In silico detection of polymorphic microsatellites in the endangered Isis tamarind, *Alectryon ramiflorus* (Sapindaceae)

Joel Nichols¹, Gabriel C. Conroy¹, Naga Kasinadhuni², Robert W. Lamont¹, and Steven M. Ogbourne¹,³

PREMISE OF THE STUDY: *Alectryon ramiflorus* (Sapindaceae) is an endangered rainforest tree known from only two populations. In this study, we identified polymorphic microsatellites, in silico, improving the effectiveness and efficiency of microsatellite development of nonmodel species. The development of genetic markers will support future conservation management of the species.

METHODS AND RESULTS: We used next-generation sequencing and bioinformatics to detect polymorphic microsatellites, in silico, reducing both the time and cost of marker development. A panel of 15 microsatellites, 12 of which were polymorphic, were subsequently characterized in 64 adult trees representing the entire species range. Mean observed heterozygosity and expected heterozygosity were 0.471 and 0.425, respectively. The polymorphism information content across loci ranged from 0.152 to 0.875.

CONCLUSIONS: The microsatellite markers developed in this study will be useful in gaining an understanding of *A. ramiflorus*’ genetic diversity, level of inbreeding, and population structure and for guiding future restoration and management efforts.

KEY WORDS *Alectryon ramiflorus*; conservation genetics; Illumina high-throughput sequencing; next-generation sequencing (NGS); reintroduction; Sapindaceae.

The Isis tamarind, *Alectryon ramiflorus* (Sapindaceae), is a dioecious tree restricted to the dry Araucarian microphyll vine forests near Childers in southern Queensland, the majority of which have been cleared for agriculture over the last century (Threatened Species Scientific Committee, 2016). Consequently, apart from several isolated roadside trees, the species currently exists in only two small populations, approximately 15 km apart (Barker and Barry, 2003; Threatened Species Scientific Committee, 2016). *Alectryon ramiflorus* is listed as endangered under the Commonwealth Environment Protection and Biodiversity Conservation Act 1999 and the Queensland Nature Conservation Act 1992, and is the subject of a Queensland Government Recovery Plan (Department of Environment and Resource Management, 2010). Building an understanding of what remains of the species’ level of genetic diversity, as well as the way in which it is partitioned throughout its now heavily fragmented former distribution, has been prioritized to support subsequent on-ground operations such as population establishment and enhancement.

Traditionally, de novo identification of microsatellites from nonmodel organisms has proven to be a time-consuming and costly process (Zalapa et al., 2012). However, advances in next-generation sequencing have made microsatellite isolation more accessible (Vieira et al., 2016). Screening microsatellites for polymorphisms can be further streamlined by sequencing two or more individuals simultaneously, prior to in silico comparison of the sequences, to identify variable loci (Bonatelli et al., 2015). Although these combined methods expedite the identification of polymorphisms, the primers designed to amplify the target microsatellite still need to be tested to confirm amplification success and their degree of polymorphism. However, the identification of numerous polymorphic loci in silico allows judicious selection of high-quality markers. In this study, we used next-generation sequencing, microsatellite selection software (QDD), and BLAST to identify polymorphic microsatellites between two individuals in silico, demonstrating how this procedure can serve as a fast and economical approach to obtain markers for genetic analysis of nonmodel species. We then characterized genetic variability at 12 polymorphic microsatellite loci in 64 *A. ramiflorus* adults to facilitate future conservation management of the species from a genetic perspective.
METHODS AND RESULTS

