

Detecting Hybridization in Hawaiian *Cyrtandra* using Genome-wide Data



Joseph A. Kleinkopf and Eric H. Roalson

School of Biological Sciences, Washington State University, Pullman, Washington, 99164-4236, USA: eric_roalson@wsu.edu

Introduction

Oceanic island systems provide ideal settings for studying the processes of evolution including community assembly and speciation due to their small sizes, young ages, and isolation. Hawaii in particular is an excellent site for studying evolution, as it has a small area, is young (<5myo), and is nearly 4000km from the nearest landmass. *Cyrtandra* (Gesneriaceae) is a genus of flowering plants with over 800 species distributed throughout Southeast Asia and the Pacific Islands. In Hawaii, 60 named species and over 62 putative hybrids exist, most of which are identified on the basis of morphology. Despite many previous studies on this lineage (Clark et al. 2008, 2009; Johnson et al. 2017), questions regarding the reconciliation of morphology and genetics remain, many of which can be attributed to the relatively young age and evidence of hybridization between species. We tested 33 *Cyrtandra* samples for species relationships using 569 loci in both concatenated and coalescent methods, and tested for hybridization using network analyses. We expected to see both a classic stepping-stone model, where species relationships group by island from oldest to youngest, and species relationships grouping by morphology. Additionally, we expected to see high levels of ILS and putative hybrids intermediate to their parent species. Concatenated and coalescent methods yielded different hypotheses of species relationships, most likely due to high levels of ILS. Species-tree methods account for ILS and are probably more accurate for our system. These trees infer progression of species from Kauai to Hawaii with instances of back-dispersal. Network analyses indicate hybridization, though not always between taxa we expected. Our study shows the utility of using genome-wide data to understand species relationships in recent plant radiations.

Methods

Identify low-copy exon regions for targeted-enrichment using the Hyb-Seq protocol (Weitemier et al. 2014)

Library-preparation using biotinylated baits to capture sequences of interest

Sequence using MiSeq 250bp PE reads

Map reads to probeset, keeping intronic regions (flanking exons) and filtering sites with <50% occupancy

Conduct concatenated and species-tree analyses in RAxML (Stamatakis 2014), ASTRAL (Mirarab et al. 2014), and ASTRID (Vachaspati and Warnow 2015)

