

# Intergeneric allopolyploidy in Funariaceae revealed through targeted sequencing

Matthew G. Johnson  
Assistant Professor  
Texas Tech University

January 13, 2019



@mossmatters

## Coauthors

**Bernard Goffinet, University of Connecticut**  
**Rafael Medina, Augustana College**





*Plagiomnium cuspidatum*  
Wyatt & Odrzykoski, 1998, *Bryologist*



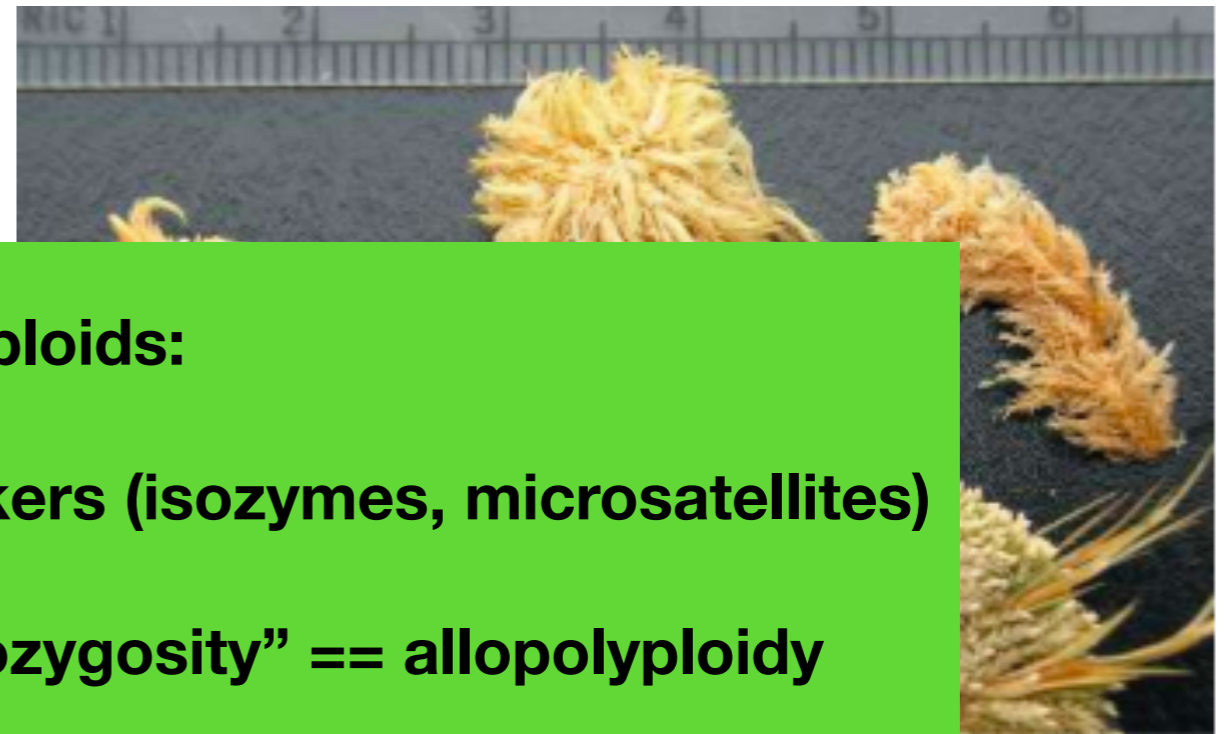
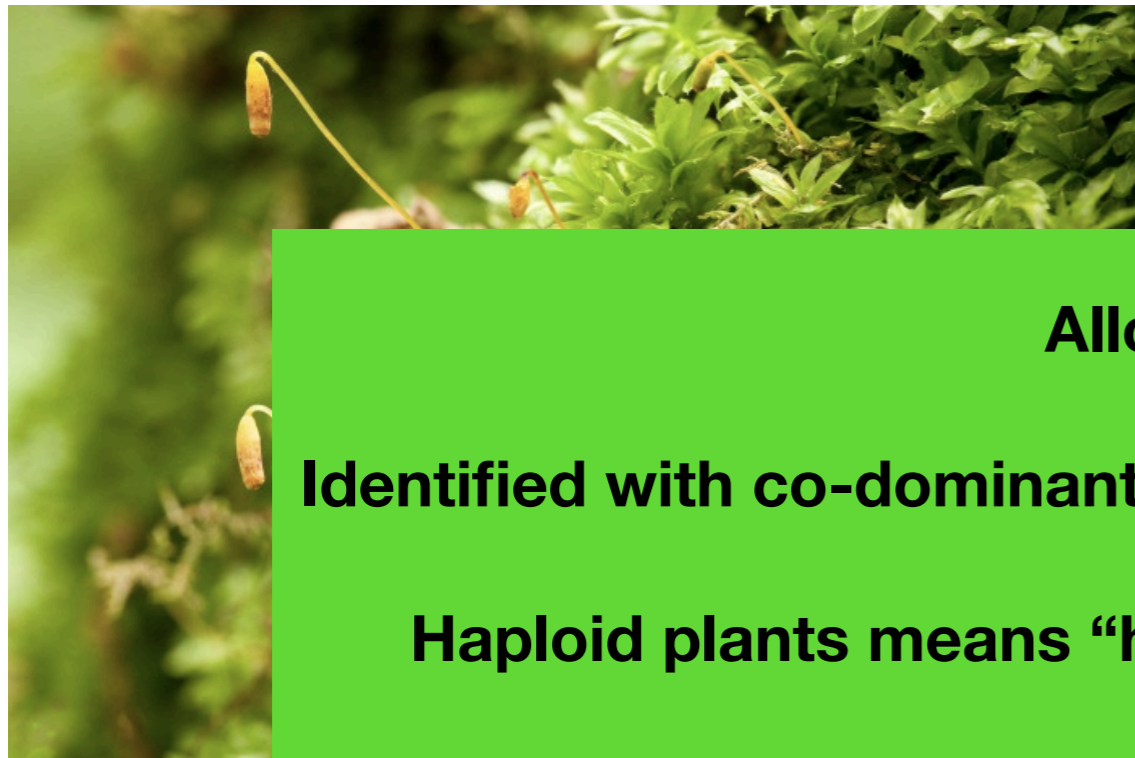
*Plagiochila britannica*  
Barbulescu et al., 2017, *Bot J Lin Soc*



*Sphagnum falcatulum*  
Karlin et al., 2009, *Molecular Ecology*



*Corsinia coriandrina*  
Boissellier-Dubayle & Bischler, 1998, *Bot Acta*



## Allopolyploids:

Identified with co-dominant markers (isozymes, microsatellites)

Haploid plants means “heterozygosity” == allopolyploidy

Hybrid parents closely related (less than 10 my divergence)

Pl  
Wyatt &

Ecology



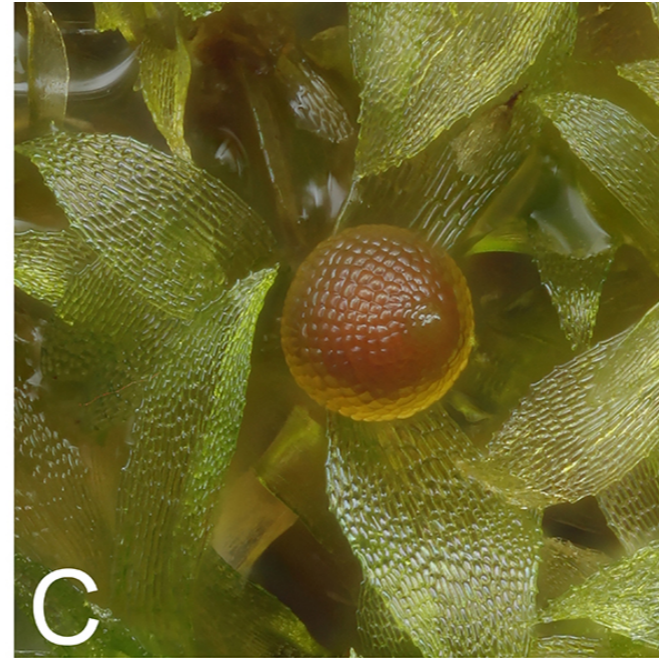
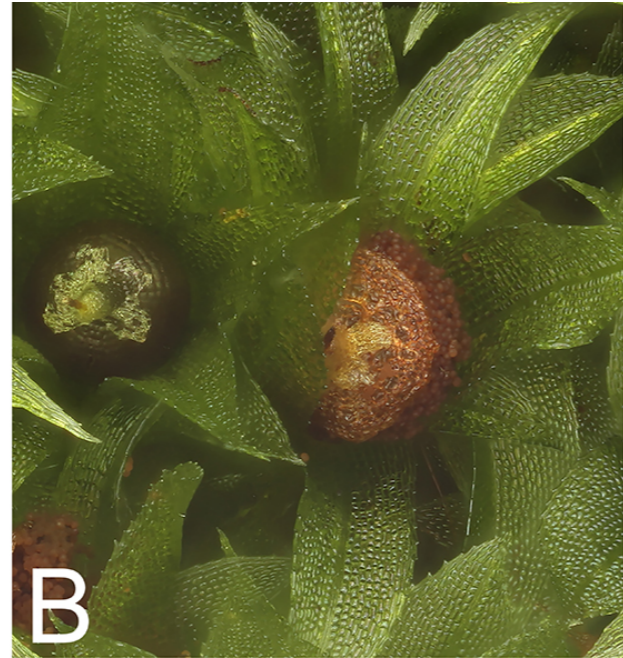
*Plagiochila britannica*

