Molecular Phylogenetics and Evolution 69 (2013) 293–298

Contents lists available at SciVerse ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Short Communication

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

Yohan Pillon [a,*], Jennifer B. Johansen [a], Tomoko Sakishima [a], Eric H. Roalson [b], Donald K. Price [a], Elizabeth A. Stacy [a]

[a] Tropical Conservation Biology and Environmental Science Program, University of Hawai'i at Hilo, 200 West Kawili Street, Hilo, HI 96720, USA
[b] School of Biological Sciences, Washington State University, 339 Abelson Hall, Pullman, WA 99164-4236, USA

ARTICLE INFO

Article history:
Received 20 January 2013
Revised 4 May 2013
Accepted 6 May 2013
Available online 15 May 2013

Keywords:
Hawai'i
Network
Next-generation sequencing
Nuclear genes
Phylogenomics
Ribosomal DNA

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.

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 the shortcomings associated with their uniparental mode of inheritance. 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

* Corresponding author.
E-mail address: pillon@hawaii.edu (Y. Pillon).

1055-7903/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ympev.2013.05.003

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-coalescence 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 species-rich 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 understory 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 on the archipelago 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 (Möller et al., 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). By developing such markers for *Hawaiian Cyrtandra*, we aimed to evaluate how phylogenomics can help to elucidate relationships within a species complex and how much data are 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)] 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). Our preliminary works on two of 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

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
Cyrt1	At3g16640	TCTP, Translationally Controlled Tumor Protein	cagctgaaggatgcgaagag	tcgttcaaaccatggccga	**80**-89-**161**-457-**105**	58	1′
Cyrt2	At2g18020	EMB2296, Embryo Defective 2296/ Ribosomal Protein L2	cgtgttcaagtctcacacgca	tcaccagaagctctagcga	**354**	61	1′
Cyrt3	At1g43170	RPL3A, cytoplasmic ribosomal protein	tatcggtgtcacgaagggt	ttgacgaagggtaaccacc	**249**-610-**133**	61	1′30
Cyrt4	At4g13940	SAHH, S-Adenosyl-L-homocysteine hydrolase	ctcttcaagccctcatgga	gaagctagggtgtccggta	**329**	61	1′
Cyrt5	At3g47470	CAB4, Chlorophyll binding protein	gactagcagaagacccgga	actcgatcacgaagagggtcg	**196**	58	1′
Cyrt6	At1g72370	40S ribosomal protein	ggaaaacgtgggagaagct	cctaggttcgctgaatgat	**206**	63	1′
Cyrt7	At1g49140	NADH-ubiquinone oxydoreductase	ttcgatccggaaaacccgt	tcgagatactggtggacca	**185**	58	1′
Cyrt8	At5g56000	HSP81.4 Heat Shock Protein 81.4	cgttttgaaagcctgacgga	caaactgcccaatcatgct	**127**-84-**98**	61	1′
Cyrt9	At1g04270	RPS15, cytosolic ribosomal protein 15	agatgatcgggagtgtcct	cttccgtgcttcactggct	**127**	61	1′
Cyrt10	At5g60360	SAG2, Senescence associated thiol protease	tgtggatcctgctggacat	aggcttgagagggtaaacc	**20**-123-**137**	63	1′
Cyrt11	At2g24200	LAP1, cytosol leucyl aminopeptidase	aagactggagcaggatgta	gcatcagctagggtgagtctt	**109**-192-**101**-700+-**41**	61	1′30
Cyrt12	At4g14880	Cytosolic O-acetylserine(thiol)lyase	gaaaatcctgccaacccaa	tccaaaacaccagggacga	**19**-220+-**138**-125-**52**-740+-**54**	61	1′30
Cyrt13	At5g50920	CLPC homolog, involved in protein transport	cgaagaagctatgatccggct	gatcaaagcctatgcgtcg	**333**	63	1′
Cyrt14	At5g61170	Ribosomal protein S19	tacgccgcccatctcaagc	ttcctcttgctcccgccat	**28**-600+-**92**-360+-**83**	61	1′30
Cyrt15	At1g14320	RPL10, Ribosomal Protein 10	tccaaaatcacggtactgc	caaagcctcctgagcatgg	**408**	61	1′
Cyrt16	At2g16600	ROC3, Cytosolic Cyclophilin/ peptidyl-prolyl cis–trans isomerase	gccggatcgtgatggagct	agaatacctggtccagtgtgc	**264**	61	1′
Cyrt17	At3g09200	RPL10, Ribosomal Protein 10	aggtgacaaagttggttcctc	atctgcttgtgggaaagag	**300**	61	1′
Cyrt18	At3g19820	DWARF1, Brassinosteroid biosynthetic protein	catgctgagacagcattgaa	gaggaacgagcatatcctg	**250**	61	1′

