A Next-generation Sequencing Approach to DNA Barcoding in Plants

& related studies of Hawaiian *Metrosideros*

Elizabeth Stacy[1], Brad Barbazuk[2], Jennifer Johansen[1], Yohan Pillon[1], Tomoko Sakishima[1,] Srikar Chamala[2]

[1]Tropical Conservation Biology and Environmental Science Program, University of Hawai`i at Hilo

[2]Department of Biology, University of Florida

Two Problems for Plant Conservationists:

1) What is a species?

2) Is this individual a hybrid?

**Hawai`i

Native Hawaiian Angiosperm Flora

- 1,030 spp. from ~270 colonizations
  - diverse sources

- $1/10^{th}$ of colonists → 1/2 of species
  - speciation >> immigration
    - isolation & heterogeneous habitats?

- Disharmonic

<u>High endemism</u>:

- 90% of 1,030 spp.
- 15% of 216 genera
- 0% of 87 families

<u>13 Most Spp-rich Families</u>

Campanulaceae	126
Asteraceae	92
Gesneriaceae	59
Rutaceae	55
Lamiaceae	54
Rubiaceae	54
Poaceae	47
Cyperaceae	45
Caryophyllaceae	34
Piperaceae	25
Malvaceae	24
Euphorbiaceae	21
Fabaceae	20

Wagner et al. 2005

Speciation of Hawaiian angiosperms ongoing

- Hybridization is common, especially on young islands

Angiosperm flora is threatened

- Historical records of 956 spp. . . .

Phyllostegia warshaurei
(n=4 wild)

Hawaiian mints:
31 spp, 18 endan'd, 3 extinct

- 107 (11 %) extinct
- 316 (33 %) threatened

→ 423 (***44 %***) extinct or threatened

→ HI's 4 counties are in US' top-5 counties for T&E plants & animals

Hibiscadelphus
6 spp; 4 endan'd
+ 2 extinct

Scheidea diffusa
(n=1 wild)

Our approach: DNA Barcoding using next-generation sequencing of pooled partial transcriptomes

WHY? → Traditional *DNA barcoding* problematic for plants

→ Ideal "Barcoding Genes"

→ easily & uniquely amplified, sequenced (single-pass)

→ species-diagnostic (appropriate variation)

→ cp genes: (rbcL), *matK, rpoB, rpoC1, ndhJ, ycf5 & accD* [Ford et al. 2009]

→ cp intergenic regions (e.g., *psbA-trnH)*: ↑ difficulty of amplification

→ poor-to-no resolution of closely related spp.

Specifically: non-normalized 454 FLX Titanium sequencing of pooled partial transcriptomes

Objectives:

a. mine for species/subspecies-diagnostic variation:

→ SNPs

→ population-level SNP genotyping & alignment w/ morphology-based species delineation

→ others (SSRs, indels)?

b. mine for candidate genes for multi-gene phylogenetic analysis

- understanding evolutionary history improves evaluation of barcoding

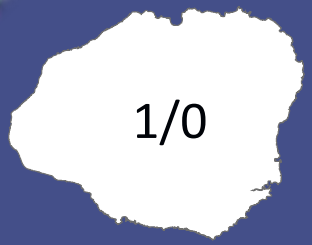
Focal Taxa: Genus *Clermontia*

- Endemic; ~5 myo (Givnish et al. 2009)
- Member of largest HI radiation
- 30 taxa (9 endangered; 1 extinct)
- Varies in floral characters
  - cospeciation with birds?
  - pollinator loss threatens 1 sp.
  - species ID requires flowers
  - double corolla **
- Center of spp. richness → young islands
  - bird-dispersed

Genus *Clermontia*

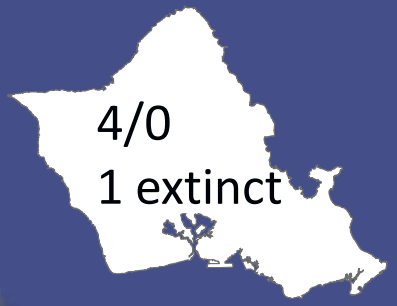
Kaua‘i & O‘ahu

C. fauriei



C. kakeana

C. oblongifolia subsp. oblongifolia



C. persicifolia

extant spp / # endangered (federal or state lists)

Genus *Clermontia*

Maui Nui & Hawai‘i Island

5/1

4/1

13/3
2 extirp; 1 extinct

C. tuberculata

C. grandiflora subsp. grandiflora

C. arborescens subsp. waihiae

C. waimeae

13/5

- 1 rediscovered

C. micrantha

C. montis-loa

Hybrids common – especially on Hawai'i Island

Maui: few hybrids in ecotones

C. grandiflora
subsp. grandiflora x micrantha ?

C. arborescens subs. waihiae x kakeana ?

Hawai'i Is.: species' boundaries fuzzy; numerous hybrids

C. calophylla
x waimeae ?

C. montis-loa x parviflora X ??

Clermontia habit: tree, treelet, (woody) epiphyte

Pat Bily (TNC) & C. arborescens subsp. arborescens

Large *C. waimea* in `Ohi`a (Kohala)

C. clermontioides subsp. clermontioides (S. Kona)

Focal taxa: Genus *Cyrtandra*

- Family Gesneraceae (African Violet)

- Genus → > 600 spp.; Pacific region
  - most spp. restricted-region endemics
  - HI: ~ 64 taxa; monophyletic
    - center of richness → O'ahu
    - 22 endangered; 2 extinct

- ~ <u>All</u> *Cyrtandra* spp. in wet forest understorey
  - vary in vegetative /calyx characters
  - min. ecological diversification
    - → *[nonadaptive radiation?]*

- Rampant hybridization in HI (except Kaua'i)
  - taxonomic nightmare (St. Johns ~600 spp.)
  - conservation status of variable multi-island spp?

K. Magnacca

- Insect-pollinated
- Bird-dispersed
- Threatened by ungulates & habitat loss

Hawaiian *Cyrtandra*

16/5

C. confertiflora var. obovata

25/9

C. platyphylla

10/3

3/2

14/4

extant spp / # endangered (federal or state lists)

Cyrtandra. sp?

10/4
+ 3 new spp.?

Methods:

- Sampling, mapping & morphological measurements:

  - 31 of 48 *Clermontia* populations (65 % complete)

  - 49 of 72 *Cyrtandra* populations (68 % complete)

  [n=10 adults/population]; no T&E spp. yet

- DNA extractions (Qiagen); for SNP genotyping

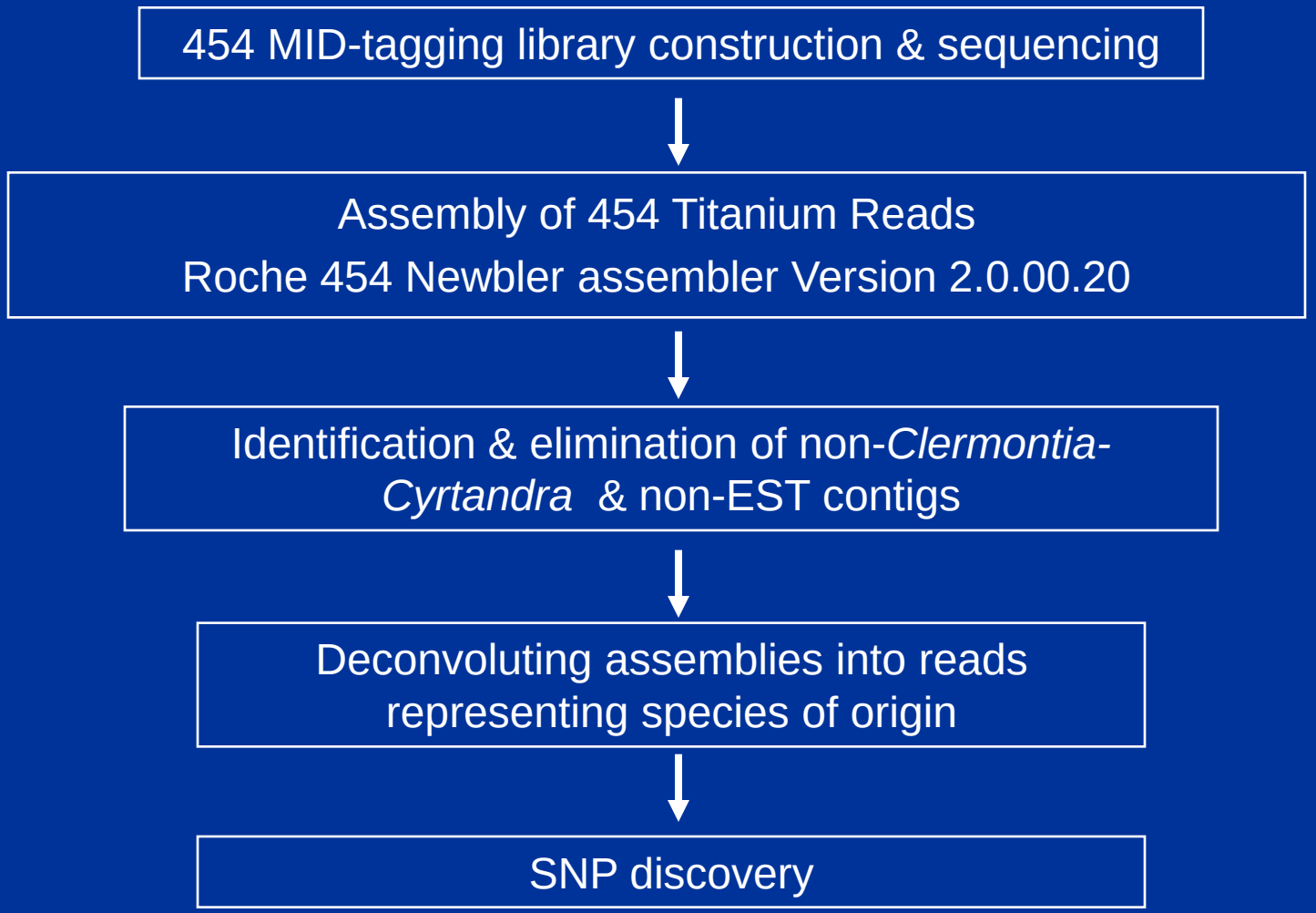
- Prep for 454:

