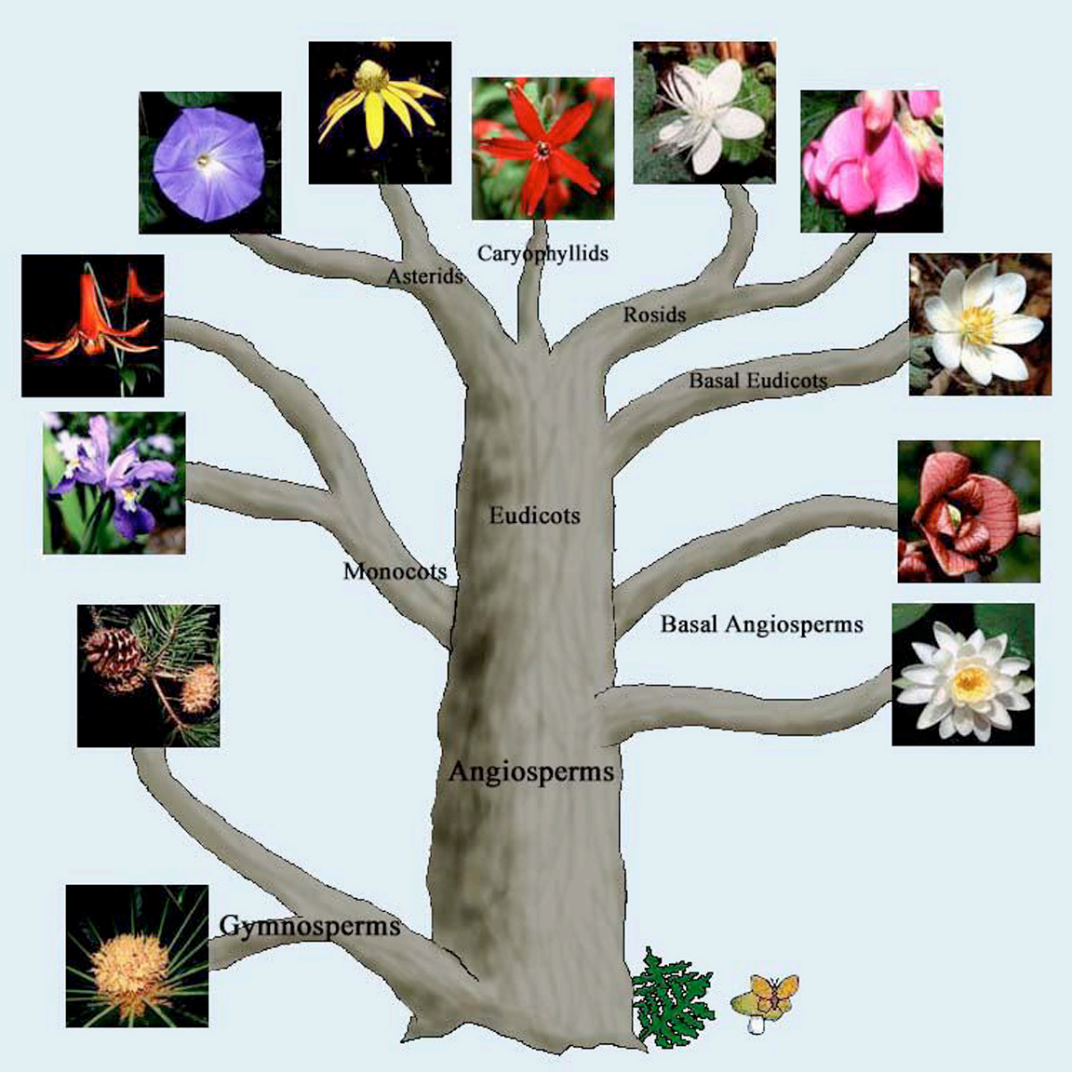
Plant Phylogenomics: Lessons from the 1KP Project



Jim Leebens-Mack

Department of Plant Biology
University of Georgia

New Methods for Phylogenomics and Metagenomics Symposium
Feb, 2013

Collaborators

Ancestral Angiosperm/ *Amborella* Genome

Vic Albert
Raj Ayyampalayam
Brad Barbazuk
John Bowers
Jim Burnette
Srikar Chamala
Andre Chanderbali
Josh Der
Claude dePamphils
Jamie Estill
Hong Ma
Doug & Pam Soltis
Stephan Schuster
Sue Wessler
Rod Wing
Kerr Wall
Norm Wickett
Eric Wafula

MonAToL

Claude dePamphilis
Tom Givnish
Cecile Ané
Raj Ayyampalayam
Sean Graham
Dennis Stevenson
Jerry Davis
Alejandra Gandolfo
Chris Pires
Norm Wickett
Wendy Zomlefer
Michael McKain
Jill Duarte

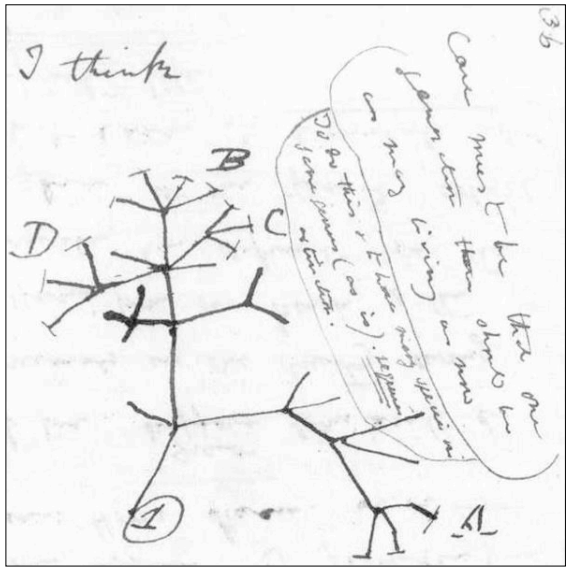
Funding: NSF, iPlant, University of Georgia, OneKP

OneKP/MSA AToL

Norman Wickett
Nam Nguyen
Siavash Mirarab
Naim Mataci
Gane Ka-Shu Wong
BGI
Eric Carpenter
Brad Ruhfel
Herve Philippe.
Gordon Burleigh
Matt Barker
Claude dePamphilis
Tandy Warnow
Jamie Estill
Raj Ayyampalayam
Doug & Pam Soltis
Sean Graham
Dennis Stevenson
Michael Melkonian
……OneKP Consortium

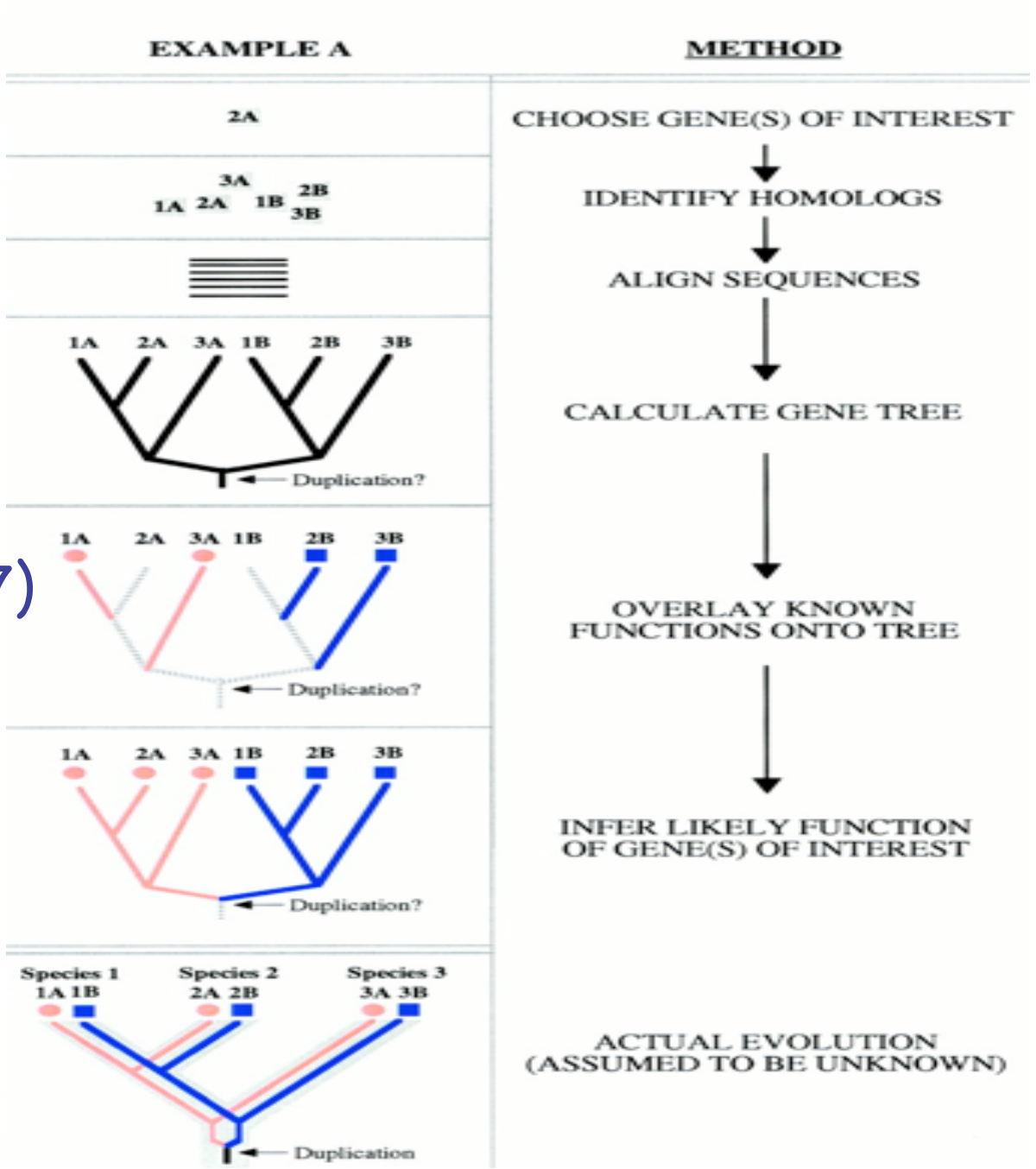
"Nothing in biology makes sense except in the light of evolution" (Theodosius Dobzhansky, 1973)

"Nothing in evolution makes sense except in the light of phylogeny"



Darwin (1837) First Notebook on Transmutation of Species

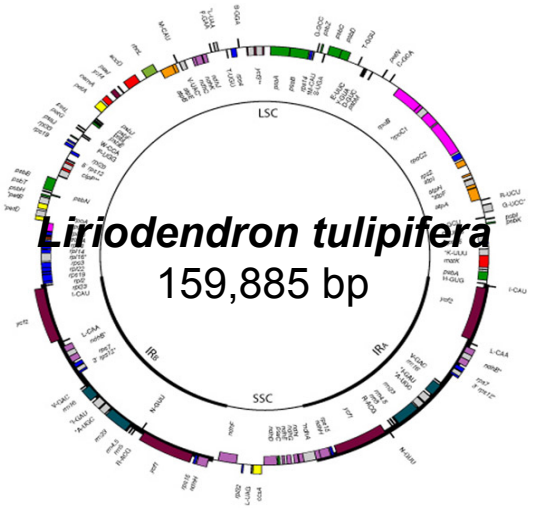
"Phylogenomics" - Jonathan Eisen (1998; Genome Research 8:163-167)