Genomic DNA from a single *A. ramiflorus* individual from each of the two known populations (Appendix 1) was extracted using the DNeasy Plant Mini Kit (QIAGEN, Hilden, Germany) and used to construct two DNA libraries at the Australian Genome Research Facility (Brisbane, Australia; http://agrf.org.au/). Libraries were constructed using the Illumina TruSeq Nano DNA Library Prep Kit (Illumina, San Diego, California, USA). Briefly, 200 ng of DNA was sheared to ~550 bp by ultra-focused sonication on the Covaris E220 Focused-ultrasonicator (Covaris, Woburn, Massachusetts, USA). Fragmented DNA underwent end-repair and size selection using SPRI beads (Beckman Coulter, Indianapolis, Indiana, USA), followed by adenylation of the 3’ ends and adapter ligation. Adapter-ligated fragments were PCR amplified for eight cycles. The final library was assessed using electrophoresis and real-time quantitative PCR (qPCR) before being sequenced using MiSeq (Illumina) according to the manufacturer’s instructions with 2 × 250-bp paired-end reads. FastQC (Andrews, 2010) was used to check adapter contamination and data quality. Paired-end reads were stitched using PEAR version 0.9.5 default parameters (Zhang et al., 2014), and subsequent merged reads were analyzed using QDD version 3.1.2 (Meglécz et al., 2014) for microsatellite detection and Primer3 for primer design (Untergasser et al., 2012). The two *A. ramiflorus* samples were processed individually, yielding approximately 20,000 microsatellites. One sample was treated as “target” and the other as “subject” for BLAST to yield similar sequences, resulting in a list of ~300 polymorphic microsatellites. A detailed description of the commands and parameters used are provided as Appendix S1.

Microsatellite loci were amplified in 12.5-μL reactions containing ~20 ng of genomic DNA, 1× reaction buffer (67 mM Tris-HCl [pH 8.8], 16.6 mM (NH₄)₂SO₄, 0.45% Triton X-100, 0.2 mg/mL gelatin), 200 μM of each dNTP, 2.0 mM MgCl₂, 0.4 μM of each primer, and 0.5 units of Taq F1 DNA polymerase (Fisher Biotec, Wembley, Australia). Touchdown PCR using 14 *A. ramiflorus* individuals was performed using an Eppendorf Mastercycler (Hamburg, Germany) with the following cycling conditions: initial denaturation at 95°C for 3 min; followed by 14 cycles of 94°C for 30 s, 63°C for 30 s (reducing by 0.5°C per cycle), and 72°C for 30 s; followed by 21 cycles of 94°C for 30 s, 56°C for 30 s, and 72°C for 30 s; and a final extension at 72°C for 10 min. PCR products were visualized on 3.0% agarose 0.6× TBE gels.

Twenty-four loci showing the clearest and most consistent amplification were selected, and forward primers were labeled with either Fam (Sigma-Aldrich, St. Louis, Missouri, USA), VIC, NED, or PET (Applied Biosystems, Foster City, California, USA) fluorescent dyes. PCR was completed for 14 *A. ramiflorus* individuals using the same reaction components described above and the following cycling conditions: initial denaturation at 95°C for 5 min; followed by 35 cycles of 94°C for 30 s, 56°C for 90 s, and 72°C for 30 s; with a final extension at 68°C for 10 min. PCR products were separated by capillary electrophoresis on a 3500 Genetic Analyzer (Applied Biosystems). Nine of the 24 loci tested revealed electrophoretic signatures that were acceptable but relatively difficult to interpret. Three of the 24 revealed monomorphic signatures (Table 1). Twelve high-quality polymorphic microsatellite loci (Table 1) were finally selected and validated in 64 individuals, 60 collected from the only two known populations (plus two individuals each from two roadside patches) located near Childers, Queensland (Appendix 1). Congeneric species could not be collected for cross-amplification testing due to permit restrictions. Fragment sizes were determined relative to a GeneScan 600 LIZ Size Standard (Applied Biosystems) using GeneMarker version 2.7.0 (SoftGenetics, State College, Pennsylvania, USA) and double-checked manually. MICRO-CHECKER version 2.2.3 (van Oosterhout et al., 2004) was used to test for scoring errors, homozygote excess, large allele dropout, and potential null alleles. Number of alleles, polymorphism information content, observed and expected heterozygosity, and null allele frequencies for each locus were calculated in CERVUS version 3.0.8 (Kalinowski et al., 2007). FSTAT version 2.9.3.2 (Goudet, 2001) was used to identify evidence of linkage disequilibrium. Tests for Hardy–Weinberg equilibrium at each locus were conducted in GenALEX version 6.501 (Peakall and Smouse, 2006, 2012) with significance levels adjusted using Bonferroni correction. Approximately 7.56 Gb (15,118,442 paired-end reads) of sequence data from two *A. ramiflorus* individuals were generated. A total of 72 alleles were resolved across the 12 loci in the 64 adult individuals tested. The number of alleles per locus, per population ranged from two to 17, with an average of 3.021 alleles per locus, per population. Mean levels of observed and expected heterozygosity per locus per population ranged from 0.375 to 0.583 and 0.313 to 0.526, respectively (Table 2). Polymorphism information content values ranged from 0.152 to 0.875 per locus, with a mean of 0.506. Six of the 12 loci exhibited an excess of homozygosity and significant deviation from conditions of Hardy–Weinberg equilibrium, which was expected given the small population sizes and extremely low number of individuals remaining in the species. Two loci (AR10 and AR30) displayed evidence of null alleles but at low frequency. No evidence of linkage disequilibrium between pairs of loci was detected. All microsatellite sequences were deposited in the National Center for Biotechnology Information’s GenBank (Table 1), and the raw sequencing reads were deposited in the Sequence Read Archive (accession no. SUB4484831).