  - → U Arizona Genetics Core : Ambion's Plant RNA Isolation Aid 8 (7 + 1) taxa individually
    - *pooled vegetative & floral buds*

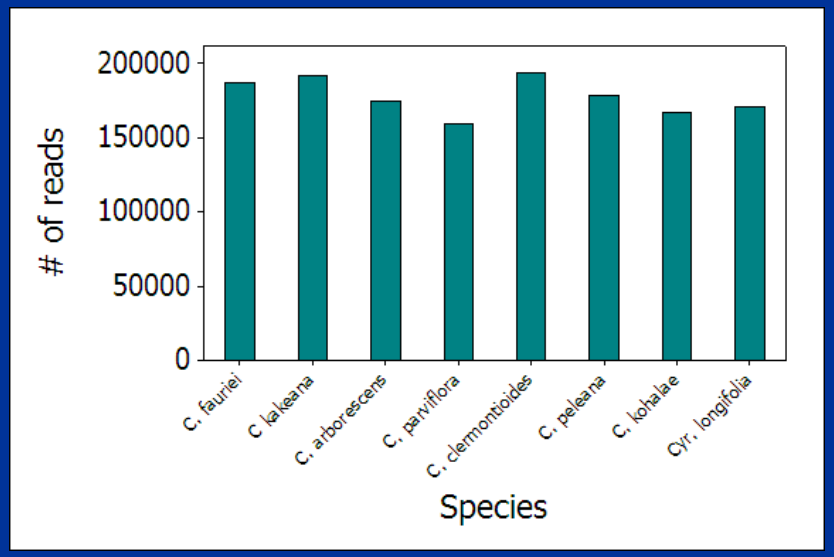
  → Roche's cDNA Rapid Library Preparation protocol

Methods & Results: Bioinformatics I → SNPs

1 week of analysis!

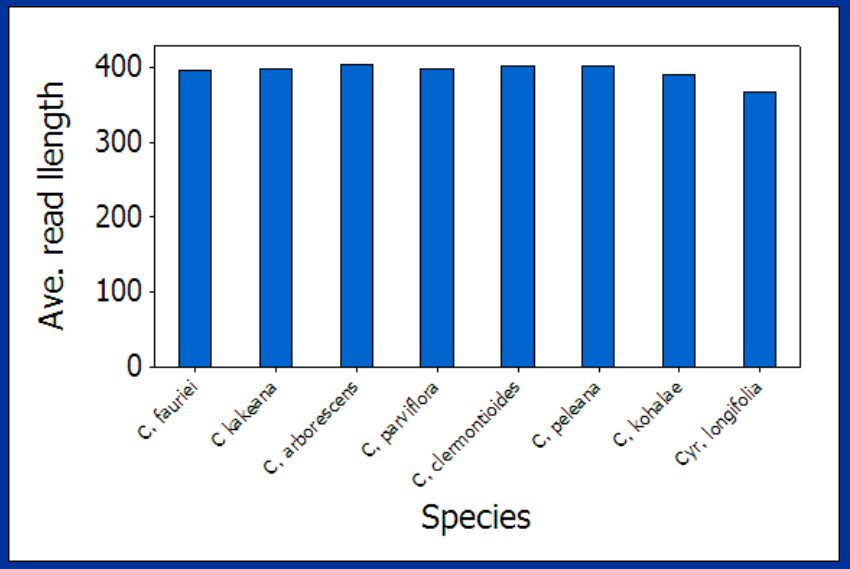


Methods & Results: Bioinformatics I → SNPs



Totals: ~1.4 million reads

~560 MB



Ave. read length = 395 bases

Deconvoluting assemblies into reads representing tissues of origin

Species-specific contigs

Contig Name	Total Reads			rborescens	Parviflora	Clermontioid	Peleana	Kohalae	Longifolia	# of species
contig00025	9			0	0	0	0	0	9	1
contig00026	135	0	0	0	0	0	0	0	135	1
contig00027	145	0	0	0	0	0	0	0	145	1
contig00028	22	0	0	0	0	0	0	0	22	1
contig00029	4	0	0	1	2	0	1	0	0	3
contig00030	67	0	0	0	0	0	0	0	67	1
contig00031	229	31	15	45	38	15	7	34	44	8
contig00032	47	0	0	0	0	0	0	0	47	1

Constitutive Expression

An approximate expression profile was made for each contig by specifying the set of MID pools that contributed to the contig.

Species-specificity: 340 contigs were enriched for a single species.

Constitutive expression: 639 contigs were expressed in all spp. at ~same level.

Lone *Cyrtandra* sp. has many more species-specific contigs:

Species	# contigs (min. 10 reads
C. fauriei	18
C. kakeana	25
C. arborescens	5
C. parviflora	9
C. clermontioides	8
C. peleana	2
C. kohalae	13
Cyrtandra longifolia	457

- Greater genetic distance caused Newbler to treat homologous alleles as from separate loci in the 2 genera?

- Bud gene expression differs between genera?

Next steps:

• Separate *Cyrtandra* reads for separate analysis.

• Assemble with transcriptome assemblers (TGICL , MIRA) with liberal alignment criteria because of the cross-species nature of the reads.

SNP discovery:

- Mosaik Aligner : uses each contig as ref. sequence for alignment of reads
- GigaBayes

Methods & Results: Bioinformatics II → Candidate genes

3 categories of candidate genes:

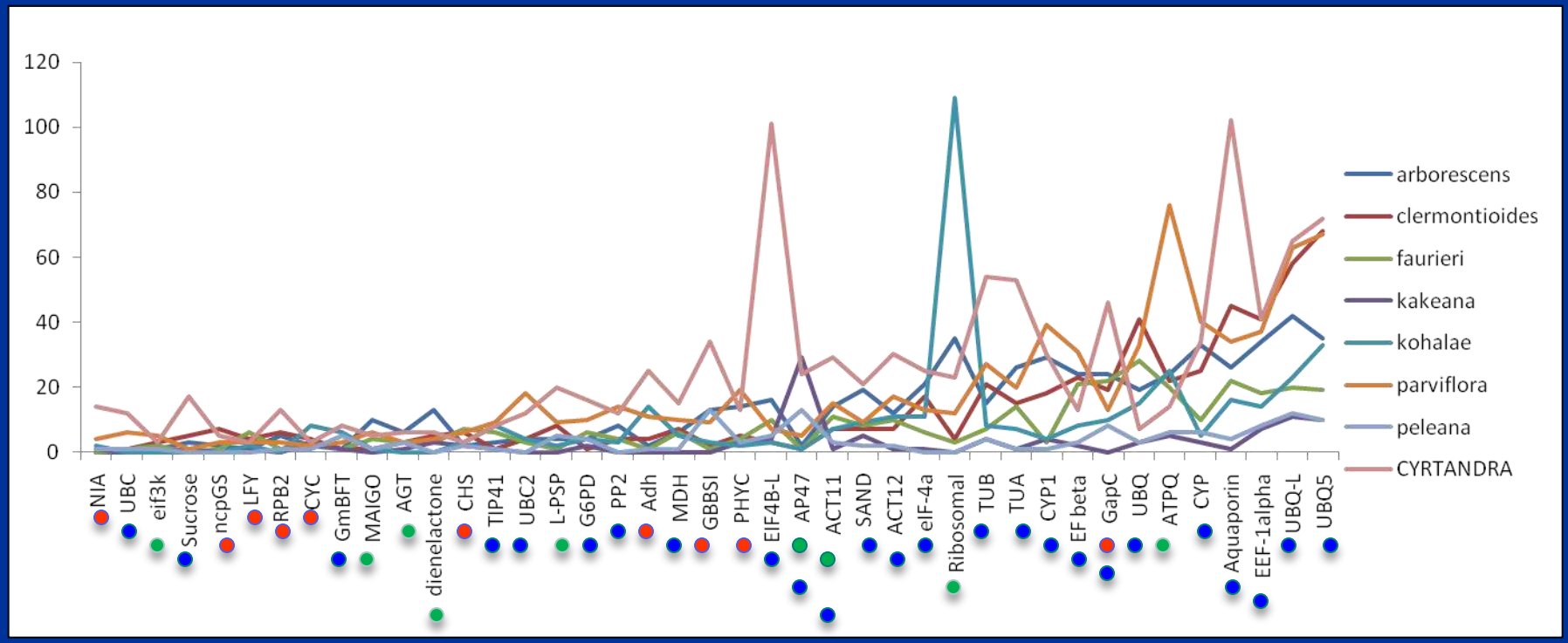
1) Low-copy number nuclear genes previously used in phylogenetic analyses