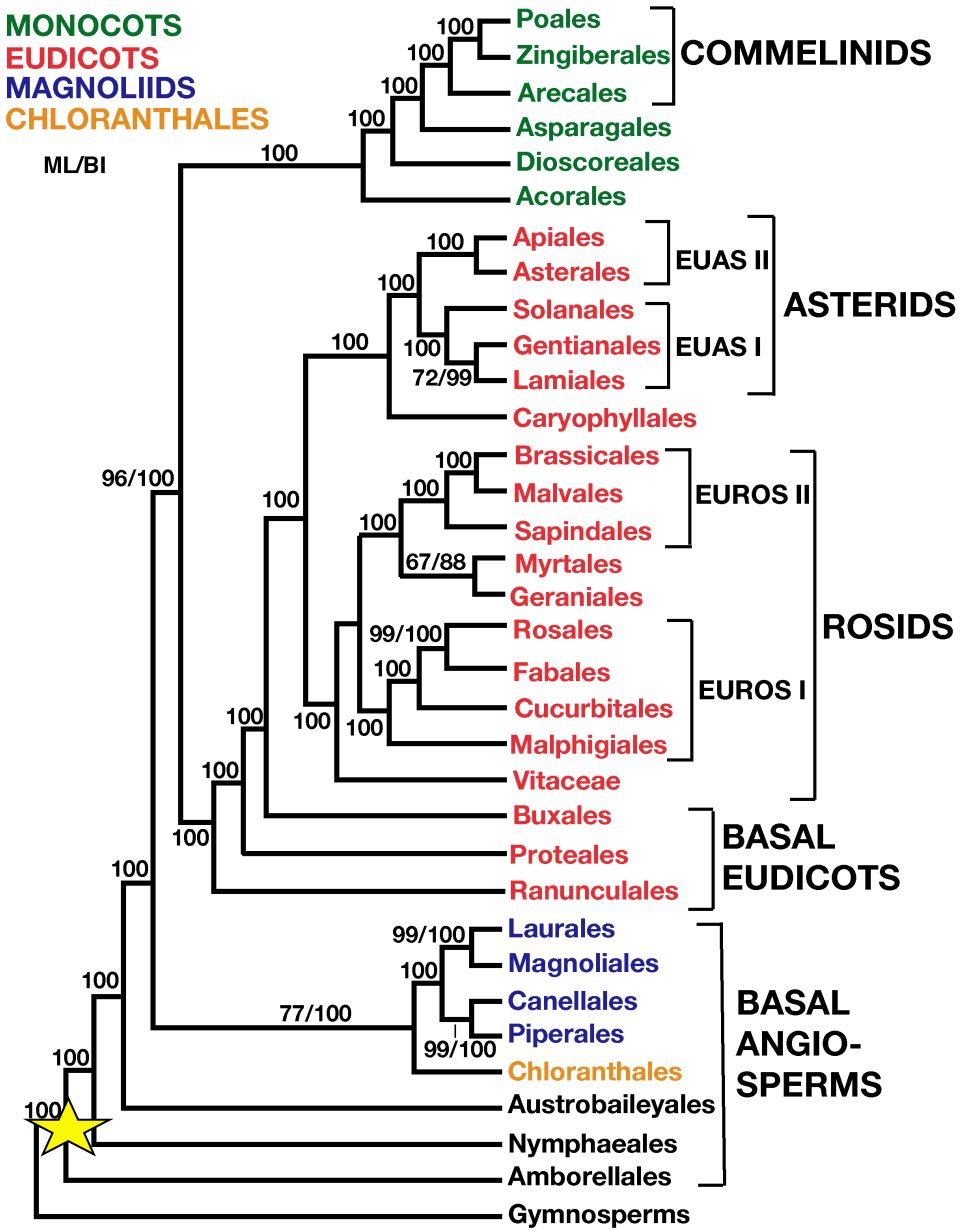
Current Usages

1. Using genome-scale data to resolve phylogentic relationships
2. Genome-Scale comparisons placed within a phylogenetic context

Phylogenomics[1]: Plastid Genome Phylogeny Resolving Many Previously Intractable Questions in Plant Systematics



Jansen et al. 2007 PNAS



What About Nuclear Gene Histories?

Grouped by Phylogeny

Angiosperms

Non Flowering

Green Algae

Grouped by Application

Agricultural

Biochemical

Medicinal

Extremophytes

www.onekp.com

Gane Ka-Shu Wong (Alberta)

Alberta

eedom To Create. Spirit To Achieve.

