Development of Microsatellite Markers for Two Australian Persoonia (Proteaceae) Species Using Two Different Techniques

Authors: Stingemore, Jessica A., Nevill, Paul G., Gardner, Michael G., and Krauss, Siegfried L.

Source: Applications in Plant Sciences, 1(10)

Published By: Botanical Society of America

URL: https://doi.org/10.3732/apps.1300023

Applications in Plant Sciences 2013 1(10): 1300023

PRIMER NOTE

DEVELOPMENT OF MICROSATELLITE MARKERS FOR TWO AUSTRALIAN *PERSOONIA* (PROTEACEAE) SPECIES USING TWO DIFFERENT TECHNIQUES[1]

JESSICA A. STINGEMORE[2,3,6], PAUL G. NEVILL[2,3], MICHAEL G. GARDNER[4,5], AND SIEGFRIED L. KRAUSS[2,3]

[2]School of Plant Biology, University of Western Australia, Nedlands, Western Australia 6009, Australia; [3]Botanic Gardens and Parks Authority, Kings Park and Botanic Garden, West Perth, Western Australia 6005, Australia; [4]School of Biological Sciences, Flinders University, GPO Box 2100, Adelaide, South Australia 5001, Australia; and [5]Evolutionary Biology Unit, South Australian Museum, North Terrace, Adelaide, South Australia 5000, Australia

- *Premise of the study:* Microsatellite markers were developed and cross-species transferability assessed for two *Persoonia* species to evaluate genetic diversity and population genetic structure of these broadly distributed southwest Australian tree species.
- *Methods and Results:* Microsatellite-enriched libraries and 454 GS-FLX shotgun sequencing were used to identity nine microsatellite loci for *P. elliptica* (one 454; eight cloning) and six for *P. longifolia* (three 454; three cloning). These loci were screened for variation in individuals from populations in southwestern Australia. In *P. elliptica*, observed and expected heterozygosities ranged from 0.46 to 0.93 and 0.42 to 0.88, respectively. For *P. longifolia*, observed and expected heterozygosities ranged from 0.04 to 0.88 and 0.04 to 0.84, respectively.
- *Conclusions:* The microsatellites identified in this study will enable the examination of population and spatial structuring of genetic diversity in *P. elliptica* and *P. longifolia*, two priority species for mine site restoration in southwestern Australia.

Key words: 454 GS-FLX; cross-species amplification; microsatellites; *Persoonia*; Proteaceae; southwest Australia.

Persoonia Sm. (Proteaceae) is an endemic Australian genus of woody perennial plants containing 100 species, 43 of which are found in Western Australia (Weston, 1995). *Persoonia elliptica* R. Br. and *P. longifolia* R. Br. are widespread, small trees within the southwestern Australian jarrah forest and are priority species for mine site restoration practitioners. These two sympatric congeners are key components of the jarrah forest and share pollinators and seed dispersers, but contrast markedly in their population densities. There are no species-specific molecular markers available for these two diploid species. Here we report the isolation and characterization of microsatellite markers for both species using (1) cloning of microsatellite-enriched libraries (*P. elliptica*) and (2) 454 GS-FLX shotgun sequencing (*P. longifolia*). This will enable the examination of genetic diversity, range-wide genetic differentiation, and mating system parameters in the two species.

METHODS AND RESULTS

Fresh leaf material was collected from *P. elliptica* plants in the southwest of Western Australia (population code: Pe-AE, 31.572°S 116.181°E, collection number: JS102; population code: Pe-AW, 31.613°S 116.151°E, collection number: JS102). Herbarium material is deposited at the Kings Park and Botanic Garden Herbarium (KPBG), Perth, Western Australia, Australia. Fresh leaf material was also collected from *P. longifolia* plants in southwestern Western Australia (population code: Pl-MD, 32.176°S 116.245°E, voucher: *R. Davis 3717* [PERTH 0481230]; and population code: Pl-CE, 33.356°S 116.345°E, voucher: *A. Gundry 486b* [PERTH 04801881]). The fresh leaf material was stored at −80°C and DNA extracted from the frozen material using the procedure described in Jobes et al. (1995) with the following modifications: 4 µL of proteinase-K was added prior to placing samples in the water bath and DNA was precipitated in the final stages by added 95% ethanol rather than isopropanol.

Genetic Identification Services (La Cañada Flintridge, California, USA; http://www.genetic-id-services.com/) was employed to develop microsatellite-enriched libraries for *P. elliptica* for four different repeat motifs (CA_n, GA_n, ACC_n, and ATG_n). Briefly, genomic DNA was restricted with seven blunt-end cutting enzymes (*Rsa*1, *Hae*III, *Bsr*B1, *Pvu*II, *Stu*I, *Sca*I, *Eco*RV). Fragments in the size range of 300–750 bp were linker adapted with oligonucleotides that contained a *Hin*dIII site and then subjected to magnetic bead capture (CpG MethylQuest DNA Isolation Kit; EMD Millipore, Billerica, Massachusetts, USA). Molecules were restricted with *Hin*dIII and ligated into the *Hin*dIII site of the pUC19 plasmid. Ligation products were introduced into *E. coli* strain DH5α (ElectroMax; Invitrogen, Carlsbad, California, USA) by electroporation. Blue-white selection was used to identify recombinant clones for sequencing on an ABI PRISM 377 DNA sequencer (Applied Biosystems, Carlsbad, California, USA) using Amersham's DYEnamic ET Terminator Cycle Sequencing Kit (Amersham Biosciences, Little Chalfont, Buckinghamshire, United Kingdom).