Conduct network analyses using gene trees and species tree from ASTRAL

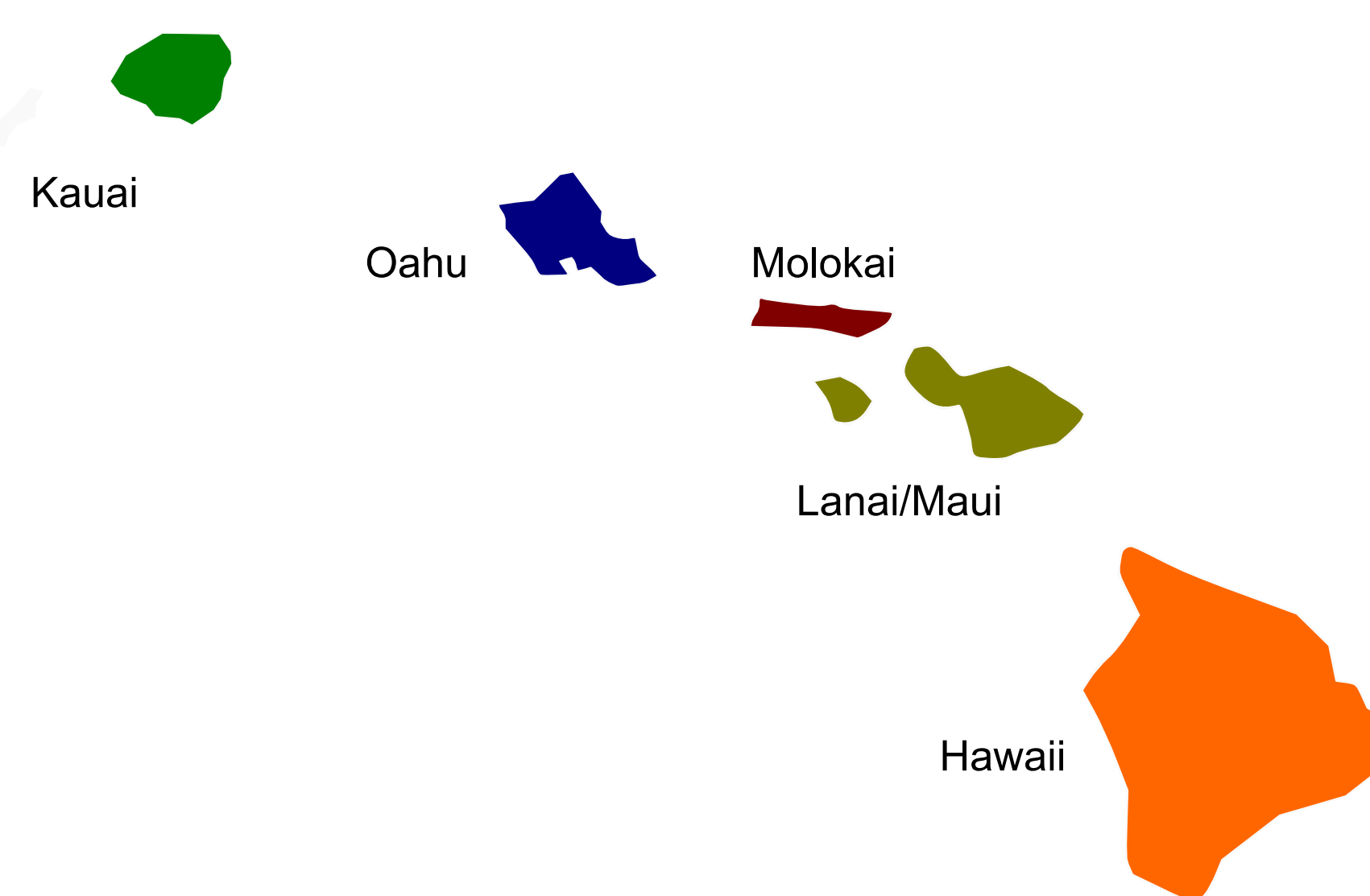


Figure 1. Map of Hawaii. Kauai is the oldest island (~5.1 My) and Hawaii is the youngest (~0.49 My).