lengths in heterozygotes, which cannot be sequenced without cloning. Therefore, avoiding introns 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.

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 μL of H_2O, 4 μL of Gotaq 5 × Buffer (PROMEGA), 2 μL of $MgCl_2$ 25 mM, 0.4 μL of dNTP 1.25 μM, 0.2 μL of each primer 10 μM, 0.1 μL of GoTaq Flexi DNA polymerase 5u/μL (PROMEGA) and 0.8 μL 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 30 s 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.

2.3. 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 the combination of loci with multiple alleles per accession (with the exeption of *BEAST, Heled and Drummond, 2010, see below) coding ambiguous sites was more appropriate than phasing alleles, and cloning was not done. DNA sequences were aligned by hand in MEGA5 (Tamura et al., 2011). The best-fit 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).

2.4. 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 and 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 α (*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 heterozygous accessions were separated (phased) by eye, using sequences from other accessions (unpublished data) as references. 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 POFAD (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 analysis programs that allows the combination of 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), and 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 and 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-tnH* 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 (Fig. 1).

A network analysis of 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*,

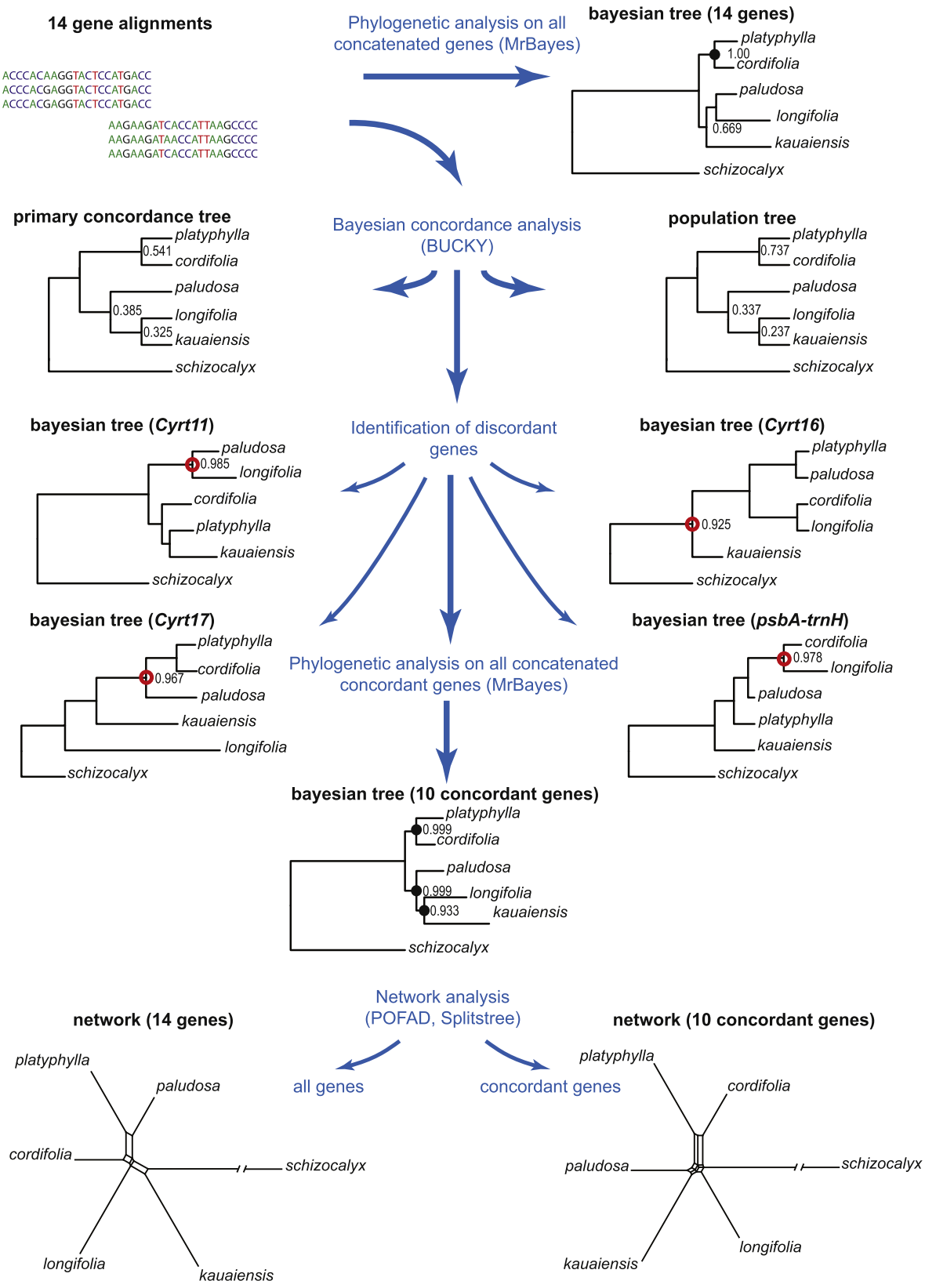

