



**HAL**  
open science

## **Gene discordance in phylogenomics of recent plant radiations, an example from Hawaiian Cyrtandra (Gesneriaceae)**

Yohan Pillon, Jennifer Johansen, Tomoko Sakishima, Eric Roalson, Donald Price,  
Elizabeth Stacy

### ► **To cite this version:**

Yohan Pillon, Jennifer Johansen, Tomoko Sakishima, Eric Roalson, Donald Price, et al.. Gene discordance in phylogenomics of recent plant radiations, an example from Hawaiian Cyrtandra (Gesneriaceae). *Molecular Phylogenetics and Evolution*, 2013, 69 (1), pp.293-298. <10.1016/j.ympev.2013.05.003>. <ird-04953044>

**HAL Id: ird-04953044**

**<https://ird.hal.science/ird-04953044v1>**

Submitted on 17 Feb 2025

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

**Gene discordance in phylogenomics of recent plant radiations,  
an example from Hawaiian *Cyrtandra* (Gesneriaceae)**

Yohan PILLON<sup>1,3</sup>, Jennifer B. JOHANSEN<sup>1</sup>, Tomoko SAKISHIMA<sup>1</sup>, Eric H. ROALSON<sup>2</sup>, Donald K. PRICE<sup>1</sup> &  
Elizabeth A. STACY<sup>1</sup>

<sup>1</sup> Tropical Conservation Biology and Environmental Science Program, University of Hawai'i at Hilo, 200  
West Kawili Street, Hilo, HI, 96720, USA

<sup>2</sup> School of Biological Sciences, Washington State University, 339 Abelson Hall, Pullman, WA 99164-4236

<sup>3</sup> Corresponding author [pillon@hawaii.edu](mailto:pillon@hawaii.edu) (+1 808 937 0400)

## ABSTRACT

Resolving species relationships within recent radiations requires analysis at the interface of phylogenetics and population genetics, where coalescence and hybridization may confound our understanding of relationships. We developed 18 new primer pairs for nuclear loci in *Cyrtandra* (Gesneriaceae), one of the largest plant radiations in the Pacific Islands, and tested the concordance of 14 loci in establishing the phylogenetic relationships of a small number of Hawaiian species. Four genes yielded tree topologies conflicting with the primary concordance tree, suggesting plastid capture and horizontal transfer via hybridization. Combining all concordant genes yielded a tree with stronger support and a different topology from the total-evidence tree. We conclude that a small number of genes may be insufficient for accurate reconstruction of the phylogenetic relationships among closely related species. Further, the combination of genes for phylogenetic analysis without preliminary concordance tests can yield an erroneous tree topology. It seems that the number of genes needed for phylogenetic analysis of closely related species is significantly greater than the small numbers commonly used, which fail to isolate coalescence, introgression and hybridization.

Keywords: Hawai'i, network, next-generation sequencing, nuclear genes, phylogenomics, ribosomal DNA

## 1. Introduction

The generation of unprecedented volumes of DNA sequences has led to novel fields in evolutionary biology, such as phylogenomics (Delsuc et al., 2005; Philippe et al., 2005), that involve the use of a large number of markers across the genome to increase the number of informative characters and therefore provide greater resolution for phylogenetic analyses. Although phylogenomics has been used more often at broad taxonomic scales, it has potential for the investigation of species complexes or young radiations as a bridge between population genetics and phylogenetics. Resolving relationships within young radiations is difficult; even in such cases where divergence is sufficient to minimize gene flow among species, incomplete lineage sorting will often create conflicting tree topologies among loci. Further, in such cases where introgression or hybrid speciation has occurred, markers from different parts of the genome could yield conflicting signals (Howarth and Baum, 2005). As such, the evolution of such groups may be better represented as a network (Huson and Bryant, 2006). Phylogenomics could have a significantly greater potential than single-gene (or few-gene) approaches for solving these problems (Delsuc et al., 2005).

Phylogenomics has yet to contribute significantly to fine-scale plant phylogenies, despite the need for improved resolution over the traditional plastid gene and more recent nuclear gene approaches. Plastid markers have been popular in plant phylogenetics because universal primers are widely available for their amplification, their haploid state facilitates combination for analysis, their physical linkage promotes congruency, and they rarely recombine (Marshall et al., 2001). This is in contrast with nuclear genes where recombination can be common even within loci (Kelly et al., 2010; Pillon et al., 2009). Despite these advantages, plastid capture seems to be common in plants (Rieseberg and Soltis, 1991) and indicates that caution should be used when inferring phylogenies among closely related species based on plastid sequences. Mitochondrial genes in contrast have been little used in

plants because of their low level of sequence variation, and they suffer the same shortcomings associated to their inheritance mode as plastid markers. Combining sequences from a large number of loci from one organelle (including plastome “phylogenomics”) has gained popularity as it allows greater precision in phylogenetic reconstruction; however, there is no associated increase in accuracy due to the tight linkage of organelle genes. Lastly, nuclear ribosomal markers, including ITS, are other popular phylogenetic markers that also have limitations (Alvarez and Wendel, 2003), including complex concerted evolution. To date, plastid and ribosomal markers have been by far the most commonly used markers in plant phylogenetics (Alvarez and Wendel, 2003; Hughes et al., 2006). In light of the above limitations, it has been argued that multiple nuclear genes should lead to the most accurate solution in phylogenetics (Cronn et al., 2003). However, not only are nuclear gene sequences more difficult to acquire due to the scarcity of universal primers, they also require more intensive and careful analyses, e.g. presence of gene recombination, deep-coalescent time, complex combination of multiple loci, etc. (Joly and Bruneau, 2006; Moody and Rieseberg, 2012; Pillon et al., 2009). Furthermore, multiple unlinked genes may have different evolutionary histories (Small et al., 2004). Thus, when using multiple independent genes for phylogenetic analysis, it is important to first test for concordance among them (Baum, 2007; Degnan and Rosenberg, 2009).