2) Genes ID'd as single copy in *Arabidopsis*, *Populus*, *Vitis* & *Oryza* (Duarte et al. 2010)

3) Genes used as controls for quantitative gene expression

→ (42 genes in total)

BLAST hits of EST library against reference sequences

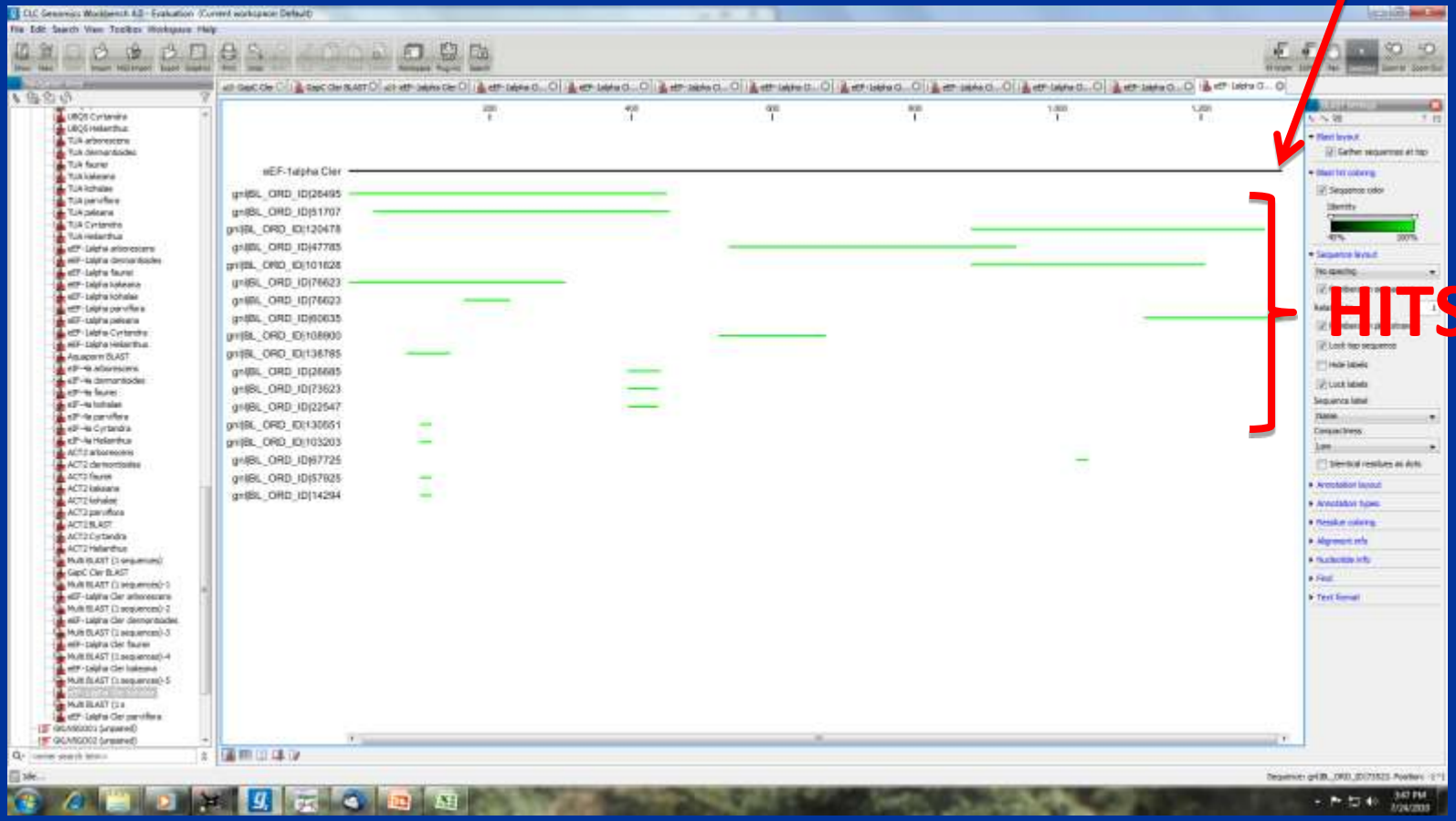


- Genes used in phylogenetic studies
- Single-copy gene
- Expression controls

- Expression levels vary across genes.
- Expression levels ~~even across taxa.
- Phylogeny genes poorly expressed.
- Expression controls best.

BLAST searches against EST libraries

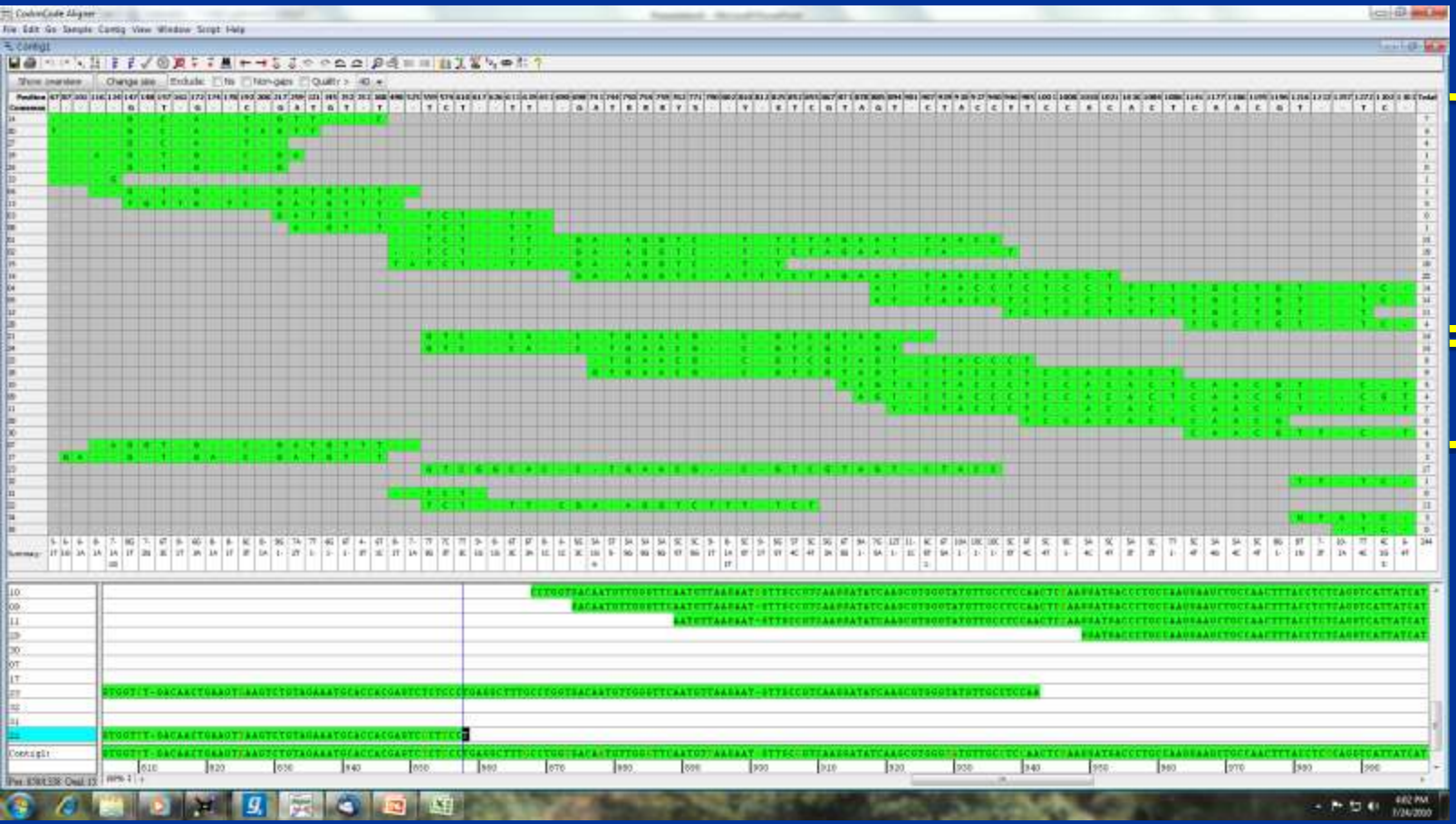
Reference sequence (e.g., *Arabidopsis, Populus*)