Fig. 1. Step-by-step procedure from gene alignments to the concordance tree and the identification of discordant genes. Numbers at nodes are posterior probabilities (Bayesian trees), sample concordance factors (primary concordance tree) or estimated coalescent units (population tree). Open red circles indicate supported incongruent nodes; solid black circles are supported nodes (posterior probabilities >0.95) congruent with the primary concordance tree. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

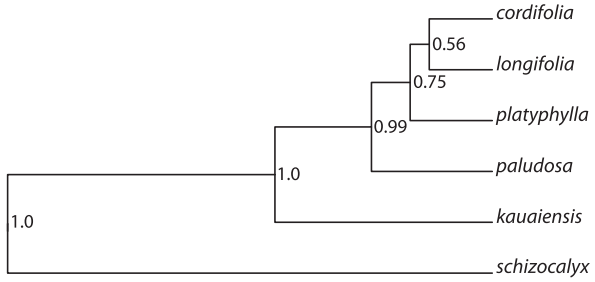

Fig. 2. Phylogenetic tree inferred with *BEAST using phased alleles. Numbers at nodes are posterior probabilities.

and one including *C. kauaiensis*, *C. longifolia* and *C. paludosa*; however, the relationship within the second group remained unclear (Fig. 1). The *BEAST phylogenetic tree yielded a tree that was different from the concordance tree and with a single well supported internal node (Fig. 2).

4. Discussion

The lack of congruence among genes observed in this study strongly suggests 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 hybrids, but: (1) our network analyses do not provide evidence of a hybrid origin of any of our taxa, and (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. The *BEAST analysis did not prove to be robust to this coalescence problem as it recovered a topology that is different from the primary concordance tree; however, it may 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 among 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.

Acknowledgments

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, S. Geib and L. Barrabé for help and advice with data analyses and two anonymous reviewers for their comments on an earlier version of this manuscript. Some phylogenetic analyses were run on the university of Oslo Bioportal. 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.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2013.05.003.

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 and 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., Kloepper, 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.
Kosakovsky Pond, 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.