Barbulescu et al., 2017, *Bot J Lin Soc*



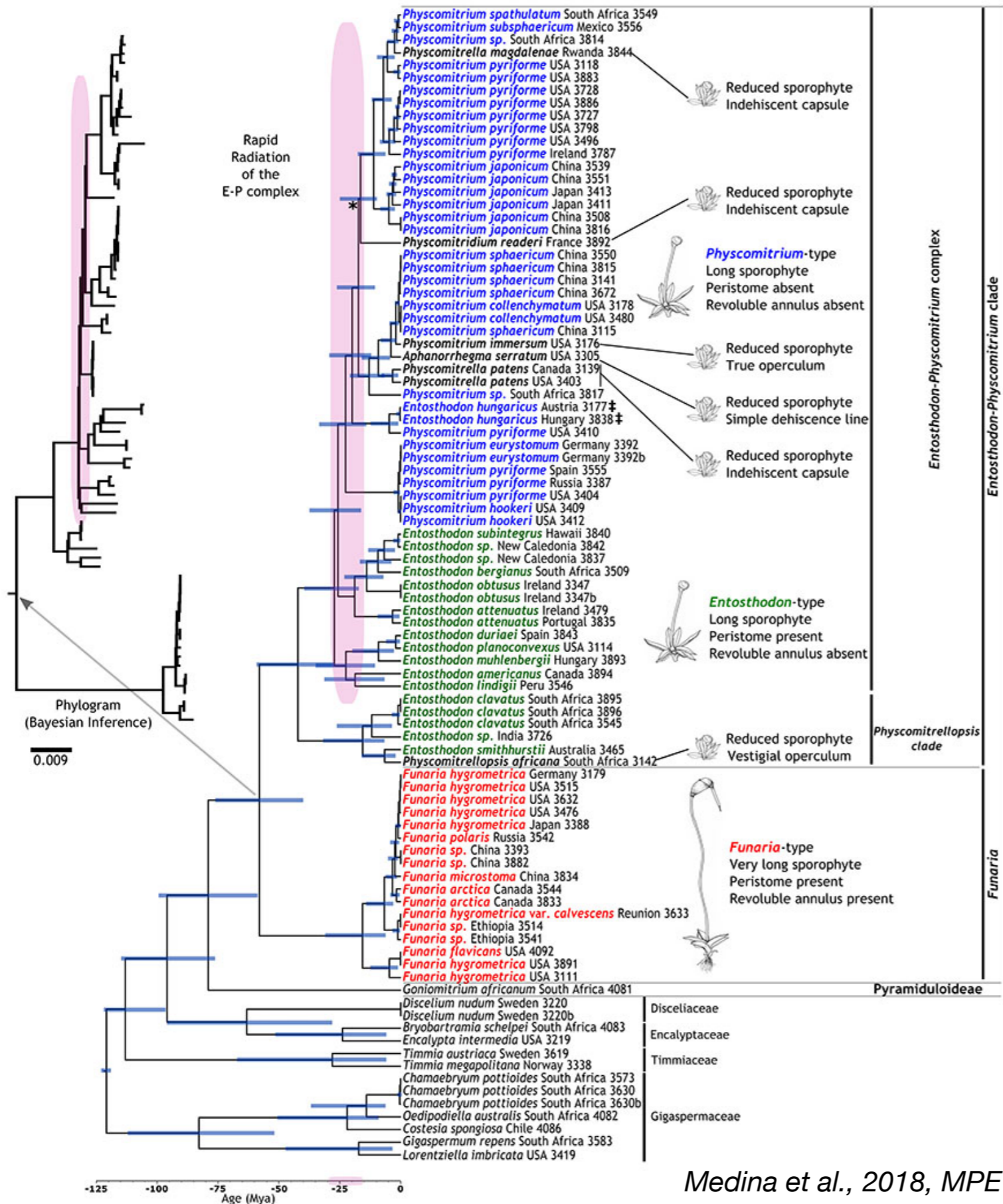
*Corsinia coriandrina*

Boissellier-Dubayle & Bischler, 1998, *Bot Acta*

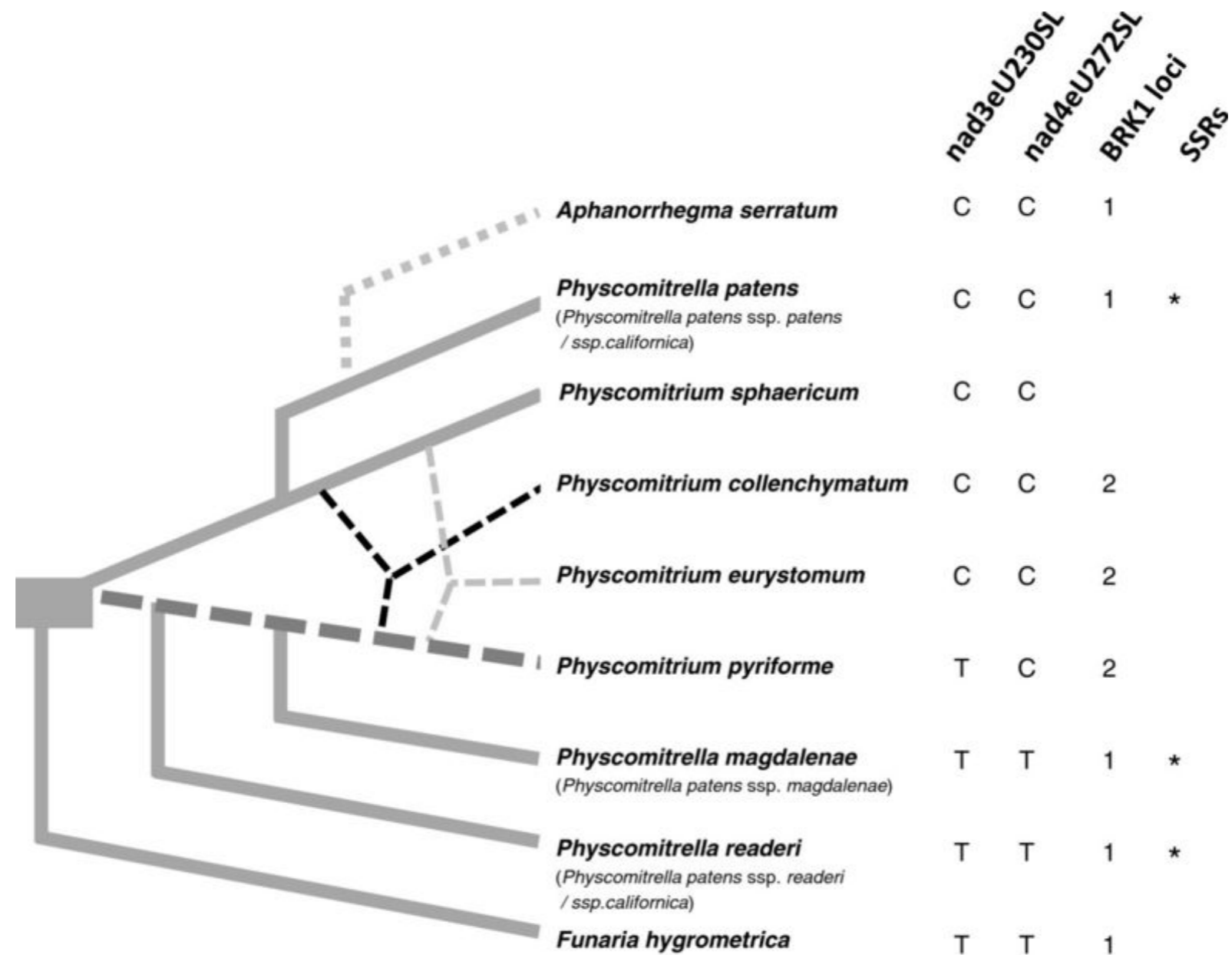


Funariaceae — family of 250+ species  
including model species *Physcomitrella patens*

Gametophytes of all species very similar  
Morphological variation in sporophyte stage



