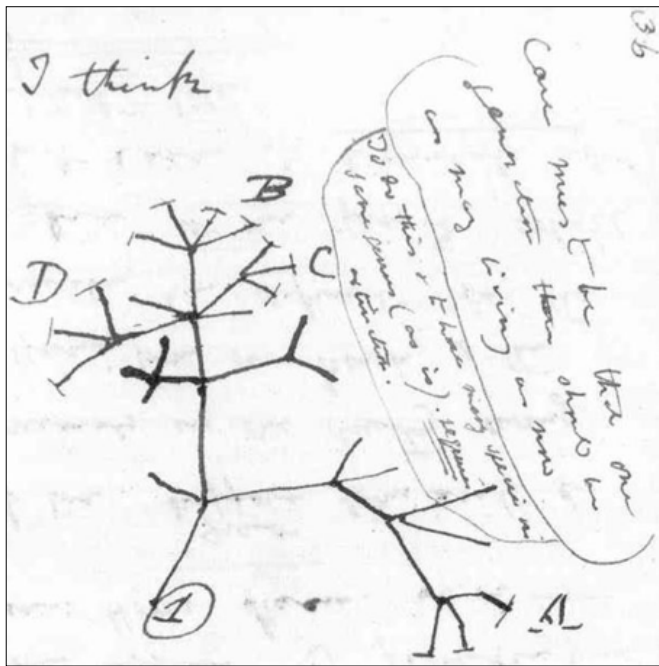


# Many Genes => Complex Interactions

Genes co-expressed with cellulose synthase gene *CESA1* in *Arabidopsis*







## Phylogenomics (Eisen et al. 1997, Philippe & Blanchette 2007):

- Using genome-scale data to resolve phylogenetic relationships
- Genome-scale/gene family analyses placed in the context of an organismal phylogeny

# One Thousand Plant Transcriptomes (1KP) Consortium ([www.onekp.com](http://www.onekp.com))



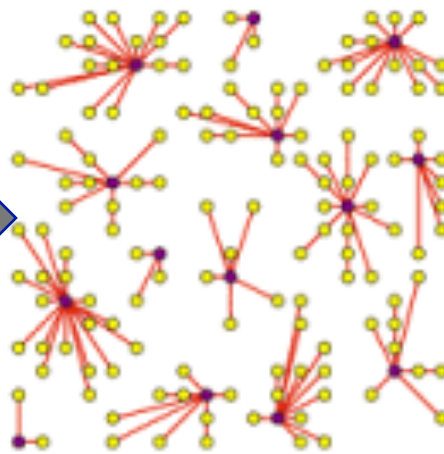
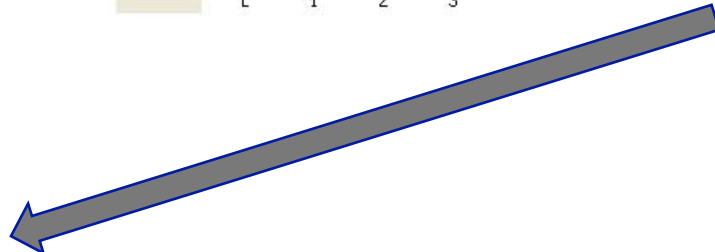
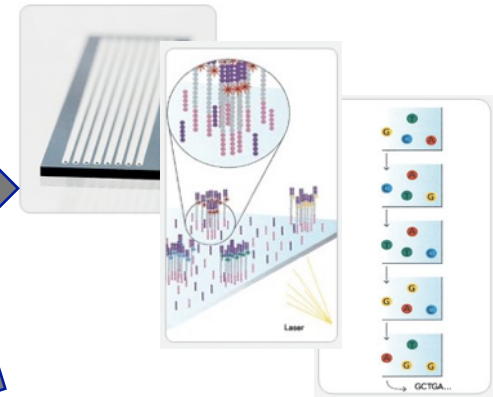
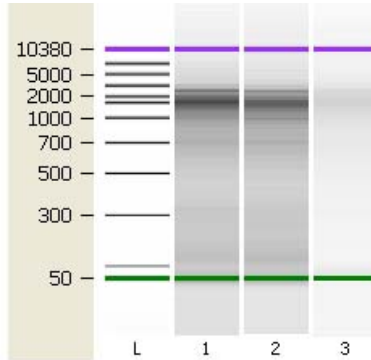
What do nuclear genes tell us about relationships across the green tree of life?



MonAToL

+ more than 100 contributors!

# Phylogenomics Workflow



http://floralgenome.org/phylomine/aln.php?action=view&aln\_id=985&plaintext=1

400 400 400 400 400

77-GI\_Eta\_4F3911-1212 EVVEAQPPELLS0Y1SCY-RKQIEVPAE  
77-Nar02\_0911-1212 EGDVDVQHALGOS-LPFYHRKGAEPALP  
H1p2277011-1212 EGVEIQHAP1GGYSSMY-RKQIEVPAE  
2a01p0870011-1212 EGVELHHEPLGGYVSSY-RKQIEVPAE  
1051d\_gene01\_e\_p01\_L6\_V113191-1212 EGVEVDHEPLGGY1SCY-RKQIEVPAE  
yoghe3.000200191-1212 EGVEVDHEPLGGYVSSY-RKQIEVPAE  
12\_e\_19790191-1212 .....VQHEPLGG51STQ-KKQIEVLANC  
12\_e\_20080191-1212 DQVEVDHEPLGGY1SSEKQKQLPAAE  
12\_e\_22488191-1212 EGVEVHHEPLGGY1ISSY-KKQIEVPAE  
12\_e\_3789101-1212 QVDAACDMYQAFQPSRNRQM...AAE  
12\_e\_3845601-1212 .....GTVSDKDRRL.....