CLC Genomics workbench 4.0

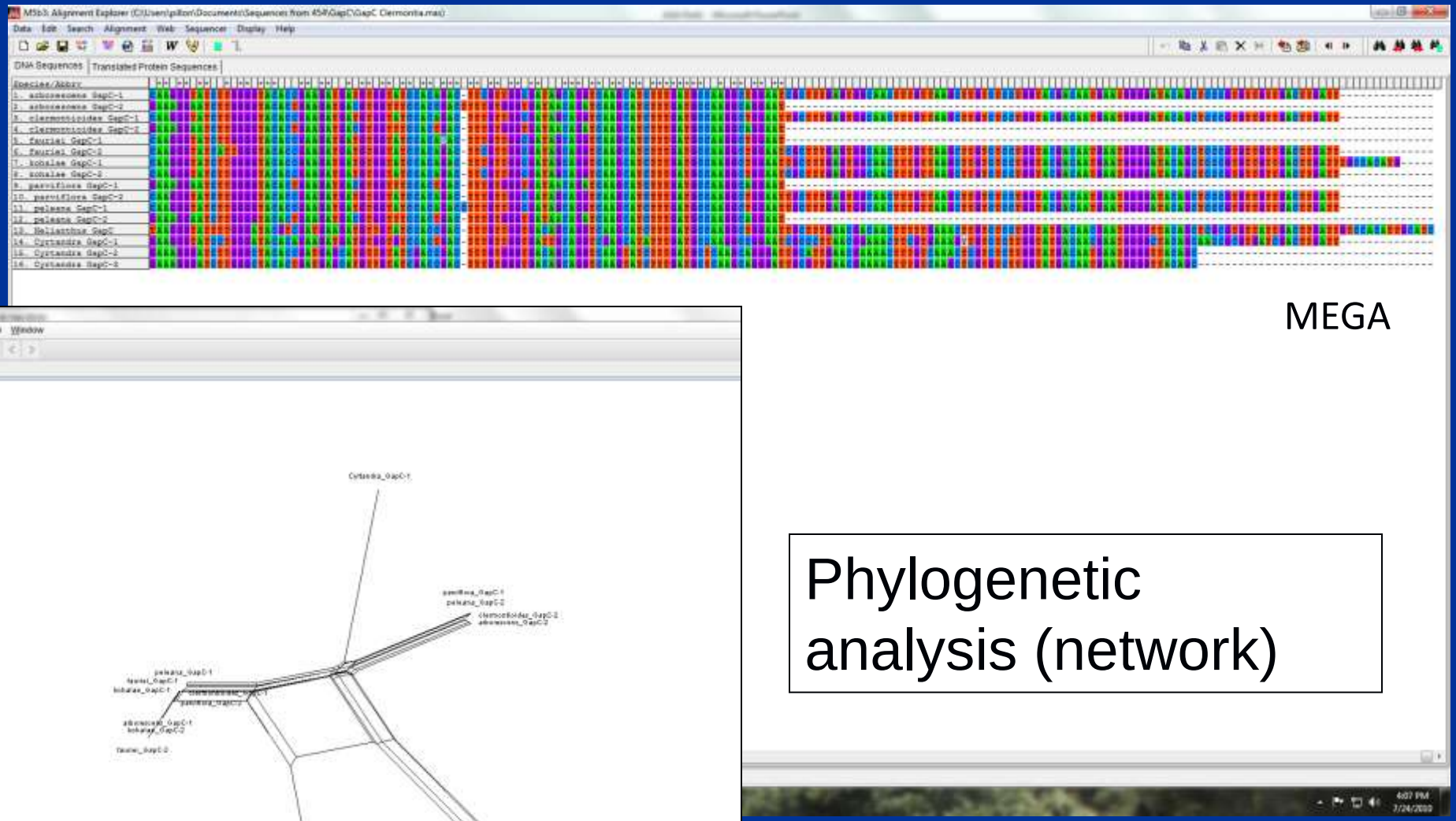
Building consensus and contigs

Alleles sorted by hand



CodonCode Aligner

Alignment across species



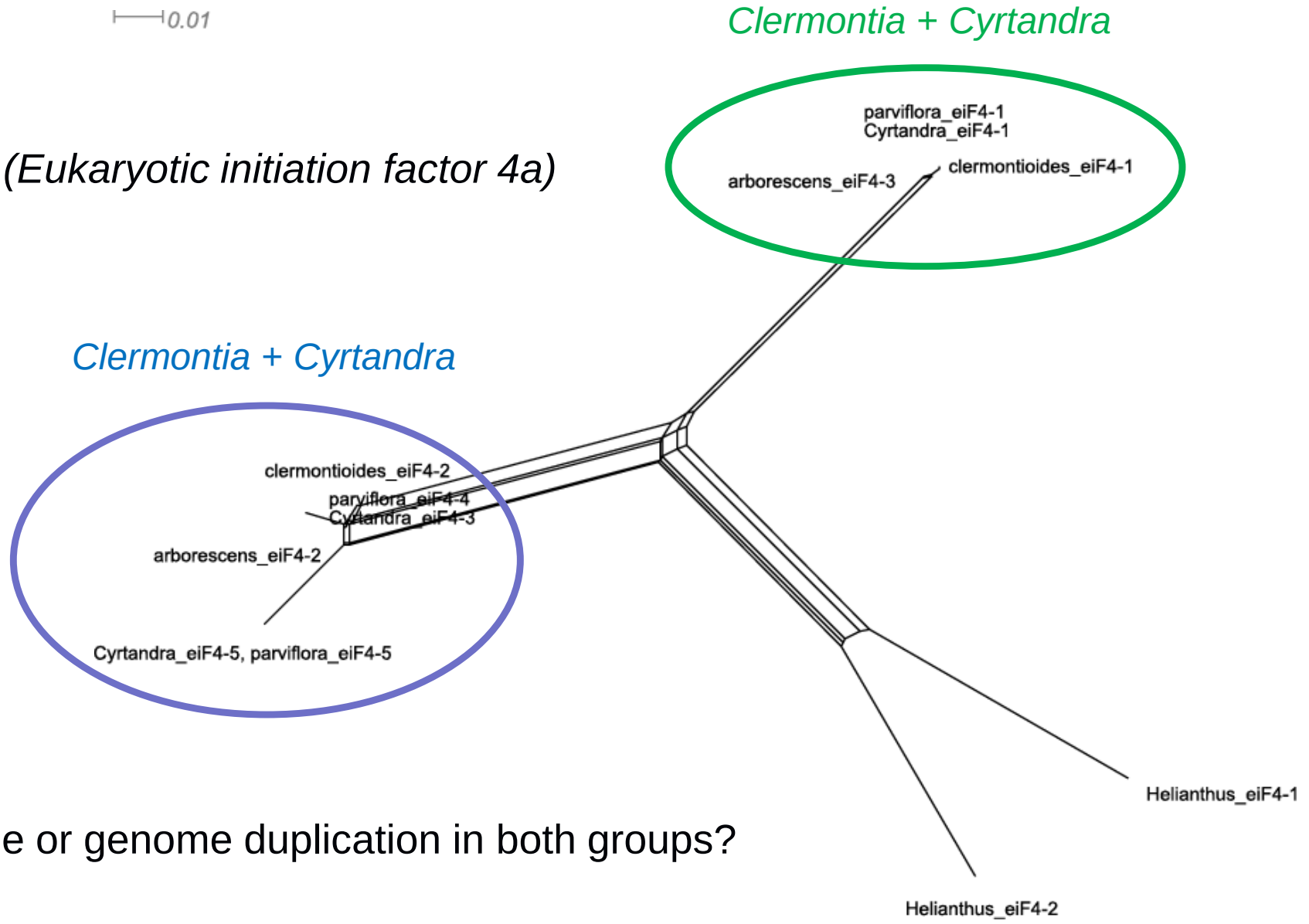
MEGA

Phylogenetic analysis (network)

