PRIMERS FOR LOW-COPY NUCLEAR GENES IN THE HAWAIIAN ENDEMIC CLERMONTIA (CAMPA JULACEAE) AND CROSS-AMPLIFICATION IN LOBELIOIDEAE

YOHAN PILLON2,5, JENNIFER JOHANSEN2, TOMOKO SAKISHIMA2, SRIKAR CHAMALA3,4, W. BRAD BARBAZUK3,4, AND ELIZABETH A. STACY2

2Tropical Conservation Biology and Environmental Science Program, University of Hawai‘i at Hilo, 200 West Kawili Street, Hilo, Hawaii 96720 USA; 3Department of Biology, University of Florida, Gainesville, Florida 32611 USA; and 4Genetics Institute, University of Florida, Gainesville, Florida 32610 USA

- Premise of the study: Primers were developed to amplify 12 intron-less, low-copy nuclear genes in the Hawaiian genus Clermontia (Campanulaceae), a suspected tetraploid.
- Methods and Results: Data from a pooled 454 titanium run of the partial transcriptomes of seven Clermontia species were used to identify the loci of interest. Most loci were amplified and sequenced directly with success in a representative selection of lobeliads even though several of these loci turned out to be duplicated. Levels of variation were comparable to those observed in commonly used plastid and ribosomal markers.
- Conclusions: We found evidence of a genome duplication that likely predates the diversification of the Hawaiian lobeliads. Some genes nevertheless appear to be single-copy and should be useful for phylogenetic studies of Clermontia or the entire Lobelioideae subfamily.

Key words: Clermontia; Hawai‘i; Lobelioideae; low-copy nuclear genes; next-generation sequencing; polyploidy.

Although phylogenetic studies of plants rely heavily on plastid and nuclear ribosomal loci, the limitations of these loci are well known. Plastid loci are uniparentally inherited and susceptible to chloroplast capture (Rieseberg and Soltis, 1991). Furthermore, their low variation is particularly problematic for investigations of recent radiations such as those in the Hawaiian Islands, and indeed these markers have provided only limited resolution there. Nuclear ribosomal loci are subject to complex concerted evolution (Alvarez and Wendel, 2003), which can be incomplete so that multiple ancient alleles are maintained within a single individual, or rapid such that traces of hybridization are quickly erased. As a result, both plastid and ribosomal markers provide only incomplete information when hybridization is common.

The genus Clermontia Gaudich. (Campanulaceae, subfamily Lobelioideae) comprises 22 species endemic to Hawai‘i, 13 of which are on the IUCN Red List of Threatened Species. Species identification in the field is often difficult, particularly in the absence of flowers, and apparent hybrids can be common. Clermontia and other Hawaiian lobeliads (six genera total) form a monophyletic clade that represents the largest plant radiation in Hawai‘i (Givnish et al., 2009). The members of this clade, as well as the members of clade 4 of Antonelli (2008), in which the Hawaiian clade is nested, are suspected paleotetraploids (Lammers, 1988). To find genetic markers that will be useful for phylogenetics and DNA barcoding within Clermontia, we used 454 data of partial cDNA libraries to design primers for single-exon nuclear loci in Clermontia and then tested their cross-amplification in Campanulaceae.