Species complexes are of particular interest for evolutionary studies as they represent ongoing speciation and of particular importance for conservation biology as they often include rare taxa. The young Hawaiian flora is well known for its striking radiations, and species delineation within groups is often difficult. *Cyrtandra* (Gesneriaceae) is one of the most diverse plant genera in the Pacific Islands and the second largest plant radiation in the Hawaiian islands, with over 53 endemic species. *Cyrtandra* comprises mostly understorey shrubs with relatively uniform white flowers and fleshy fruits; most variation among species occurs in inflorescence structure and vegetative characters. They represent a

monophyletic group in Hawai'i with a crown age of 4.4 My (Clark et al., 2009), consistent with a first colonization onto Kaua'i, the oldest and westernmost of the main islands. Individual *Cyrtandra* species are usually restricted to a single island, and hybridization is considered common (Smith et al., 1996). Almost all species of the genus investigated cytologically (36 species from Peninsular Malaya to Hawai'i) have  $2n=34$  chromosomes (RBGE 'WebCyte' – Gesneriaceae cytology database, accessed 21 March 2013), and there is no indication of polyploidy in the group. Its high species diversity but low ecological diversity suggests that *Cyrtandra* has undergone a non-adaptive radiation. Next generation Sequencing offers the possibility to identify many novel nuclear single-copy genes in non-model organisms (Zimmer and Wen, 2012); and by developing such markers for *Hawaiian Cyrtandra*, we aimed at evaluating how phylogenomics can help relationships in a species complex and how much data is needed for this purpose.

## 2. Materials and methods

### 2.1 Development of novel markers

We obtained a pooled, partial transcriptome library from leaf and floral buds [fixed in the field in RNA later (QIAGEN, Maryland, USA)] for two Hawaiian *Cyrtandra*: *C. longifolia* (Kaua'i) and a purported hybrid *C. hawaiiensis* × *calpidicarpa* (O'ahu). RNA isolation, cDNA synthesis and 454 sequencing were done at the University of Arizona Genetics Core Lab. We conducted BLAST searches of the 400 most highly expressed genes in *Arabidopsis* (C. Fizames, pers. comm.) against our data in CLC DNA Workbench (using default parameters) to identify a set of genes with maximum coverage in both samples. We favoured loci (generally only a small portion of a gene) that comprised a single, long exon (200 bp) with matches in both species, and designed primers with FastPCR (Kalendar et al., 2011) using default

settings. The presence of introns was tested by comparison with genomic and cDNA sequences in *Arabidopsis* ([www.arabidopsis.org](http://www.arabidopsis.org)). Our preliminary works on two those genes indicated a high frequency of heterozygotes (34 % in *Cyrt2* and 36% in *Cyrt4*, Pillon et al., 2013). Introns often contain indels which result in the common occurrence of alleles of different lengths in heterozygotes, which cannot be sequenced without cloning, therefore avoiding intron allowed the direct sequencing of most accessions.

Each of 18 novel primer pairs (table 1) for putatively single-copy nuclear genes were tested on six accessions: *C. schizocalyx* (Vanuatu, outgroup), *C. kauaiensis* (endemic to Kaua'i), *C. longifolia* (endemic to Kaua'i), *C. cordifolia* (endemic to O'ahu), *C. platyphylla* (Hawai'i, also present on Maui) and *C. paludosa* (Hawai'i, also present on Kaua'i, O'ahu and Maui). Chromosome counts ( $2n=34$ ) are available for all five Hawaiian species (Webcyte, 2012).

## 2.2 Sequencing procedure

Leaf material was collected in the field and dried in silica gel, and genomic DNA was extracted using the Nucleospin® Plant II Kit (MACHEREY-NAGEL). The nuclear regions were amplified using the following mix: 12.3  $\mu$ L of H<sub>2</sub>O, 4  $\mu$ L of Gotaq 5  $\times$  Buffer (PROMEGA), 2  $\mu$ L of MgCl<sub>2</sub> 25 mM, 0.4  $\mu$ L of dNTP 1.25  $\mu$ M, 0.2  $\mu$ L of each primer 10  $\mu$ M, 0.1  $\mu$ L of GoTaq Flexi DNA polymerase 5u/ $\mu$ L (PROMEGA) and 0.8  $\mu$ L of DNA template. The following amplification program was used: 2 min at 94°C, 38 cycles of 1 min at 94°C, 1 min at 58-63°C, 1 min to 1 min 30s at 72°C and a final extension of 5 min at 72°C (table 1). We followed Clark et al. (2009) for amplification of ETS, ITS and *psbA-trnH*. PCR products were sequenced directly with no subsequent cloning. The identity of each amplified gene was validated through BLAST or tBLASTx searches in GenBank.

## 2.4 Sequence analyses

Forward and reverse sequences were assembled in Sequencer 4.9 paying particular attention to sites with multiple signals that were encoded using the IUAC nucleotide ambiguity codes. Because most phylogenetic analyses currently available do not allow to combine loci with multiple alleles per accession (with the exception of \*BEAST, Heled and Drummond, 2010, see below) coding ambiguous sites was more appropriate than phasing alleles, and we did not perform cloning for the same reason. DNA sequences were aligned by hand in MEGA5 (Tamura et al., 2011). The best evolution model for each sequence was determined using jModeltest (Posada, 2008), and the presence of recombination was tested using GARD (Kosakovsky Pond et al., 2006) assuming no site-to-site rate variation (most sensitive setting). A 31pb stem-loop region in *psbA-trnH* was excluded from analyses (see Clark et al., 2009).

Single-gene and concatenated phylogenetic analyses—We conducted Bayesian phylogenetic analysis on direct sequences using MrBayes (Ronquist and Huelsenbeck, 2003). We ran 720,000 generations of MCMC sampling one tree every 60 generations and with a burnin of 120,000 generations, using the gene specific substitution model and default settings otherwise. Loci were unlinked in the concatenated analysis so that each locus could follow its own evolution model. Convergence was checked using Tracer.