Conservation  
.....T1000TT100TOT:31111...30

Quality

Consensus

Amborella  
Liriodendron  
Persea  
Scaevola  
Hordeum  
Triplaris  
Oryza  
Zea  
Saccharum  
Asparagus  
Allium2  
Yucca  
Medicago  
Lotus  
Glycine  
Populus  
Trifolium  
Mimulus  
Arabidopsis  
Eschscholzia  
Solanum  
Lantana  
Nicotiana

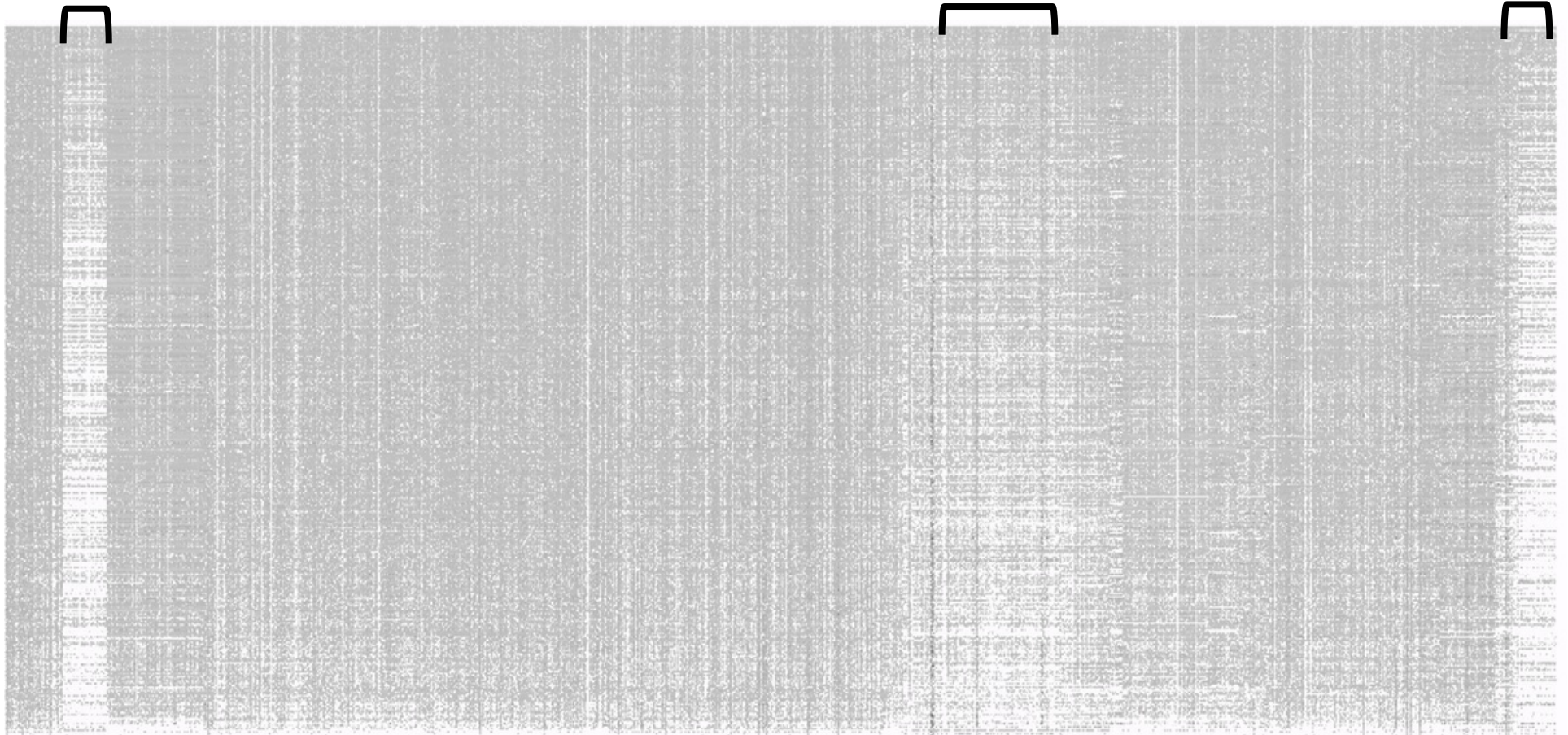