Phylogeny based on coding regions from all 82 chloroplast genes



Beike et al., 2014, BMC EvBio

Hybrid origin and large genome size in Physcomitrium-Entosthodon complex

# PHYSCOMITRIUM PHYLOGENY

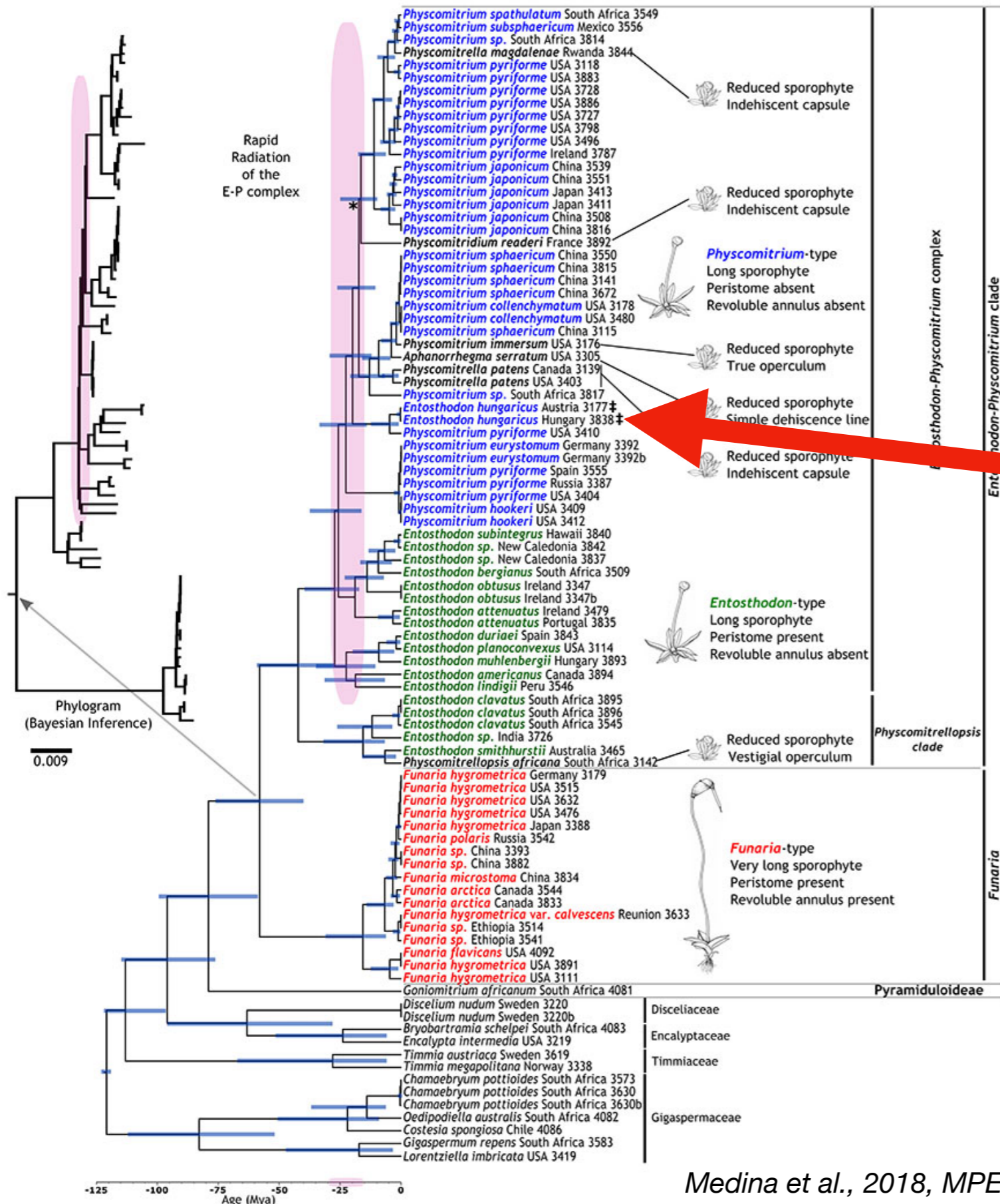


Photo: M. Ahrens

*Entosthodon hungaricus*

European endemic

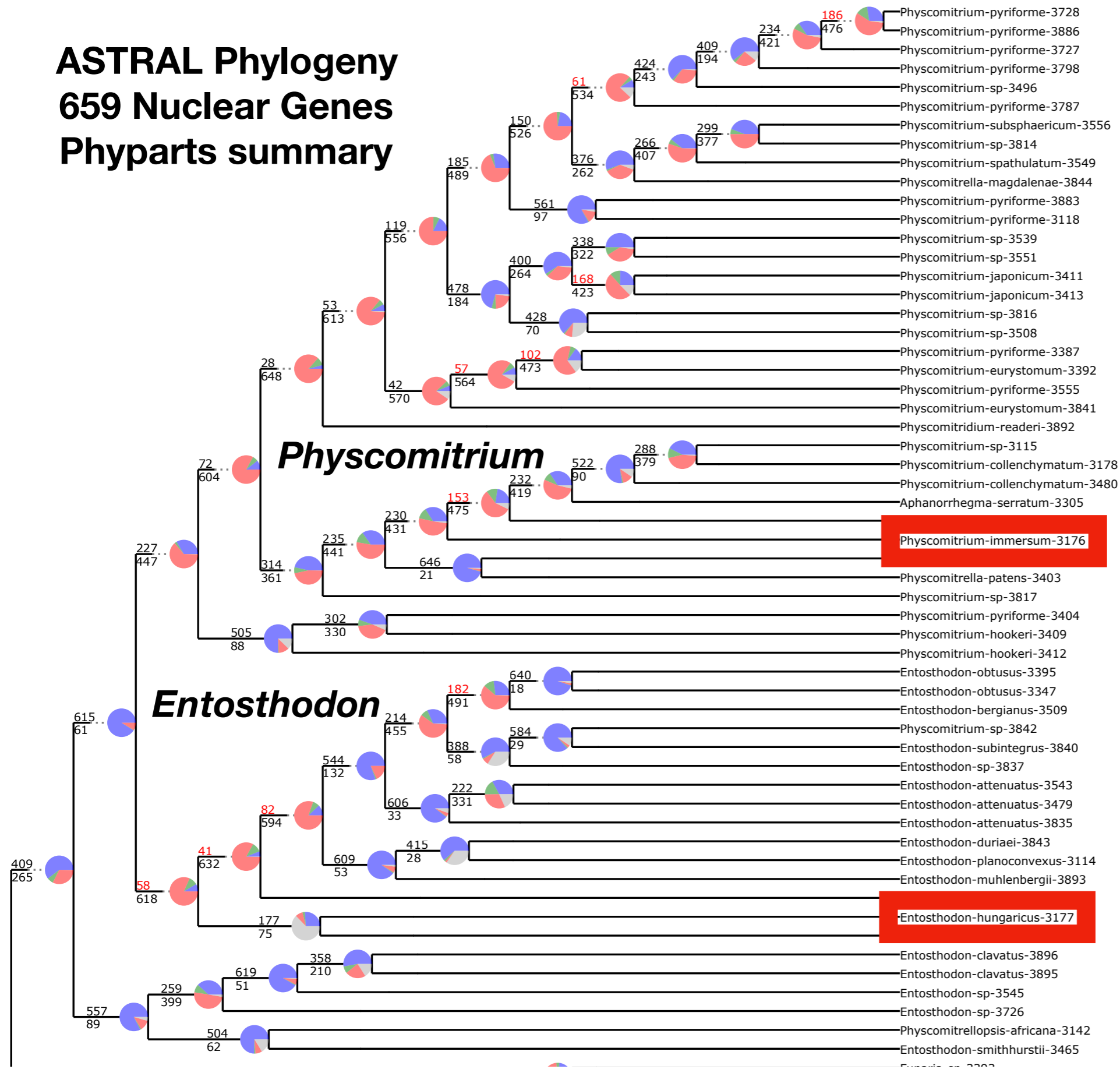
Not with other *Entosthodon*  
In organellar phylogeny



# ORGANELLAR VS NUCLEAR CONFLICT



## ASTRAL Phylogeny 659 Nuclear Genes Phyparts summary



Samples with high numbers of paralogs generate gene conflict:

*E. hungaricus*  
*P. immersum*





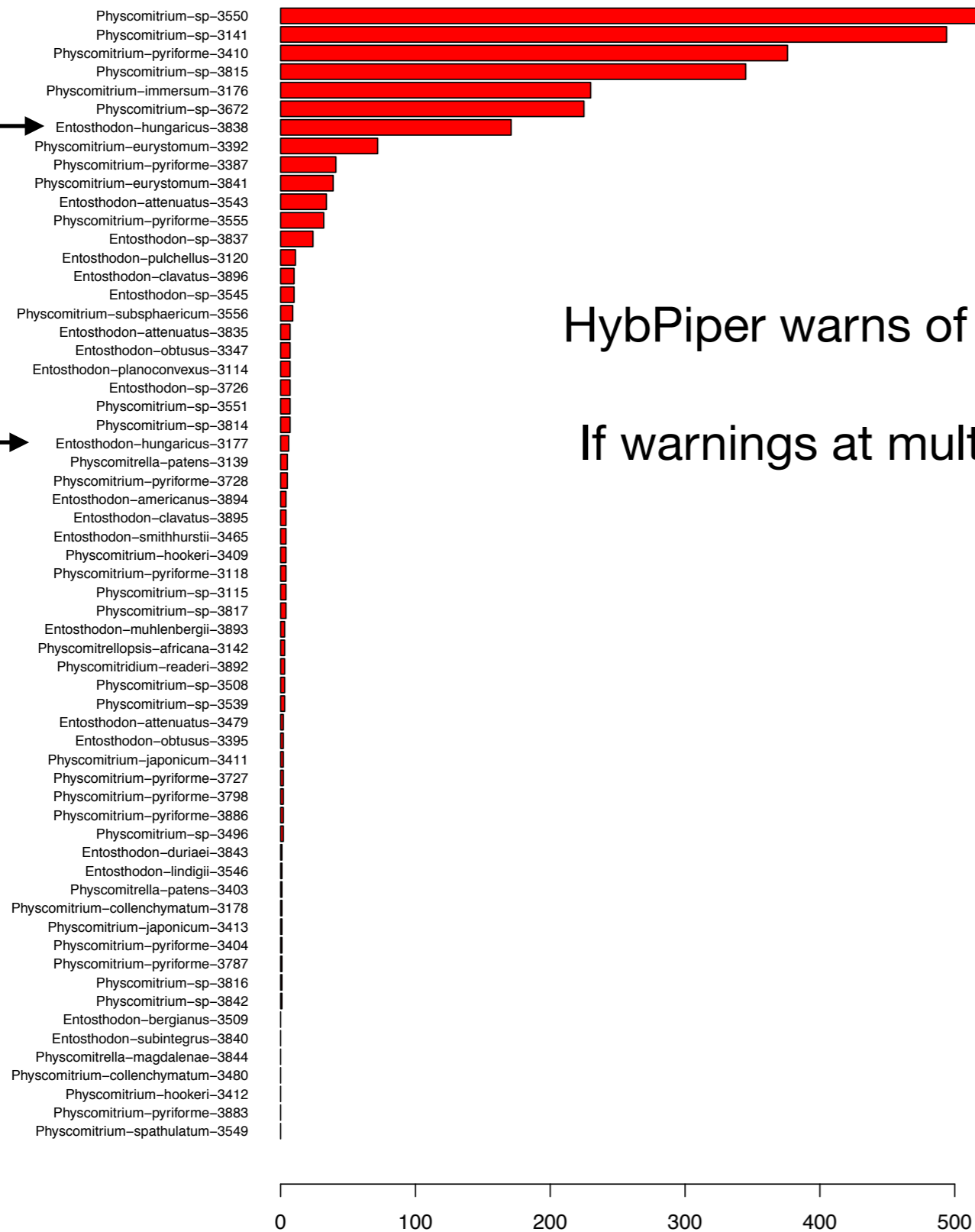
# RESEARCH GOALS



1. Identify samples with heterozygosity in target sequencing data
2. Extract phased DNA sequences to include homeologs in gene tree inference
3. Identify potential parents for allopolyploid taxa



# PARALOG WARNINGS



HybPiper warns of multiple sequences assembled for one gene

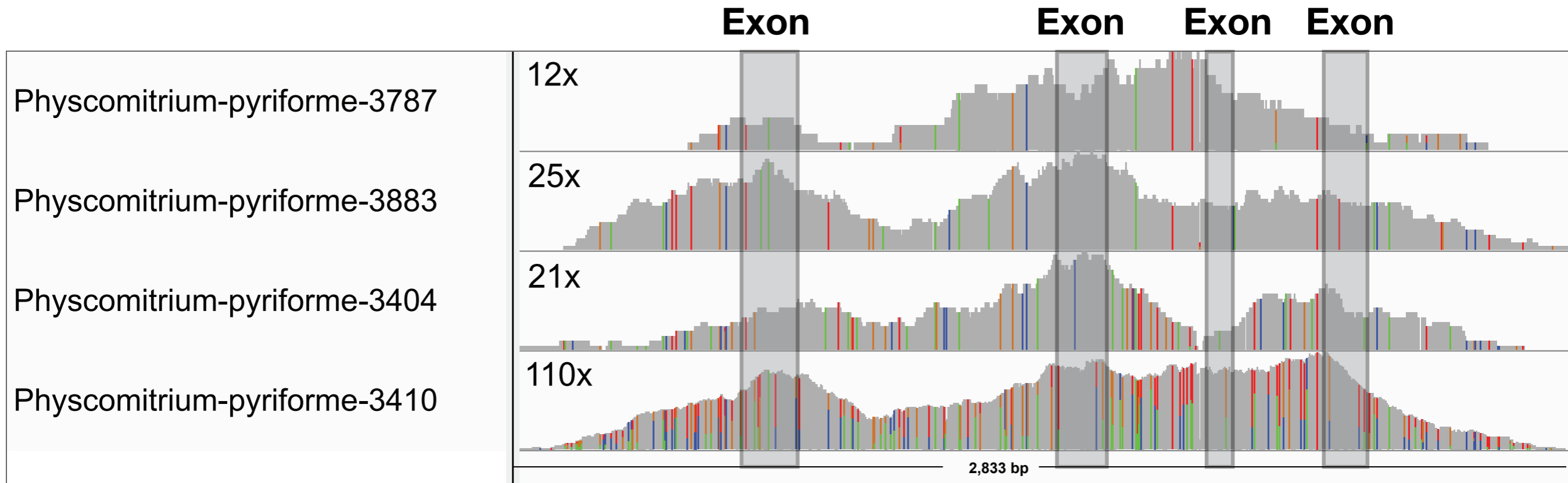
If warnings at multiple genes, could indicate polyploid sample



Number of genes with paralog warnings



# IDENTIFYING ALLOPOLYPLOIDS



## Workflow

Reads aligned to HybPiper “supercontigs”