musea ventures

Alberta Innovates Technology Futures

华大基因 BGI

AAGP **Ancestral Angiosperm Genome Project**

ancangio.uga.edu

Acorus to Zingiber MonATol

MonATol

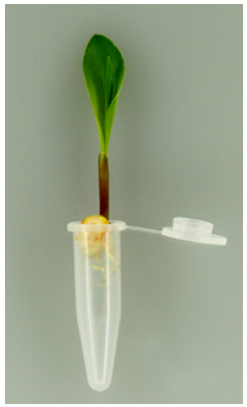
monatol.uga.edu

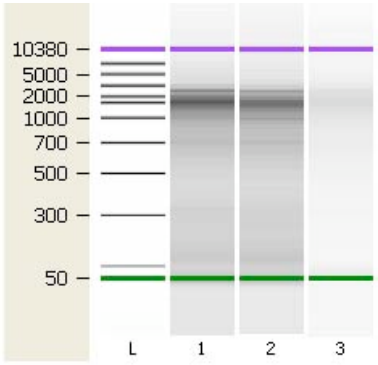
iPlant Collaborative™

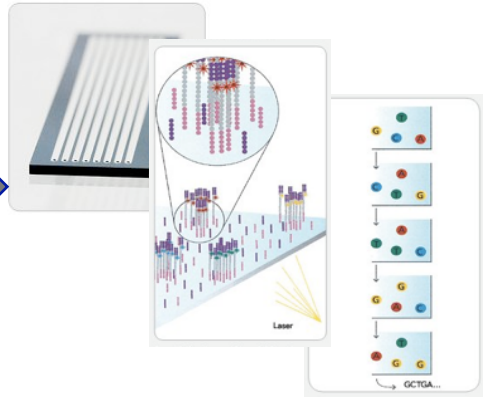
iPlant Tree of Life

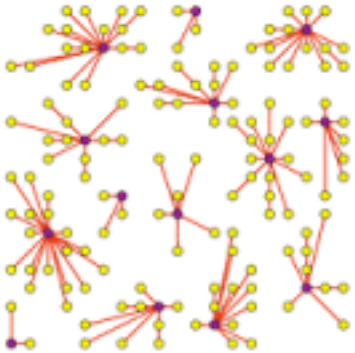
NSF

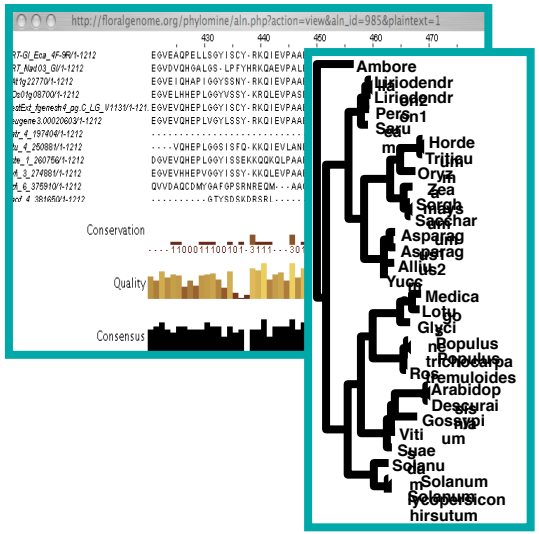
Phylo-Transcriptomics…



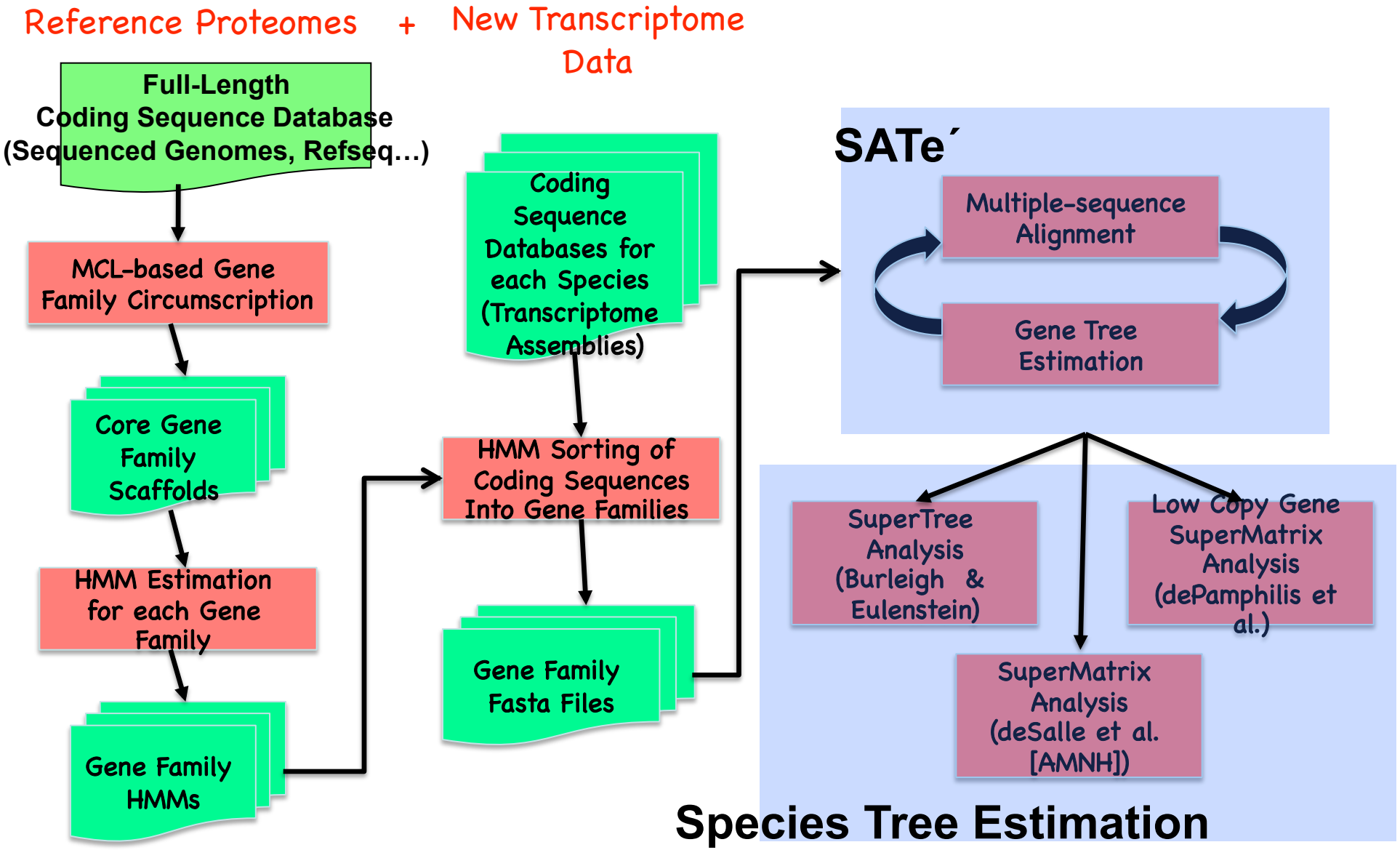




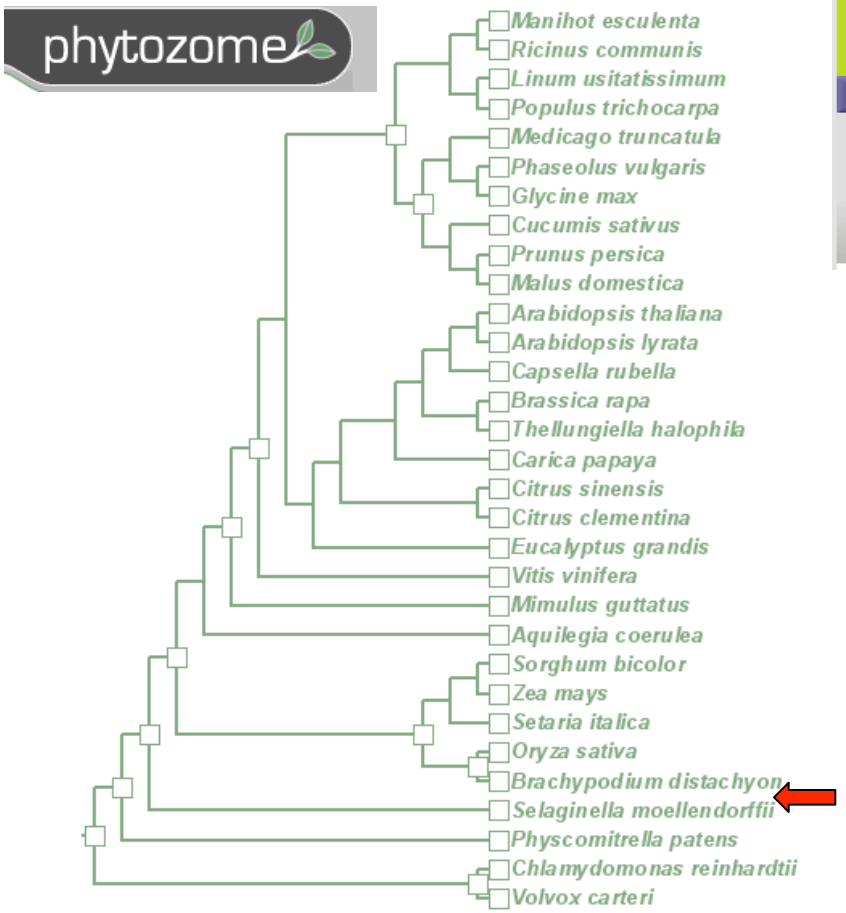


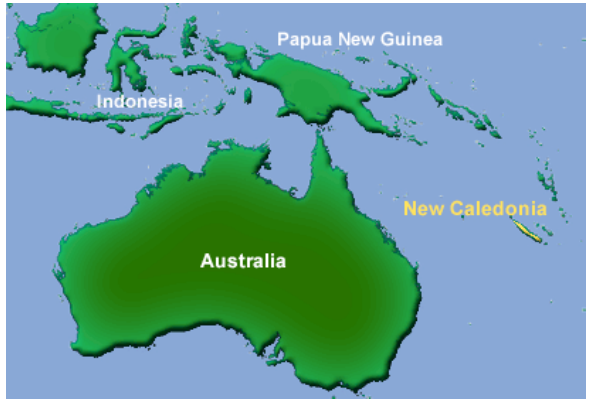


Gene Family Circumscription, Sequence Alignment, Gene Tree Estimation and Species Tree Estimation

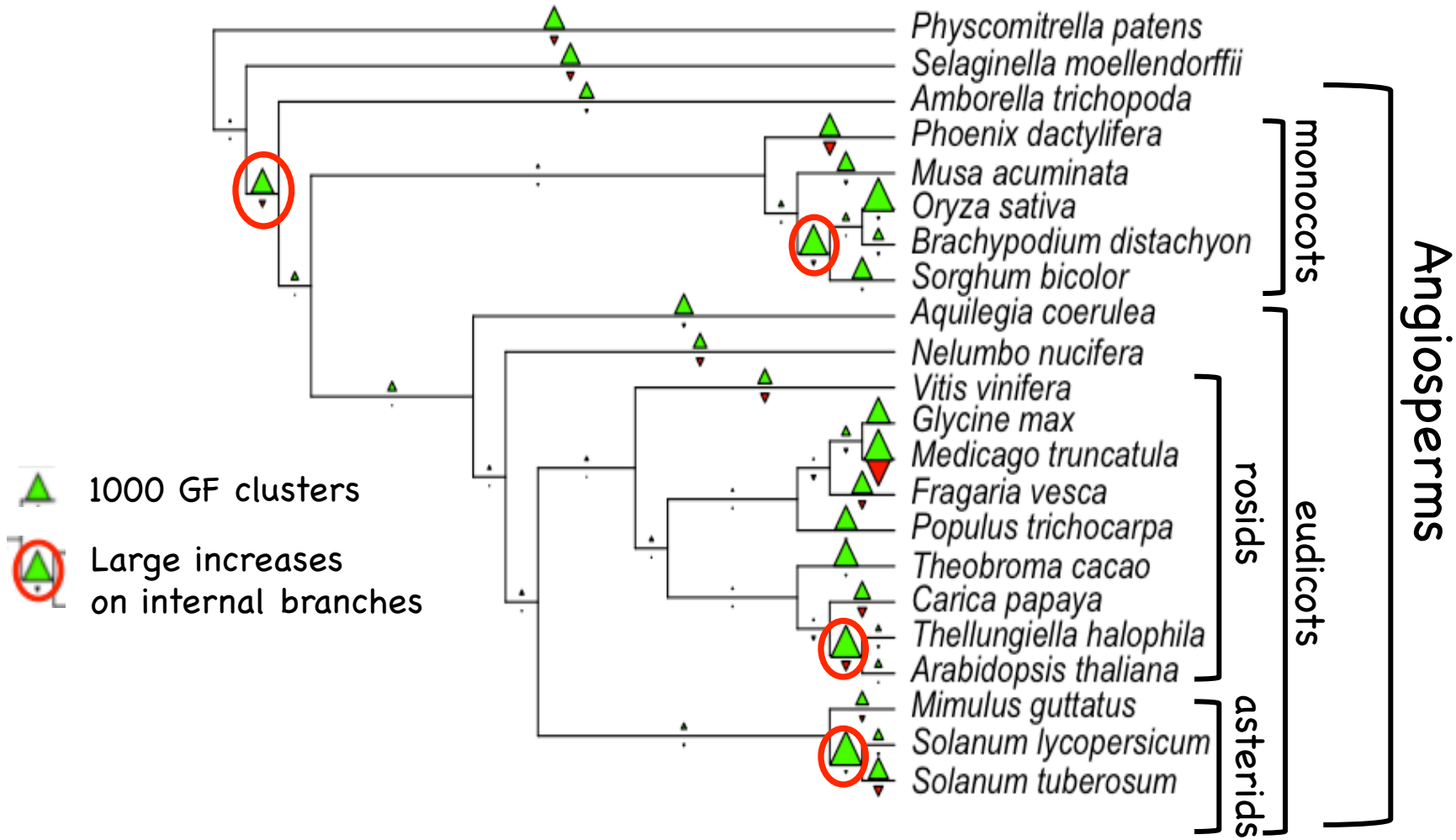


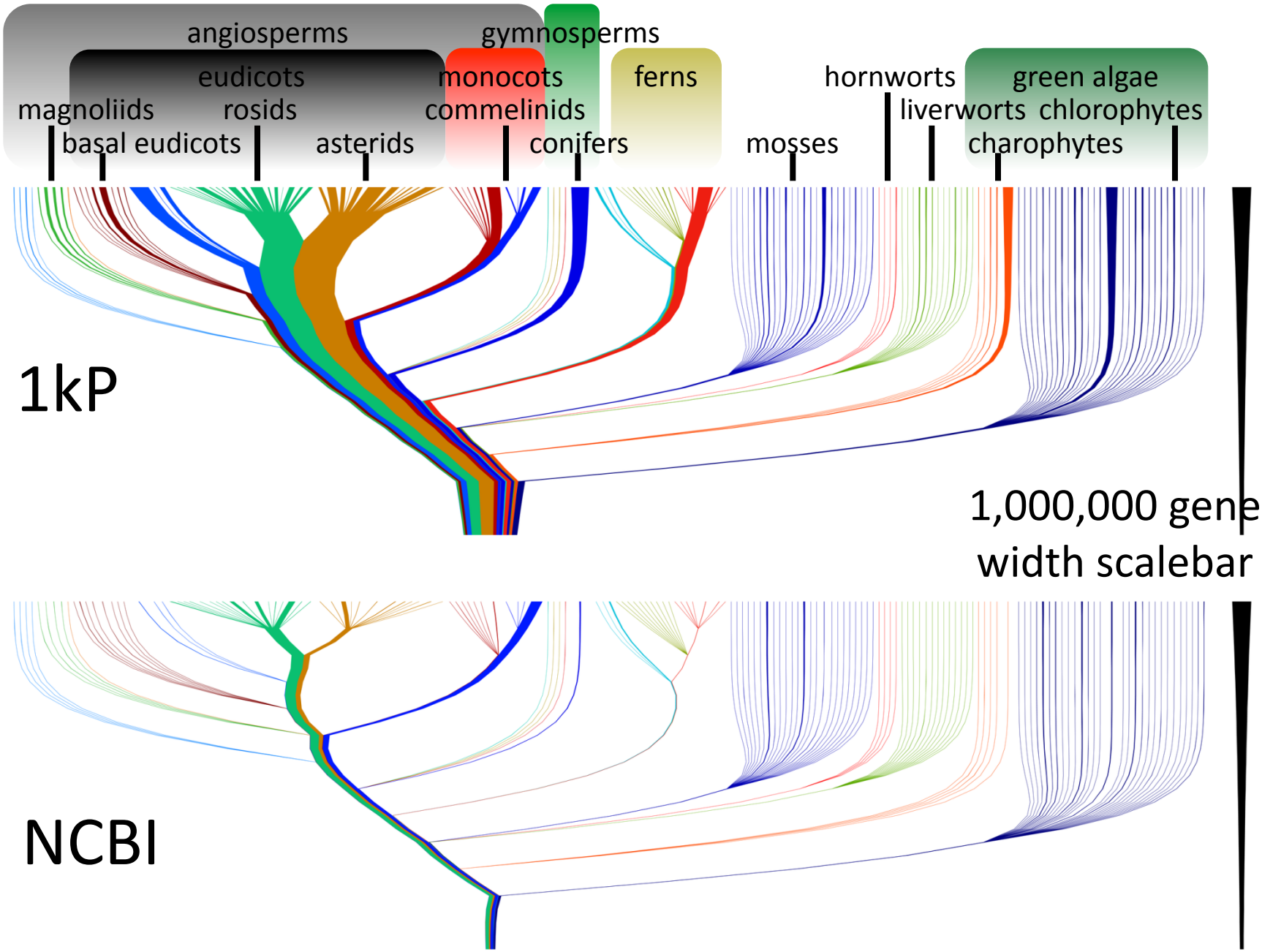
Gene Family Classification for Full-Length Genes from Sequenced Genomes – orthoMCL clustering





Gain and Loss of "Gene Families"