## 2.5 Incongruence analysis

We tested for incongruence in our dataset using BUCKY (Ané et al., 2007; Larget et al., 2010), which also estimates the number of clusters of loci sharing the same tree topology. It builds a primary concordance tree, that which is true for the largest proportion of genes, and a population tree with nodes with the highest concordance factors based on the full gene datasets. We used the MrBayes output for the single-gene analysis and set  $\alpha$  (*a priori* level of discordance among loci) to 3 (yielding a

level of discordance closer to our observation of the number of gene groups using the BUCKY website) and default settings.

## 2.6 Network analysis

Alleles of heterozygotes accessions were separated (phased) by eyes, using sometime sequences from other accessions (unpublished data). We measured the pairwise genetic distance (number of substitutions) between phased alleles of each gene dataset in MEGA5 (Tamura et al., 2011). A single pairwise non-standardized distance matrix between taxa was obtained using POFAAD (Joly and Bruneau, 2006), which was then used to build a network using the NeighborNet analysis in Splitstree (Huson and Bryant, 2006; Huson and Kloepper, 2005).

## 2.7 \*BEAST analysis

\*BEAST (Heled and Drummond, 2010) is currently one of the few phylogenetic software that allows to combine loci with multiple sequences/alleles per terminal. We phased all alleles (including homozygotes), and used the specific substitution model for each locus, a strict molecular clock (to reduce calculation time), a Yule Process for tree prior with a nuclear ploidy type set for all loci except for *psbA-trnH* which was set as mitochondrial. We ran four times a MCMC chain of 200,000,000 generations with a 20,000,000 burnin and checked convergence with Tracer.

## 3. Results

Eleven loci (Appendix A & B) were successfully sequenced in all accessions along with the nuclear ribosomal ETS and ITS loci as well as the plastid locus *psbA-trnH*. Two loci were apparently duplicated (*Cyrt6* and *Cyrt13*), and one (*Cyrt15*) was duplicated in all Hawaiian accessions but not in the

accession from Vanuatu. Amplification was most problematic for longer fragments, and sequencing was difficult for loci that included introns. We came across multiple occurrences of heterozygotes with alleles of different lengths (due to the presence of indels) that could not be sequenced directly, and we did not attempt to clone them. Most nuclear loci had a percentage of variable sites between those of *psbA-trnH* and ITS. No recombination was detected in any of the datasets.

The phylogenetic analysis of the 14 concatenated genes yielded a tree with a single strongly supported node: *C. cordifolia* + *C. platyphylla* (Fig 1.). An analysis of concordance suggests the existence of five groups of genes yielding different tree topologies. The primary concordance tree and the population tree were the same but differed from the total 14-gene concatenated tree with respect to the placement of *C. longifolia*. The individual gene analyses of 10 of the loci (concordant genes) did not yield any hard conflicts with the concordance tree. Four genes yielded different topologies with a least one supported node (posterior probability >0.92) not found in the concordance tree: *Cyrt11* (indicating a sister relationship between *C. longifolia* and *C. paludosa*), *Cyrt16* (*C. kauaiensis* sister to all other Hawaiian species), *Cyrt17* (*C. paludosa* sister to *C. cordifolia* + *C. platyphylla*), and the plastid *psbA-trnH* (*C. longifolia* sister to *C. cordifolia*). An analysis of the 10 concatenated concordant genes yielded a tree identical to the concordant-gene tree with all nodes with posterior probabilities above 0.93 (figure 1).

A network analysis on phased alleles of all 14 loci produced a highly reticulated network with ambiguities for most relationships. When only the 10 concordant genes were included, the level of reticulation was reduced and distinction between two groups became more obvious, one including *C. cordifolia* and *C. platyphylla*, and one including *C. kauaiensis*, *C. longifolia* and *C. paludosa*; however, the relationship within the second group remained unclear (figure 1). The \*BEAST phylogenetic tree yielded a tree that was different from the concordance tree and with a single well-supported internal node.

#### 4. Discussion

The lack of congruence among genes observed in this study strongly suggest that spurious relationships may be commonly inferred from single-gene phylogenies, as has been noted previously in *Gossypium* (Cronn et al., 2003). In our study of six species of *Cyrtandra*, four out of 14 genes yielded topologies that were incongruent with the primary concordance tree. Soltis et al. (2008) observed that polymorphic sites can affect phylogenetic placement of hybrid but: 1) our network analyses do not provide evidence of a hybrid origin of any of our taxa, 2) polymorphic sites were observed only twice in parsimony informative sites (*Cyrt4* and *Cyrt7*) and were not associated with discordance. Therefore we do not think that IUPAC coding of ambiguous sites is the source of the different topologies observed. Incongruence may be explained by deep coalescence in some cases (*Cyrt11*, *Cyrt16*), although we believe that in some cases horizontal gene transfer through introgression seems more likely (*Cyrt17*, *psbA-trnH*). The sister relationship between *C. longifolia* and *C. cordifolia* recovered with *psbA-trnH* may be an example of plastid capture. In contrast, the discordance in *Cyrt17* seems better explained by introgression into *C. paludosa* sometime between its origin on Kaua'i and its arrival on Hawai'i island; at this locus individuals of *C. paludosa* from Kaua'i are more closely related to other Kaua'i species than they are to conspecific individuals from O'ahu or Hawai'i (Pillon et al. unpublished). The relationships among *C. kauaiensis*, *C. longifolia* and *C. paludosa* may not be considered fully resolved, as the sister relationship between *C. longifolia* and *C. kauaiensis* is only moderately supported (PP=0.93) and because conflicts among genes are observed. Gene discordance may be explained by simultaneous divergence of these two species making coalescence a major confounding factor, or by hybridization between those species. The \*BEAST analysis did not turn to be robust to this coalescence problem as it recovered a

topology that is different from the primary concordance tree, however, it might have been more accurate if multiple accessions had been sampled within each species.