METHODS AND RESULTS

We obtained a pooled, partial transcriptome library from leaf and floral buds (fixed in the field in RNAlater [QIAGEN, Gaithersburg, Maryland, USA] and stored at −80°C) of seven taxa: Clermontia arborescens (H. Mann) Hillebr., Clermontia cerniioides (Gaudich.) A. Heller, Clermontia fauriei Rock, Clermontia parviflora Gaudich. ex A. Gray, and Clermontia peleana Rock. RNA isolation, cDNA synthesis, and 454 sequencing were done at the University of Arizona Genetics Core Laboratory. The 454 run provided 1.4 million reads with an average length of 395 bp. 454 adapters, ribosomal RNA, and low-quality and low-complexity sequences were removed/trimmed using SeqClean (http://compbio.dfci.harvard.edu/tgi/software/), and each taxon was assembled separately by the TGI Clustering tools (TGICL; Pertea et al., 2003), using default settings. We conducted BLAST searches of the 400 most highly expressed genes in Arabidopsis (C. Fizames, personal communication) against our data in CLC DNA Workbench (CLC bio, Aarhus, Denmark) to identify a set of genes with high coverage within each of all or most of the species. We selected loci (generally only a small portion of a gene) that comprised a single, long exon (200 bp) with matches in multiple species, and designed primers with FastPCR (PrimerDigital Ltd., Helsinki, Finland; http://www.primerdigital.com/fastpcr.html) for their amplification using default settings. The presence of introns was tested by comparison with genomic and cDNA sequences in the Arabidopsis Information Resource database (www.arabidopsis.org). Avoiding introns allowed the direct sequencing of accesses even in the case of gene duplications; introns often contained indels.

doi:10.3732/apps.1200450

Applications in Plant Sciences 2013 1(6): 1200450; http://www.bioone.org/loi/apps © 2013 Botanical Society of America
which often resulted in alleles of different lengths in heterozygotes or among copies of duplicated genes. Twelve exon regions were identified (Table 1, Appendix 1) and were tested on seven accessions: *C. fauriei* (the earliest diverging species within the genus), *C. arborescens*, *C. kakeana*, *Cyanea asplenifolia* Hillebr. (*Cyanea* is a Hawaiian endemic genus and putative sister group of *Clermontia*; Givnish et al., 2009), *Hippobroma longiflora* (L.) G. Don (belonging to a different major clade of Lobelioideae and a likely tetraploid; Antonelli, 2008), *Lobelia erinus* L. (one of the earliest diverging Lobelioideae; Antonelli, 2008), and *Campanula persicifolia* L. (Campanuloideae). 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, Düren, Germany). The nuclear regions were amplified using the following mix: 12.3 μL of H2O, 4 μL of GoTaq 5× Buffer (Promega Corporation, Madison, Wisconsin, USA), 2 μL of MgCl2, 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 5 U/μL (Promega Corporation), 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 63°C, and 1 min at 72°C, and a final extension of 5 min at 72°C. PCR products were sequenced directly at the Core Genetics Laboratory at the University of Hawai‘i Hilo. The identity of each amplified gene was validated through BLAST or BLASTx (*Clerm4, Clerm10*) searches in GenBank. In every case, the 10 best matches (identities >80%) were either the same gene from a different species or a gene that was not yet annotated.

All 12 nuclear regions were successfully amplified and sequenced in *Clermontia*, *Cyanea*, and *Hippobroma*; a single gene was not amplified in *Lobelia erinus*, and three could not be sequenced in *Campanula* (Table 1). A high number of ambiguous bases were found consistently in the forward and reverse sequences of some accessions, suggesting the presence of multiple gene copies. In five genes (*Clerm1, Clerm6, Clerm10, Clerm11, Clerm12*), ambiguous sites were identical across the three *Clermontia* species and *Cyanea* but absent in the other genera (example in Fig. 1). To confirm the hypothesis of gene duplication, we separated alleles computationally from the direct sequences using PHASE (default settings) within the software DnaSP (Librado and Rozas, 2009), and built a neighboring joining tree of the alleles in SeaView (Gouy et al., 2010) with default settings. In each of these five cases, we recovered two clades, each comprising a single allele from each of the four Hawaiian lobelia species examined (example in Appendix S1). This pattern strongly suggests a genome duplication event that predates the divergence of *Clermontia* and *Cyanea*. *Clerm5* was duplicated in *Clermontia* but apparently not in *Cyanea*. This is the only gene for which direct sequences turned out to be difficult to read due to the divergence of the two gene copies, which may be due to the presence of an intron not present in *Arabidopsis*. Five genes (*Clerm2, Clerm4, Clerm7, Clerm8, and Clerm9*) behaved as single-copy genes in *Clermontia*. No recombination was detected in those genes using genetic algorithms for recombination detection (GARD; Kosakovsky Pond et al., 2004).