# 410 (mostly) single copy genes sampled from 1021 taxa

*Chromista*

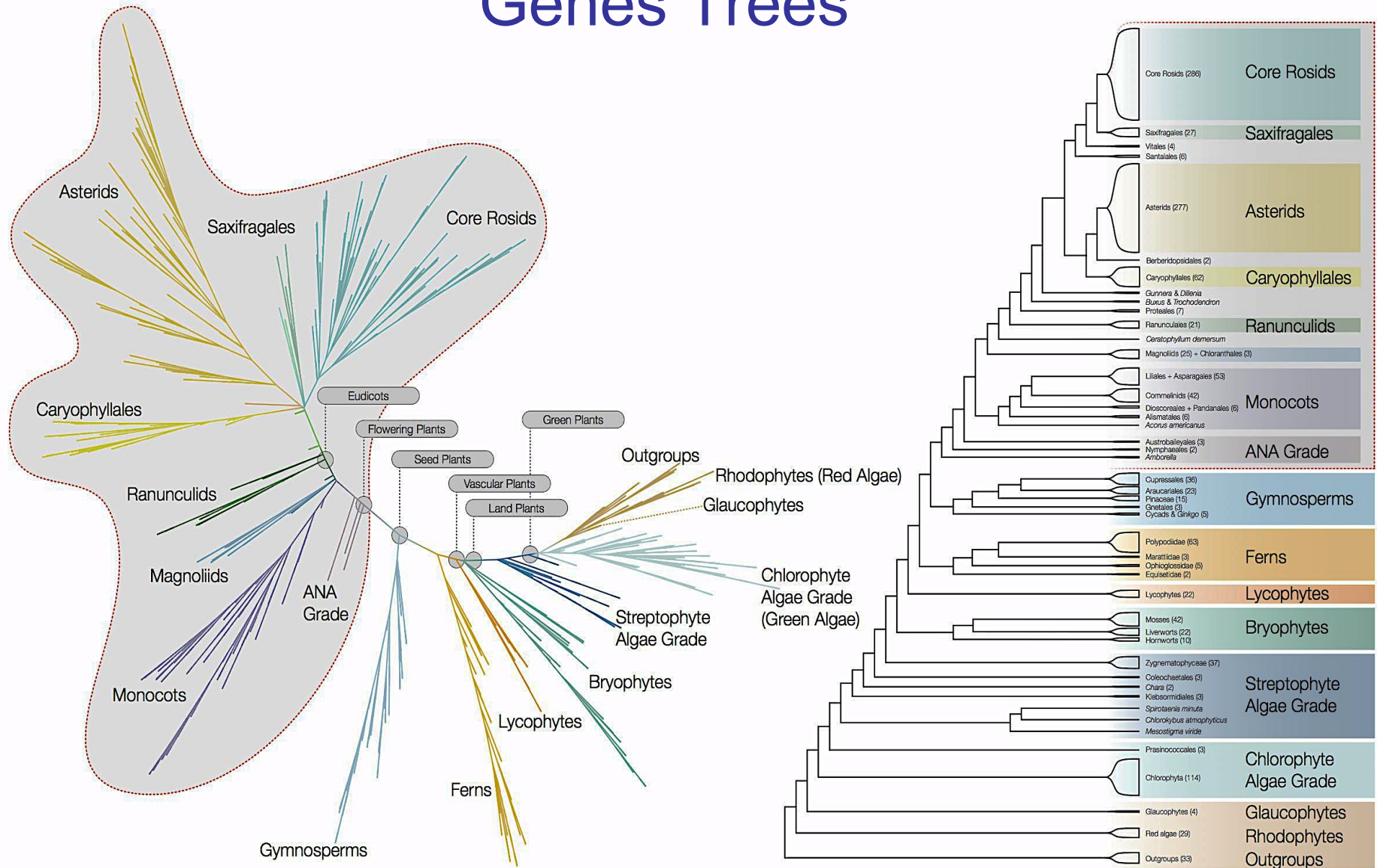
*Chlorophyta*

*Rhodophyta*



[www.onekp.com/public\\_data.html](http://www.onekp.com/public_data.html)

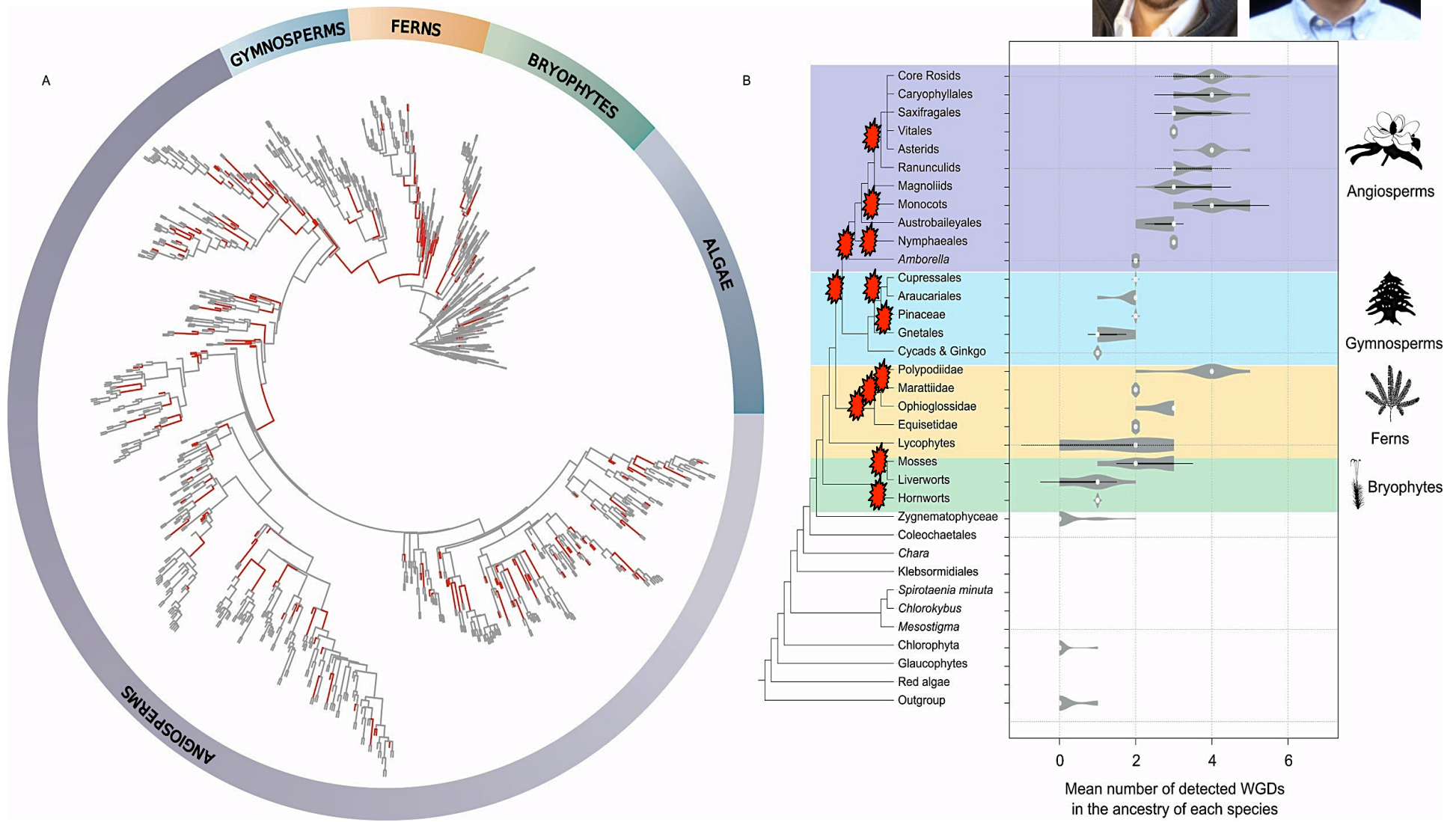
# Coalescent-Based Analysis of 410 Single-Copy Genes Trees