The resolution of species in young radiations may require a potentially very large number of independent genes. Although our sampling was limited to a small number of morphologically distinct Hawaiian species that are not expected to be each other's closest relatives, resolving their relationships by way of multiple nuclear genes nonetheless proved difficult. The number of genes we used (14), despite being significantly greater than that typically used in plant phylogenetic studies, was still insufficient to clarify the phylogeny of this young radiation with high confidence. These results demonstrate that discordance among single-copy nuclear genes is a common phenomenon that needs to be taken into account when investigating recent plant radiations. Concordance analysis using BUCKY (Larget et al., 2010) proved efficient for reconstruction of the primary concordance tree and the population tree that better reflect the relationships between the species as inferred by a majority of genes. This approach allowed detection of discordant genes and subsequent phylogenetic and network analyses based on concordant genes alone. Despite the relatively greater resolution and support of the tree based on concordant genes, some species relationships had limited support. As observed in *Helianthus* (Moody and Rieseberg, 2012), a significantly greater number of genes would be needed to clarify species relationships and to infer or exclude hybrid speciation in this young group.

### **Acknowledgements**

The authors wish to thank the following for facilitating the collection of plant specimens: DLNR DOFAW, DLNR State Parks Division, K. Cassell and J. Lau. The sample from Vanuatu was collected during the expedition Santo 2006 (IRD/MNHN/Pro-Natura). We thank H. Issar (University of Arizona

Genetics Core), and A. Veillet (Hilo Genetics Core Facilities) for technical assistance, M. Lebrun and C. Fizames for information on nuclear genes, C. Ané and L. Barrabé for help and advice with data analyses and two anonymous reviewers for their comments on an earlier version of this manuscript. This work was supported by the Gordon and Betty Moore Foundation, the NSF CREST Program (HRD: 0833211), the US Fish and Wildlife (grant to EHR), and the National Tropical Botanical Garden McBryde Research Endowment for Hawaiian and Pacific Botany.

## REFERENCES

- Alvarez, I., Wendel, J.F., 2003. Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution* 29, 417-434.
- Ané, C., Larget, B., Baum, D.A., Smith, S.D., Rokas, A., 2007. Bayesian estimation of concordance among gene trees. *Molecular Biology and Evolution* 24, 412-426.
- Baum, D.A., 2007. Concordance trees, concordance factors, and the exploration of reticulate genealogy. *Taxon* 56, 417-426.
- 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.
- Cronn, R., Small, R.L., Haselkorn, T., Wendel, J.F., 2003. Cryptic repeated genomic recombination during speciation in *Gossypium gossypioides*. *Evolution* 57, 2475-2489.
- Degnan, J.H., Rosenberg, N.A., 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution* 24, 332-340.
- Delsuc, F., Brinkmann, H., Philippe, H., 2005. Phylogenomics and the reconstruction of the tree of life. *Nature review Genetics* 6, 361-375.
- Heled, J., Drummond, A.J., 2010. Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution* 27, 570-580.
- Howarth, D.G., Baum, D.A., 2005. Genealogical evidence of homoploid hybrid speciation in an adaptive radiation of *Scaevola* (Goodeniaceae) in the Hawaiian islands. *Evolution* 59, 948-961.
- Hughes, C.E., Eastwood, R.J., Bailey, C.D., 2006. From famine to feast? selecting nuclear DNA sequence loci for plant species-level phylogeny reconstruction. *Philosophical Transactions of the Royal Society of London B* 361, 211-225.
- Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* 23, 254-267.
- Huson, D.H., Klopper, T.H., 2005. Computing recombination networks from binary sequences. *Bioinformatics* 21, ii159-ii165.
- Joly, S., Bruneau, A., 2006. Incorporating allelic variation for reconstructing the evolutionary history of organisms from multiple genes: an example from *Rosa* in North America. *Systematic Biology* 55, 623-636.
- Kalendar, R., Lee, D., Schulman, A.H., 2011. Java web tools for PCR, in silico PCR, and oligonucleotide assembly and analysis. *Genomics* 98, 137-144.
- Kelly, L.J., Leitch, A.R., Clarkson, J.J., Hunter, R.B., Knapp, S., Chase, M.W., 2010. Intragenic recombination events and evidence for hybrid speciation in *Nicotiana* (Solanaceae). *Molecular Biology and Evolution* 27, 787-799.
- Kosakovskiy, S.L., Posada, D., Gravenor, M.B., Woelk, C.H., Frost, S.D.W., 2006. GARD: a genetic algorithm for recombination detection. *Bioinformatics* 22, 3096-3098.
- Larget, B.R., Kotha, S.K., Dewey, C.N., Ané, C., 2010. BUCKy: gene tree/species tree reconciliation with Bayesian concordance analysis. *Bioinformatics* 26, 2910-2911.
- Marshall, H.D., Newton, C., Ritland, K., 2001. Sequence-repeat polymorphisms exhibit the signature of recombination in lodgepole pine chloroplast DNA. *Molecular Biology and Evolution* 18, 2136-2138.
- Moody, M.L., Rieseberg, L.H., 2012. Sorting through the chaff, nDNA gene trees for phylogenetic inference and hybrid identification of annual sunflowers (*Helianthus* sect. *Helianthus*). *Molecular Phylogenetics and Evolution* 64, 145-155.

