Primers for Low-Copy Nuclear Genes in the Hawaiian Endemic Clermontia (Campanulaceae) and Cross-Amplification in Lobelioideae

Authors: Pillon, Yohan, Johansen, Jennifer, Sakishima, Tomoko, Chamala, Srikar, Barbazuk, W. Brad, et al.

Source: Applications in Plant Sciences, 1(6)

Published By: Botanical Society of America

URL: https://doi.org/10.3732/apps.1200450

Applications in Plant Sciences 2013 1(6): 1200450

Primer Note

Primers for low-copy nuclear genes in the Hawaiian endemic *Clermontia* (Campanulaceae) and cross-amplification in Lobelioideae[1]

Yohan Pillon[2,5], Jennifer Johansen[2], Tomoko Sakishima[2], Srikar Chamala[3,4], W. Brad Barbazuk[3,4], and Elizabeth A. Stacy[2]

[2]Tropical Conservation Biology and Environmental Science Program, University of Hawai'i at Hilo, 200 West Kawili Street, Hilo, Hawaii 96720 USA; [3]Department of Biology, University of Florida, Gainesville, Florida 32611 USA; and [4]Genetics 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 RNA*later* [QIAGEN, Gaithersburg, Maryland, USA] and stored at −80°C) of seven taxa: *Clermontia arborescens* (H. Mann) Hillebr., *C. clermontioides* (Gaudich.) A. Heller, *C. fauriei* H. Lév., *C. kakeana* Meyen, *C. kohalae* Rock, *C. parviflora* Gaudich. ex A. Gray, and *C. 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 accessions even in the case of gene duplications; introns often contained indels,

[1]Manuscript received 27 August 2012; revision accepted 27 November 2012.

The authors thank the following for facilitating the collection of plant specimens: Hawaii's Department of Land and Natural Resources–Division of Forestry and Wildlife, Maui Land and Pineapple (R. Bartlett), The Nature Conservancy (E. Naboa and P. Bily), and the Volcano Rare Plant Facility (P. Moriyasu and J. Enoka). The authors thank H. Issar and A. Veillet for technical assistance, and M. Lebrun and C. Fizames for information on nuclear genes. Funding was provided by the Gordon and Betty Moore Foundation.

[5]Author for correspondence: pillon@hawaii.edu

doi:10.3732/apps.1200450

Table 1. Identity of the 12 intron-less, low-copy nuclear genes identified in this study, with primer sequences, results from cross-amplification tests, and inference of putative gene duplication.[a]

Gene	Putative *Arabidopsis* homolog	Putative product	Forward primer sequence (5′–3′)	Reverse primer sequence (5′–3′)	Length (bp)	Failed sequences	Duplication in Hawai'i
Clerm1	At1g14320	RPL10 (ribosomal protein)	GACCTGCTAGGTGTTACCGT	ATCTTCTGACGACCAGGGA	484		yes
Clerm2	At1g61520	Lhca*3 (chlorophyll-binding protein)	TTGGATTCGACCCACTTGG	ACTTAAGGCTGGTCAGCAC	594	*Campanula*	no
Clerm3	At2g05070	Lhcb2.2 (chlorophyll-binding protein)	GGTGACTACGGATGGGACAC	CATCAGCCAGTCCAAGTGGGT	382		yes
Clerm4	At3g26520	TIP2 (tonoplast intrinsic protein)	GATGACATTCGGGCTGGTA	CAGTAGACCCAGTGGTTGG	210	*Campanula*	no
Clerm5	At4g18100	RPL32e (ribosomal protein)	ATTGATTCCCGGGTGAGAA	GTCACCACAACATCAAGCT	248	*Campanula*, *Lobelia*	*Clermontia*
Clerm6	At3g13920	EIF4-1 (translation initiation factor)	TTTGAGAAGCCCTCTGCAA	TTGCCTACGCAACATGTCAA	351		yes
Clerm7	At4g14880	ATCYS3-A (acetylserine(thiol)lyase)	TCTTATCGAACCAACAAGTGG	TCGGATTGGCAGGATTCTC	257		no
Clerm8	At5g05170	CESA3 (cellulose synthase)	AGAGCCATCGCAATTGGCTG	GCGAAATACCACTCTGGAGC	264		no
Clerm9	At3g23810	SAHH2 (*S*-adenosyl-L-homocysteine hydrolase)	CATGTCCTTAGCCGACTTCG	ACCATTAGCCTGCATTTGG	608		no
Clerm10	At1g79550	PGK (phosphoglycerate kinase)	GTCAAGATGGCAAATGATTGC	CCTTCAGTGGAAGCATGAGC	212		yes
Clerm11	At1g66580	SAG24 (senescence-associated gene/ribosomal protein L10e)	GACCTGCTAGGTGTTACCGT	CCCTCATACCAGTCTGGAGC	339		yes
Clerm12	At3g56940	Magnesium protoporphyrin IX monomethyl ester cyclase	TCTACACGACGGATTTCGAGG	TGAGGCGTCTACCAAGCTC	264		yes

[a] Missing sequences are due to failed amplification (*Clerm5*), weak amplification (*Clerm4*), or amplification of a different gene (*Clerm2*).