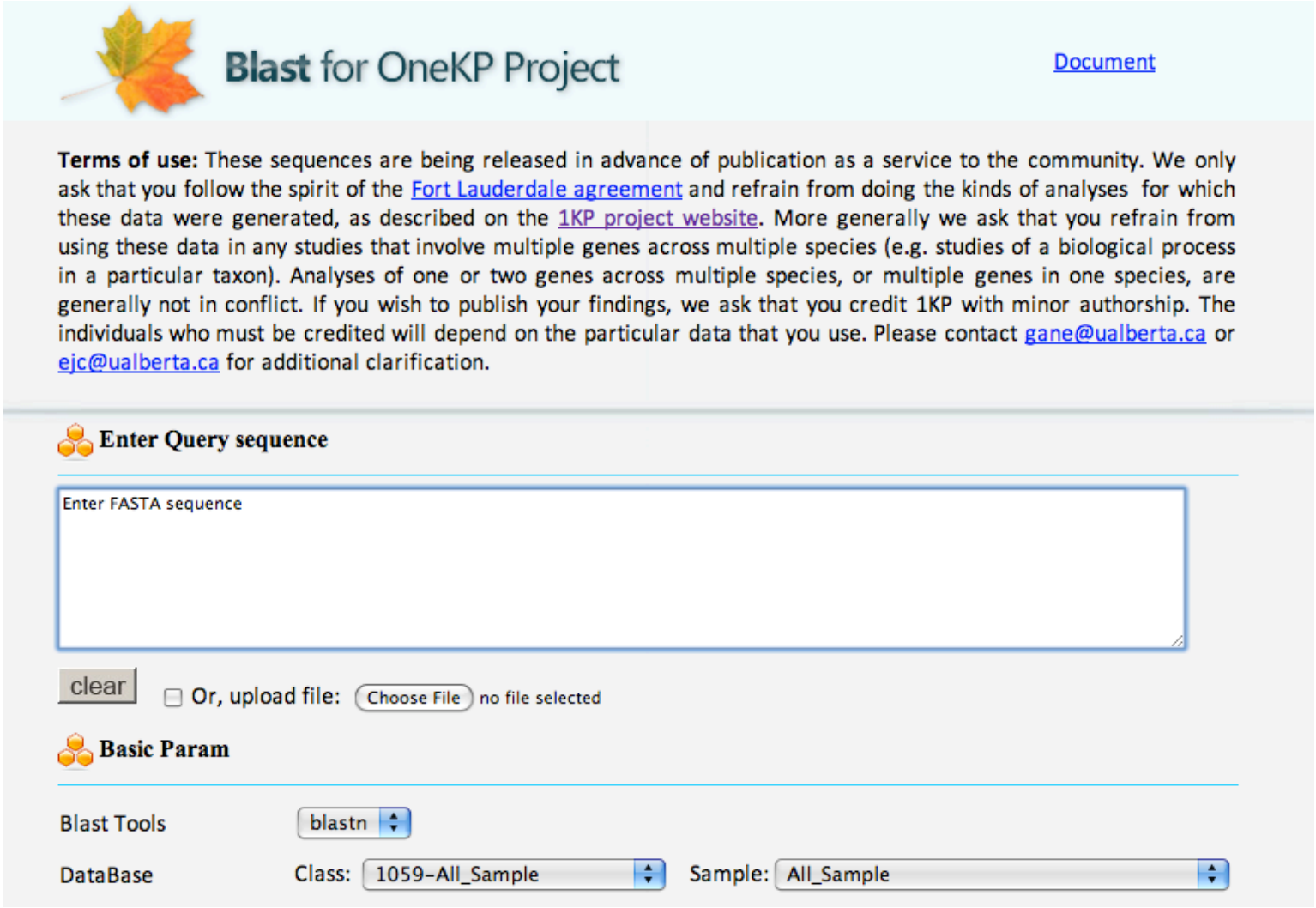



angiosperms
gymnosperms
eudicots
monocots
ferns
hornworts
green algae
magnoliids
rosids
commelinids
liverworts
chlorophytes
basal eudicots
asterids
conifers
mosses
charophytes
1kP
1,000,000 gene
width scalebar
NCBI

1kp –Questions to be Addressed through Estimation of Gene Trees and Species Relationships

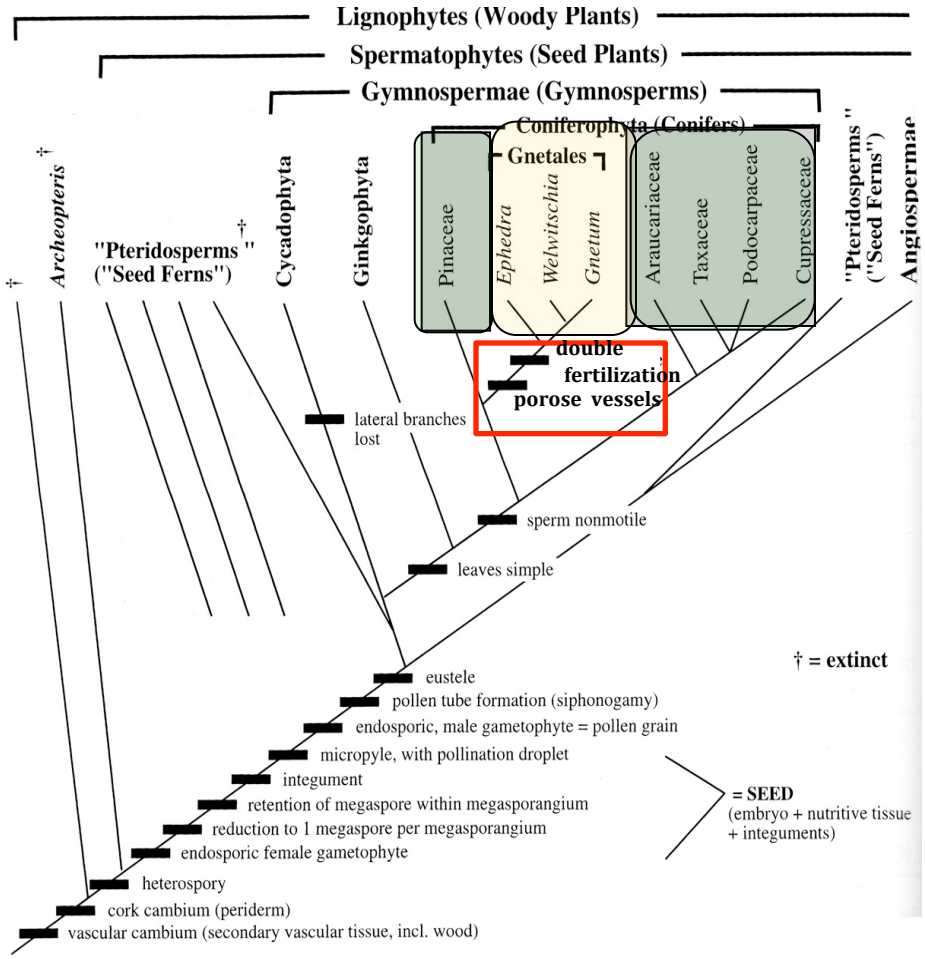
- What are the relationships among lineages across the green plant tree of life?
- What was the nature and gene composition of the likely ancestor of the Viridiplantae?
- What was the impact of lateral gene transfer (from bacteria) on the early evolution of the Viridiplantae and the Chlorophyta/ Streptophytra?
- Were gene and/or genome duplication events associated with innovations in green plant evolution including the origin of the flower, the seed, the vascular cambium (wood), alternation of generations, the shift from a life history dominated by the haploid phase to a dominant diploid life stage, colonization of land and shifts from single cells to multicellularity,
- Has polyploidy played a role in the diversification of angiosperms?

Check out OneKP data

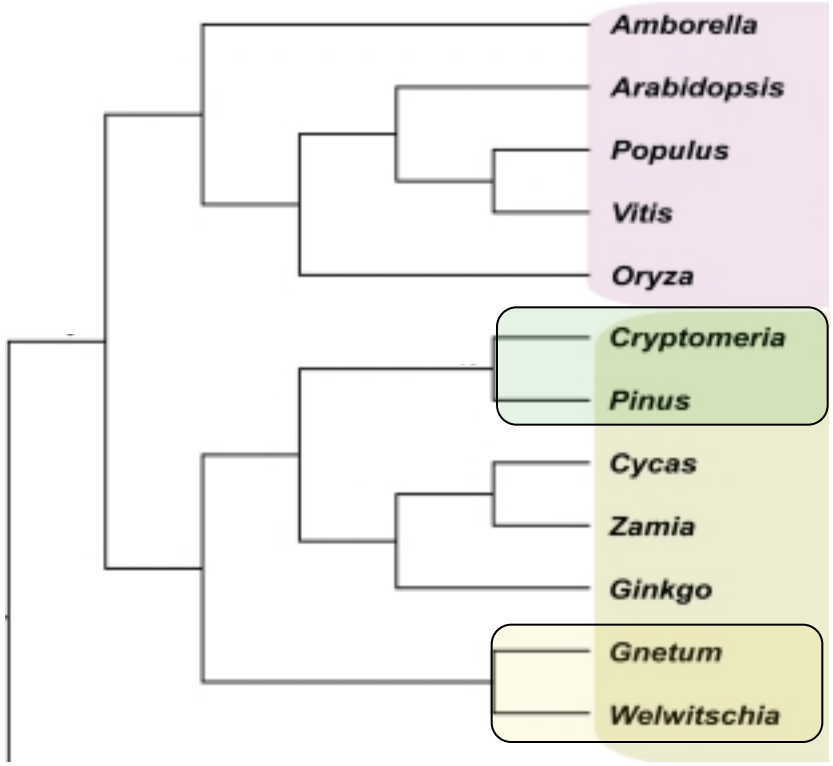


Surprises from Phylogenomics: Monophyly of Conifers?