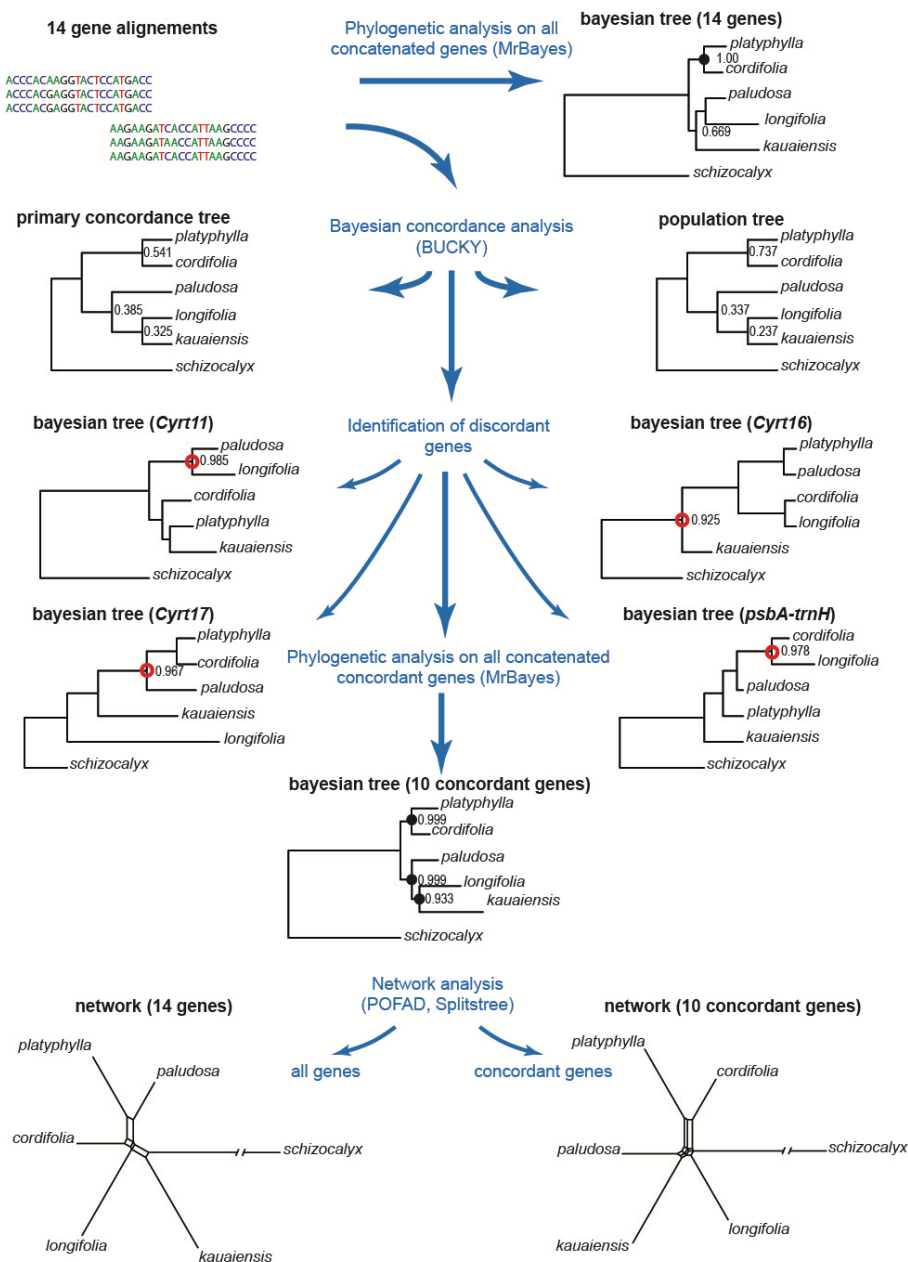
- Philippe, H., Delsuc, F., Brinkmann, H., Lartillot, N., 2005. Phylogenomics. *Annual Reviews of Ecology, Evolution and Systematics* 36, 541-562.
- Pillon, Y., Hopkins, H.C.F., Munzinger, J., Amir, H., Chase, M.W., 2009. Cryptic species, gene recombination and hybridization in the genus *Spiraeanthemum* (Cunoniaceae) from New Caledonia. *Botanical Journal of the Linnean Society* 161, 137-152.
- Pillon, Y., Johansen, J., Sakishima, T., Chamala, S., Barbazuk, W.B., Roalson, E.H., Price, D.K., Stacy, E.A., 2013. Potential use of low-copy nuclear genes in DNA barcoding: a comparison with plastid genes in two Hawaiian plant radiations. *BMC Evolutionary Biology* 13, 35.
- Posada, D., 2008. jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution* 25, 1253-1256.
- Rieseberg, L.H., Soltis, D.E., 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. *Evolutionary Trends in Plants* 5, 65-84.
- Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 1572-1574.
- Small, R.L., Cronn, R.C., Wendel, J.F., 2004. Use of nuclear genes for phylogeny reconstruction in plants. *Australian Systematic Botany* 17, 145-170.
- Smith, J.F., Burke, C.C., Wagner, W.L., 1996. Interspecific hybridization in natural populations of *Cyrtandra* (Gesneriaceae) on the Hawaiian islands : evidence from RAPD. *Plant Systematics and Evolution* 200, 61-77.
- Soltis, D.E., Mavrodiev, E.V., Doyle, J.J., Rauscher, J., Soltis, P.S., 2008. ITS and ETS sequence data and phylogeny reconstruction in allopolyploids and hybrids. *Systematic Botany* 33, 7-20.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: Molecular Evolutionary Genetic Analysis using maximum likelihood, evolutionary distance, and maximum parsimony distance. *Molecular Biology and Evolution* 28, 2731-2739.
- Zimmer, E.A., Wen, J., 2012. Using nuclear gene data for plant phylogenetics: Progress and prospects. *Molecular phylogenetics and evolution* 65, 774-785.

**Table 1.** Identity and PCR conditions of the 18 low-copy nuclear genes identified in this study with primer sequences.