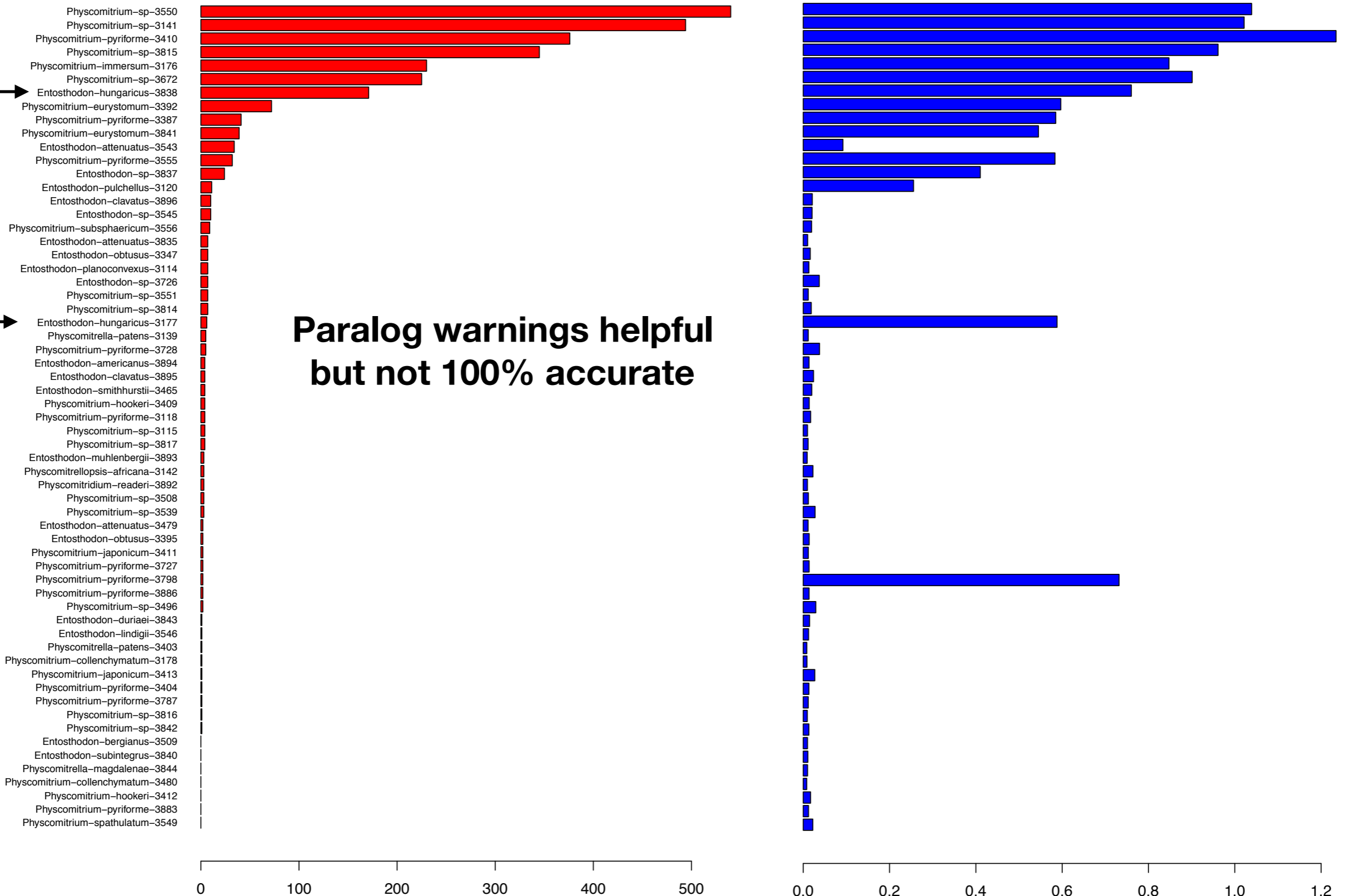
Picard: Remove PCR duplicates

GATK: Call Variant Sites within individuals

**Code: [github.com/mossmatters/phyloscripts](https://github.com/mossmatters/phyloscripts)**



# IDENTIFYING ALLOPOLYPLOIDS



**Paralog warnings helpful  
but not 100% accurate**

**Number of genes with paralog warnings**

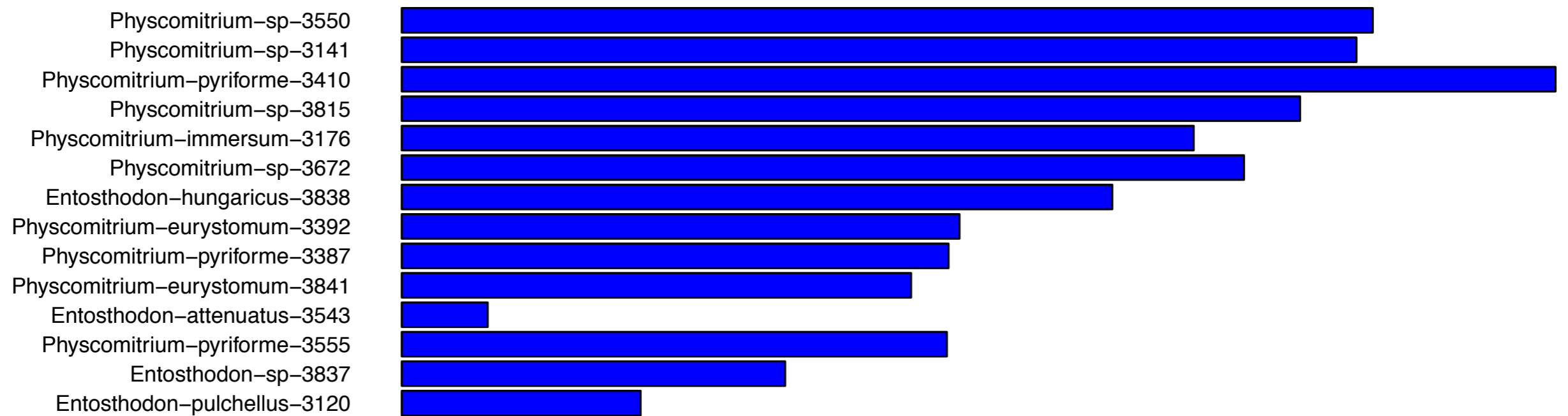
**Heterozygosity**



# CONFIRMING HETEROZYGOSITY



## Percent Heterozygosity

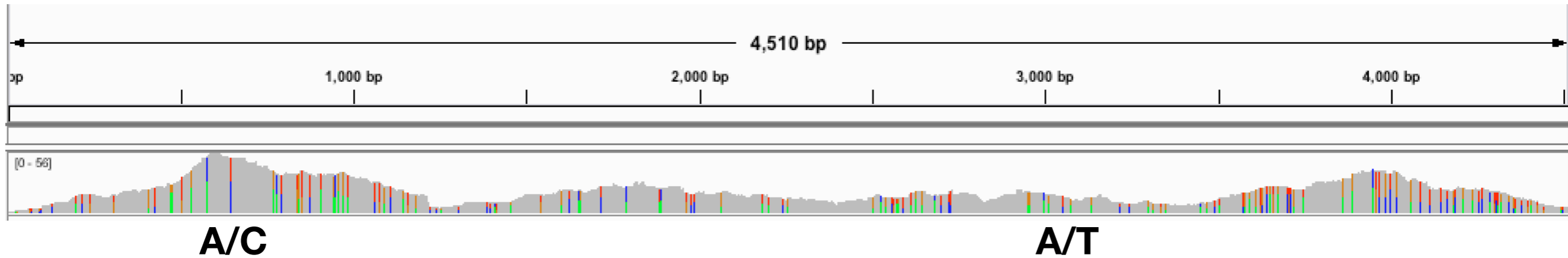


Initial project:

*Entosthodon hungaricus*  
*Physcomitrium pyriforme* (3410)  
*Physcomitrium immersum*



# ALLELE PHASING



How to generate two sequences for gene tree analysis?

Use the sequencing reads (300 bp paired-end) to generate haplotypes.

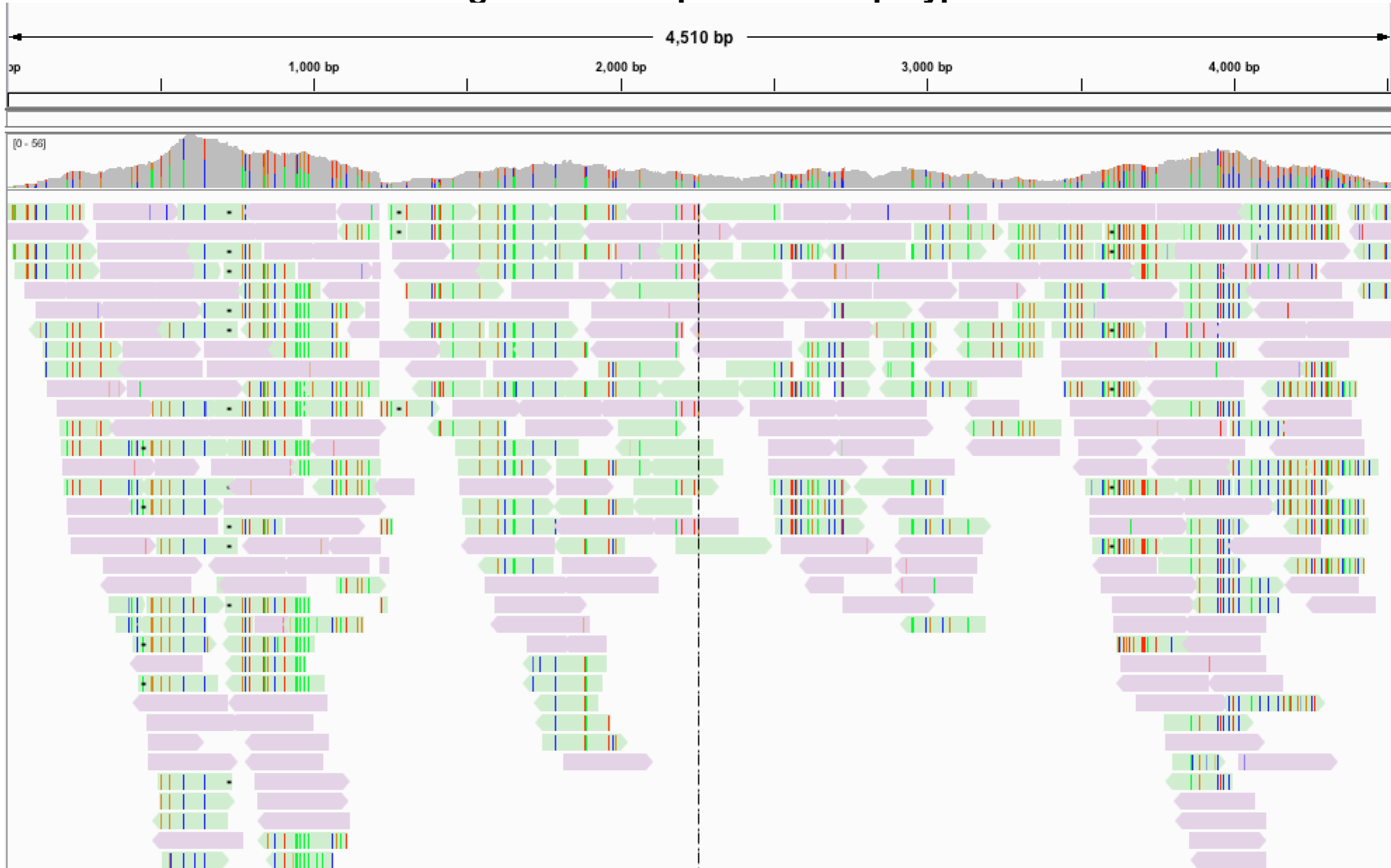
Read-backed phasing using GATK and WhatsHap