Plastid Genome Analyses

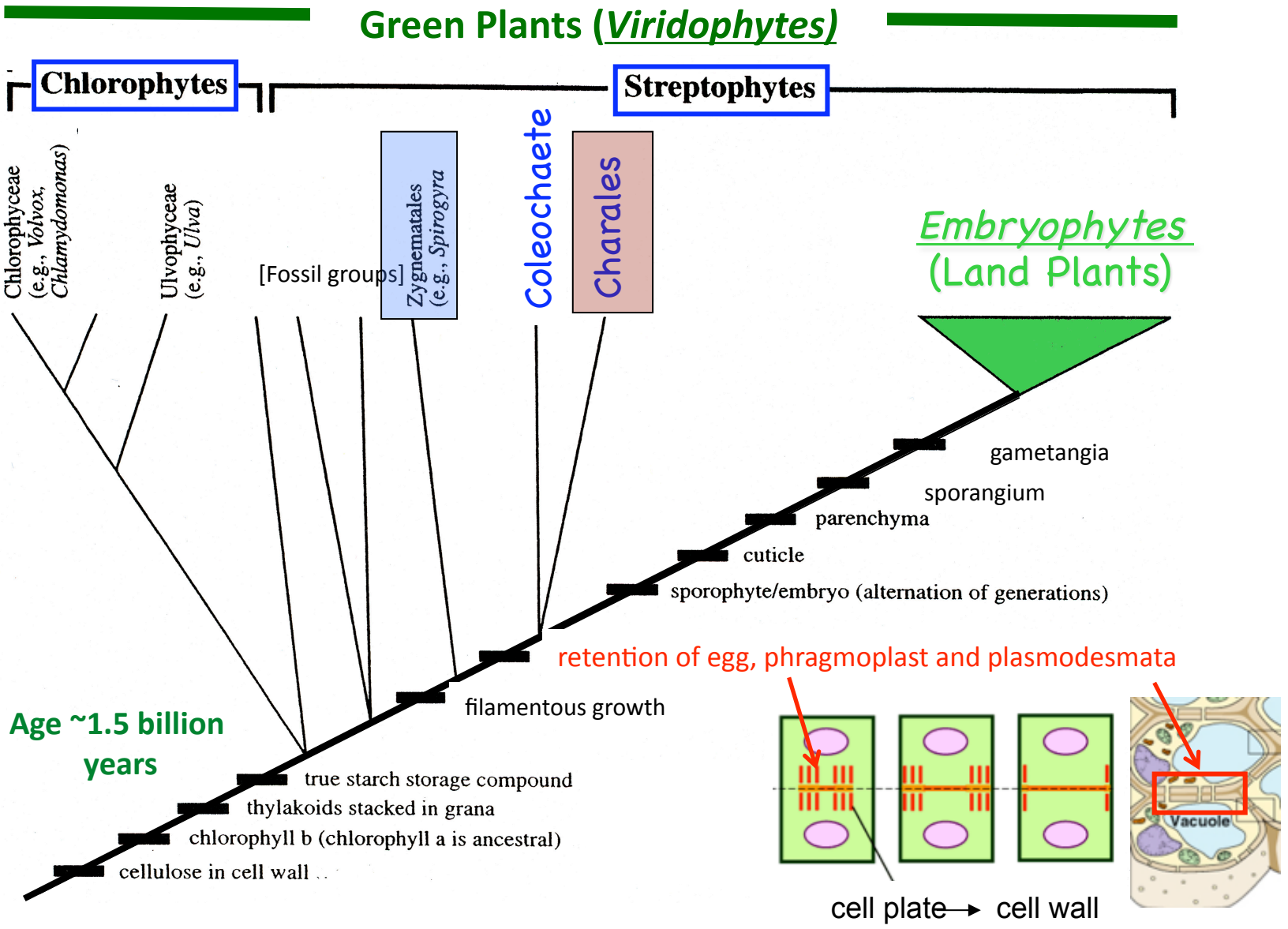


Large Nuclear Gene Analyses

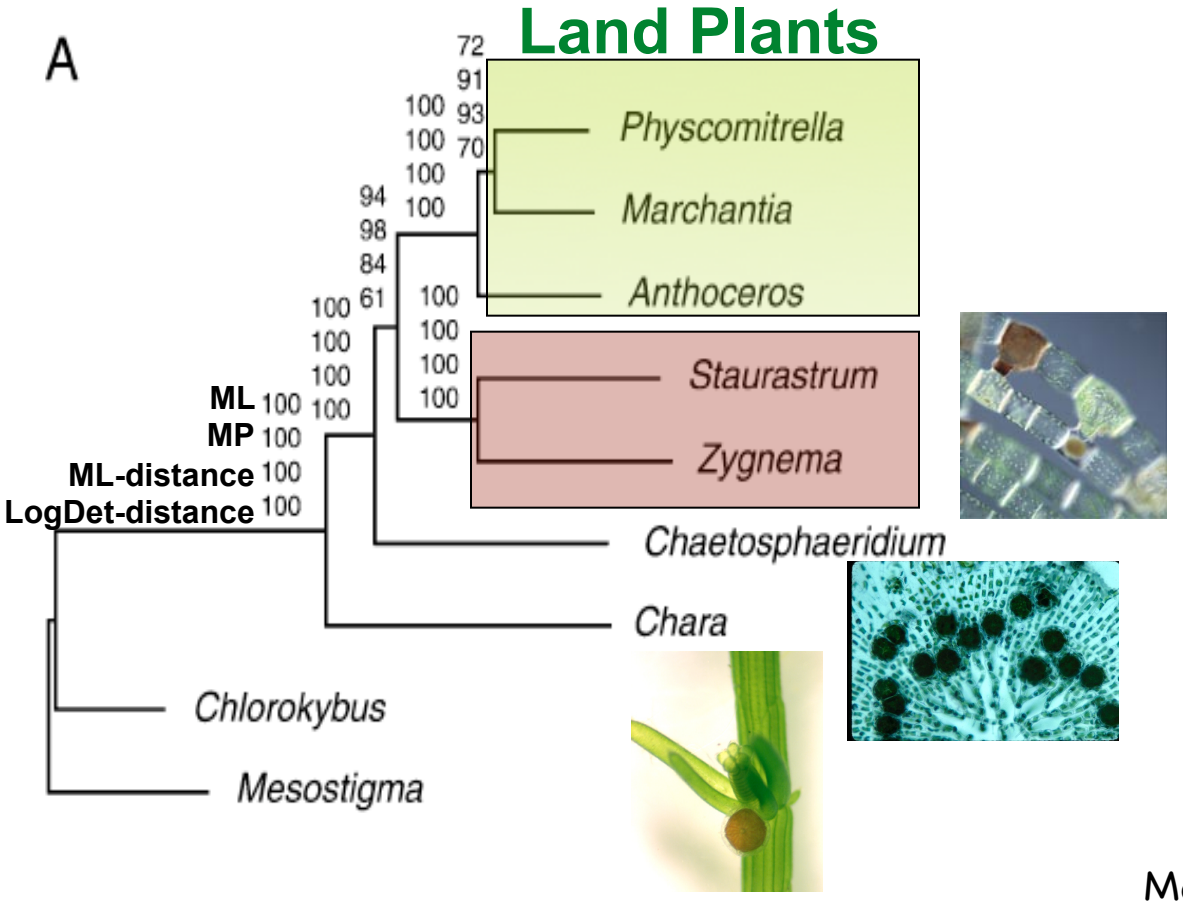


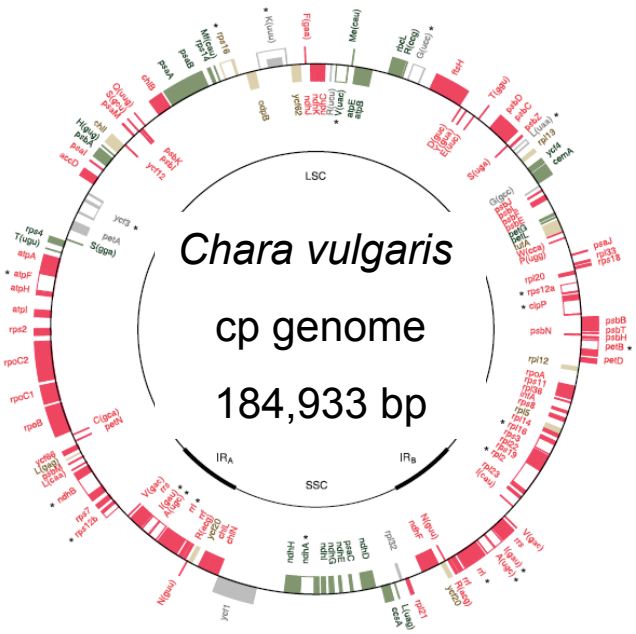
Cibrián et al. 2010, Lee et al 2011, OneKp unpublished

Origin of Land Plants: Retention of egg, phragmoplast and plasmodesmata preadaptations for colonization of land



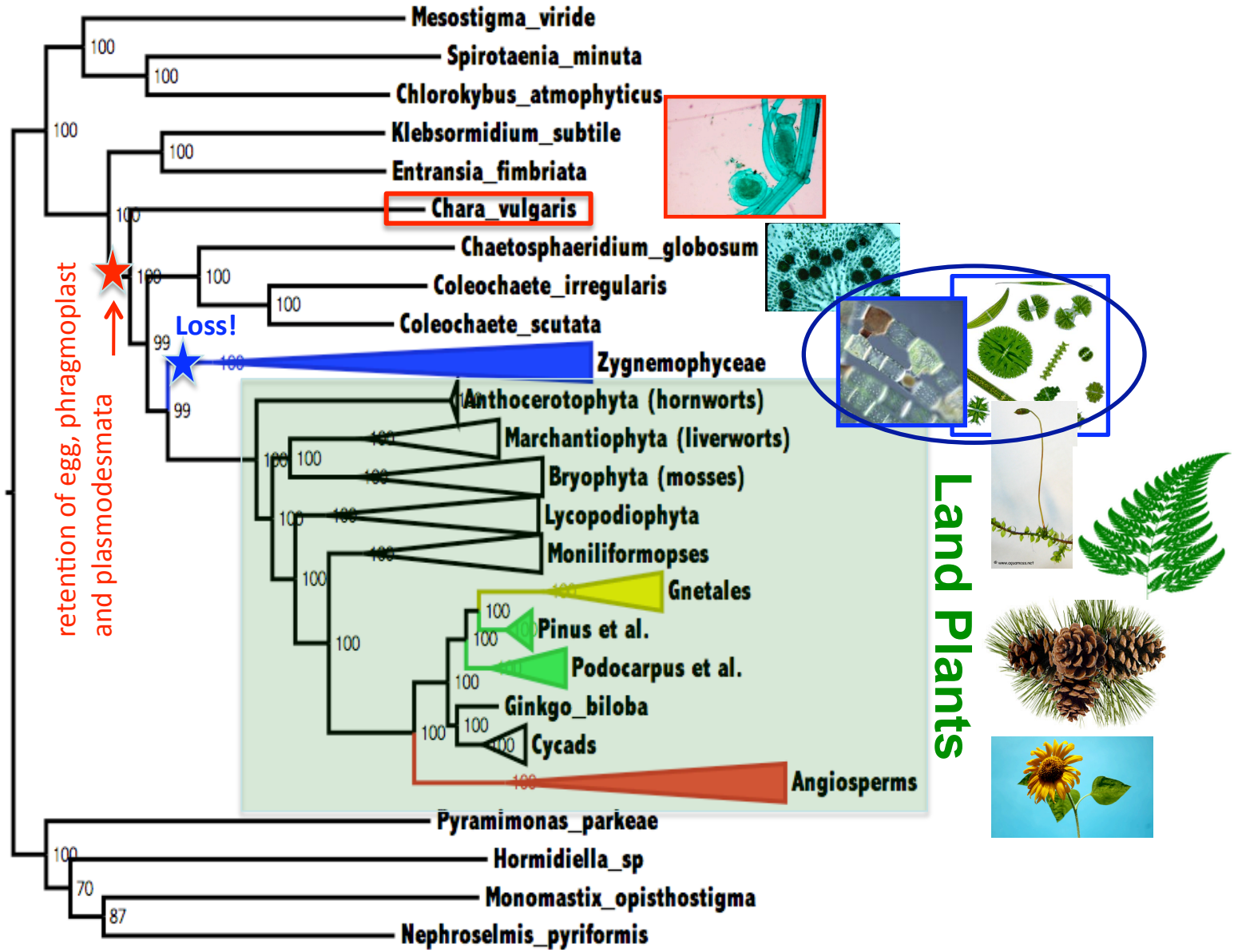
Plastid phylogenomic analyses yielding surprising results for the origin of land plants – filamentous rather than more complex algal lineages inferred sister to Land Plants!