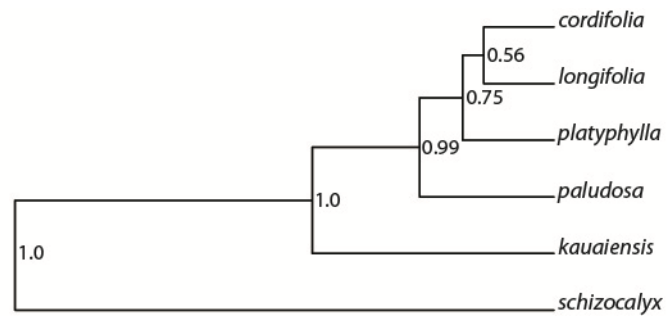
	Putative Arabidopsis homolog	Putative product	Primer F	Primer R	Length (bp) Exon-intron	Annealing Temperature (°C)	Extension time
<i>Cyrt1</i>	At3g16640	TCTP, Translationally Controlled Tumor Protein	cagctgaaggatgcgaagag	tcgttcaaaccatggccga	<b>80-89-161-457-105</b>	58	1'
<i>Cyrt2</i>	At2g18020	EMB2296, Embryo Defective 2296/ Ribosomal Protein L2	cgtgttcaagtctcacacgca	tcaccagaagctctagcga	<b>354</b>	61	1'
<i>Cyrt3</i>	At1g43170	RPL3A, cytoplasmic ribosomal protein	tatcgggtgcacgaagggt	ttgacgaaggtaaccacc	<b>249-610-133</b>	61	1'30
<i>Cyrt4</i>	At4g13940	SAHH, S-Adenosyl-L-homocysteine hydrolase	ctcttcaagccctcatgga	gaagctagggtgtccgga	<b>329</b>	61	1'
<i>Cyrt5</i>	At3g47470	CAB4, Chlorophyll binding protein	gactagcagaagaccgga	actcgatcacgaagagggtcg	<b>196</b>	58	1'
<i>Cyrt6</i>	At1g72370	40S ribosomal protein	ggaaaacgtgggagaagct	cctaggttcgctgaatgat	<b>206</b>	63	1'
<i>Cyrt7</i>	At1g49140	NADH-ubiquinone oxidoreductase	ttcgatccgaaaaccgt	tcgagatactggtggacca	<b>185</b>	58	1'
<i>Cyrt8</i>	At5g56000	HSP81.4 Heat Shock Protein 81.4	cgttttgaaagcctgacgga	caaactgcccaatcatgct	<b>127-84-98</b>	61	1'
<i>Cyrt9</i>	At1g04270	RPS15, cytosolic ribosomal protein 15	agatgatcgggagtgctct	cttccgtgcttactggct	<b>127</b>	61	1'
<i>Cyrt10</i>	At5g60360	SAG2, Senescence associated thiol protease	tgtggatcctgctggacat	aggcttgagagggtaaacc	<b>20-123-137</b>	63	1'
<i>Cyrt11</i>	At2g24200	LAP1, cytosol leucyl aminopeptidase	aagactggagcaggatgta	gcatcagctagggtgagtctt	<b>109-192-101-700+-41</b>	61	1'30

<i>Cyrt12</i>	At4g14880	Cytosolic O-acetylserine(thiol)lyase	gaaaatcctgccaacccaa	tccaaacaccaggacga	<b>19-220+-138-125-52-740+-54</b>	61	1'30
<i>Cyrt13</i>	At5g50920	CLPC homolog, involved in protein transport	cgaagaagctatgatccggct	gatcaaagcctatgcgtcg	<b>333</b>	63	1'
<i>Cyrt14</i>	At5g61170	Ribosomal protein S19	tacgccgccatctcaagc	ttcctctgtctccgcat	<b>28-600+-92-360+-83</b>	61	1'30
<i>Cyrt15</i>	At1g14320	RPL10, Ribosomal Protein 10	tcctaaatcacggtactgc	caaagcctcctgagcatgg	<b>408</b>	61	1'
<i>Cyrt16</i>	At2g16600	ROC3, Cytosolic Cyclophilin/peptidyl-prolyl cis-trans isomerase	gccggatcgtgatggagct	agaatacctggtccagtgtgc	<b>264</b>	61	1'
<i>Cyrt17</i>	At3g09200	RPL10, Ribosomal Protein 10	aggtgacaaagttggttcctc	atctgcttgtggaaagag	<b>300</b>	61	1'
<i>Cyrt18</i>	At3g19820	DWARF1, Brassinosteroid biosynthetic protein	catgctgagacagcattgaa	gaggaacgagcatatcctg	<b>250</b>	61	1'

**Fig. 1.** Step by step procedure from gene alignments to the concordance tree and the identification of discordant genes. Numbers at nodes are posterior probabilities (Bayesian trees), sample concordance factors (primary concordance tree) or estimated coalescent units (population tree). Red circles indicate supported incongruent nodes; black rounds are supported nodes (posterior probabilities > 0.95) congruent with the primary concordance tree.



**Fig.2.** Phylogenetic tree inferred with \*BEAST using phased alleles. Numbers at nodes are posterior probabilities.



**Appendix A.** Comparison of variation of the 11 low-copy nuclear genes, one plastid gene (*psbA-trnH*) and two nuclear ribosomal loci (ITS, ETS)

successfully sequenced in all six *Cyrtandra* accessions.