CONCLUSIONS

Here we have demonstrated that partial genomic sequencing and in silico identification of polymorphic loci using bioinformatics can serve as an effective and efficient approach to isolate and design microsatellite markers for genetic analysis in non-model species. The increased efficiency and effectiveness is observed via a reduction in the time and cost of the laboratory phase of primer and loci validation and via an increase in quality of the developed loci. The approach is potentially even more valuable for identification of polymorphic microsatellites in species with low genetic diversity such as *A. ramiflorus*. The 12 markers validated in this study will be employed in the conservation of this endangered species by providing important tools to inform the future management of *A. ramiflorus*.

ACKNOWLEDGMENTS

This work was financed by the Burnett Mary Regional Group (BMRG; MNES 16/18) and received support from the
TABLE 1. Characteristics of 15 microsatellite loci developed in Alectryon ramiflorus.

| Locus  | Primer sequences (5′–3′) | Repeat motif | Allele size range (bp) | GenBank accession no. |
|--------|--------------------------|--------------|------------------------|-----------------------|
| AR10   | F: ACACACATGATCCTCAACTCTAGT<br>R: TCGCATATAAGCACAAGCTCTCA | (AT)_{13} | 274–284 | MH347325 |
| AR13   | F: CTGTCTTGTGTCAGCAGCTCA<br>R: AGTGCTTCACGGGAGCAGTTC | (AG)_{6} | 323–325 | MH347326 |
| AR16   | F: TGCAAAATTGACGAGGTGGG<br>R: CGGCTGACAGCTGGTGCTT | (AT)_{10} | 266–278 | MH347327 |
| AR17   | F: AGTGCTCCATTACCTAACCAAGG<br>R: AGGAGAAGGTTGGCAGATTTGA | (AG)_{10} | 228–230 | MH347328 |
| AR19   | F: GAATTCGCTAAGTTGGGCCT<br>R: CCAGGCCATTTGTTTGGCTT | (AT)_{15} | 201–207 | MH347329 |
| AR23   | F: AAGACCGATGCTCAAGGGGA<br>R: TCAACATCTTACCTCACAACATG | (AT)_{19} | 132–158 | MH347330 |
| AR30   | F: GCGCTTCATATTCAATCTGTC<br>R: GTGGGTCTTGGAGGCTTT | (AG)_{12} | 238–240 | MH347331 |
| AR32   | F: GGCAGCTTACTGTCAGGCT<br>R: TGTCAGCTTCTGTTGAGCCT | (AT)_{10} | 314–348 | MH347332 |
| AR34   | F: GAGTCAACTGATTCTGTGTGT<br>R: GAACCGAATGACCCAATCCCT | (AG)_{10} | 236–240 | MH347333 |
| AR35   | F: GTCGGCCATCTCACAAGGAT<br>R: GGAAGCAGATGATGAGAGAGG | (AT)_{12} | 341–411 | MH347334 |
| AR36   | F: GCCGGAAATGGGAGACACCC<br>R: ACCCTAATTCTCCTTCCCATCC | (AC)_{10} | 351–353 | MH347335 |
| AR38   | F: CTCCGCGCAATCTGAGTCTTT<br>R: GAGCAGCTGTACAGTTTCCTT | (AT)_{12} | 187–193 | MH347336 |
| AR18   | F: TGACCAAAACAAAGAGGGACAGC<br>R: AGACATTCTGTTGATATCCACAGA | (AG)_{12} | 239 | MH634503 |
| AR43   | F: GACCAATGCTACATCAACAGC<br>R: GCAGATACGTTCTGGGAGACG | (AT)_{10} | 149 | MH634504 |
| AR48   | F: TCCATTTGGTGCTCAGGAGCG<br>R: GTGCTGTGTTGCTGACTGTT | (AT)_{10} | 319 | MH634505 |

