 doi: 10.1111/j.1471-8286.2007.01928.x

PRIMER NOTE

Novel nuclear intron-spanning primers for Arecaceae evolutionary biology

CHRISTINE D. BACON,* F. ALEX FELTUS,† ANDREW H. PATERSON† and C. DONOVAN BAILEY*
Department of Biology, MSC 3AF, New Mexico State University, PO Box 30001, Las Cruces, NM 88003, USA, †Plant Genome Mapping Laboratory, University of Georgia, 111 Riverbend Road, Athens, GA 30602, USA

Abstract

In this study, 96 nuclear 'conserved intron-scanning primers' were screened across subfamilies the Arecaceae (palms) for potential use in research focused on palm evolutionary biology. Primers were evaluated based on their ability to amplify single polymerase chain reaction products in Arecaceae, the clarity of sequencing reads, and the interspecific variability observed. Ultimately, the results suggest that: (i) seven of the loci are likely to be suitable when comparing non-Arecaceae outgroups and Arecaceae ingroups; (ii) seven loci may be of use when comparing subfamilies of Arecaceae; and (iii) four of the loci may be of use when comparing closely related genera.

Keywords: Arecaceae, CISP, conserved intron-scanning primers, intron, nuclear gene, Palmae

Received 23 March 2007; revision accepted 26 June 2007

The palm family (Arecaceae) represents a large and economically important group of monocots whose systematic relationships remain inadequately understood. Obstacles to the use of DNA sequences in studies of palm evolutionary biology have been attributed to low rates of substitution in palm chloroplast genomes (Wilson *et al*. 1990; Clegg & Zurawski 1992; Gaut *et al*. 1992, 1996) and incomplete concerted evolution resulting in intragenomic heterogeneity in nrDNA internal transcribed spacer sequence data (Baker *et al*. 2000; Lewis & Namoff 2006).

In this study, monocot-specific conserved intron-scanning primers (CISPs; Feltus *et al*. 2006; Lohithaswa *et al*. 2007) were used to amplify loci in a broad sample of Arecaceae to screen for variable nuclear encoded regions. The primers were originally developed following a general 'comparative anchor tagged sequences approach' (reviewed by Hughes *et al*. 2006). Specifically, comparisons of 2074 annotated *Oryza* (rice) bacterial artificial chromosome (BAC) sequences to 19 719 *Allium* (onion) and 15 661 *Musa* (banana) expressed sequence tags (ESTs; Lohithaswa *et al.* 2005) facilitated the development of 4868 20-bp candidate primers anchored in exons that span introns in low-copy regions of the rice genome (Lohithaswa *et al*. 2007). Ninety-six of these CISP primers were randomly chosen for refinement. In this way, 'universal' monocot-specific primers were developed that minimize false inferences of orthology while attempting to exploit the potentially divergent nature of some introns.

DNA samples were selected to enable comparisons of sequence variation between (i) non-Arecaceae outgroups and Arecaceae ingroups, (ii) subfamilies of Arecaceae, and (iii) closely related members of the tribe Chamaedoreeae. The sampling thus included *Oryza sativa*, *Allium cepa*, and *Tradescantia virginiana* as outgroups and positive controls, a representative from each subfamily of Arecaceae (Asmussen *et al*. 2006), and four representatives of the tribe Chamaedoreeae (Table S1, Supplementary material). The results derived from polymerase chain reaction (PCR) amplifications, sequencing reads, and sequence comparisons allowed us to estimate the potential utility of each primer set following a suite of previously suggested criteria (e.g. Strand *et al*. 1997; Bailey *et al*. 2004; Small *et al*. 2004; Syring *et al*. 2005). These included: (i) successful amplification in at least four accessions (for comparative purposes); (ii) amplification of a single band (to minimize potential problems with mistaken orthology); (iii) intron length (such that longer noncoding regions were given priority); (iv) high quality sequencing reads with limited evidence of polymorphism (to reduce issues of mistaken orthology and the need to develop more specific primers); and (v) ultimate levels of variability observed.