	<i>Cyrt2</i>	<i>Cyrt4</i>	<i>Cyrt5</i>	<i>Cyrt7</i>	<i>Cyrt8</i>	<i>Cyrt9</i>	<i>Cyrt10</i>	<i>Cyrt11</i>	<i>Cyrt16</i>	<i>Cyrt17</i>	<i>Cyrt18</i>	<i>psbA-trnH</i>	ITS	ETS
Length (bp)	354	329	198	185	309	127	274-280	1050+	264	300	250	433-471	881-893	516-521
Number of variable sites	11	7	1	2	10	4	3	30	5	11	4	8	35	35
Percent variable sites	3.1	2.1	0.5	1.1	3.2	3.1	1.1	2.9	1.4	3.7	1.6	1.7	3.9	6.7
1st position (non-synonymous)	3 (3)	3 (3)	0	0	0	1 (1)	1 (1)	0	1 (1)	3 (3)	1 (1)	-	-	-
2nd position (non-synonymous)	0	1 (1)	0	0	0	0	0	0	3 (3)	4 (4)	0	-	-	-
3rd position (non-synonymous)	8 (1)	3 (1)	1 (1)	2 (0)	6 (0)	4 (0)	1 (0)	3 (0)	1 (0)	5 (0)	3 (0)	-	-	-
Intron	-	-	-	-	2	-	-	27	-	-	-	-	-	-
Indels	0	0	0	0	0	0	1	5	0	0	0	3	8	2
Number of heterozygotes	1	2	0	3	0	0	1	3	0	3	1	0	2	1
Model of evolution	F81	F81	JC	F81	HKY	JC	K80	HKY	K80	HKY	K80	HKY	GTR + $\gamma$	HKY

**Appendix B.** Genbank accession numbers for the genes included in this study. Because sequences shorter than 200 bp are not accepted in Genbank, the full sequences for *Cyrt5*, *Cyrt7* and *Cyrt9* are provided after the table. Direct sequences of *Cyrt6* and *Cyrt13* revealed two copies and were not submitted to Genbank.

	<i>Cyrtandra schizocalyx</i>	<i>Cyrtandra kauaiensis</i>	<i>Cyrtandra longifolia</i>	<i>Cyrtandra paludosa</i>	<i>Cyrtandra cordifolia</i>	<i>Cyrtandra platyphylla</i> var. <i>parviflora</i>
	G117	G60	G47	G199	G258	G72
Island	Vanuatu, Santo	Kaua'i	Kaua'i	Hawai'i	O'ahu	Hawai'i
Locality	Butmas	Koke'e	Koke'e	Waiakea Forest Reserve	Kulu'i Gulch	Waiakea Forest Reserve
Voucher	Pillon 534	Johansen 19	Johansen 17	Johansen 44	No voucher	Johansen 27
<i>Cyrt1</i>	—	KC435349	KC435347	KC435346	KC435348	—
<i>Cyrt2</i>	KC283100	JX845927	JX845925	JX845944	JX845953	JX845929
<i>Cyrt3</i>	KC435343	KC435342	KC435341	KC435345	KC435344	—
<i>Cyrt4</i>	KC283099	JX845985	JX845983	JX846002	JX846011	JX845987
<i>Cyrt6</i>	—	—	KC813144*	—	—	—
<i>Cyrt8</i>	KC435317	KC435319	KC435318	KC435320	KC435321	KC435322
<i>Cyrt10</i>	KC435311	KC435312	KC435313	KC435314	KC435315	KC435316
<i>Cyrt11</i>	KC435308	KC435306	KC435305	KC435309	KC435310	KC435307
<i>Cyrt12</i>	—	—	—	KC435354	KC435355	—
<i>Cyrt13</i>	—	—	KC813146*	—	—	—
<i>Cyrt14</i>	KC435350	KC435353	—	KC435351	KC435352	—
<i>Cyrt15</i>	KC435356	—	—	—	—	—
<i>Cyrt16</i>	KC435338	KC435336	KC435335	KC435339	KC435340	KC435337
<i>Cyrt17</i>	KC435323	KC435325	KC435324	KC435326	KC435327	KC435328
<i>Cyrt18</i>	KC435330	KC435332	KC435329	KC435331	KC435333	KC435334
ETS	KC283092	KC283088	KC283090	KC283093	KC283089	KC283091
ITS	KC283087	KC283082	KC283081	KC283083	KC283084	KC283080
<i>psbA</i>	KC283096	JX846043	JX846041	JX846059	JX846068	JX846045

\*These sequences are 454 sequences of cDNA obtained from a different individual (G58) from the same population of *Cyrtandra longifolia*.

**Cyrt5 sequences***Cyrtandra schizocalyx*

GACTAGCAGAAGACCCGGAGAACTTGAGATGGTATGTCCAGGCTGAGCTTGTGAACGGCCGATGGGCAATGCTC  
 GGTGTTGCAGGCATGCTGTTGCCTGAAGTTTTTCGTCCATCGGAATCATCAATGCTCCGAAATGGTACGACGCAG  
 GGAAAGCGGATTACTTCGCCTCCTCATCGACCCTCTTCGTGATCGAGT

*Cyrtandra longifolia*

??  
 gACTAGCAGAAGACCCGGAGAACTTGAGATGGTATGTCCAGGCTGAGCTTGTGAACGGCCGATGGGCaATGCTCGGTGTTGCA  
 GGCATGCTGTTGCCTGAAGTATTTTCGTCCATCGGAATCATCAATGCTCCGAAATGGTACGACGCAGGGAAAGCG  
 GATTACTTCGCCTCCTCATCGACCCTCTTCGTGATCGAGT

*Cyrtandra kauaiensis*

gACTAGCAGAAGACCCGGAGAACTTGAGATGGTATGTCCAGGCTGAGCTTGTGAACGGCCGATGGGCAATGCTC  
 GGTGTTGCAGGCATGCTGTTGCCTGAAGTATTTTCGTCCATCGGAATCATCAATGCTCCGAAATGGTACGACGCAG  
 GGAAAGCGGATTACTTCGCCTCCTCATCGACCCTCTTCGTGATCGAGT

*Cyrtandra cordifolia*

GACTAGCAGAAGACCCGGAGAACTTGAGATGGTATGTCCAGGCTGAGCTTGTGAACGGCCGATGGGCAATGCTC  
 GGTGTTGCAGGCATGCTGTTGCCTGAAGTATTTTCGTCCATCGGAATCATCAATGCTCCGAAATGGTACGACGCAG  
 GGAAAGCGGATTACTTCGCCTCCTCATCGACCCTCTTCGTGATCGAGT