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 H_2O, 4 µL of GoTaq 5× Buffer (Promega Corporation, Madison, Wisconsin, USA), 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 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 tBLASTx (*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 neighbor-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 lobeliad 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.,

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)	1 (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	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 Hawai'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. Lobelioideae: Molecular signal overshadows morphology. *Molecular Phylogenetics and Evolution* 46: 1–18.

Barrier, M., B. G. Baldwin, R. H. Robichaux, and M. D. Purugganan. 1999. Interspecific hybrid ancestry of a plant adaptive radiation: Allopolyploidy of the Hawaiian silversword alliance (Asteraceae) inferred from floral homeotic gene duplications. *Molecular Biology and Evolution* 16: 1105–1113.

Givnish, T. J., K. C. William, A. R. Mast, T. B. Paterson, T. J. Theim, A. L. Hipp, J. M. Henss, et al. 2009. Origin, adaptive radiation and diversification of the Hawaiian lobeliads (Asterales: Campanulaceae). *Proceedings of the Royal Society of London, Series B, Biological Sciences* 276: 407–416.

Gouy, M., S. Guindon, and O. Gascuel. 2010. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution* 27: 221–224.

Kosakovsky Pond, S. L., D. Posada, M. B. Gravenor, C. H. Woelk, and S. D. W. Frost. 2006. GARD: A genetic algorithm for recombination detection. *Bioinformatics* 22: 3096–3098.

Lammers, T. G. 1988. Chromosome numbers and their systematic implications in Hawaiian Lobelioideae (Campanulaceae). *American Journal of Botany* 75: 1130–1134.

Librado, P., and J. Rozas. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics (Oxford, England)* 25: 1451–1452.

Pertea, G., X. Q. Huang, F. Liang, V. Antonescu, R. Sultana, S. Karamycheva, Y. Lee, et al. 2003. TGIR Gene Indices clustering tools (TGICL): A software system for fast clustering of large EST datasets. *Bioinformatics (Oxford, England)* 19: 651–652.

Pillon, Y., J. Johansen, T. Sakishima, S. Chamala, W. B. Barbazuk, E. H. Roalson, and E. A. Stacy. 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.

Rieseberg, L. H., and D. E. Soltis. 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. *Evolutionary Trends in Plants* 5: 65–84.

Appendix 1. Location information, voucher specimens, and GenBank accession numbers for *Clerm1*, *Clerm2*, *Clerm3*, *Clerm4*, *Clerm5*, *Clerm6*, *Clerm7*, *Clerm8*, *Clerm9*, *Clerm10*, *Clerm11*, *Clerm12*, *ETS*, *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.

***Clermontia arborescens* (H. Mann) Hillebr. subsp. *waihiae* (Wawra) Lammers:** Maui, Pu'u Kukui, 20°55′26″N, 156°36′11″W, no voucher: JX500281, JX500282, JX500287, JX500291, JX500293, JX500298, JX500300, JX500305, JX500312, JX500318, JX500324, JX500327, —, JX500333, JX500337, JX500341, JX500345, JX500349.

***Clermontia fauriei* H. Lév.:** Kaua'i, Alaka'i swamp trail, 22°9′12″N, 159°37′52″W, *Johansen 1*: JX500283, JX500288, —, JX500294, —, JX500301, JX500306, JX500313, JX500319, —, —, JX500330, JX500334, JX500338, JX500342, JX500346, JX500349.

***Clermontia kakeana* Meyen:** Maui, Pu'u Kukui, 22°56′3″N, 156°36′53″W, *Johansen 5*: —, JX500286, —, JX500292, —, —, JX500304, JX500311, JX500317, —, —, —, JX500332, JX500336, JX500340, JX500344, JX500348.

***Cyanea asplenifolia* Hillebr.:** Maui, no voucher (Volcano Rare Plant Facility): —, JX500289, —, JX500295, JX500299, —, JX500307, JX500314, JX500320, —, —, —, JX500335, JX500339, JX500343, JX500347, JX500351.

***Hippobroma longiflora* (L.) G. Don:** Hawai'i (introduced), *Pillon 1426*: —, —, —, JX500296, —, —, JX500308, JX500315, —, JX500324, —, —, —, —, —, —, —.

***Lobelia erinus* L.:** Hawai'i (cultivated), *Pillon 1427*: JX500284, JX500290, —, JX500297, —, JX500302, JX500309, JX500316, JX500321, JX500325, JX500328, —, —, —, —, —, —.

***Campanula persicifolia* L.:** France (cultivated), no voucher: JX500285, —, —, —, —, JX500303, JX500310, —, JX500322, JX500326, JX500329, JX500331, —, —, —, —, —.