Correspondence: Christine D. Bacon, Fax: (970) 491 0649; E-mail: cbacon@simla.colostate.edu

Table 1 Primer sequence information

No.	Name	Sequence	Tm	Rice chromosome
1	BRSC2_001_FOR/REV	TTTGCTGGTGATGATGCACC/CCACCAATCCAGACACTGTA	57.3	2
2	BRSC3_011_FOR/REV	AAGAGGTCAGTTGCTGAGGA/ACCACGGTACTTCTTGCCAG	57.3	3
3	ORSC7_004_FOR/REV	CCATATTCATCAGCAAGTGC/ATCTGGGATACAGCTGGTCA	55.3	7
4	ORSC2_007_FOR/REV	GGAGATCAACGTCTTCTTGT/GCATACTCAGGAGCACAATA	55.3	2
5	ORSC8_001_FOR/REV	CACAGAGAAGTTAAGTGCCA/ATCAAGGAGTACCGTGGCAA	55.3	8
8	ORSC2–009_FOR/REV	GACATCTGCTTCCGCACTCT/AGGCCTTCCTCCTGAACATG	59.4	2
11	BRSC1_008_FOR/REV	GCTGAAATCACTGACAAGCT/GATGTAGCACAAGTGCTCTG	55.3	1
12	BRSC10_002_FOR/REV	TTCGAGCTCAACGACGTCTT/CATCCCGAACAGCGTAATCC	57.3	10
13	ORSC10_003_FOR/REV	AGATGATTCTCCAAGCTTCC/CCAGAAGGIIIICAGGGTCA	55.3	10
17	BRSC4_006_FOR/REV	GGTTGGTCACCAAAGAMGG/GTGCTGGAACAACAGCTGCT	57.3	4
18	BRSC3_004_FOR/REV	AGGATGAGCAAGGAGATCAC/CATTTCCTGTGCACAATGGA	57.3	3
19	ORSC3_004_FOR/REV	TACATGAGGCAGCAGGCCAT/ACTGGTGCTGGGAAGCATGT	59.4	3
20	ORSC5_005_FOR/REV	ACAGTGTGGCTCACACCATC/GCTGGTGATGATGCACCAAG	59.4	5
21	ORSC7_006_FOR/REV	GATCTTCATCTTGATCCTGC/CTTGATCAACCATGGCAAGG	55.3	7
22	BRSC3_007 FOR/REV	AGTGAGGATTACCCTAGCAA/AGGCTGATCAAGCAAGTCCT	55.3	3

PCRs for the 96 primer pairs screened in this study included 1.5 U of *Taq*, 165 μM of each dNTP, 1× PCR buffer, 0.5 μM of each primer, and 50 ng of gDNA. The generalized PCR profile applied to all loci included a 3-min, 94 °C denaturation followed by 35 cycles of 30 s at 94 °C, 30 s at 55 °C, 1 min at 72 °C, followed by a final extension of 7 min at 72 °C. PuReTaq Ready-To-Go PCR beads (Amersham Biosciences) were used in secondary attempts to amply samples that failed using standard *Taq* and buffers. PCR products were cleaned using QIAquick PCR purification Kits (QIAGEN). For a few primers that exhibited two clean bands in a subset of accessions amplified, PCR products were cleaned using QIAquick Gel Extraction Kits. BLASTN searches (Altschul *et al*. 1990) against The Institute for Genome Research (TIGR) Rice Genome Annotation Project database (http://tigrblast.tigr.org/euk-blast) confirmed the respective chromosomal location in rice and the size of introns (Table 1 and Table S2, Supplementary material).

Forty-five of the 96 primer pairs amplified products from at least four accessions. Twenty of these 45 amplified multiple bands and were abandoned from further screening. Fifteen (Table 1) of the remaining primer pairs produced useable sequence data across a minimum of two ingroups and one outgroup sample resulting in alignable data sets (Table 2). No attempts to resolve amplification or sequencing failures were performed due to the scope of the study and additional conditions and optimization is recommended.

Eight of the 15 primer sets appeared to have the greatest potential for the studies involving both non-Arecaceae outgroups and subfamilial representatives of the Arecaceae. These included primers 1, 2, 4, 5, 8, 12, 17, and 18, which amplified loci in a minimum of five taxa, generated usable alignments of at least 260 bp, and which had between 4.5% and 23% ($\bar{x}$ = 13.3%) parsimony-informative characters (Table 2). In addition, topologies supported by these loci resolved the Arecaceae as monophyletic and supported relationships within the family (Fig. S1, Supplementary material). Only the locus amplified by primer set 8 supported topologies suggestive of problems with obvious gene-tree/species-tree issues (e.g. Pamilo & Nei 1988; Doyle 1992; Wendel & Doyle 1998). Of the primers that looked particularly useful, 4, 5, 12, 17, and 18 generated well-resolved trees with high bootstrap support values (Fig. S1). Furthermore, these five loci correspond to independent linkage groups in rice (Table 1), suggesting that they may also represent independent loci in the Arecaceae.