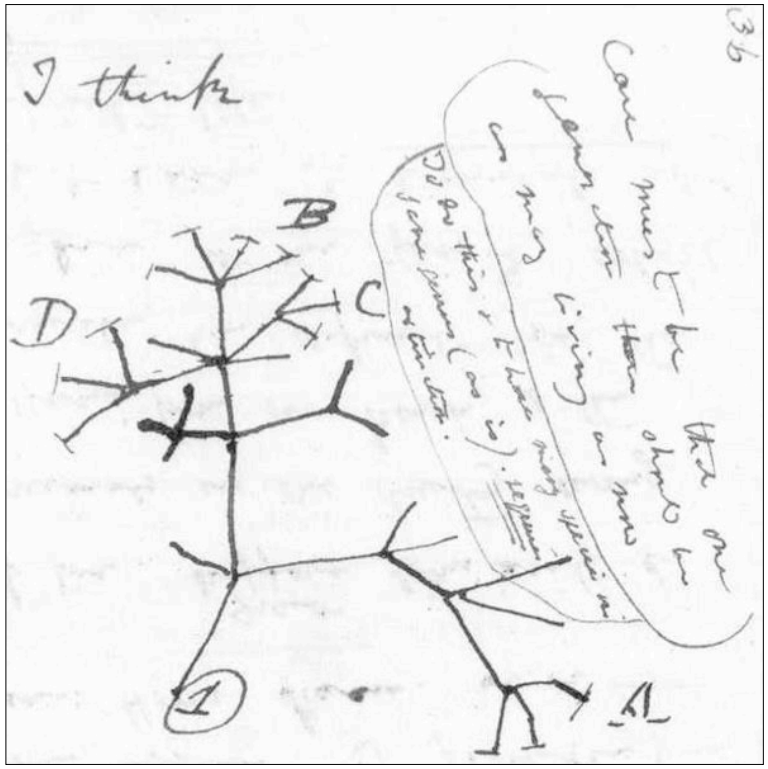
Monique Turmel et al 2006 MBE

Analysis of 604 single-copy nuclear genes agrees with plastome analysis suggesting that retention of egg, plasmodesmata, and phragmoplast **lost** in diverse Zygnemophyceae



Why should you believe these results?

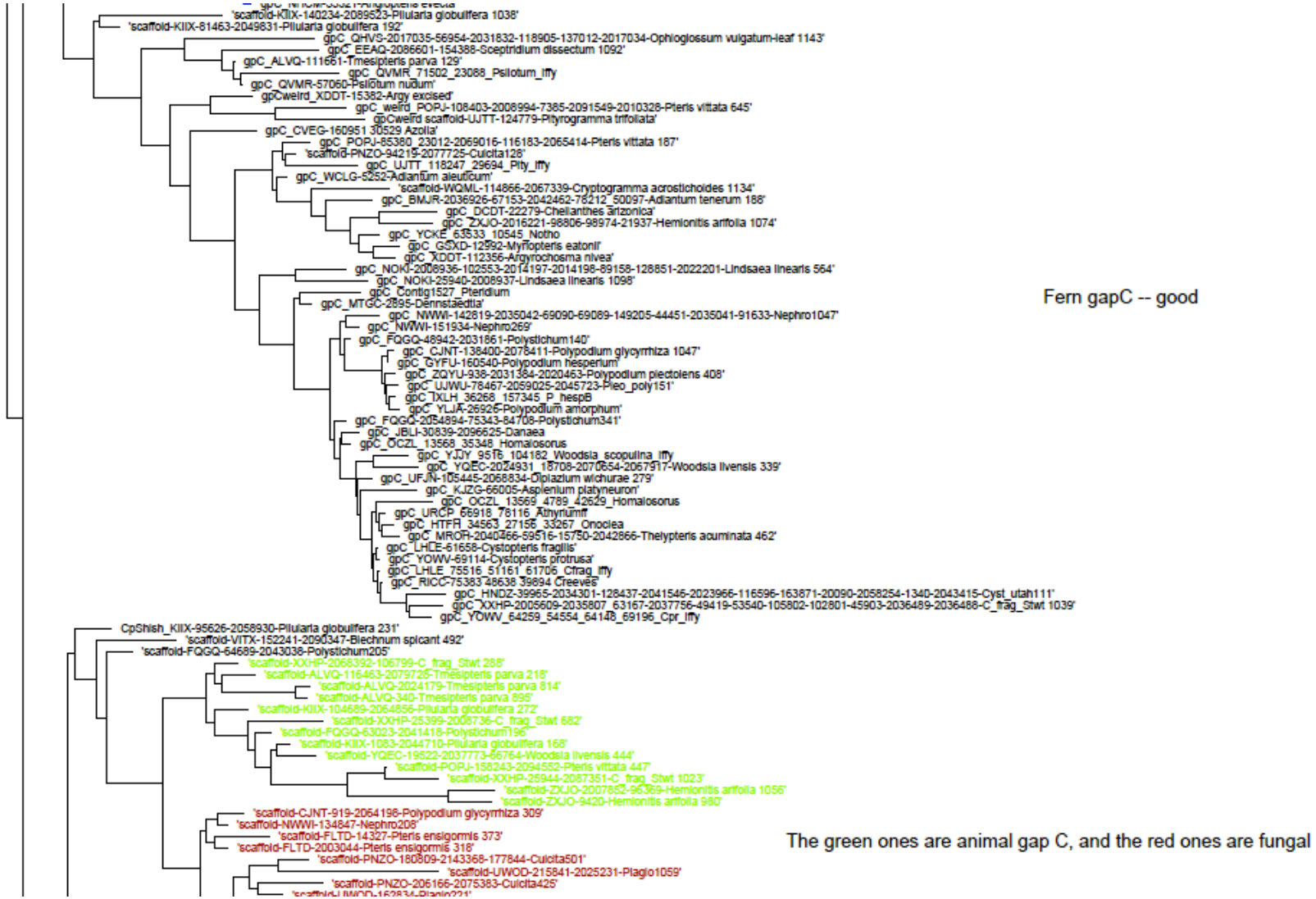
- You shouldn't…. Inferred trees are hypotheses!



Consider Possible Complications

- Lots of missing data!
  - Remove genes/taxa with lots of missing data
- Ortholog identification
  - Focus on genes/genomes that tend not to be duplicated
- Contamination
  - BLAST, Sequence placement on reference tree (SEPP?), Long branch trimming
- Model misspecification
  - Model validation – simulations
  - Test robustness of hypothesis among inferences derived from alternative models

Contamination is an issue!