**Code: [github.com/mossmatters/phyloscripts](https://github.com/mossmatters/phyloscripts)**



# ALLELE PHASING

## Phasing success in putative allopolyploids



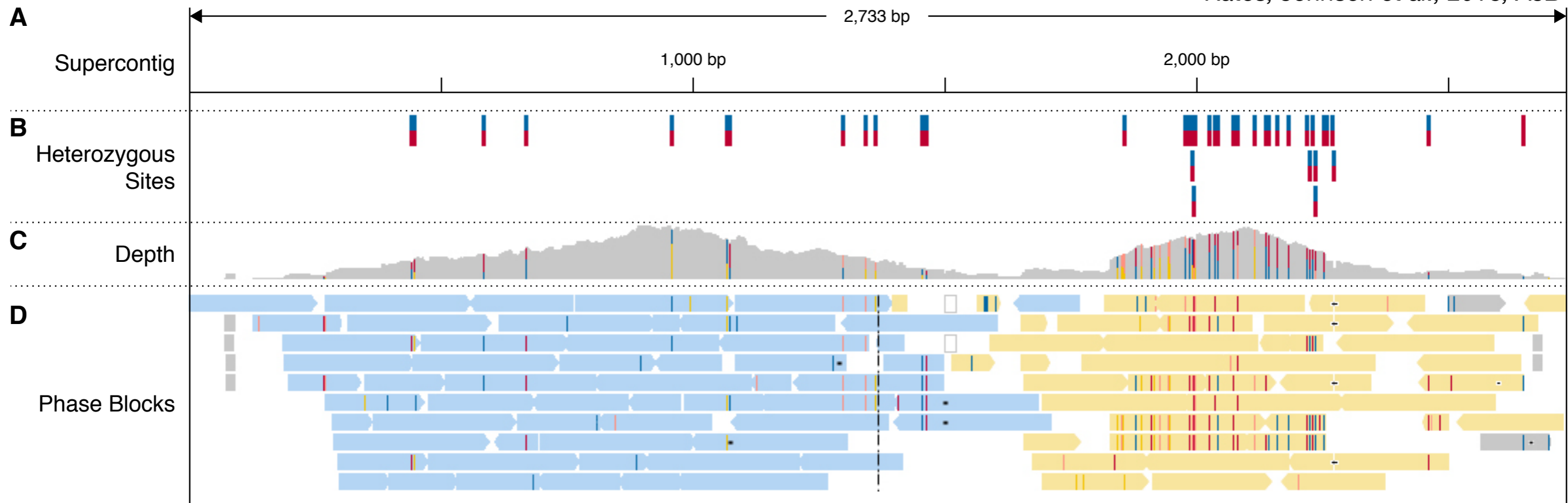
Reads colored by haplotype



# ALLELE PHASING

## Read-backed Phasing using GATK and WhatsHap

*Kates, Johnson et al., 2018, AJB*

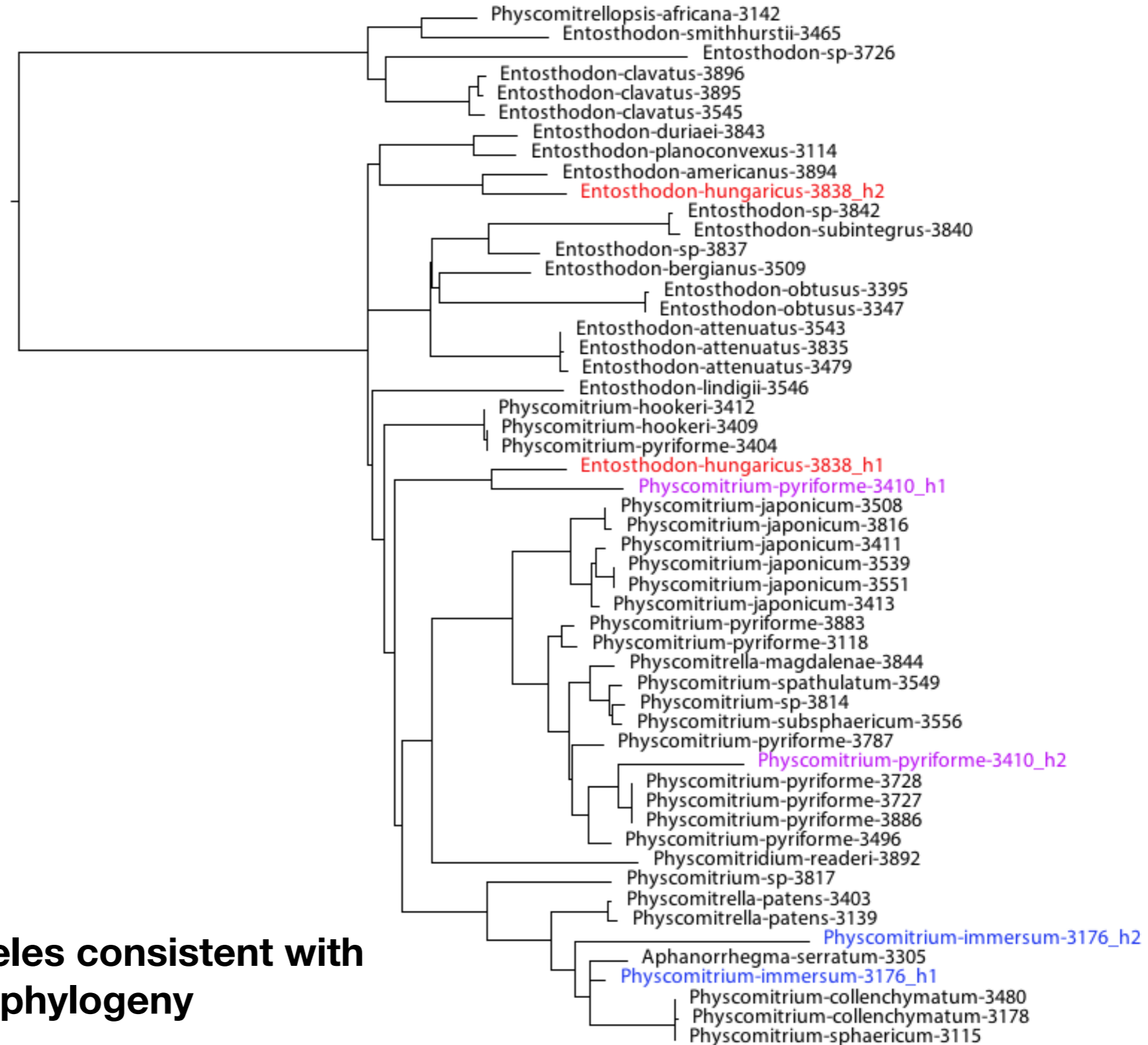


For homeologs, only the phased sequence in the largest block was retained





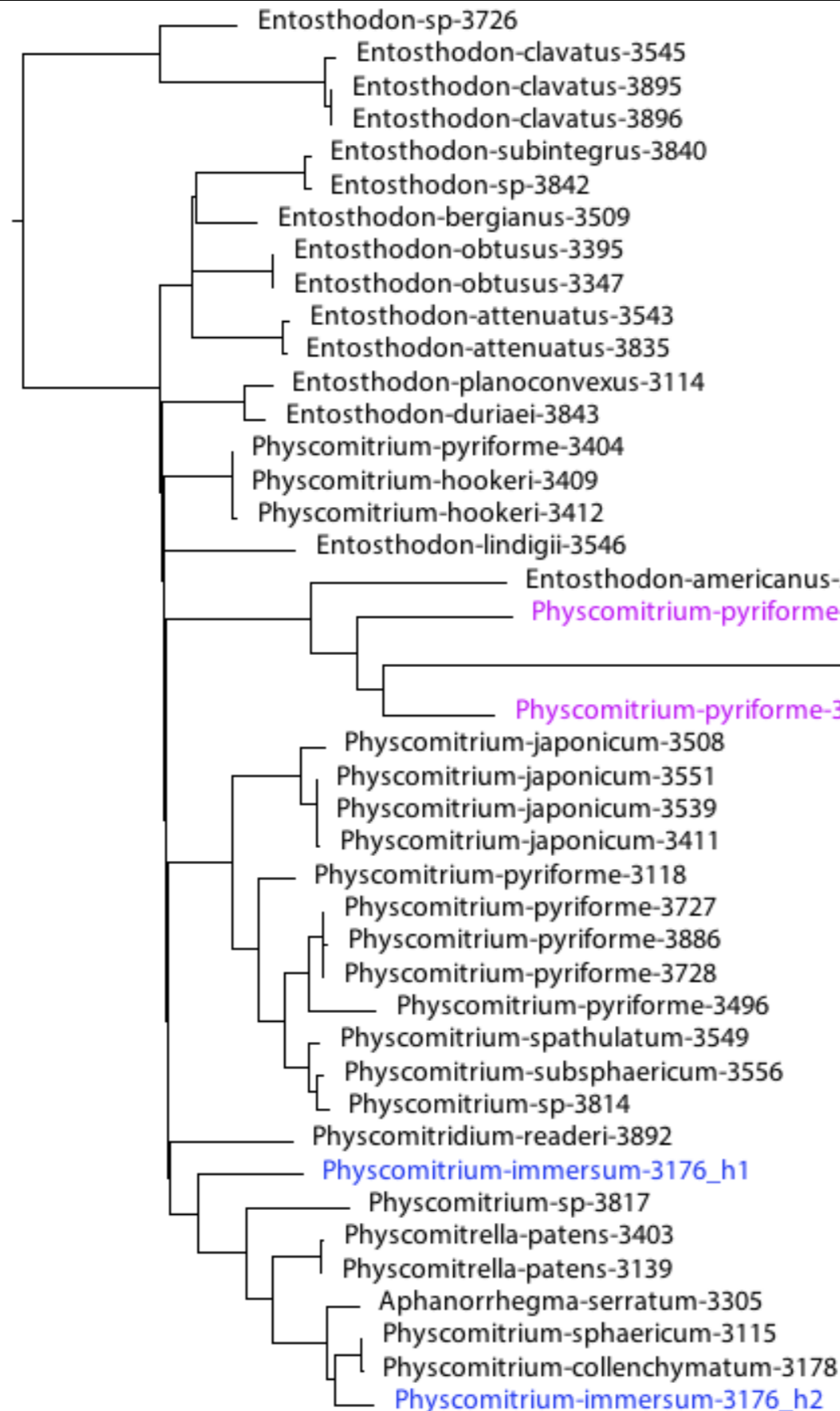
# GENE TREES WITH HOMEOLOGS



Many genes have alleles consistent with chloroplast phylogeny



# GENE TREES WITH HOMEOLOGS



Other genes are not consistent with allopolyploidy, but these are more rare

Evidence of tetraploid hybrids?  
Duplicated genes?

Need a rigorous way to summarize gene trees...