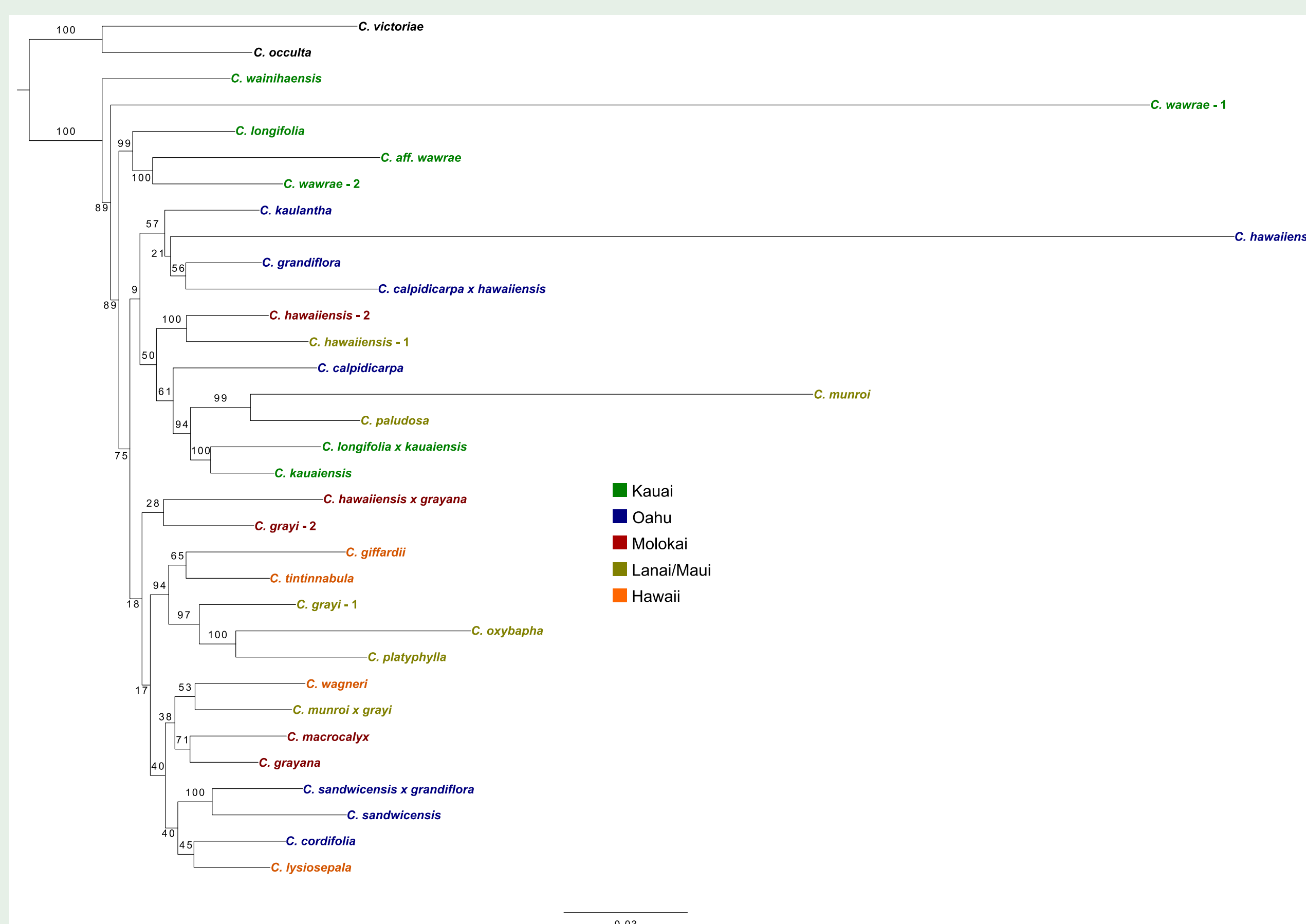


Figure 2. Maximum likelihood tree inferred in RAxML from the concatenated dataset with all taxa included. Bootstrap values are plotted on branches. Scale represents branch lengths in per-site substitution rate.