Splitstree

Constructed networks for each of 10 genes: example…

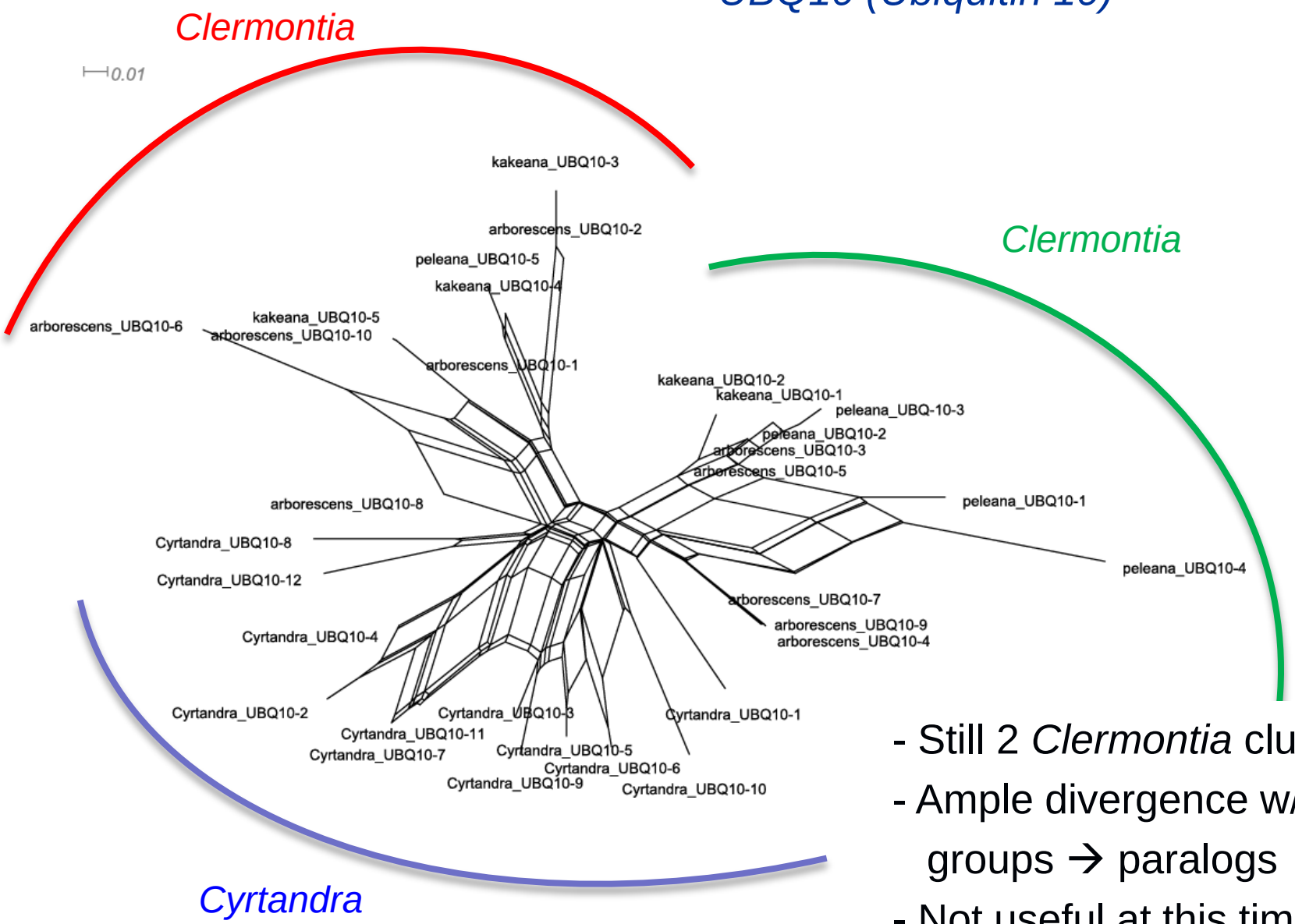
eIF4a (Eukaryotic initiation factor 4a)



- Gene or genome duplication in both groups?

another…

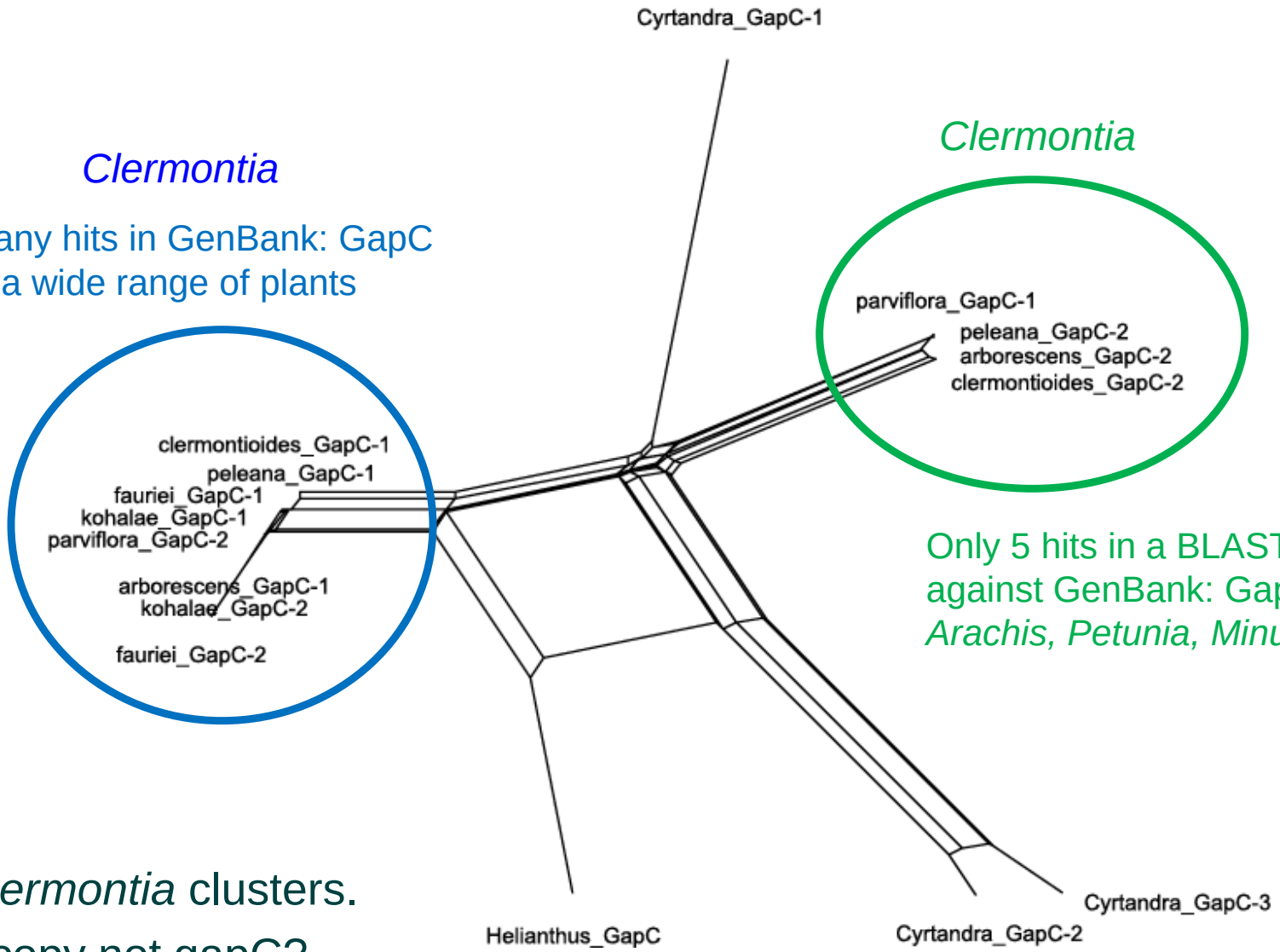
UBQ10 (Ubiquitin 10)



- Still 2 *Clermontia* clusters
- Ample divergence w/in groups → paralogs
- Not useful at this time