### Table 1.

| Gene   | Putative product                                                                 | Forward primer sequence (5′→3′) | Reverse primer sequence (5′→3′) | Length (bp) | Failed sequences | Duplicate in Hawai‘i? | Failed amplification a |
|--------|----------------------------------------------------------------------------------|---------------------------------|---------------------------------|-------------|------------------|-----------------------|------------------------|
| *Clerm1* | RPL10 (ribosomal protein)                                                        | CTACCAGTCTGTTACCGT              | CCCTCATACCAGTCTGGAGC            | 248         | yes              | yes                   | yes                    |
| *Clerm2* | Lhca*3 (chlorophyll-binding protein)                                             | TTGGATTCGACCCACTTGG             | ACTTAAGGCTGGTCAGCAC             | 594         | yes              | no                    | no                     |
| *Clerm4* | TIP2 (tonoplast intrinsic protein)                                               | GATGACATTCGGGCTGGTA             | CAGTAGACCCAGTGGTTGG             | 210         | yes              | no                    | no                     |
| *Clerm5* | RPL32e (ribosomal protein)                                                       | TTTGAGAAGCCCTCTGCAA             | TTTGCCTACGCAACATGTCAA           | 382         | no               | yes                   | yes                    |
| *Clerm6* | EIF4-1 (translation initiation factor)                                           | TCTTATCGAACCAACAAGTGG           | TCTACACGACGGATTTCGAGG           | 257         | no               | no                    | no                     |
| *Clerm7* | ATCYS3-A (acetylserine(thiol)lyase)                                              | AGAGCCATCGCAATTGGCTG            | CATGTCCTTAGCCGACTTCG           | 339         | no               | yes                   | yes                    |
| *Clerm8* | CESA3 (cellulose synthase)                                                       | GACCTGCTAGGTGTTACCGT            | CCCTCATACCAGTCTGGAGC           | 339         | yes              | yes                   | yes                    |
| *Clerm9* | SAHH2 (S-adenosyl- L -homocysteine hydrolase)                                    | TCTACACGACGGATTTCGAGG           | TGAGGCGTCTACCAAGCTC             | 212         | yes              | yes                   | yes                    |
| *Clerm10* | PGK (phosphoglycerate kinase)                                                  | GACCTGCTAGGTGTTACCGT            | CCCTCATACCAGTCTGGAGC           | 339         | yes              | yes                   | yes                    |
| *Clerm11* | RPL32e (ribosomal protein)                                                       | TTTGAGAAGCCCTCTGCAA             | TTTGCCTACGCAACATGTCAA           | 257         | no               | no                    | no                     |
| *Clerm12* | RPL10 (ribosomal protein)                                                        | CTACCAGTCTGTTACCGT              | CCCTCATACCAGTCTGGAGC            | 248         | yes              | yes                   | yes                    |

*Missing sequences are due to failed amplification (Clerm4) or amplification of a different gene (Clerm2).*

---

http://www.bioone.org/loi/apps

---

[Clermontia fauriei](http://www.bioone.org/loi/apps)