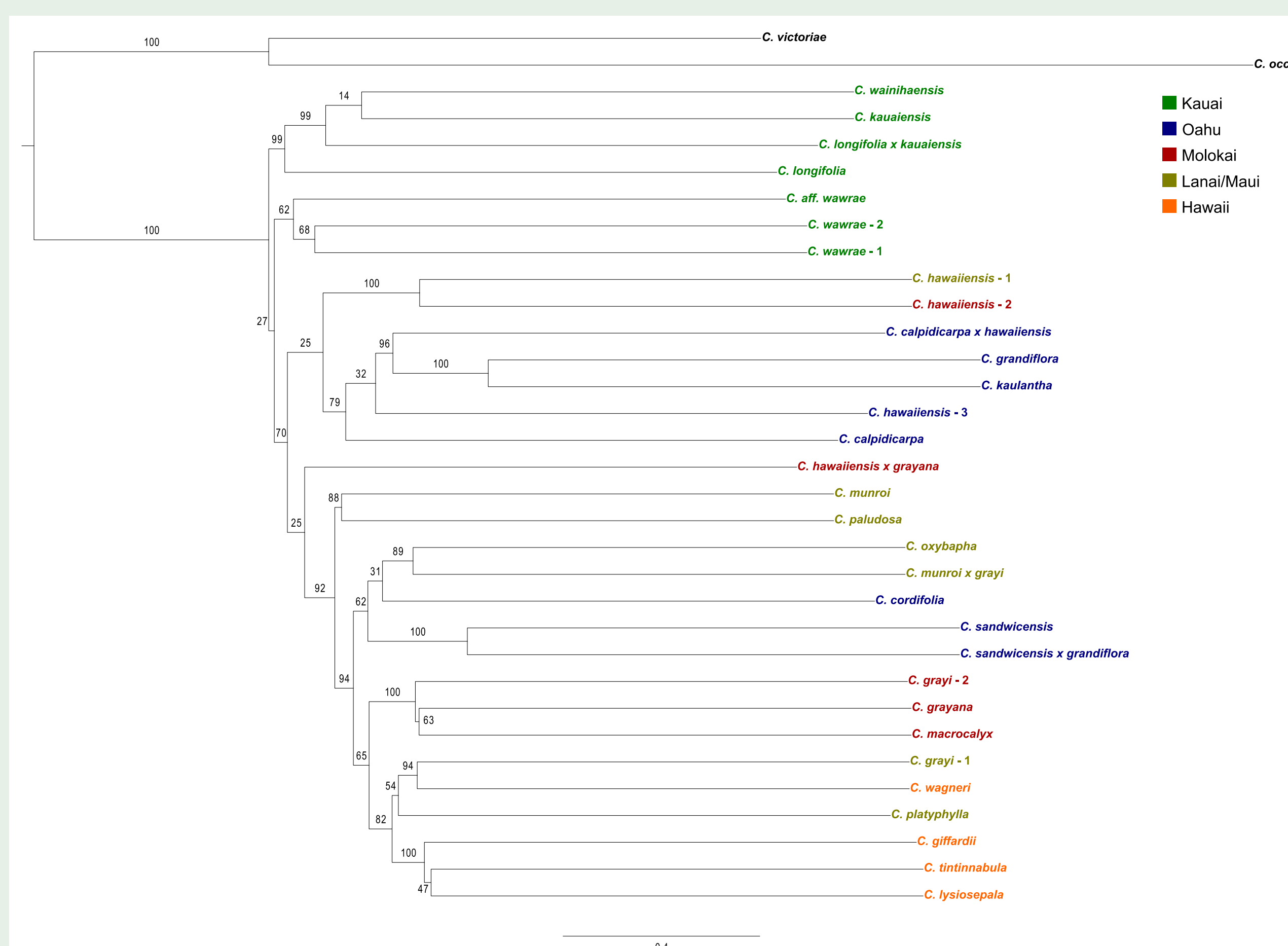


Figure 3. Species tree inferred in ASTRAL from 569 gene trees created in RAxML using the dataset with all taxa included. Bootstrap values are plotted on branches. Scale represents branch lengths in per-site substitution rate.