*Cyrtandra paludosa*

GACTAGCAGAAGACCCGGAGAACTTGAGATGGTATGTCCAGGCTGAGCTTGTGAACGGCCGATGGGCAATGCTC  
 GGTGTTGCAGGCATGCTGTTGCCTGAAGTATTTTCGTCCATCGGAATCATCAATGCTCCGAAATGGTACGACGCAG  
 GGAAAGCGGATTACTTCGCCTCCTCATCGACCCTCTTCGTGATCGAGT

*Cyrtandra platyphylla* var. *parviflora*

GACTAGCAGAAGACCCGGAGAACTTGAGATGGTATGTCCAGGCTGAGCTTGTGAACGGCCGATGGGCAATGCTC  
 GGTGTTGCAGGCATGCTGTTGCCTGAAGTATTTTCGTCCATCGGAATCATCAATGCTCCGAAATGGTACGACGCAG  
 GGAAAGCGGATTACTTCGCCTCCTCATCGACCCTCTTCGTGATCGAGT

**Cyrt7 sequences***Cyrtandra schizocalyx*

TTCGATCCGGAAAACCCGTACAAGGATCCGGTGGCGATGCTGGAGATGAGGGAGCAYCTGGTTAGGGAGAAGTG  
 GATCGACATCGAGAAGGCCAAGATTCTGCGGGAGAAGGTCAAGTGGTGCTACCGCGTGGAGGGTGTCAACCACC  
 TCCAGAAATGCCGCCACCTGGTCCACCAGTATCTCGA

*Cyrtandra longifolia*

TTCGATCCGGAAAACCCGTACAAGGATCCGGTGGCGATGCTGGAGATGAGGGAGCAYCTGGTTAGGGAGAAGTG  
 GATCGACATCGAGAAGGCCAAGATTCTGCGGGAGAAGGTCAAGTGGTGCTACCGCGTGGAGGGTGTCAACCACC  
 TCCAGAAATGCCGSCACCTGGTCCACCAGTATCTCGA

*Cyrtandra kauaiensis*

TTCGATCCGGAAAACCCGTACAAGGATCCGGTGGCGATGCTGGAGATGAGGGAGCACCTGGTTAGGGAGAAGTG  
 GATCGACATCGAGAAGGCCAAGATTCTGCGGGAGAAGGTCAAGTGGTGCTACCGCGTGGAGGGTGTCAACCACC  
 TCCAGAAATGCCGCCACCTGGTCCACCAGTATCTCGA

*Cyrtandra cordifolia*

TTCGATCCGGAAAACCCGTACAAGGATCCGGTGGCGATGCTGGAGATGAGGGAGCAYCTGGTTAGGGAGAAGTG  
 GATCGACATCGAGAAGGCCAAGATTCTGCGGGAGAAGGTCAAGTGGTGCTACCGCGTGGAGGGTGTCAACCACC  
 TCCAGAAATGCCGSCACCTGGTCCACCAGTATCTCGA

*Cyrtandra paludosa*

TTCGATCCGGAAAACCCGTACAAGGATCCGGTGGCGATGCTGGAGATGAGGGAGCACCTGGTTAGGGAGAAGTG  
 GATCGACATCGAGAAGGCCAAGATTCTGCGGGAGAAGGTCAAGTGGTGCTACCGCGTGGAGGGTGTCAACCACC  
 TCCAGAAATGCCGCCACCTGGTCCACCAGTATCTCGA

*Cyrtandra platyphylla* var. *parviflora*

TTCGATCCGGAAAACCCGTACAAGGATCCGGTGGCGATGCTGGAGATGAGGGAGCACCTGGTTAGGGAGAAGTG  
 GATCGACATCGAGAAGGCCAAGATTCTGCGGGAGAAGGTCAAGTGGTGCTACCGCGTGGAGGGTGTCAACCACC  
 TCCAGAAATGCCGCCACCTGGTCCACCAGTATCTCGA

**Cyrt9 sequences***Cyrtandra schizocalyx*

AGATGATCGGGAGTGTCTTGGAGTATAACAATGGCAAGACCTTCAATCCGGTTGAAATCAAGCCCGAAATGATTG  
GCCACTATCTTGCTGAATTCTCGATTCATACAAGCCAGTGAAGCACGGAAG

*Cyrtandra kauaiensis*

AGATGATCGGGAGTGTCTTGGAGTATAACAATGGCAAGACCTTCAATCCGGTTGAAATCAAGCCCGAAATGATTG  
GCCACTATCTTGCTGAATTCTCTATTCATACAAGCCAGTGAAGCACGGAAG

*Cyrtandra longifolia*

AGATGATCGGGAGTGTCTTGGAGTATAACAATGGCAAGACCTTCAATCCGGTTGAAATCAAGCCCGAAATGATTG  
GCCACTATCTTGCTGAATTCTCTATTCATAcAAGCCAGTGAAGCACGGAAG

*Cyrtandra cordifolia*

AGATGATCGGGAGTGTCTTGGAGTATAACAATGGCAAGACCTTCAATCCGGTTGAAATCAAGCCCGAAATGATCG  
GCCACTATCTTGCTGAATTCTCTATTCATACAAGCCAGTGAAGCACGGAAG

*Cyrtandra paludosa*

AGATGATCGGGAGTGTCTTGGAGTATAACAATGGCAAGACCTTCAATCCGGTTGAAATCAAGCCCGAAATGATTG  
GCCACTATCTTGCTGAATTCTCTATTCATACAAGCCAGTGAAGCACGGAAG

*Cyrtandra platyphylla* var. *parviflora*

AGATGATCGGGAGTGTCTTGGAGTATAACAATGGCAAGACCTTCAATCCGGTTGAAATCAAGCCCGAAATGATCG  
GCCACTATCTTGCTGAATTCTCTATTCATATAAGCCAGTGAAGCACGGAAG