The comparison of data from Arecaceae-only matrices included information from seven primer sets (2, 4, 5, 11, 12, 13, and 21) that produced data from at least five of the eight representatives of Arecaceae, alignments with at least 168 aligned bases, and included 5.4–13.6% ($\bar{x}$ = 6.6%) parsimony-informative sites (Table 2). Gene trees supported by these data suggest that the loci amplified by primer sets from 4, 5, 11, 12, 13, 21 can produce resolved and supported subfamilial-level phylogenies (Fig. S2, Supplementary material).

Within the tribe Chamaedoreeae, primers 4, 5, 12, and 21 amplified useable sequence data from all four representatives and each included variable sites across the alignment (Table 2). For these intratribal comparisons, the percentage of parsimony-informative sites ranged from 0.5 to 1.1% ($\bar{x}$ = 0.86%), representing a paucity of potential characters among the species of *Chamaedorea*. However, the variable sites ranged from 6.6 to 13.6% ($\bar{x}$ = 8.5%), identifying that the variation occurred between genera and not between the three species of *Chamaedorea*. However, increased intratribal sampling may boost the proportion of parsimony-informative sites and the utility of the primers at interspecific

Table 2 Alignment and sequence divergence data including all comparisons: all palms including outgroups, all palms excluding outgroups, and Chamaedoreeae

Primer no.	No. of terminals	Aligned no. of bp	Total gaps	Parsimony-informative substitutions	Variable sites	Parsimony-informative indels	Total parsimony-informative characters	Average % G-C content	Range of pairwise distribution
Sequence divergence including outgroups*									
1	10	966	12	44	456	1	45	47	0–0.17
2	10	320	8	37	231	4	41	42	0.05–0.34
3	4	744	25	17	439	0	17	37	0.11–0.47
4	9	1040	58	209	711	38	247	37	0.1–0.53
5	8	680	34	55	390	5	60	39	0–0.40
8	9	649	21	111	414	8	119	54	0.07–0.33
12	7	293	15	34	139	4	38	58	0.01–0.33
17	5	263	10	26	187	2	28	39	0–0.57
18	6	407	23	30	215	2	32	42	0.05–0.33
19	5	275	7	2	112	1	3	49	0.03–0.33
22	5	216	8	11	101	7	18	42	0.08–0.39
Sequence divergence excluding outgroups									
1	8	965	7	15	390	0	15	48	0–0.01
2	8	316	6	35	210	2	37	41	0.05–0.34
3	3	623	13	0	243	0	0	39	0.11–0.38
4	7	868	18	64	199	9	73	37	0.02–0.13
5	7	680	19	33	293	4	37	39	0–0.17
8	6	492	14	41	200	4	45	56	0.06–0.15
11	7	168	9	15	106	8	23	41	0–0.4
12	6	293	13	18	106	3	21	57	0.01–0.25
13	5	339	21	19	117	10	29	40	0.03–025
18	5	405	17	15	170	1	16	41	0.04–0.19
19	4	191	6	0	83	0	0	49	0.03–0.25
20	8	389	9	11	121	1	12	50	0–0.11
21	7	331	13	16	150	6	22	29	0–0.4
22	4	215	3	3	72	1	4	43	0.07–0.15
Sequence divergence including only Chamaedoreeae									
4	4	843	28	8	56	1	9	36	0.01–0.06
5	4	660	5	3	37	0	3	40	0–0.03
12	4	272	6	1	37	2	3	57	0–0.09
21	4	330	7	2	28	1	3	27	0–0.08

*Outgroups consisted of *Oryza sativa*, *Tradescantia virginiana*, *Allium cepa*.

levels. At present, these markers appear variable enough for studies involving closely related genera, but not closely related species of *Chamaedorea*. The topologies recovered from these matrices are all congruent with one another (unpublished data) and are consistent with previously published relationships within the tribe and within the *Chamaedorea* subgenus *Stephanostachys* (Hodel 1992; Uhl & Dransfield 1987; Thomas *et al*. 2006).

Acknowledgements

This research represents a partial fulfilment of the requirements for the degree of Master of Science in biology at NMSU and was supported by NSF EF-0542228 (C.D. Bailey). We thank William Baker and Conny Asmussen-Lange for unpublished phylogenetic trees of Arecaceae and Carl Lewis (FTG) for providing DNA samples.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–410.

Asmussen CB, Dransfield J, Deickmann V *et al*. (2006) A new subfamily classification of the palm family (Arecaceae): evidence from plastid DNA. *Botanical Journal of the Linnean Society*, **151**, 15–38.

Bailey CD, Hughes CE, Harris SA (2004) Using RAPDs to identify DNA sequence loci for species level phylogeny reconstruction: an example from *Leucaena* (Fabaceae). *Systematic Botany*, **29**, 4–14.