One Thousand Plant Transcriptomes Initiative. *In review*



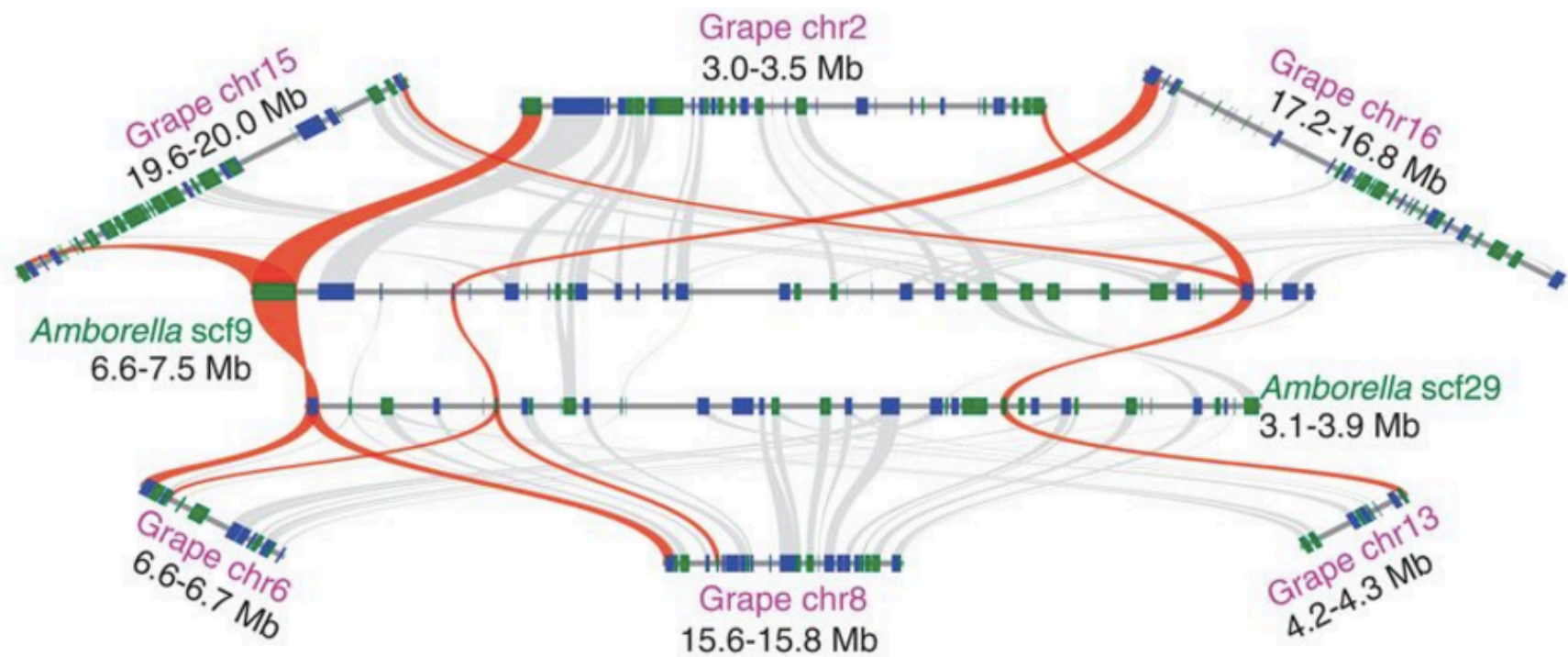
# Inferred history of genome duplications (Barker Lab, Arizona)