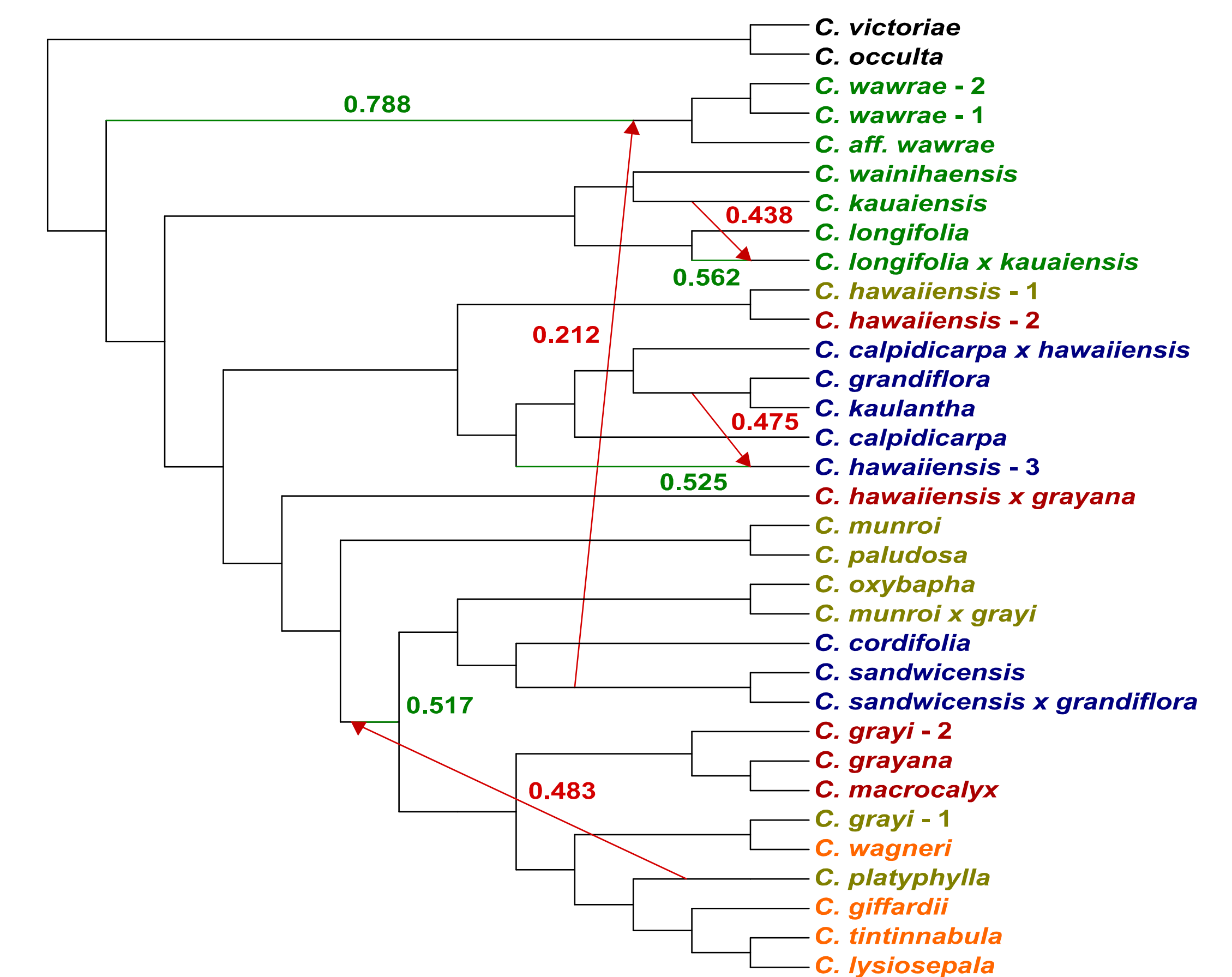


Figure 4. Network analyses conducted in SNaQ using the species tree from ASTRAL and concordance factors from gene trees generated in RAxML. Red edges indicate gene flow and gamma values indicate percentage of genes inferred to be introgressed. Hybridization events (h) ranged from 2-5 depending on the dataset.

Summary

- High-throughput sequencing and modern phylogenomic methods can provide insight into species relationships of recent plant radiations.
- Species relationships hypothesized here largely agree with prior studies (Clark et al. 2008, 2009; Johnson et al. 2017), including monophyly of Hawaiian *Cyrtandra* and dispersal/back-dispersal events.
- For Hawaiian *Cyrtandra*, species-tree methods (ASTRAL) are more accurate than concatenated methods, most likely due to high levels of incomplete lineage sorting.
- Putative hybrids formed relationships we expected based on their parent species. However, except in the case of *C. longifolia x kauaiensis*, network analyses show that these taxa are not necessarily hybrids.
- Network analyses show that *C. wawrae* may be a hybrid, with parents from Kauai and a back-dispersed taxa from Oahu or Maui Nui.
- Hybrid edges in incorrect directions indicate ancient gene flow or gene flow with unsampled lineages; hybridization in Hawaiian *Cyrtandra* is likely a natural process and not a consequence of human activity in Hawaii.
- Sampling is important to understanding the species relationships and hybridization events in Hawaiian *Cyrtandra*. Continued studies of this lineage will benefit from sequencing more widely between and among species.

Works Cited

- Clark J., Ree R., Alfaro M., King M., Wagner W., Roalson E. 2008. A Comparative Study in Ancestral Range Reconstruction Methods: Retracing the Uncertain Histories of Insular Lineages. *Systematic Biology*. 57:693–707.
- Clark J.R., Wagner W.L., Roalson E.H. 2009. Patterns of diversification and ancestral range reconstruction in the southeast Asian–Pacific angiosperm lineage *Cyrtandra* (Gesneriaceae). *Molecular Phylogenetics and Evolution*. 53:982–994.
- Johnson M.A., Clark J.R., Wagner W.L., McDade L.A. 2017. A molecular phylogeny of the Pacific clade of *Cyrtandra* (Gesneriaceae) reveals a Fijian origin, recent diversification, and the importance of founder events. *Molecular Phylogenetics and Evolution*. 116:30–48.
- Mirarab S., Reaz R., Bayzid M.S., Zimmermann T., Swenson M.S., Warnow T. 2014. ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics*. 30:i541–i548.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 30:1312–1313.
- Vachaspati P., Warnow T. 2015. ASTRID: accurate species trees from internode distances. *BMC genomics*. 16:S3.
- Weitemier K., Straub S.C.K., Cronn R.C., Fishbein M., Schmickl R., McDonnell A., Liston A. 2014. Hyb-Seq: Combining Target Enrichment and Genome Skimming for Plant Phylogenomics. *Applications in Plant Sciences*. 2:1400042.
- Thank you to Dr. Joanna Kelley and Corey Quackenbush for insight regarding library preparation, and Wade Roberts for advice on analyses.

Table 1. Probeset and sequencing statistics.

	Total Sequences	Max Sequence	Min Sequence	Average Sequence	Total Length
Probe Set	570	2053 bp	120 bp	317.2 bp	180784 bp
Filtered Data	569	2590 bp	233 bp	819.3 bp	466201 bp

Table 2. Concatenated dataset statistics.

	Total Length	Constant Sites	Parsimony informative	Parsimony uninformative	Missing Data
Filtered Data	466201 bp	334617 bp	58683 bp	72901 bp	8.31%
No Outgroups	466201 bp	343083 bp	52029 bp	71089 bp	8.37%