Cyanea asplenifolia

Hippobroma longiflora

**Fig. 1.** Example of electropherograms for *Clerm12* comparing *Clermontia fauriei*, *Cyanea asplenifolia*, and *Hippobroma longiflora*. Only the forward sequence is shown; asterisks indicate ambiguous sites. The matching positions of multiple ambiguous sites in *Clermontia* and *Cyanea* implicate an ancestral gene duplication prior to the divergence of these two genera rather than heterozygosity. *Note:* black = G; green = A; red = T; blue = C.
Table 2. Comparison of variation of the five apparently nonduplicated nuclear genes with three plastid and two nuclear ribosomal loci.a

| Characteristic | Clerm2 | Clerm4 | Clerm7 | Clerm8 | Clerm9 | rbcL | matK | psbA-trnH | ITS | ETS |
|----------------|--------|--------|--------|--------|--------|------|------|-----------|------|-----|
| Length (bp)    | 594    | 210    | 257    | 264    | 608    | 599  | 889  | 460–462   | 874–877 | 594–600 |
| No. of variable sites | 12 | 5 | 4 | 5 | 5 | 10 | 16 | 9 | 17 | 12 |
| Percent variable sites | 2.0 | 2.4 | 1.6 | 1.9 | 0.8 | 1.7 | 1.8 | 1.9 | 1.9 | 2.0 |
| 1st position (nonsynonymous) | 2 (2) | 4 (1) | 0 | 1 (1) | 0 | 3 (1) | 9 (8) | NA | NA | NA |
| 2nd position (nonsynonymous) | 2 (2) | 0 | 0 | 1 (1) | 0 | 0 | 3 (3) | NA | NA | NA |
| 3rd position (nonsynonymous) | 8 (2) | 4 (0) | 4 (0) | 3 (0) | 5 (0) | 7 (3) | 4 (1) | NA | NA | NA |
| Indels          | 0      | 0      | 0      | 0      | 0      | 0    | 2    | 1    | 2    | |

Note: NA = not applicable.

a For each gene, variation was measured on a sample that included Clermontia fauriei, C. kakeana, C. arborescens, and Cyanea asplenifolia.

2006; http://www.datamonkey.org/). The percentage of variable sites within each of these genes was comparable to those of the plastid loci rbcL, matK, and psbA-trnH and the nuclear ribosomal ITS and ETS (Table 2). Genotyping of a broader taxonomic sample of Clermontia revealed a much greater number of variants at these newly described nuclear genes and a different pattern of evolution compared to plastid genes (Pillon et al., 2013).

CONCLUSIONS

The selection of intron-less regions proved successful for the amplification and direct sequencing of several nuclear loci and the detection of duplicated genes. The large number of gene duplications shared between Clermontia and Cyanea strongly supports the hypothesis of whole-genome duplication that predates the diversification of the lobeliads in Hawaii'i. Whole genome duplication has similarly been demonstrated in Hawaiian silverswords (Barrier et al., 1999). Despite the genome duplication, we nevertheless identified a number of apparently single-copy genes, whether due to the loss of one copy in each case or the selectivity of our primers for one copy. Geographic and taxonomic patterns of variations of two of these markers within Clermontia are examined in a study of their potential use as DNA barcodes (Pillon et al., 2013).

LITERATURE CITED

ÁLVAREZ, I., and J. F. WENDEL. 2003. Ribosomal ITS sequences and plant phylogenetic inference. Molecular Phylogenetics and Evolution 29: 417–434.

ANTONELLI, A. 2008. Higher level phylogeny and evolutionary trends in Campanulaceae subfam. Lobelineae: Molecular signal overshadows morphology. Molecular Phylogenetics and Evolution 46: 1–18.

Appendix 1. Location information, voucher specimens, and GenBank accession numbers for Clerm1, Clerm2, Clerm3, Clerm4, Clerm5, Clerm6, Clerm7, Clerm8, Clerm9, Clerm10, Clerm11, Clerm12, ITS, ITS, matK, psbA-trnH, and rbcL. For duplicated genes, only cDNA sequences were submitted when available. Voucher specimens were deposited at the Herbarium of the University of Hawai'i (HAW). The vouchers for Clermontia arborescens and Campanula persicifolia have been lost, and vouchers were not collected for Cyanea asplenifolia because it is an endangered plant. — signifies that no sequence is available for the particular locus for that accession.

http://www.bioone.org/loi/apps