last one…

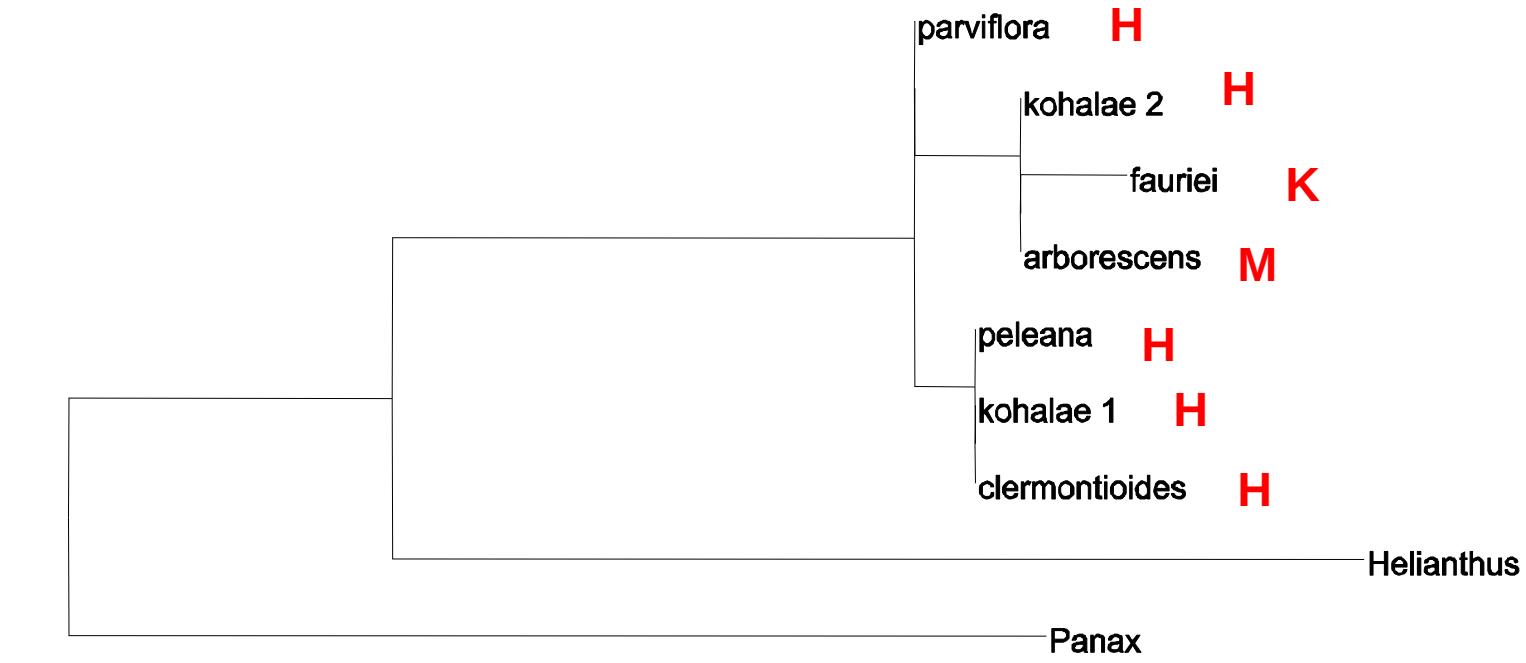
GapC (Glyceraldehyde-3-phosphate dehydrogenase)



- 2 *Clermontia* clusters.
- 2nd copy not gapC?
- Genome duplication not certain.

Phylogenetic analysis of the “true” *GapC gene*

Maximum parsimony [227-base fragment]



Other candidate genes?

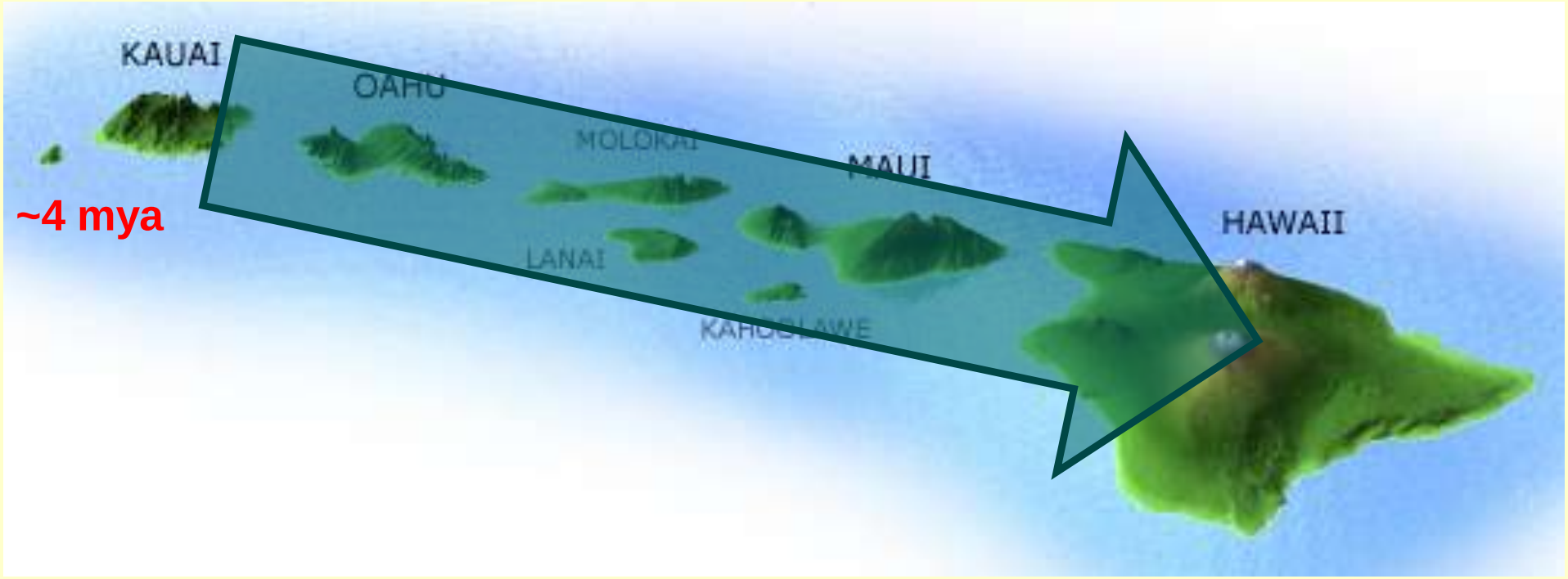



Reproductive isolation & genetic divergence within Hawaii's dominant tree, *Metrosideros polymorpha*






Hawaiian *Metrosideros* – an emerging model for speciation studies



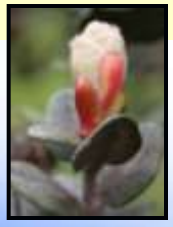
Percy D M et al. 2008. Proc. R. Soc. B

- 10 cpDNA loci

Hawaiian *Metrosideros* – an emerging model for speciation studies





- 5 species ○

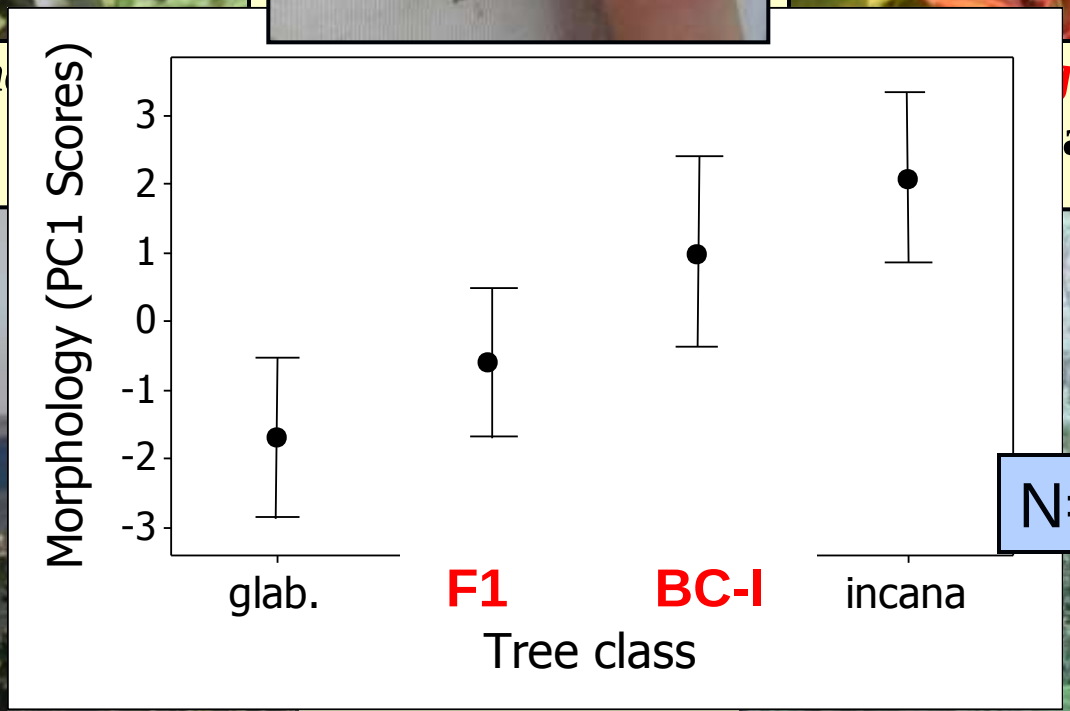
- *M. polymorpha* △
  - highly abundant species
  - on all main islands
  - 8 varieties (vegetative characters)

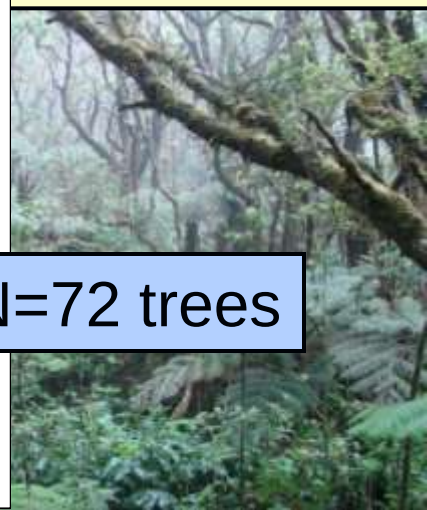
- Degree of natural hybridization varies

Reproductive isolation (RI) between two successional varieties

M. p. var. incan **(pubescent)**

laberrima
abrous)