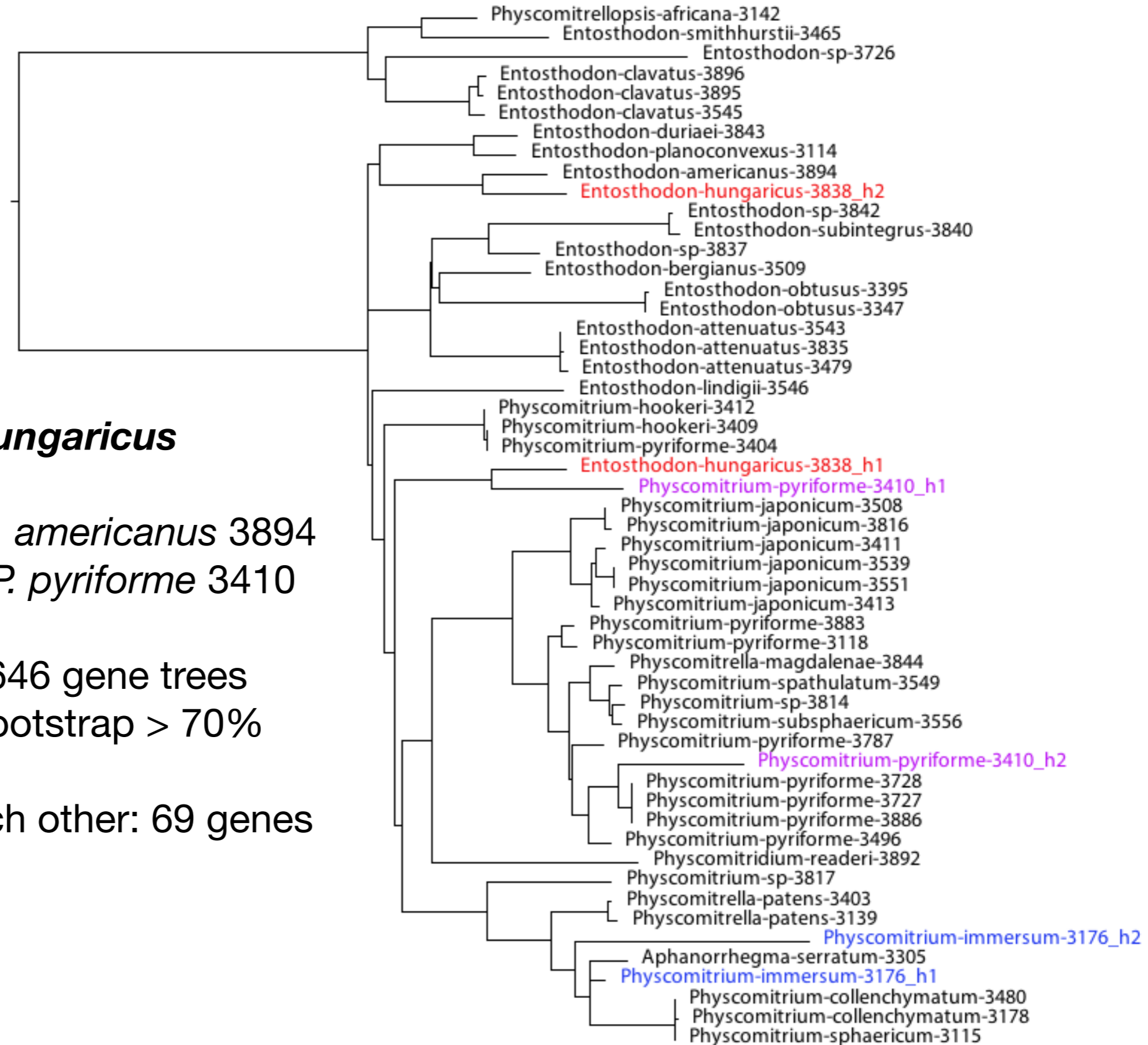
## *Entosthodon hungaricus*

Homeolog 1: Sister to *E. americanus* 3894

Homeolog 2: Sister to *P. pyriforme* 3410

Pattern found: 416/646 gene trees  
399 with gene tree bootstrap > 70%

Homeologs sister to each other: 69 genes



# INTERGENERIC HYBRID?

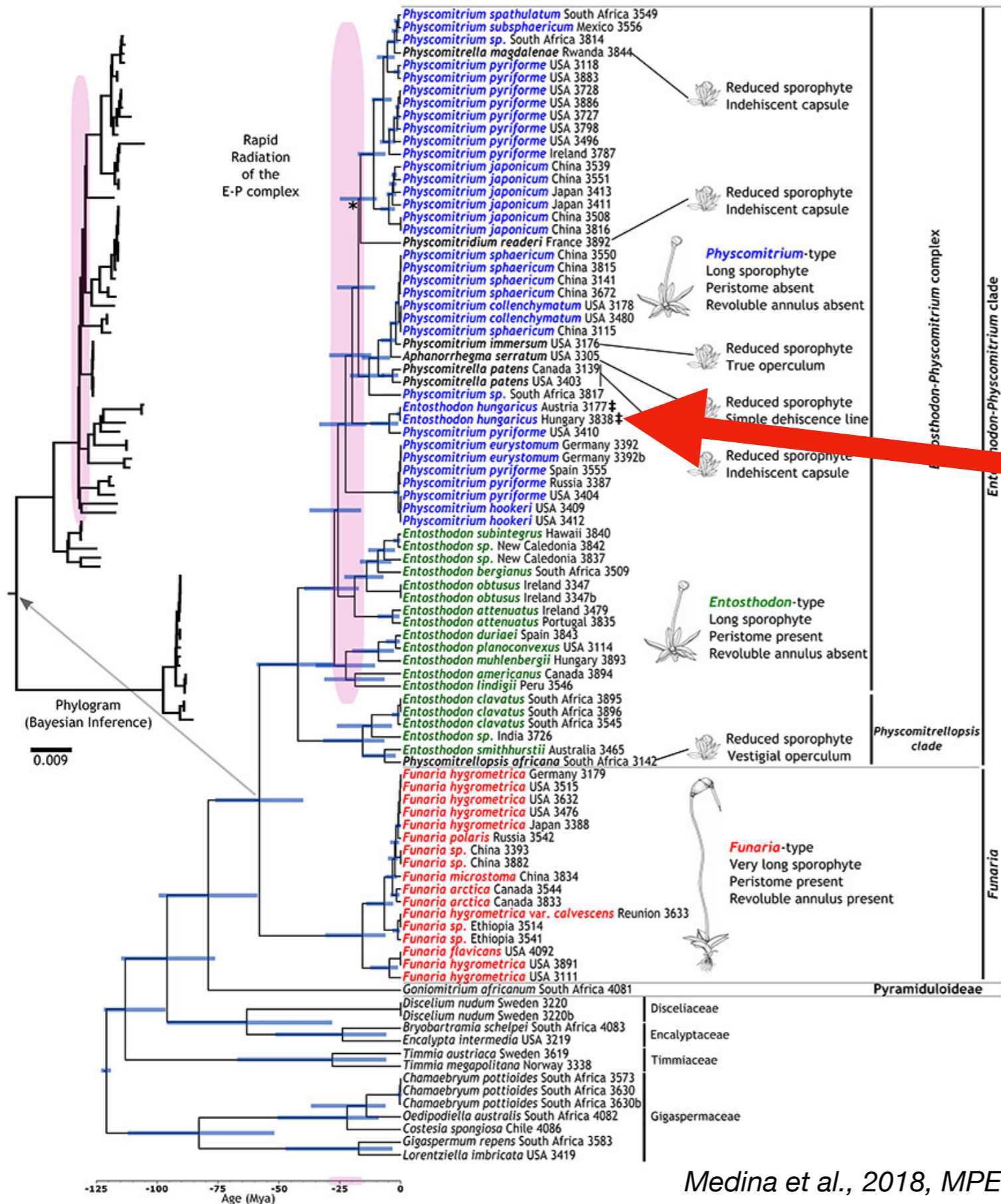


Photo: M. Ahrens

*Entosthodon hungaricus*

**Maternal parent: *Physcomitrium***

**Paternal parent: *Entosthodon***

**Parental divergence: > 25 mya**



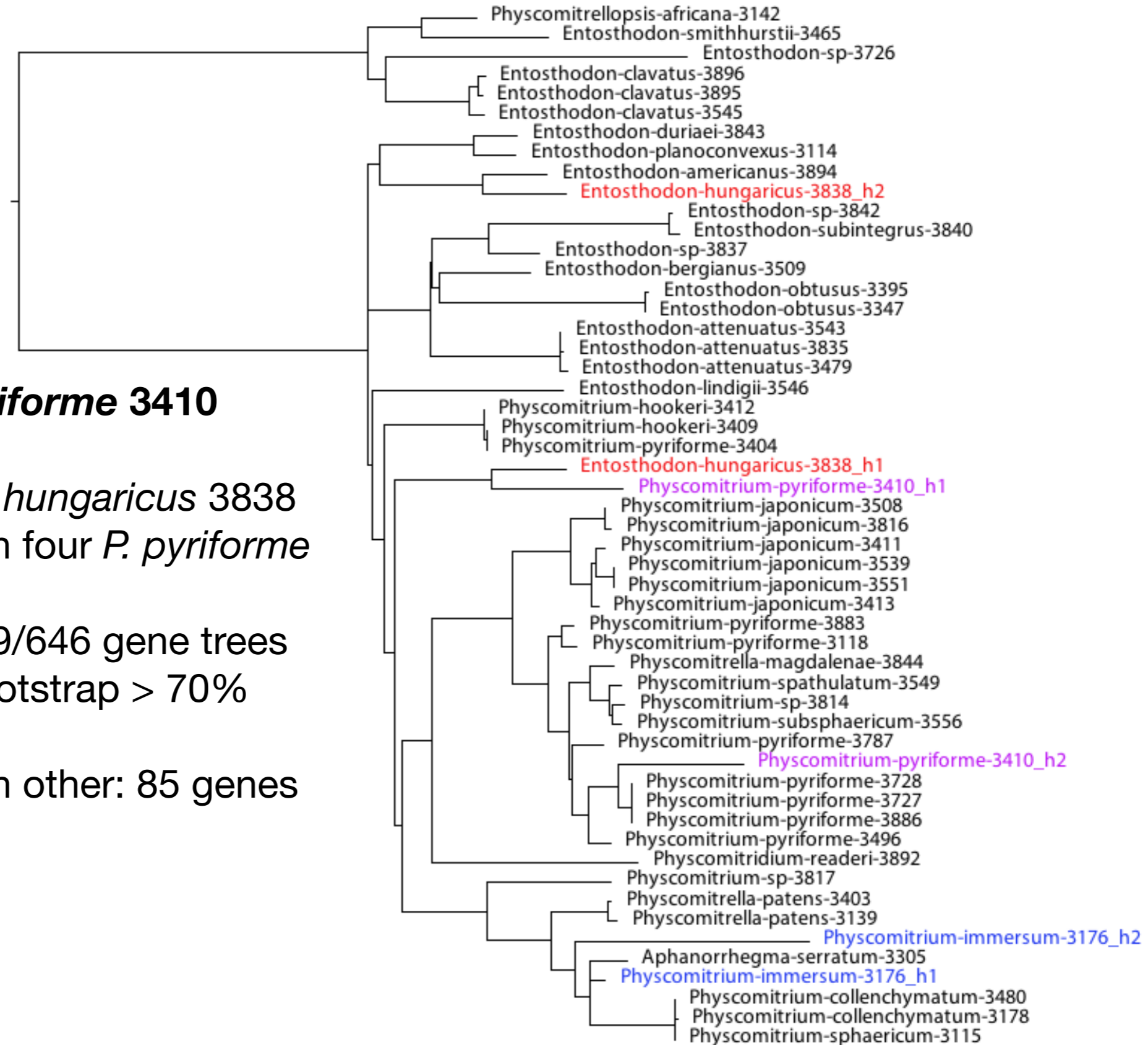
## *Physcomitrium pyriforme* 3410

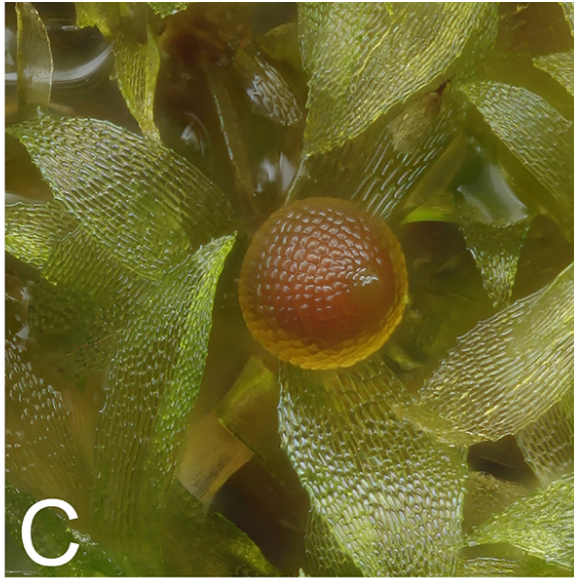