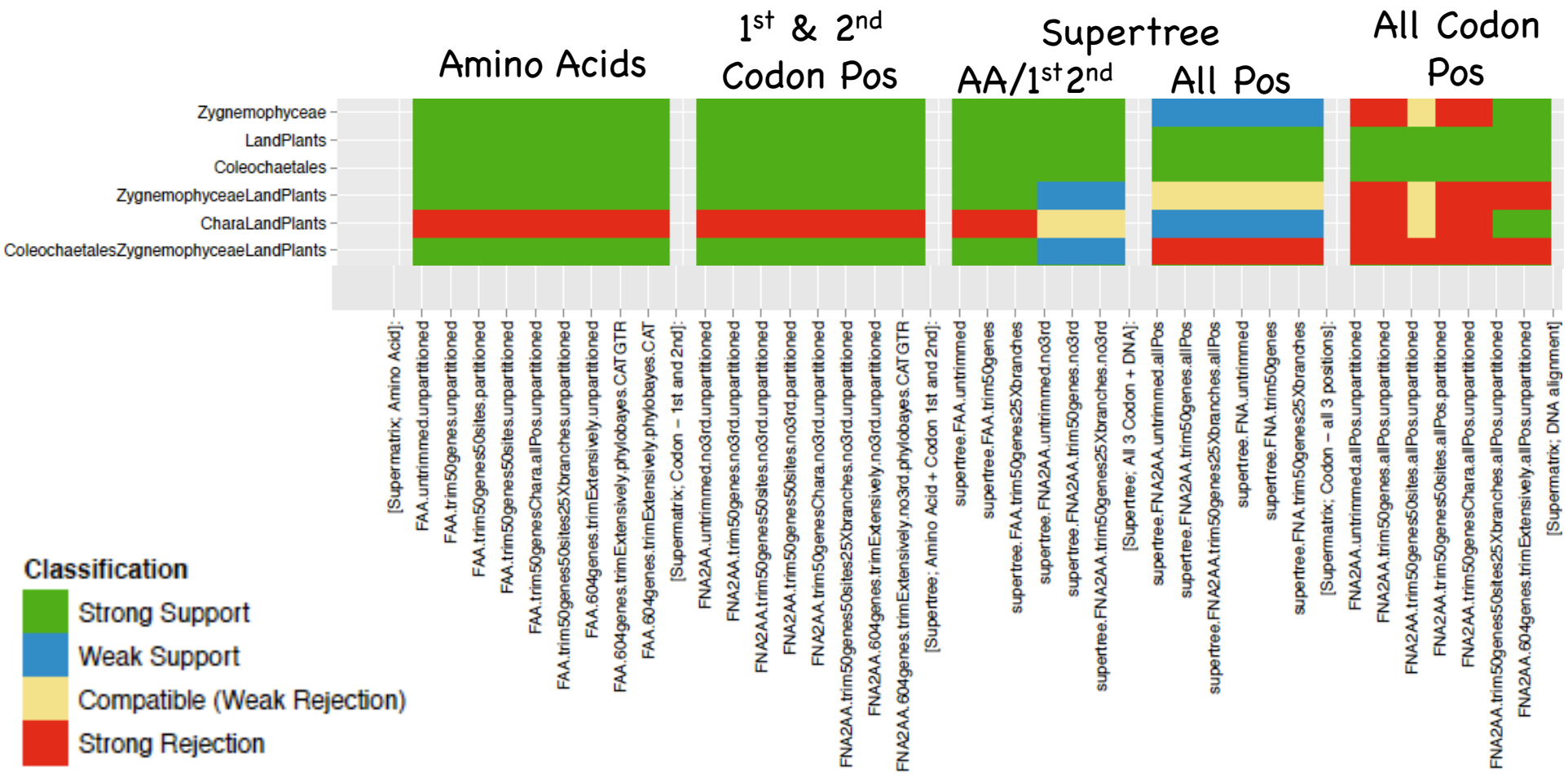


Data Matrix/Analysis Perturbation

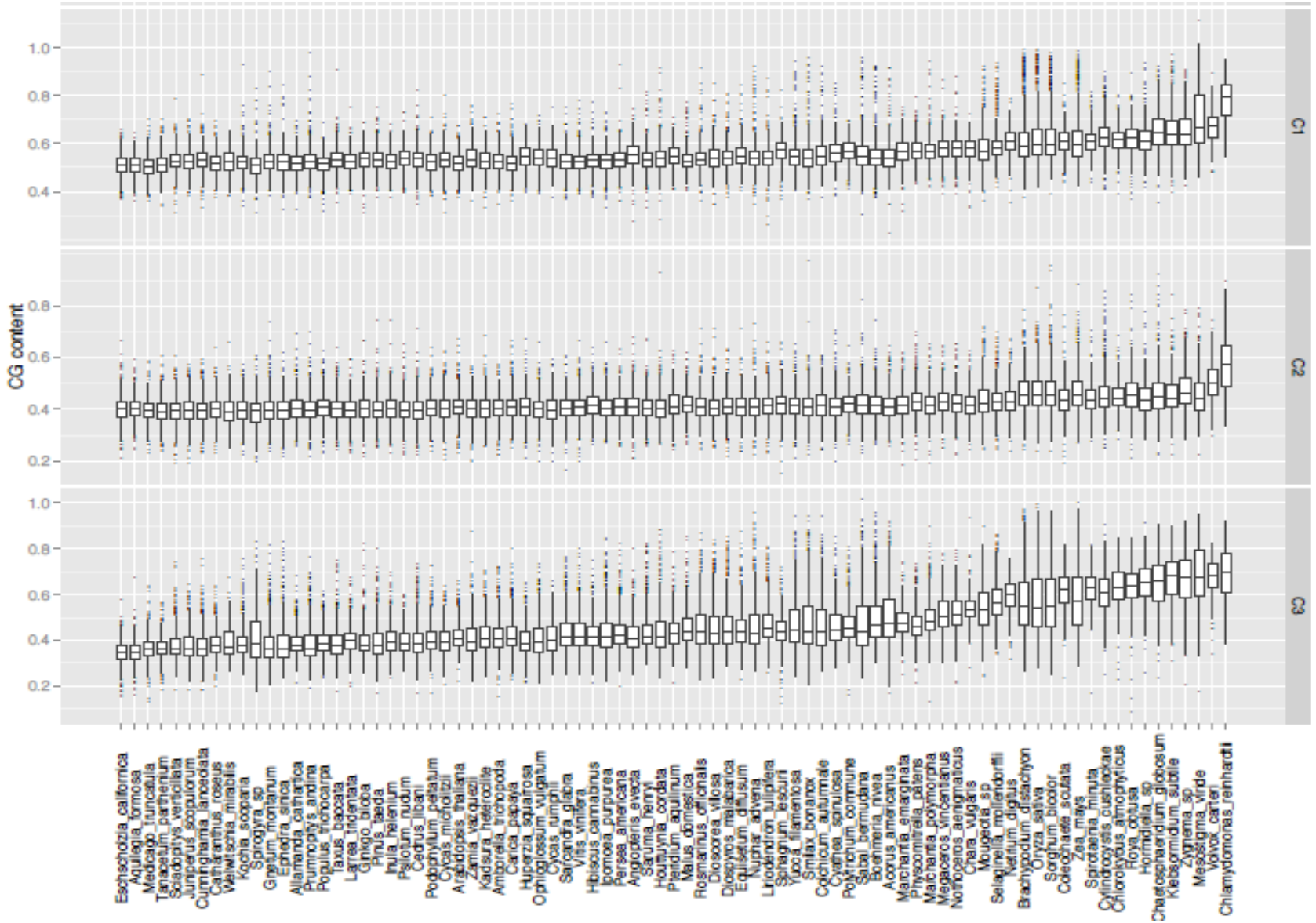
- DNA vs Amino Acids
- All codon positions vs remove third position
- Remove gappy genes or sites
- Remove genes on long branches in gene trees
- Remove genes where contamination seems to persist.
- Supermatrix and Supertree estimation

Good News: Some results seem to be very robust

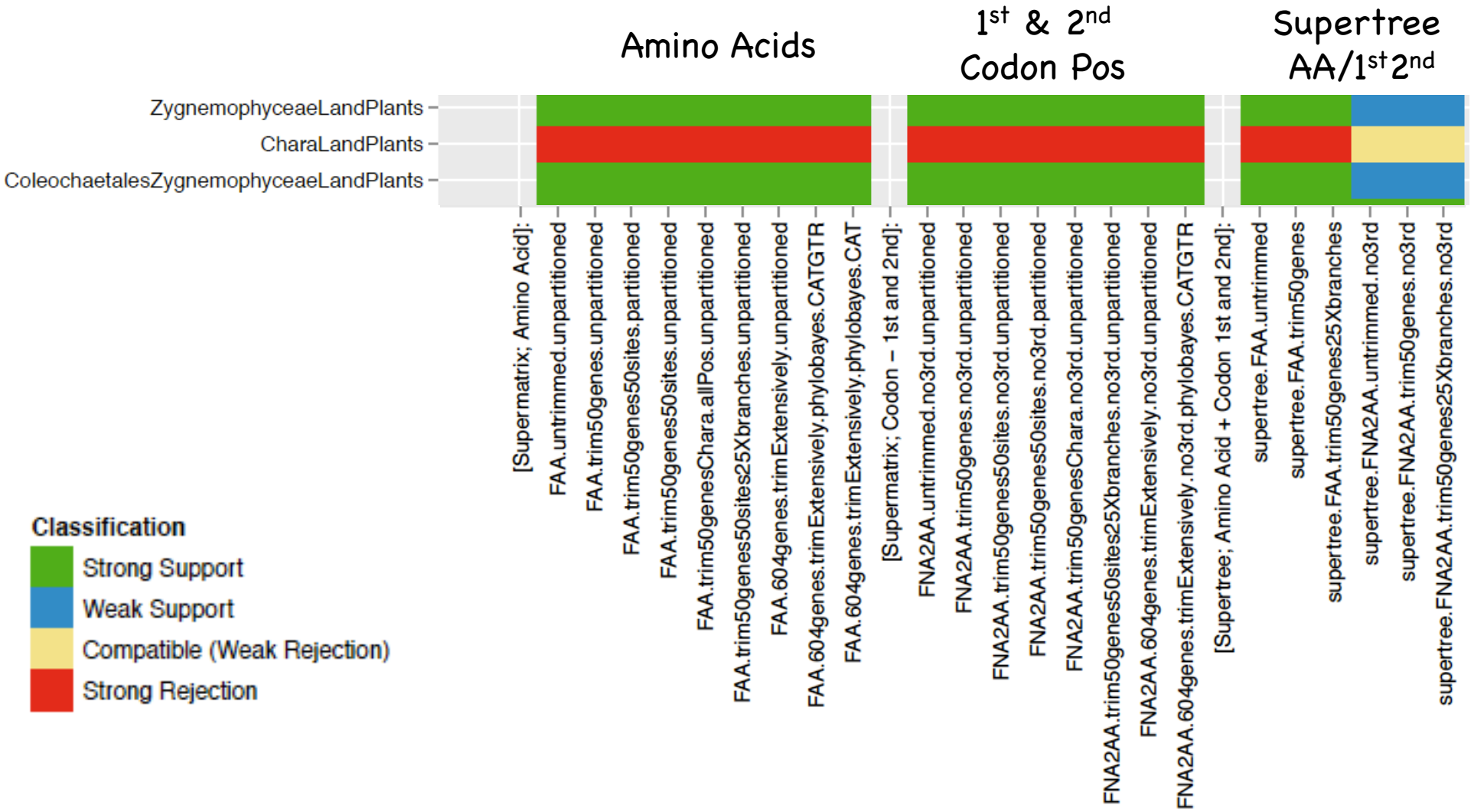
Issues with inclusion of 3rd codon position?

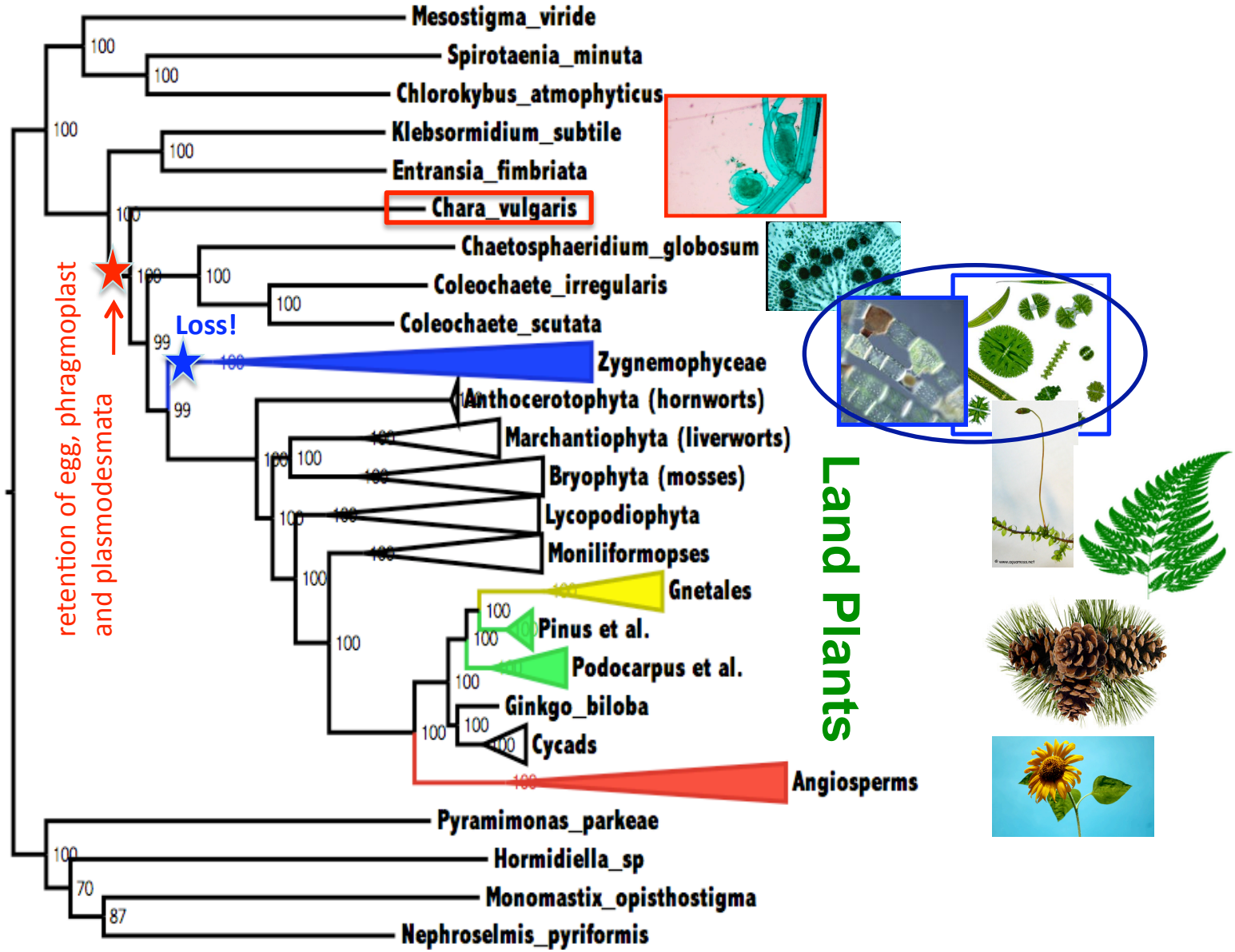


Problems due to heterogeneous substitution process?