N=72 trees

RI: & ♂ Fitness:

2-way ANOVA →

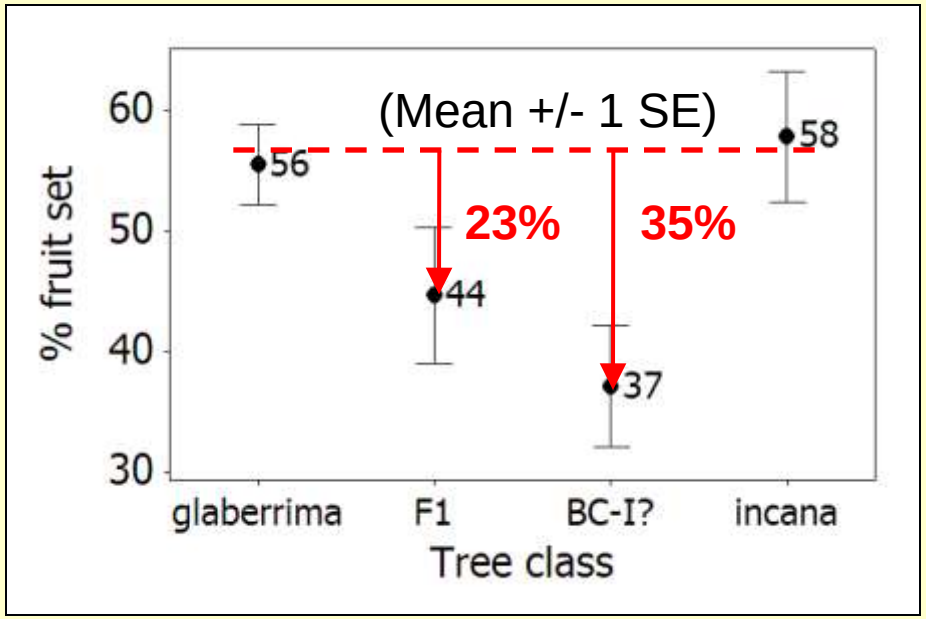
Sour					p
					036
♂					882
* ♂	9	0.3905	0.043	0.57	0.822
Error	124	9.482	0.076		
Total	139	10.773			

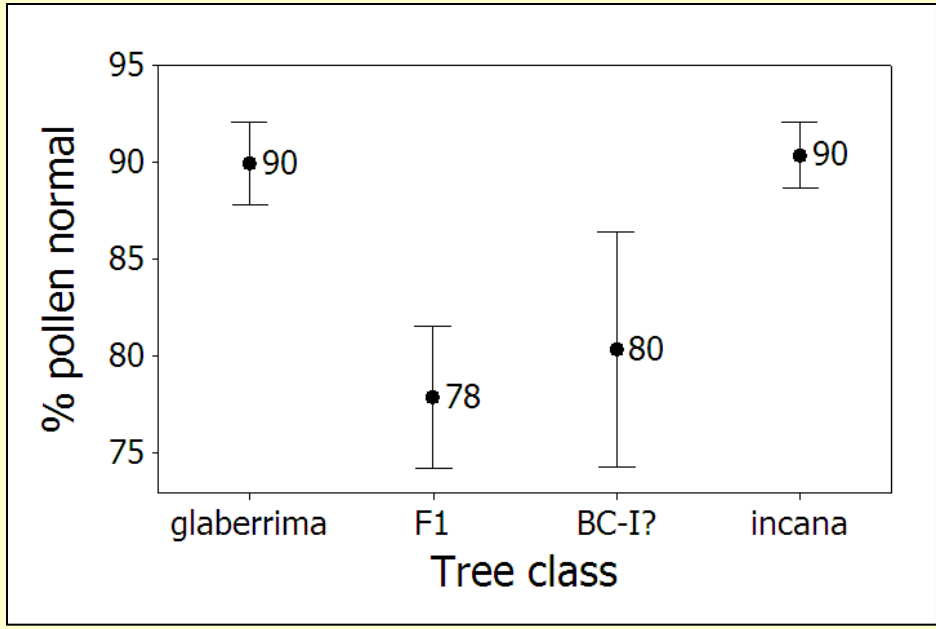
1-way ANOVA → % FS = :
F=3.91; df=3,136; **p=0.010**)

Kruskal-Wallis:
(H=10.38, df=3,
p=0.016)

Cotton blue
(300 grains)

♂





Conclusions of RI Study

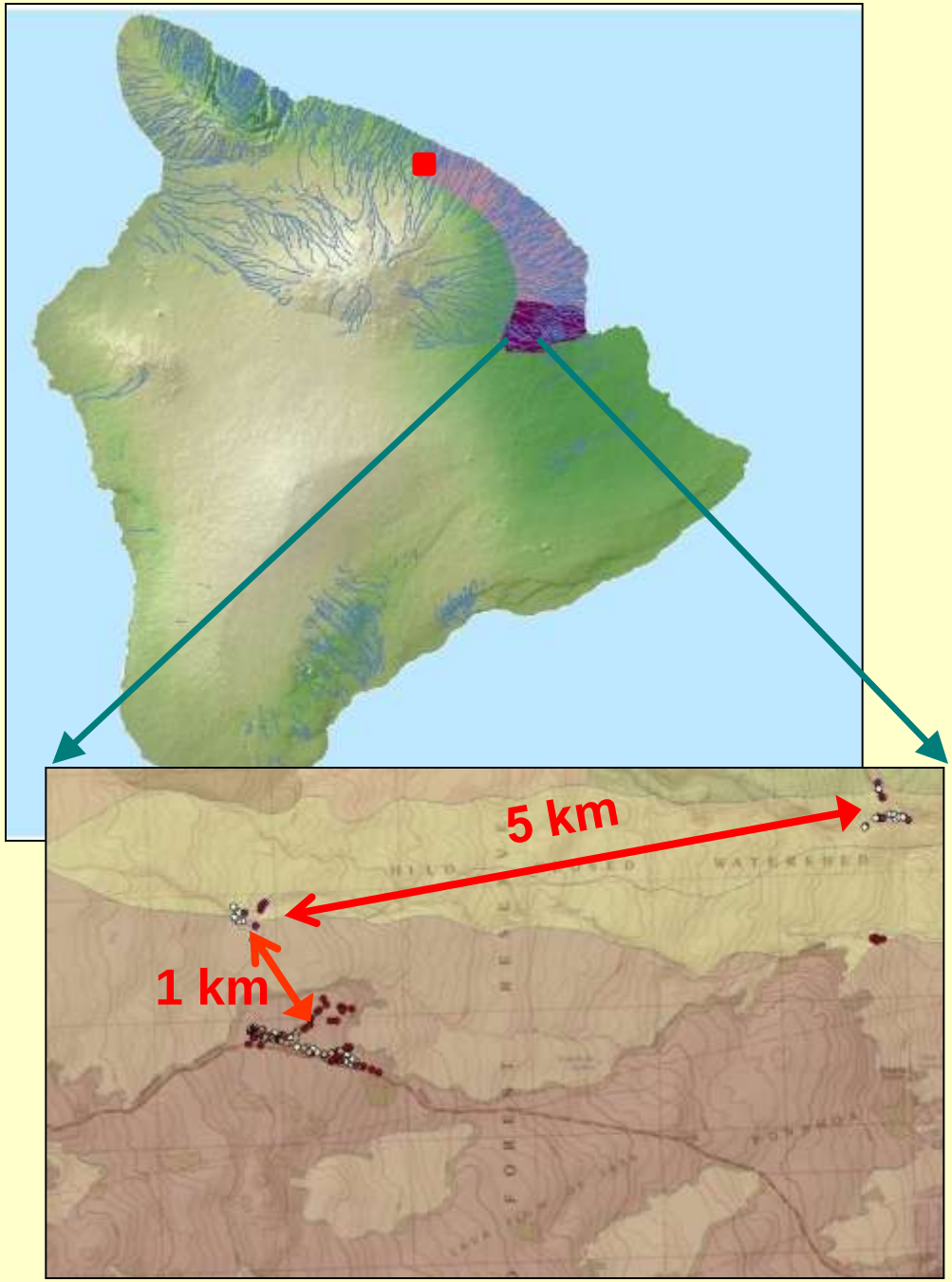
- Partial, late-acting RI between successional varieties on east HI Island fits a model of negative epistasis
  - 1st report of epistasis causing partial RI *within* a tree species
  - Beginning to characterize early stages of RI in trees.