One Thousand Plant Transcriptomes Initiative. *In review*

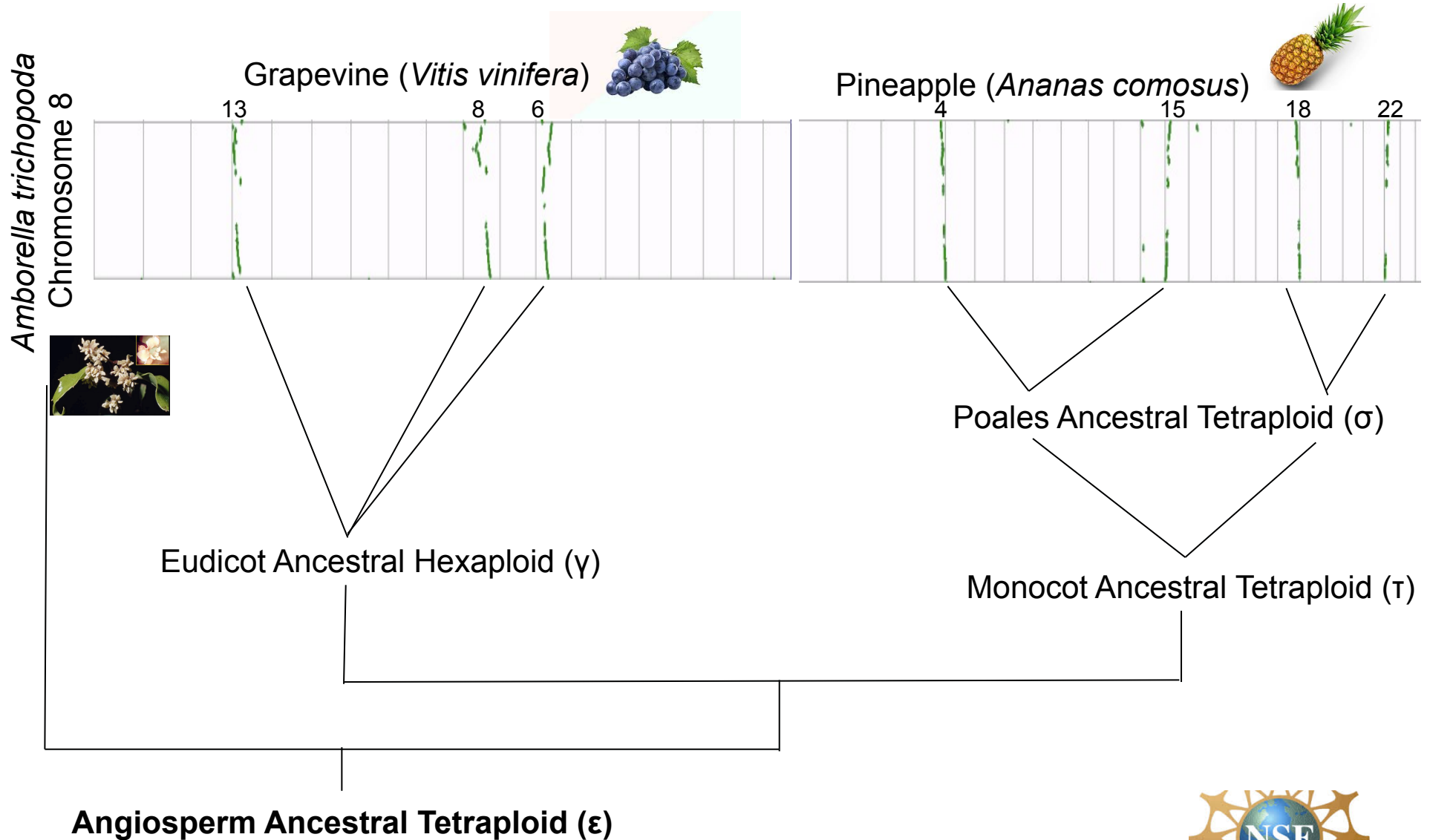


# Polyploid Ancestral Angiosperm Supported by Synteny Analysis of *Amborella* Genome



*Amborella* Genome Project *Science* 2013 ;342:1241089; *in prep.*

# Ancestral Angiosperm Genome Reconstruction



## Phytozome quick search (advanced)

Flagships

All genomes and families

Early Release Genomes



All released species



*Amaranthus hypochondriacus*  
v1.0



*Amborella trichopoda*  
v1.0



*Ananas comosus*  
v3



Angiosperm



*Aquilegia coerulea*  
v3.1



*Arabidopsis halleri*  
v1.1



*Arabidopsis thaliana*  
v2

Search in

for

GO

## About Phytozome

12.1.5

Phytozome, the Plant Comparative Genomics portal of the Department of Energy's Joint Genome Institute, provides JGI users and the broader plant science community a hub for accessing, visualizing and analyzing JGI-sequenced plant genomes, as well as selected genomes and datasets that have been sequenced elsewhere. As of release v12.1.5, Phytozome hosts 92 assembled and annotated genomes, from 81 Viridiplantae species. More than half of these genomes have been sequenced, assembled and/or annotated with JGI Plant Science program resources. By integrating this large collection of plant genomes into a single resource and performing comprehensive and uniform annotation and analyses, Phytozome facilitates accurate and insightful comparative genomics studies.

## News (details...)

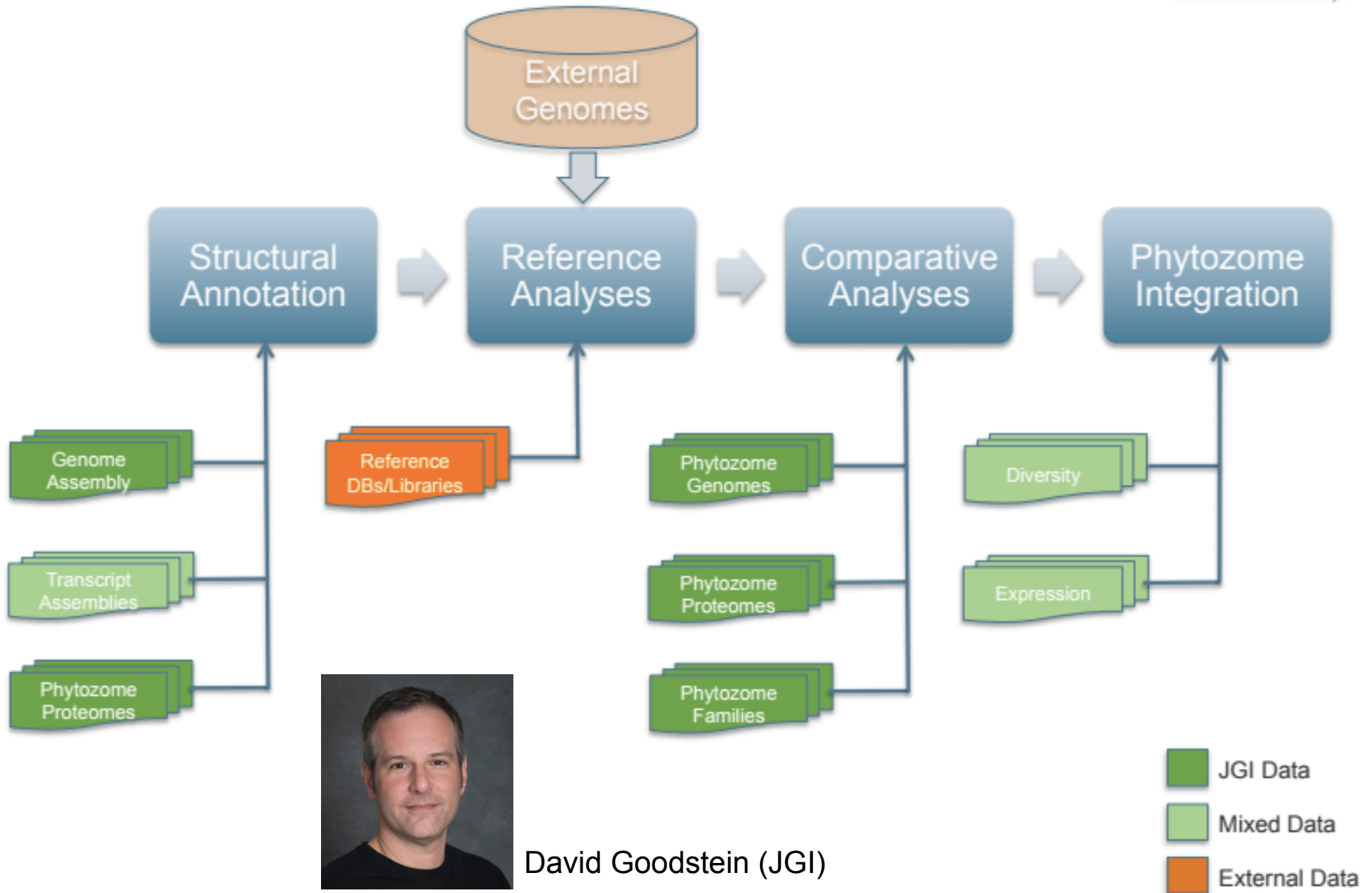


(2017-12-20) **Five genomes added**  
 (2017-10-11) **Olive genome released**  
 (2017-09-28) **2 *P. hallii* genomes + sunflower**

*Phytozome, the **Plant Comparative Genomics** portal of the Department of Energy's Joint Genome Institute, provides JGI users and the broader plant science community a hub for accessing, visualizing and analyzing JGI-sequenced plant genomes as well as selected genomes and datasets that have been sequenced elsewhere.....*