Baker WJ, Hedderson TA, Dransfield J (2000) Molecular phylogenetics of subfamily Calamoideae (Palmae) based on nrDNA ITS and cpDNA *rps*16 intron sequence data. *Molecular Phylogenetics and Evolution*, **14**, 195–217.

Clegg MT, Zurawski G (1992) Chloroplast DNA and the study of plant phylogenetics: present status and future prospects. In:

Molecular Systematics of Plants (eds Soltis PS, Soltis DE, Doyle JJ), pp. 1–13. Chapman & Hall, New York.

Doyle JJ (1992) Gene trees and species trees: molecular systematics as one-character taxonomy. *Systematic Botany*, **17**, 144–163.

Feltus FA, Singh HP, Lohithaswa HC *et al*. (2006) A comparative genomics strategy for targeted discovery of single-nucleotide polymorphisms and conserved-noncoding sequences in orphan crops. *Plant Physiology*, **140**, 1183–1191.

Gaut BS, Morton BR, McCaig BC, Clegg MT (1996) Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbc*L. *Proceedings of the National Academy of Sciences, USA*, **93**, 10374–10279.

Gaut BS, Muse SV, Clark WD, Clegg MT (1992) Relative rates of nucleotide substitution at the *rbc*L locus of monocotyledonous plants. *Journal of Molecular Evolution*, **35**, 292–303.

Hodel DR (1992) *Chamaedorea Palms: The Species and Their Cultivation*. Allen Press, Lawrence, Kansas.

Hughes CE, Eastwood RJ, Bailey CD (2006) From famine to feast? Selecting nuclear DNA sequence loci for plant species-level phylogeny reconstruction. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **361**, 211–225.

Lewis CE, Namoff S (2006) *DNA Barcoding for Identification of Palms: Sequencing Is Not the Best Strategy*. Botany 2006 meeting, Chico, California.

Lohithaswa HC, Feltus FA, Singh HP, Paterson AH (2005) *Conserved PCR Primers from Rice–Banana and Rice–Onion Alignments*. XIII. Plant-Animal Genome Conference, San Diego, California.

Lohithaswa HC, Feltus FA, Singh HP *et al.* (2007) Leveraging the rice genome sequence for monocot comparative and translational genomics. *Theoretical and Applied Genetics*, **115**, 237–243.

Pamilo P, Nei M (1988) Relationships between gene-trees and species-trees. *Molecular Biology and Evolution*, **5**, 568–583.

Small RL, Cronn RC, Wendel JF (2004) L.A.S. Johnson review, 2: use of nuclear genes for phylogeny reconstruction in plants. *Australian Systematic Botany*, **17**, 145–170.

Strand AE, Leebens MJ, Milligan BG (1997) Nuclear DNA-based markers for plant evolutionary biology. *Molecular Ecology*, **6**, 113–118.

Syring J, Willyard A, Cronn RC, Liston A (2005) Evolutionary relationships among *Pinus* (Pinaceae) subsections inferred from multiple low-copy nuclear loci. *American Journal of Botany*, **92**, 2086–2100.

Thomas MM, Bateman RM, Garwood NC *et al*. (2006) Molecular phylogeny of the palm genus *Chamaedorea*, based on the low-copy nuclear genes PRK and RPB2. *Molecular Phylogenetics and Evolution*, **38**, 398–415.

Uhl N, Dransfield J (1987) *Genera Palmarum*. Allen Press, Lawrence, Kansas.

Wendel J, Doyle JJ (1998) Phylogenetic incongruence: window into genome history and molecular evolution. In: *Molecular Systematics of Plants II. DNA Sequencing* (eds Soltis DE, Soltis PS, Doyle JJ), pp. 265–296. Kluwer Academic Publishers, Boston, Massachusetts.

Wilson MA, Gaut B, Clegg MT (1990) Chloroplast DNA evolved slowly in the palm family (Arecaceae). *Molecular Biology and Evolution*, **7**, 303–314.

Supplementary material

The following supplementary material is available for this article:

Fig. S1 Strict consensus bootstrap values for individual loci at the interfamilial level and including outgroups.

Fig. S2 Strict consensus bootstrap values for individual loci at the interfamilial level and excluding outgroups. Bootstrap values from 1000 replicated and tree-bisection–reconnection branch swapping are shown near nodes.

Table S1 Analysed species, voucher, and GenBank Accessions numbers

Table S2 Locus information on intron size, putative function, and name based on the TIGR Rice Database

This material is available as part of the online article from:
http://www.blackwell-synergy.com/doi/abs/10.1111/j.1471-8286.2007.01928.x
(This link will take you to the article abstract).

Please note: Blackwell Publishing are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.