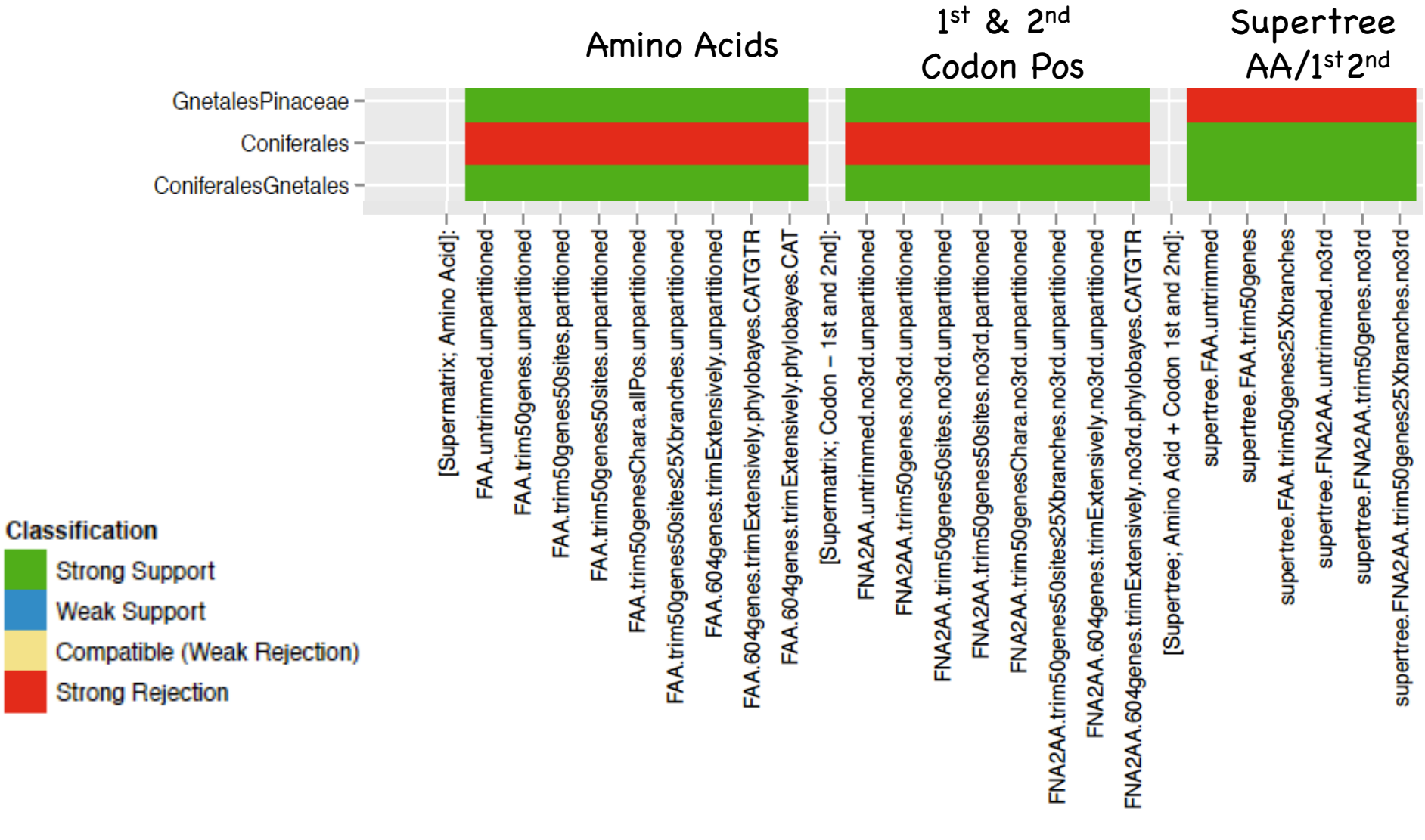


Filamentous Algae + Land Plant Hypothesis Seems to be Robust

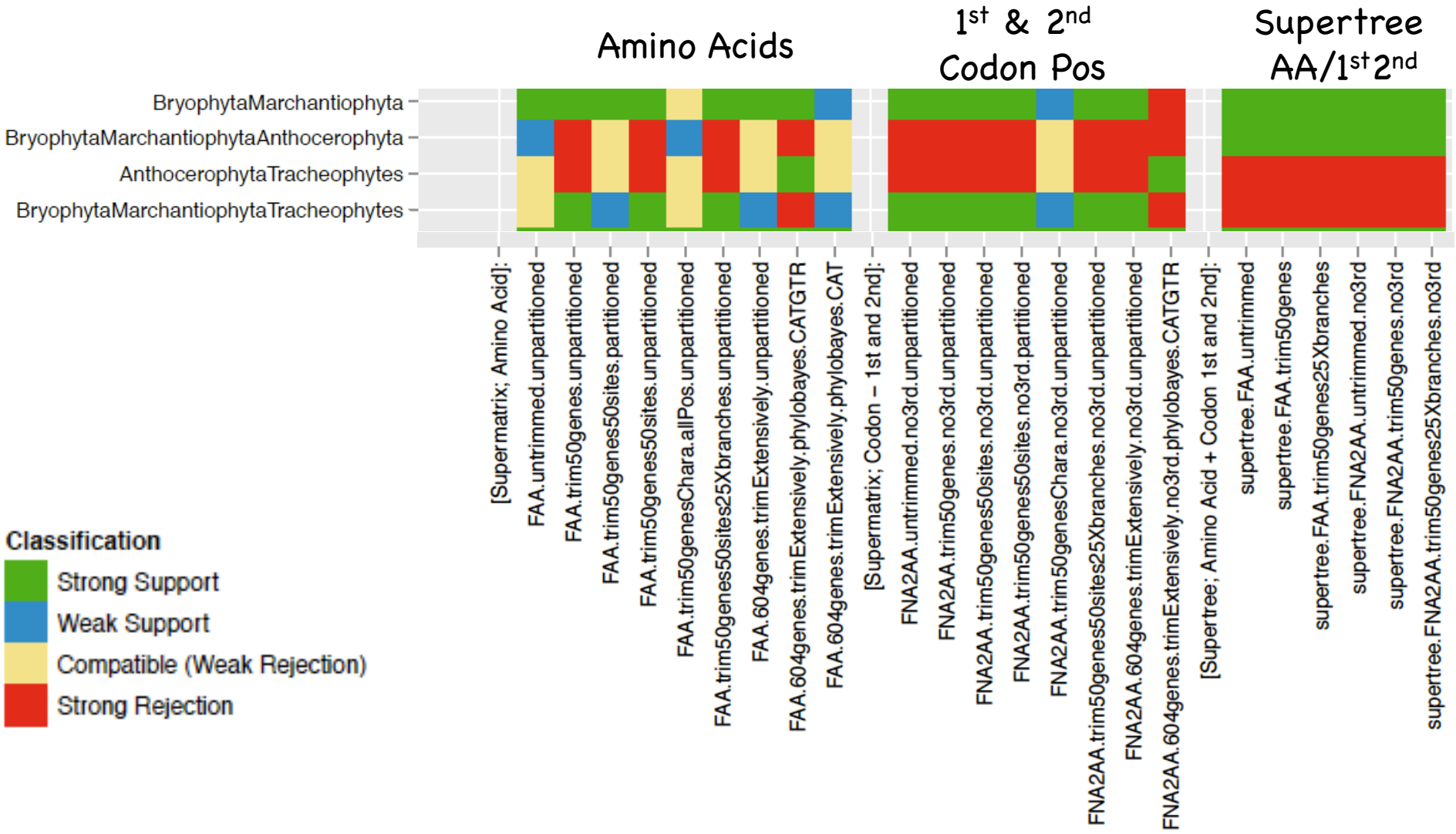



Mesostigma_viride
Spirotaenia_minuta
Chlorokybus_atmophyticus
Klebsormidium_subtile
Entransia_fimbriata
Chara_vulgaris
Chaetosphaeridium_globosum
Coleochaete_irregularis
Coleochaete_scutata
Zygnemophyceae
Anthocerotophyta (hornworts)
Marchantiophyta (liverworts)
Bryophyta (mosses)
Lycopodiophyta
Moniliformopses
Gnetales
Pinus et al.
Podocarpus et al.
Ginkgo_biloba
Cycads
Angiosperms
Pyramimonas_parkeae
Hormidiella_sp
Monomastix_opisthostigma
Nephroselmis_pyriformis
Loss!
retention of egg, phragmoplast and plasmodesmata
Land Plants

Supertree and Supermatrix Analyses Provide Different Inferences Concerning the Monophyly of Conifers



Mosses Sister + Liverworts Clade Sister to Vascular Plants?



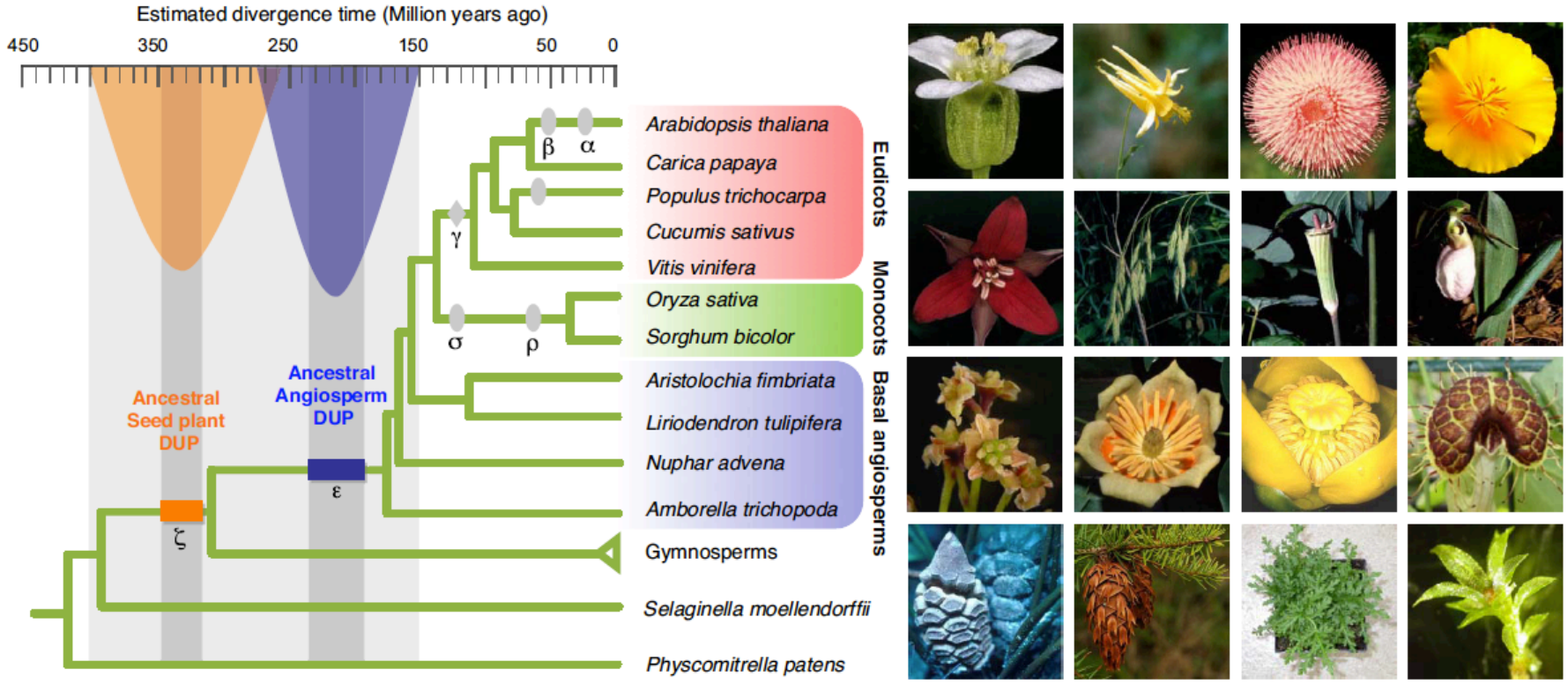
Conclusions

- Genome-scale phylogenetic inference is complex!
- Model mis-specification can results in statistical inconsistency: more data -> stronger support for the wrong answer.
- Proposal: Apply multiple analysis strategies and explore/understand basis of conflicts among resulting trees

Phylogenomics???

1. Using genome-scale data to resolve phylogentic relationships
2. Genome-Scale comparisons placed in a phylogenetic context

Phylogenetic Analysis and Molecular Dating of 100's of Gene Families Implicates Genome Duplications Associated with Origin of Angiosperms and Seed Plants



Jiao et al. 2011

Thank You!