\*All values are based on 64 samples representing the entire known species range.
\*An annealing temperature of 56°C was used for all loci.
\*These loci were found to be monomorphic in Alectryon ramiflorus and are therefore not reported in Table 2.

TABLE 2. Genetic properties of the 12 newly developed polymorphic microsatellites of Alectryon ramiflorus.

| Locus | Isis Mill (n = 30) | Cordalba NP (n = 30) | Franceys Rd. (n = 2) | Helms Scrub (n = 2) |
|-------|-------------------|---------------------|---------------------|-------------------|
|       | A | H_{e} | H_{o} | A | H_{e} | H_{o} | A | H_{e} | H_{o} | A | H_{e} | H_{o} |
| AR10  | 2   | 0.207 | 0.185 | 3   | 0.533 | 0.580 | 1   | 0.000 | 0.000 | 2   | 1.000 | 0.500 |
| AR13  | 2   | 0.500 | 0.411 | 4   | 0.433 | 0.469 | 1   | 0.000 | 0.000 | 2   | 0.500 | 0.375 |
| AR16  | 2   | 0.867 | 0.728 | 4   | 0.571 | 0.497 | 2   | 1.000 | 0.500 | 3   | 0.500 | 0.625 |
| AR17  | 2   | 0.500 | 0.357 | 3   | 0.367 | 0.438 | 1   | 0.000 | 0.000 | 2   | 0.000 | 0.500 |
| AR19  | 2   | 0.500 | 0.406 | 4   | 0.500 | 0.618 | 2   | 1.000 | 0.500 | 3   | 0.500 | 0.625 |
| AR23  | 2   | 0.103 | 0.212 | 3   | 0.567 | 0.499 | 1   | 0.000 | 0.000 | 2   | 0.100 | 0.500 |
| AR30  | 2   | 0.300 | 0.255 | 3   | 0.567 | 0.639 | 2   | 1.000 | 0.500 | 3   | 0.500 | 0.625 |
| AR34  | 2   | 0.103 | 0.212 | 3   | 0.500 | 0.794 | 3   | 1.000 | 0.625 | 3   | 0.500 | 0.625 |
| AR35  | 2   | 0.300 | 0.255 | 3   | 0.567 | 0.639 | 3   | 1.000 | 0.625 | 3   | 0.500 | 0.625 |
| AR36  | 2   | 0.300 | 0.255 | 3   | 0.567 | 0.639 | 3   | 1.000 | 0.625 | 3   | 0.500 | 0.625 |
| Mean  | 3.67 | 0.453 | 0.455 | 4.67 | 0.473 | 0.526 | 1.75 | 0.583 | 0.313 | 2.00 | 0.375 | 0.406 |

Note: A = number of alleles; H_{e} = expected heterozygosity; H_{o} = observed heterozygosity; n = number of individuals sampled.
Locality and voucher information are available in Appendix 1.

Queensland Herbarium (BRI; Brisbane, Queensland, Australia) and the University of the Sunshine Coast (USC; Sunshine Coast, Queensland, Australia). The authors would like to acknowledge Rod Buchanan (BMRG), Paul Forster (BRI), and Mike Johnson from the Bundaberg and District Urban Landcare Association for assistance with population identification and sample collection (collection permit no. WITK18499917).