Homeolog 1: Sister to *E. hungaricus* 3838

Homeolog 2: In clade with four *P. pyriforme*

Both patterns found: 299/646 gene trees  
212 with gene tree bootstrap > 70%

Homeologs sister to each other: 85 genes





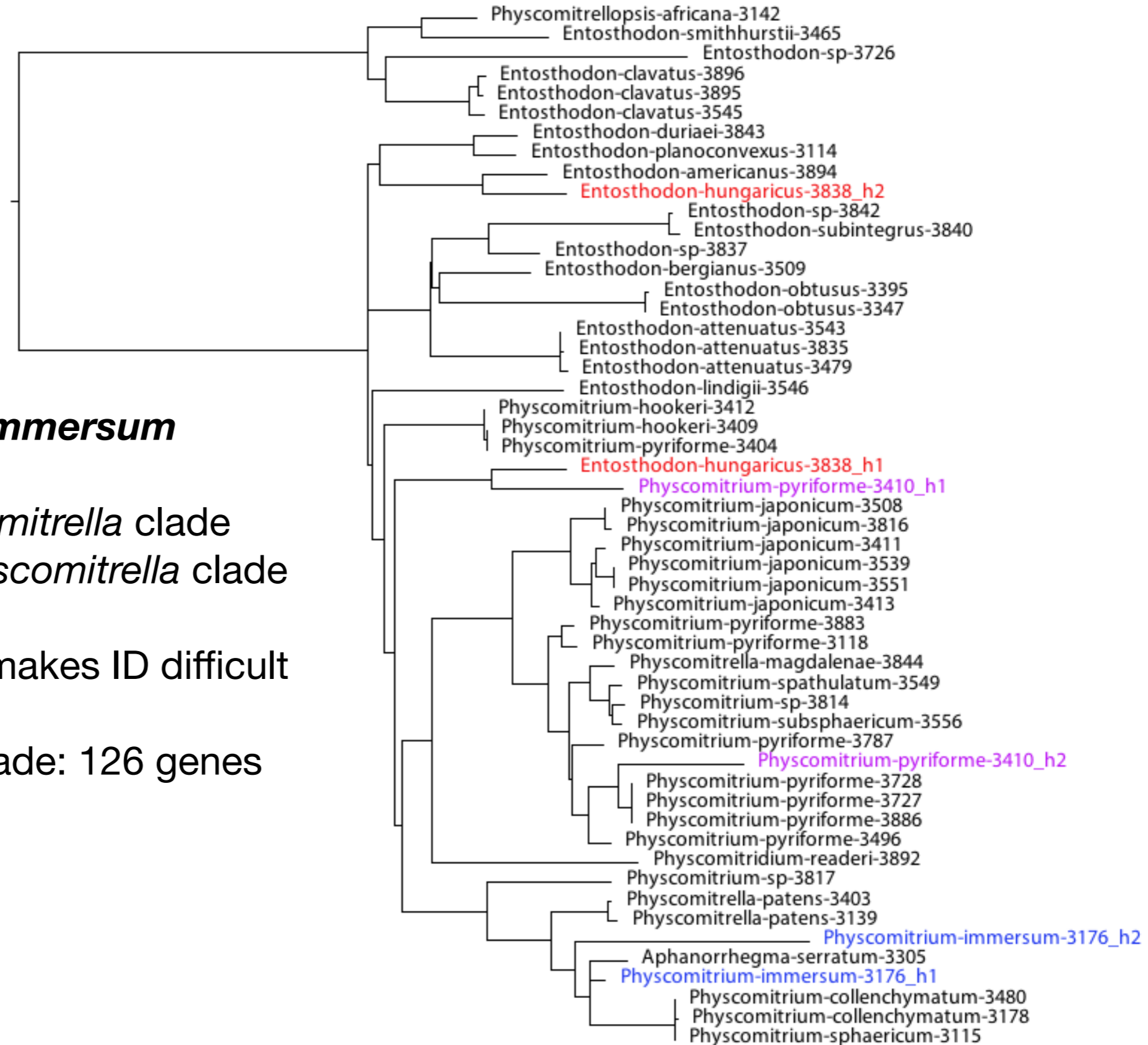
## *Physcomitrium immersum*

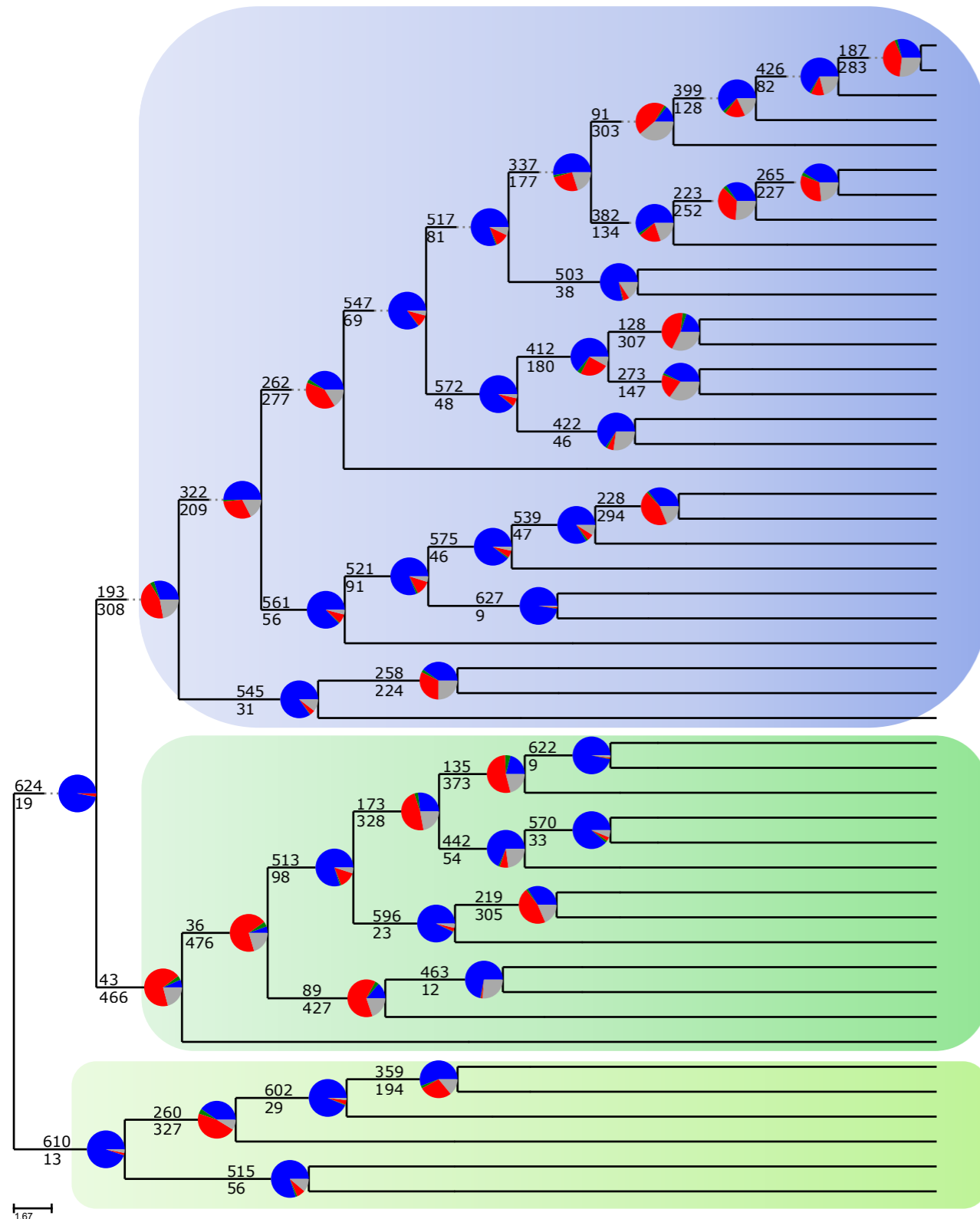
Homeolog 1: *Physcomitrella* clade

Homeolog 2: Not in *Physcomitrella* clade

Gene tree discordance makes ID difficult

Homeologs in same clade: 126 genes





Identifying ancient allopolyploid parents in mosses made easier with target sequencing