M. p. var. polymorpha

Monotypic at high elevations in HI

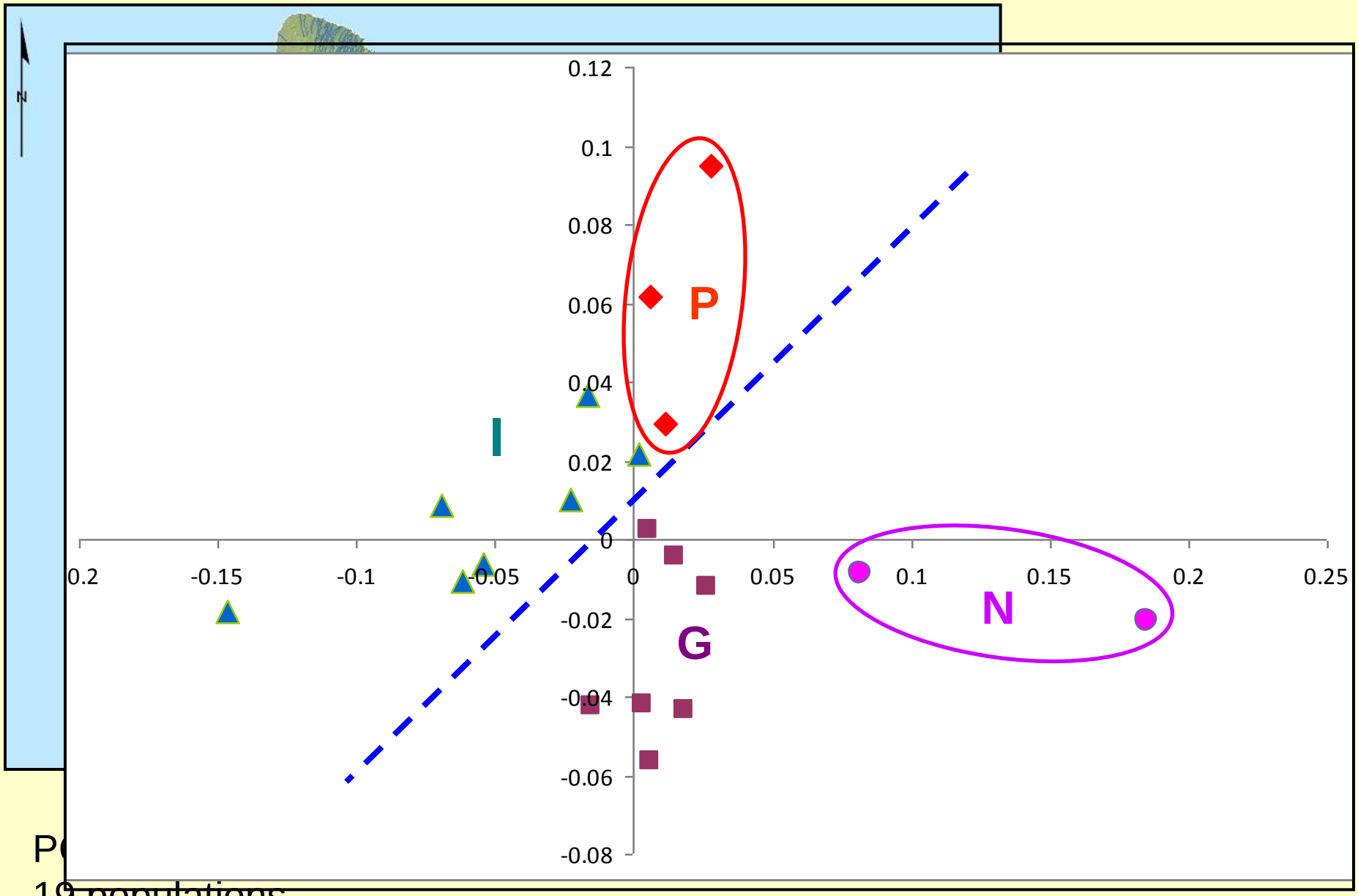
M. polymorpha var. newellii

Newellii has a Highly Restricted Habitat & Geographic Range



- Range of historical collections (Bishop Museum) = 11,000 ha
- Potential geographic range = 46,000 ha
- 2 populations
  - Separate rivers/streams

Extreme-habitat varieties on the edge…



P
19 populations

Var. newellii ~originated on Hawai`i Island

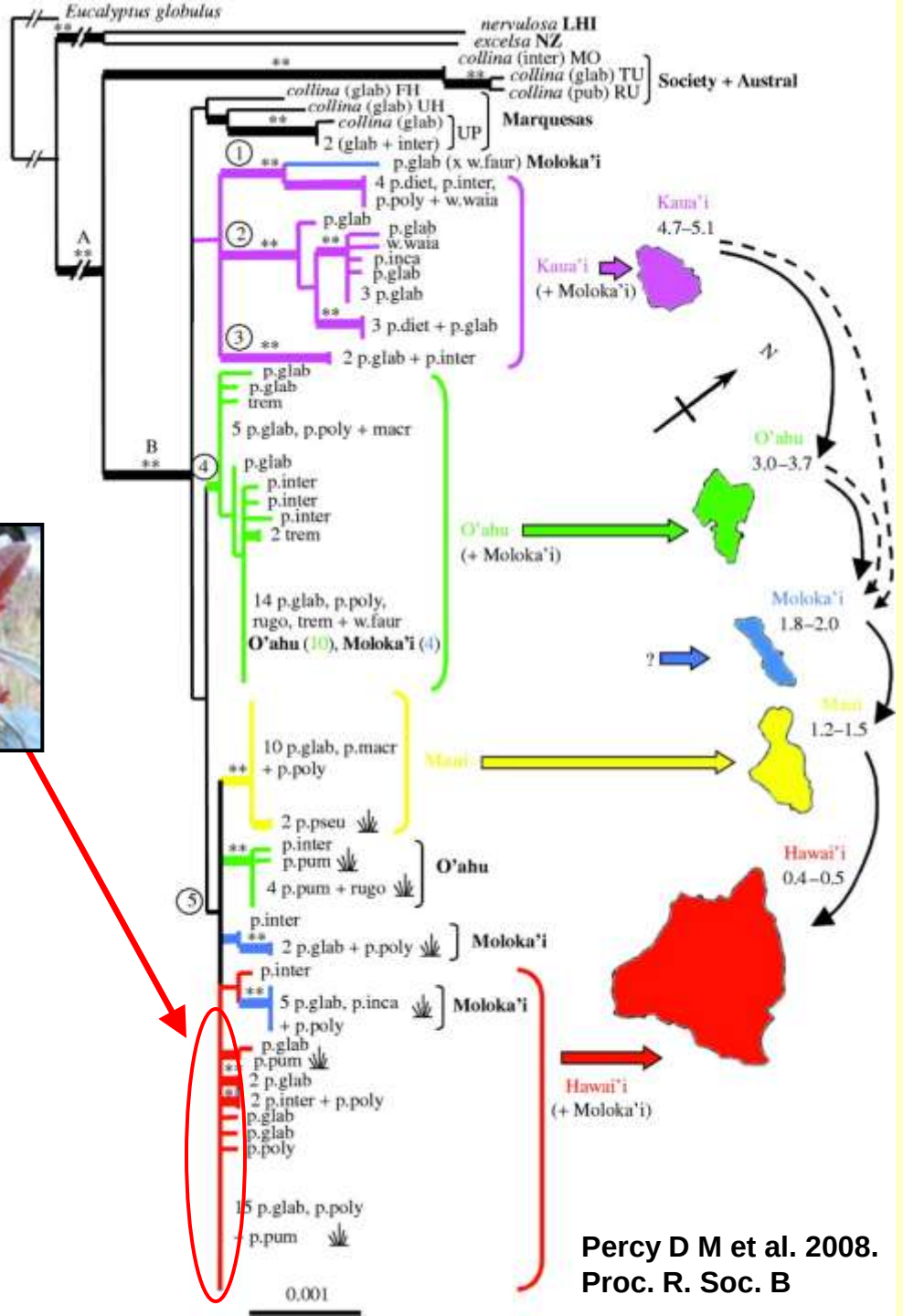
Based on 4 cp regions:

- rpl16
- rps16
- trnS-trnG
- trnD-trnE

→ F_{ST} = 0.045-**0.33** (mean = **0.14**) within a long-lived, ubiquitous, wind-dispersed, woody sp.

→ ≤ 0.5 MY (J. Price)

→ ≤ 200 mi^2



Percy D M et al. 2008. Proc. R. Soc. B

Conclusions – *M. polymorpha*

- Hybrid zone:

  → Starting to characterize the genetic basis of early RI in trees.

  - Morphology & fitness variation indicate 2 hybrid classes.
  - SSRs consistent with hybridization.

- SSR studies:

  → Strongest population differentiation is associated with *extreme-habitat varieties*

  → *var. newellii:* *restricted geographic area* (& small N) important?

  → Conservation status of *var. newellii*?

Acknowledgments

- Nick DeBoer — DLNR-DOFAW
- Heather Issar — NARS – Betsy Gagne
- Melissa Johnson — Maui Land & Pineapple
- Ala Malia Leka — Nature Conservancy
- Karl Magnacca
- Keenan Morrison
- Don Price
- Eric Roalson — NSF RIG, CREST, CAREER
- Steve Rounsley — Gordon and Betty Moore Foundation
- Anne Veillet
- Court Warr