DATA ACCESSIBILITY

Raw sequencing reads were deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (accession no. SUB4484831). Sequence information for the developed primers has been deposited to NCBI’s GenBank; accession numbers are provided in Table 1.
SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

APPENDIX S1. Bioinformatics steps for detecting polymorphic microsatellites between samples.

LITERATURE CITED

Andrews, S. 2010. FastQC: A quality control tool for high throughput sequence data. Website http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ [accessed 30 June 2016].

Barker, M., and S. Barry. 2003. Recovery plan for the Isis tamarind Alectryon ramiflorus 2003–2007. Report to Environment Australia, Canberra. Queensland Parks and Wildlife Service, Brisbane, Queensland, Australia.

Bonatelli, I. A., B. C. Carstens, and E. M. Moraes. 2015. Using next generation RAD sequencing to isolate multispecies microsatellites for Pilosocereus (Cactaceae). *PLoS ONE* 10: e0142602.

Department of Environment and Resource Management. 2010. Burnett Mary Natural Resource Management Region Back on Track Actions for Biodiversity. Department of Environment and Resource Management, Brisbane, Queensland, Australia.

Goudet, J. 2001. FSTAT, version 2.9.3, a program to estimate and test gene diversities and fixation indices. Lausanne University, Lausanne, Switzerland.

Kalinowski, S. T., M. L. Taper, and T. C. Marshall. 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology* 16: 1099–1106.

Meglécz, E., N. Pech, A. Gilles, V. Dubut, P. Hingamp, A. Trilles, R. Grenier, and J.-F. Martin. 2014. QDD version 3.1: A user-friendly computer program for microsatellite selection and primer design revisited. Experimental validation of variables determining genotyping success rate. *Molecular Ecology Resources* 13: 1302–1313.

van Oosterhout, C., W. F. Hutchinson, D. P. M. Wills, and P. Shipley. 2004. MICROCHECKER (version 2.2.3): Software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology* 4: 535–538.

Peakall, R., and P. E. Smouse. 2006. GenAlEx 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Resources* 6: 288–295.

Peakall, R., and P. E. Smouse. 2012. GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* 28: 2537–2539.

Threatened Species Scientific Committee. 2016. Commonwealth conservation advice for *Alectryon ramiflorus* (Isis tamarind). *Environment Protection and Biodiversity Conservation Act 1999*. Department of the Environment, Canberra, Australian Capital Territory, Australia.

Vieira, M. L. C., L. Santini, A. L. Diniz, and C. D. F. Munhoz. 2016. Microsatellite markers: What they mean and why they are so useful. *Genetics and Molecular Biology* 39: 312–328.

Zalapa, J. E., H. Cuevas, H. Zhu, S. Steffan, D. Senalik, E. Zeldin, B. McCown, et al. 2012. Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. *American Journal of Botany* 99: 193–208.

Zhang, J., K. Kobert, T. Flouri, and A. Stamatakis. 2014. PEAR: A fast and accurate Illumina Paired-End read mergeR. *Bioinformatics* 30: 614–620.

APPENDIX 1. Voucher information for *Alectryon ramiflorus* used in this study.a

| Voucher specimen accession no. | Collection localityb | Geographic coordinates | n |
|-------------------------------|----------------------|------------------------|---|
| BRI ACO0755000c,d            | Isis Mill            | 25.19221°S, 152.18361°E | 30 |
| BRI ACO0743408c,d            | Cordalba NP          | 25.12948°S, 152.06201°E | 30 |
| BRI ACO0603630d              | Franceys Rd.         | 25.203471°S, 152.276182°E | 2 |
| BRI ACO0659311d              | Helms Scrub          | 25.24813°S, 152.262560°E | 2 |

Note: n = number of individuals sampled.

*aVoucher samples for each location have been provided to the Queensland Herbarium (BRI), Brisbane, Australia.

bCollection localities are located near Childers, Queensland, Australia.

*cOne individual from this population used for library construction.

dAll individuals used to test amplification and polymorphism.

http://www.wileyonlinelibrary.com/journal/AppsPlantSci