# Annotation/Integration Workflow



# Current Phytozome Content



PZ v12	Species	Genomes	Expression (species)	Diversity (species)
Eudicots	47	49	7	7
Monocots	18	23	6	2
Other Land Plants	4	4	1	-
Chlorophytes	9	9	1	1
<b>Specialty Sites</b>				
BMAP	20	20	-	-
BrachyPan	1	54 + PG	-	-
BioMart Archive	83	124	-	-

The screenshot shows the Phytozome v12 homepage. At the top, there are navigation links for 'Species', 'Tools', 'Help', 'Download', 'Help', 'Cart', and 'Subscribe'. Below this is a 'Phytozome quick search' section with tabs for 'Flagships', 'All genomes and families', and 'Early Release Genomes'. A row of genome thumbnails is displayed, including 'All released species', 'Amaranthus hybridizans v1.0', 'Arabidopsis thaliana v1.0', 'Arabidopsis lyrata v1.0', 'Arabidopsis lyrata v1.0', 'Arabidopsis lyrata v1.0', 'Arabidopsis lyrata v1.0', and 'Arabidopsis lyrata v1.0'. A search bar is located below the thumbnails. The 'About Phytozome' section on the left provides an overview of the resource, and the 'News' section on the right lists recent updates.

This screenshot displays a detailed phylogenetic tree on the Phytozome v12 website. The tree is rooted and shows the relationships between various plant species. The interface includes a search bar at the top and a list of species names on the right side of the tree.

This screenshot shows another view of a phylogenetic tree on the Phytozome v12 website. The tree is more complex, showing a larger number of species and their relationships. The interface includes a search bar and a list of species names.

This screenshot shows a list of genomes available on the Phytozome v12 website. The list includes various plant species and their corresponding genome versions. The interface includes a search bar and a list of species names.

PZ, BMAP, BrachyPan will be integrated in 2018

David Goodstein (JGI)



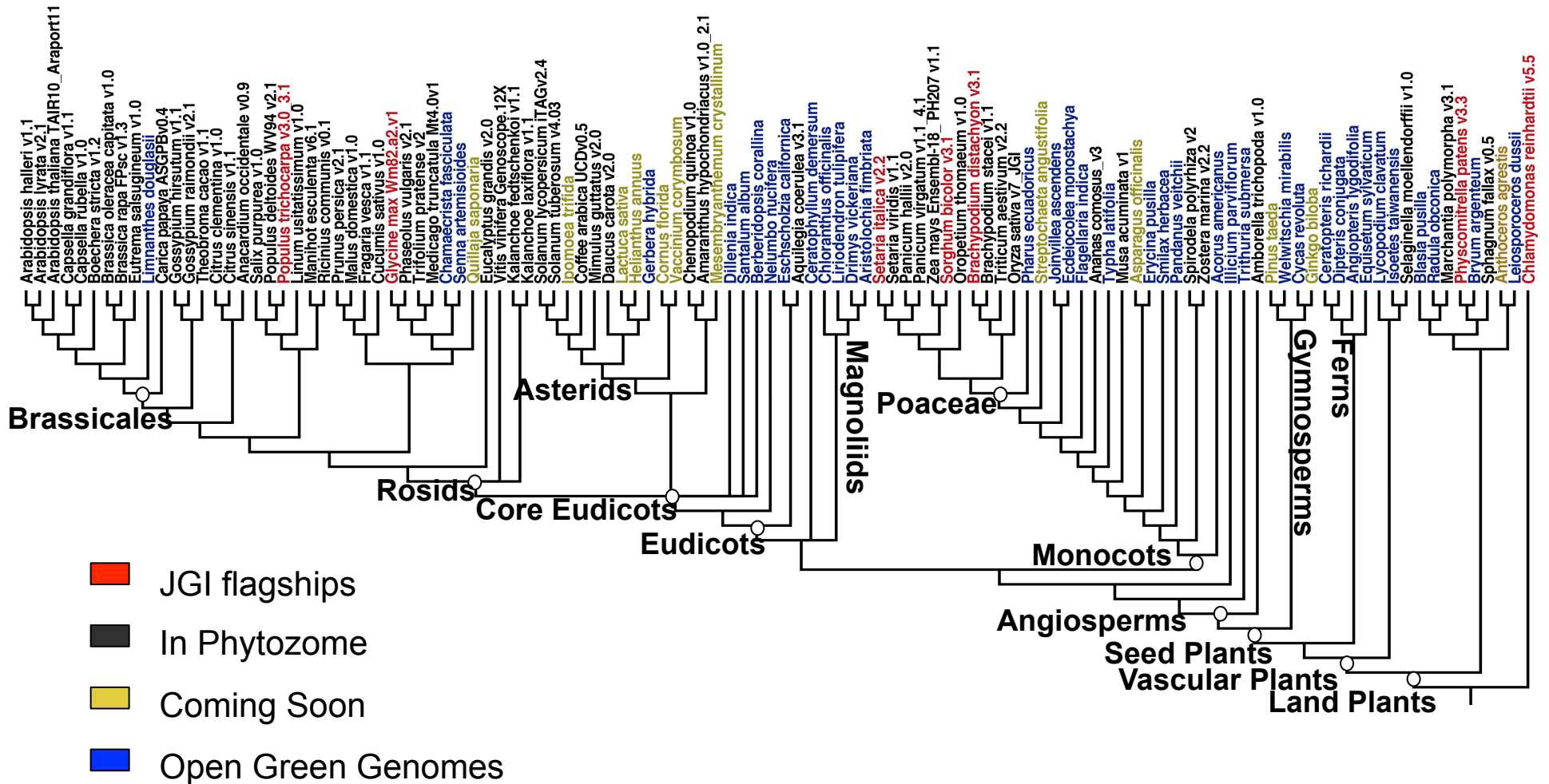


# Primary Objectives of OGG

- Improve understanding of genome structure and gene function in plant species relevant to DOE's mission through improved resources for comparative genome analyses
- Improve understanding of the diversity of gene and genome content, structure and function across all land plant lineages

-> Fill phylogenetic gaps in available genome sequences in order to refine ancestral genome reconstruction and inferences about the evolution of genome content and gene function.