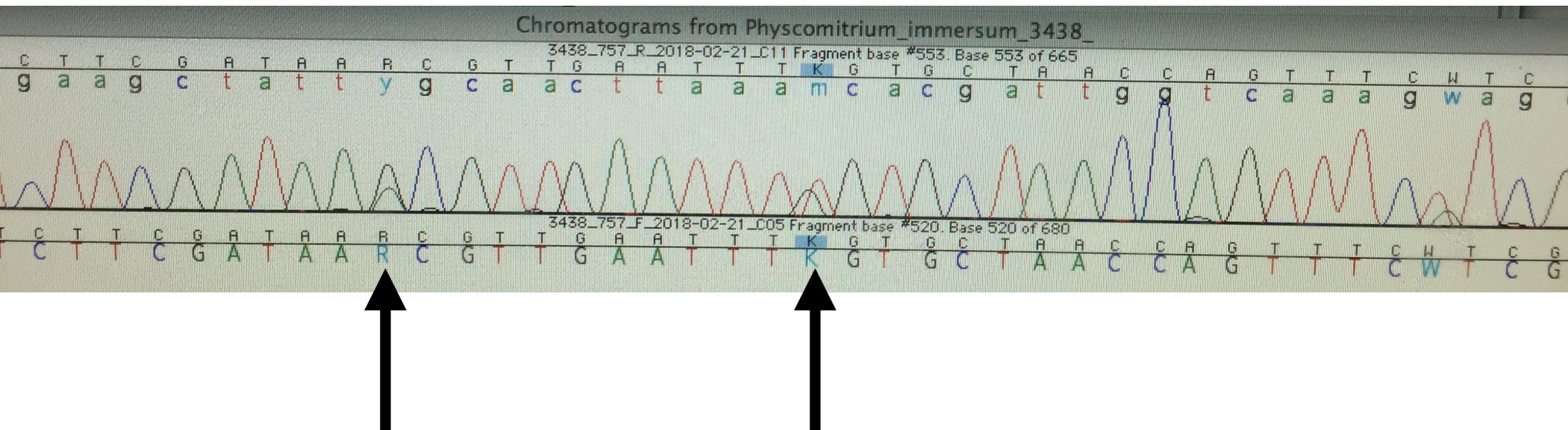
Confirmed hybrid origin of at least two species in *Entosthodon-Physcomitrium* complex

Disagreement among nuclear genes in haploid taxa complicate assessment of allopolyploid hybrid parents

Future: Identify all homeologs to include in a single ASTRAL analysis to more accurate placement of parental genomes

## Future Directions

Large phylogeny of *E. hungaricus* and *P. immersum* using 2-step PCR (5 genes)



Sanger sequencing of HybSeq genes reveals “double peaks” in SNP locations for other samples of *P. immersum* and *E. hungaricus*



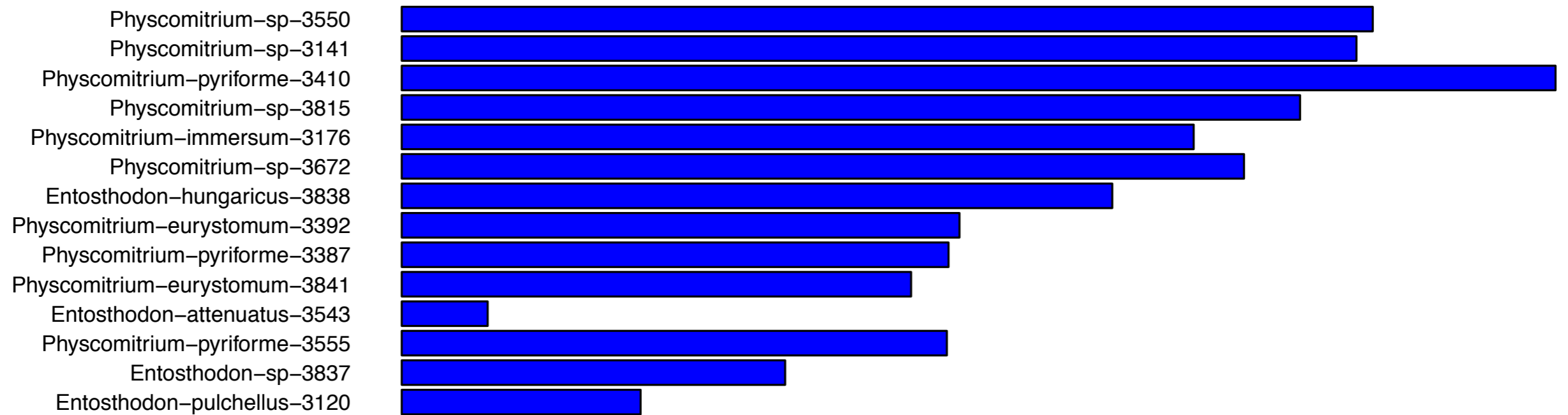


# CONFIRMING HETEROZYGOSITY



## Future Directions

Expand analysis to cover all 15 samples with high heterozygosity





# ACKNOWLEDGEMENTS



## Moss Tree of Life

Nico Devos  
Yang Liu  
Jon Shaw  
Norm Wickett

## Allele Phasing

Elliot Gardner  
Heather Rose Kates  
Nyree Zerega

*Kates et al., AJB, 2018 10.1002/ajb2.1068*

## Funding



## Join Us!

The Goffinet (UConn) and Johnson (Texas Tech)  
Labs are recruiting graduate students!

More info:

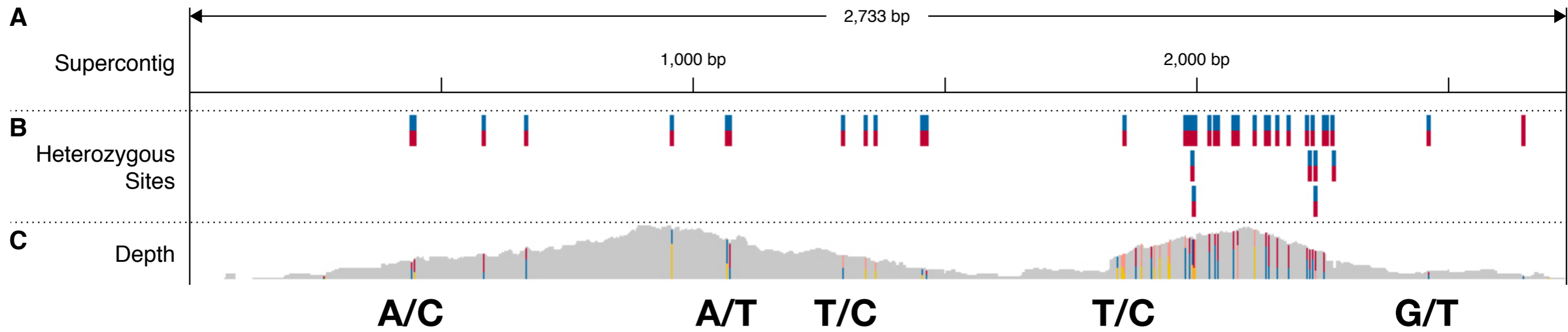
[bryology.uconn.edu](http://bryology.uconn.edu)  
[mossmatters.com](http://mossmatters.com)



# ALLELE PHASING



## Read-backed Phasing using GATK and WhatsHap



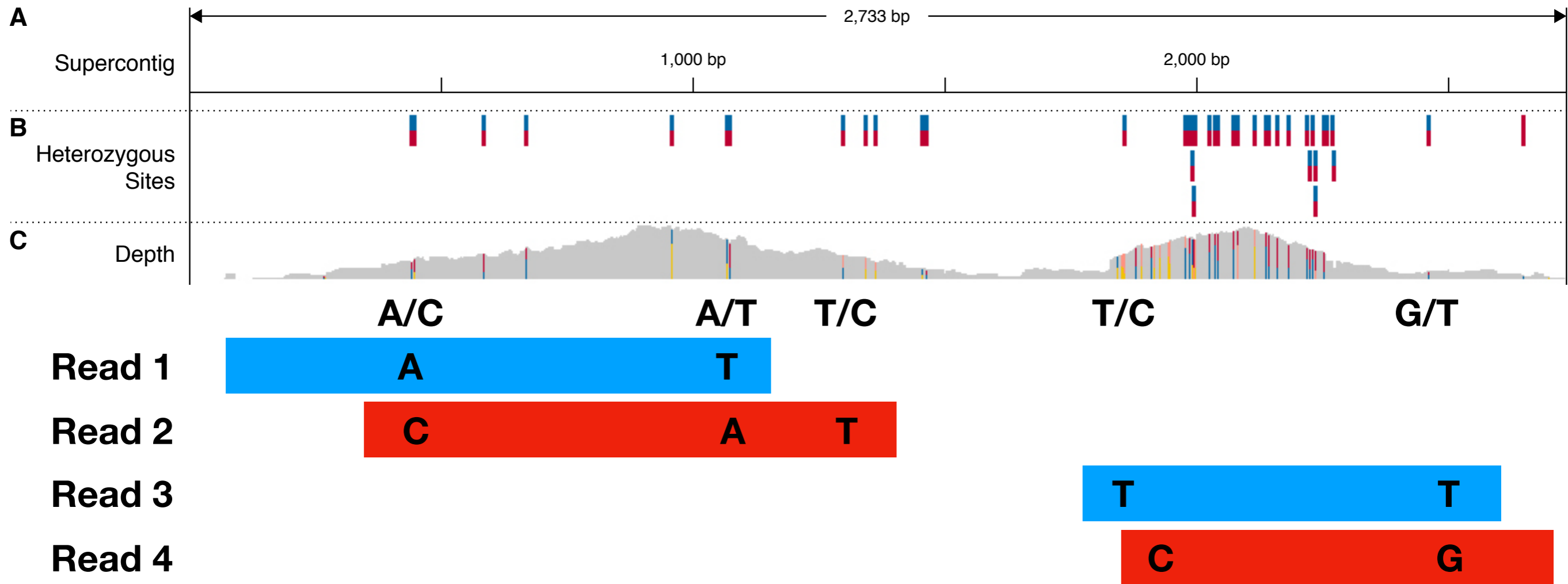
**How to build two sequences?**



# ALLELE PHASING

## How to build two sequences?

*Kates, Johnson et al., 2018, AJB*



**If read depth is low, may not be able to connect all SNPs**

**For example: Introns**



# ALLELE PHASING



## Read-backed Phasing using GATK and WhatsHap

*Kates, Johnson et al., 2018, AJB*

<b>E</b>	<i>Consensus</i>	AAGAACCTTGTCACCGGGGACATGTCAGAGCTGAAGGTGT..... TGAGTCCCTTATATTCAACAGATTAGTTTGACA
	<i>Ambiguity</i>	AAGAACSTTGTCAC <b>Y</b> GGGG <b>MRTKTCR</b> GAGCTGAAGGTGT..... TGAGTC <b>Y</b> CTT <b>AY</b> ATTCAA <b>Y</b> AGAK <b>Y</b> AGTTTGACA
	<i>Phased Allele 1</i>	AAGAACGTTGTCAC <b>C</b> GGGG <b>CA</b> T <b>G</b> T <b>C</b> AGAGCTGAAGGTGT..... TGAGTC <b>Y</b> CTT <b>AY</b> ATTCAA <b>Y</b> AGAK <b>Y</b> AGTTTGACA
	<i>Phased Allele 2</i>	AAGAACCTTGTCAC <b>T</b> GGGG <b>AGT</b> <b>TTC</b> GAGCTGAAGGTGT..... TGAGTC <b>Y</b> CTT <b>AY</b> ATTCAA <b>Y</b> AGAK <b>Y</b> AGTTTGACA

Options if supercontig cannot be fully phased:

1. Split into two genes
2. Retain consensus sequence
3. Retain ambiguity characters
4. Delete sequence outside largest phase block

Code available: [github.com/mossmatters/phyloscripts](https://github.com/mossmatters/phyloscripts)