# Open Green Genomes Initiative

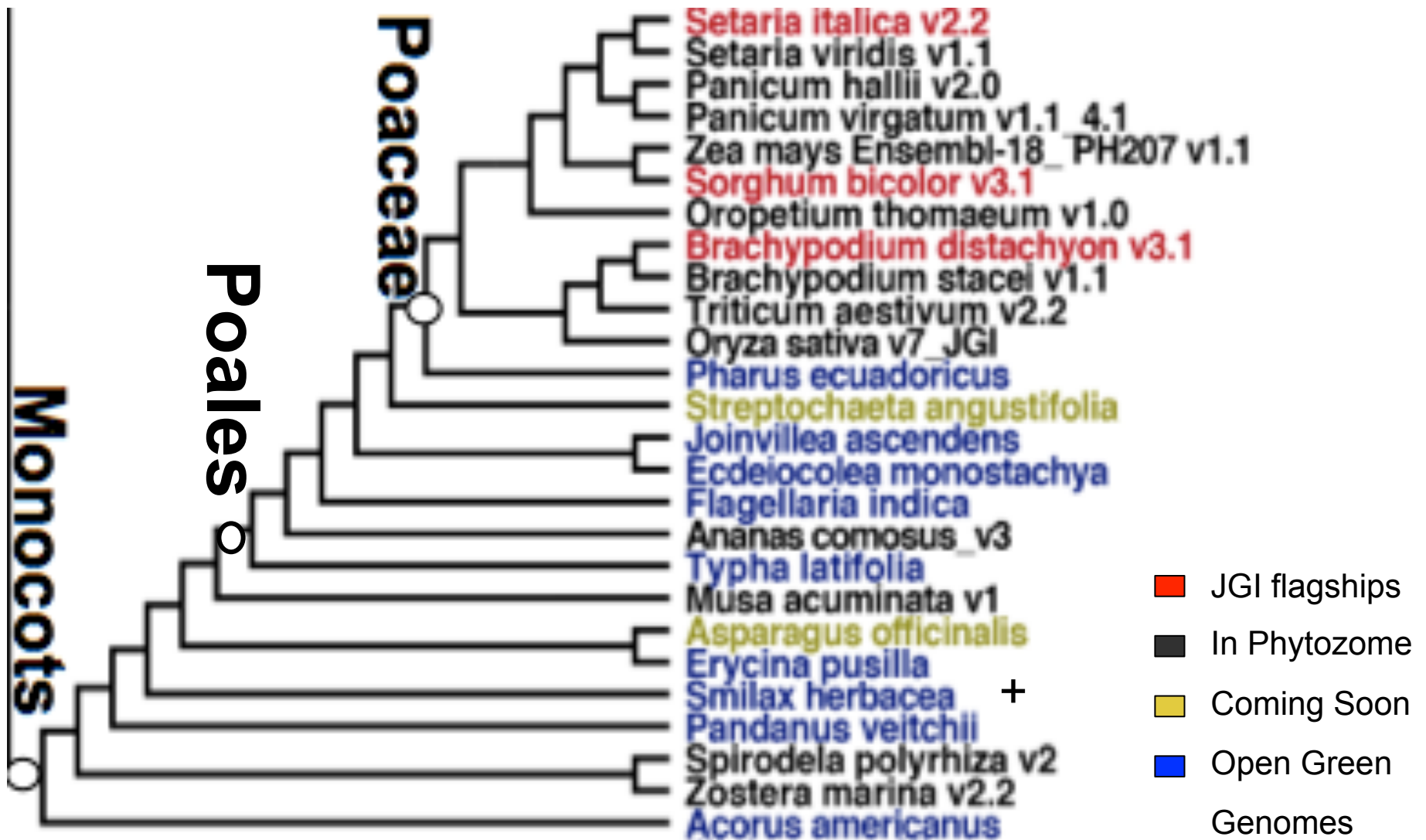


# OGG Sampling Rationale

- Phylogenetic position – minimize phylogenetic distance among well assembled and annotated genomes
- Genome size and complexity – ease of genome sequencing and assembly negatively correlated with genome size, heterozygosity and repeat content
- Experimental tractability – availability of material and ability to propagate plants



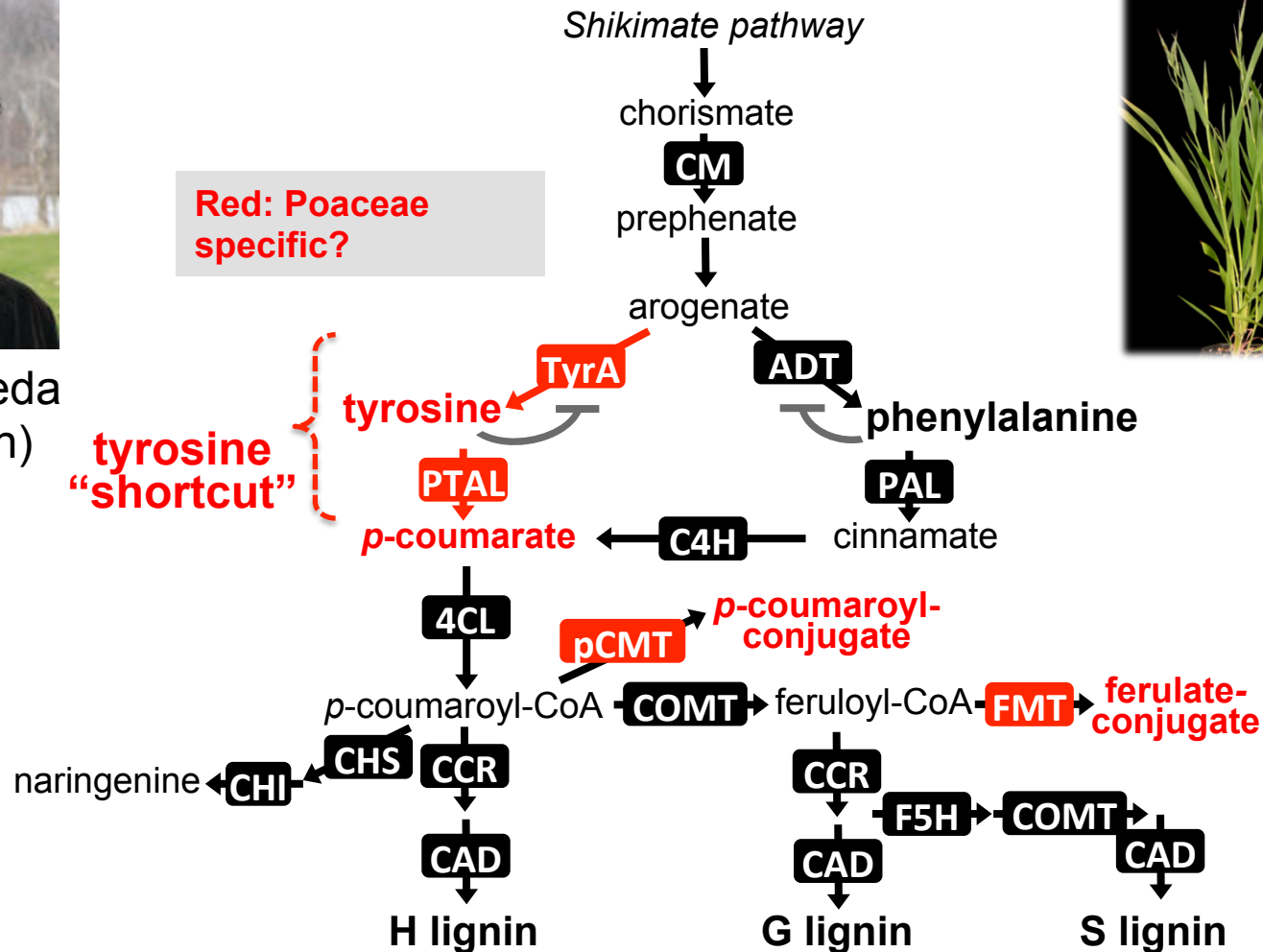
# Increasing Diversity of Monocot Reference Genomes



# Tyrosine-derived lignin biosynthetic pathway in grasses



Hiroshi Maeda  
(Wisconsin)



*When and how did the grass-specific tyrosine-derived lignin metabolic network evolve?*

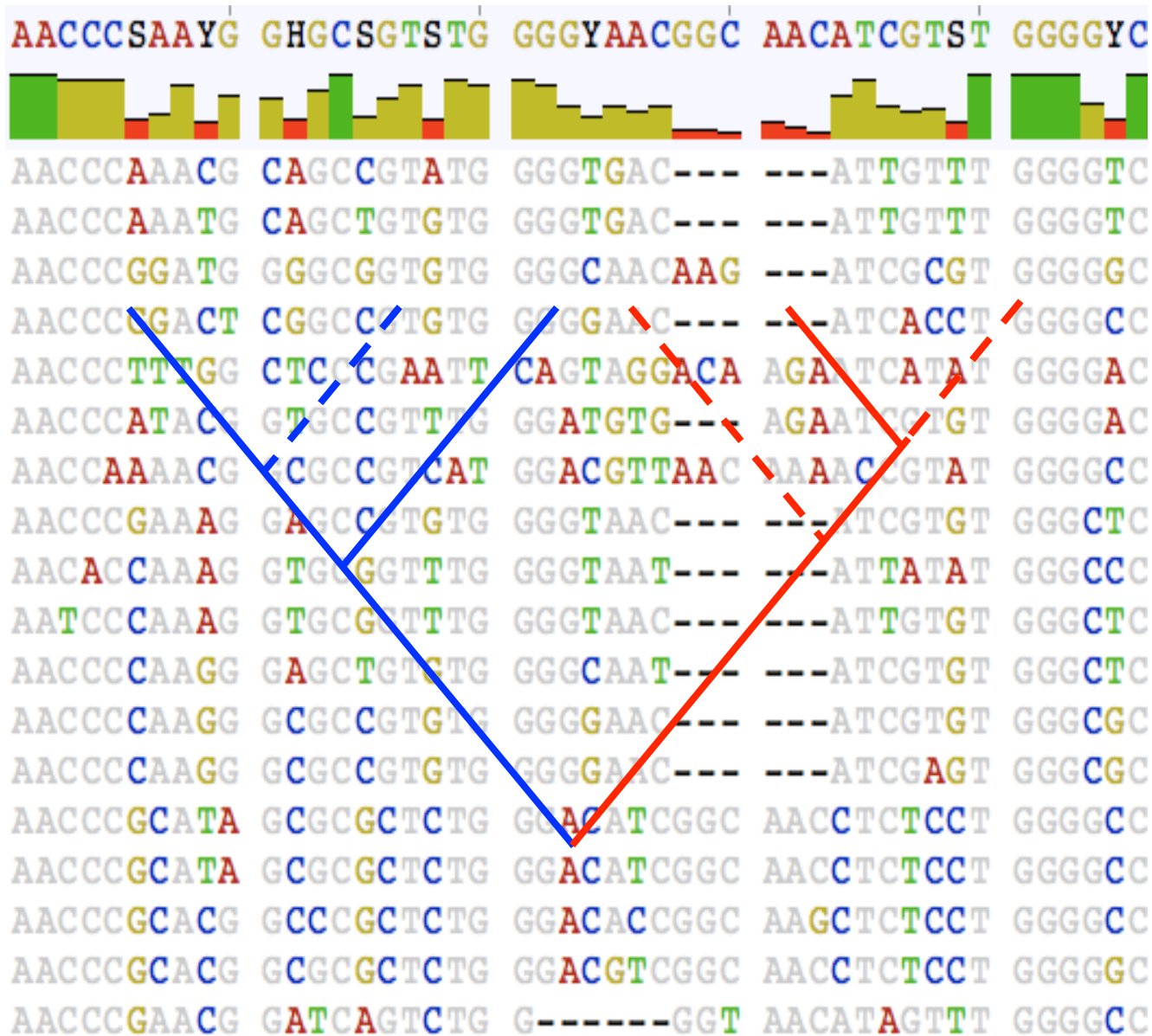




# Take home points

- Open Green Genomes initiative is a community driven effort to generate reference quality genome assemblies and annotations for un/undersampled lineages in the land plant phylogeny.
- Improved taxon sampling will yield new insights in the evolution of gene/gene family content and genome structure.
- Associated gene expression atlases and diversity panel resequencing for 35 species will facilitate gene annotation, and initial investigations of gene function and evolutionary constraint.

# Thank You !



# Phylogenetic Definition of a Gene Family/Subfamily: Last common ancestor of gene set and all descendants