One hundred and six clones were sequenced, including 22 from the CA library, 22 from the GA library, 20 from the AAC library, and 22 from the ATG library. Eighty-two different microsatellite-containing clones were identified from the four libraries, and primers were designed for 72 sequences using DesignerPCR version 1.03 (Research Genetics Inc., www.lifetechnologies.com) and synthesized for 24. These primer pairs were initially tested to verify amplification, determine the optimum annealing temperature, and to establish size ranges for later PCR multiplexing, using DNA from eight individuals of *P. elliptica*. PCR was carried out in a total volume of 10 µL, containing approximately

[1]Manuscript received 27 March 2013; revision accepted 23 May 2013.

The authors thank the Australian Research Council through the Discovery Grant Scheme (DP1095297) for funding this work.

[6]Author for correspondence: jstingemore@bgpa.wa.gov.au

doi:10.3732/apps.1300023

10 ng genomic DNA template, 1× PCR Polymerization Buffer containing dNTPs (Fisher Biotech, Wembley, Western Australia, Australia), 0.2 μM each of unlabeled forward and reverse primer (GeneWorks, Hindmarsh, South Australia, Australia), 0.5 units *Taq* DNA polymerase (Fisher Biotech), and 2 mM of $MgCl_2$ (Fisher Biotech). PCR was carried out in a Veriti 96-Well Thermal Cycler (Applied Biosystems) using the following reaction conditions: an initial activation step at 95°C for 15 min; followed by 35 cycles of 95°C for 30 s, annealing at 59°C for 90 s, extension at 72°C for 90 s; followed by a final extension at 72°C for 15 min. PCR products were separated on a 2% agarose gel stained with SYBR Safe DNA Gel Stain (Invitrogen), and fragment sizes determined by comparison to a Low DNA Mass Ladder (Invitrogen).

The nine loci that amplified successfully were then screened on eight individuals of *P. elliptica* to test for polymorphism and trial multiplexing groups. Multiplexing was performed on four groups of primers (Table 1) using the QIAGEN Multiplex Kit (QIAGEN, Valencia, California, USA) in 12.5-μL reaction volumes containing 5–30 ng DNA template, 6.25 μL QIAGEN Multiplex PCR Master Mix, 1.25 μL Q solution, 0.1 μM of each forward primer (labeled; unique to primer), 0.1 μM of each reverse primer (unlabeled), and sterile H_2O to 12.5 μL. The multiplex PCR was conducted in a Veriti 96-Well Thermocycler (Applied Biosystems) with the following conditions: an initial activation step at 95°C for 15 min; followed by 35 cycles of 95°C for 30 s, annealing at 59°C for 90 s, extension at 72°C for 90 s; followed by a final extension at 72°C for 15 min. Following PCR, samples were diluted 1 : 30 in sample loading solution (Beckman Coulter, Brea, California, USA) with the addition of 0.4 μL fluorescently labeled 400-bp size standard per sample (Beckman Coulter) for capillary electrophoresis on a CEQ 8800 Genetic Analysis System (Beckman Coulter). Fragment peaks were visualized using CEQ Genetic Analysis System software (Beckman Coulter), and fragment (allele) sizes were scored manually. Eight individuals of *P. longifolia* were also genotyped using the above protocol, to test for cross-species amplification for these nine loci.

Genomic DNA of *P. longifolia* was sent to the Australian Genomic Research Facility (AGRF), Adelaide, Australia, for shotgun sequencing on a Titanium GS-FLX (454 Life Sciences, a Roche Company, Branford, Connecticut, USA), following Gardner et al. (2011). The sample occupied 6.25% of a plate and produced 108,806 individual sequences, with an average read length of 367 bp. The average GC content of these data was 37.82%. The program QDD version 1 (Meglécz et al., 2010) was used to screen the raw sequences for eight or greater di-, tri-, tetra-, or pentabase repeats, remove redundant sequences, and design primers using Primer3 (Rozen and Skaletsky, 2000). Software running parameters were set to default values, except PCR product lengths, which were set to 80–480 bp. We identified 14.2% of all reads containing microsatellite loci. Dinucleotide motifs were the most frequent (79,604), followed by tri-, tetra-, and then pentanucleotide motifs (14,550, 1072, and 445, respectively) (Meglécz et al., 2012). Primer pairs were designed for 108 loci and, from these, we excluded all loci that contained imperfect repeats, had a >2°C difference between the forward and reverse primer annealing temperature, short repeat motifs within the flanking region or primer sequence, or had poly A/T runs of >7 bp as there is an association between a high degree of poly A/T and instability (Li et al., 2002).

From the remainder, we arbitrarily selected 28 loci and then followed guidelines from Gardner et al. (2011) for further development; the loci were initially trialed for amplification. Initial PCR amplification was carried using the same method as for cloned microsatellites, mentioned above.

The 12 loci that amplified successfully were then screened on eight individuals of *P. longifolia* to test for polymorphism and trial multiplexing groups. Multiplexing was performed on four groups of primers (Table 1) using the QIAGEN Multiplex Kit in 12.5-μL reaction volumes containing 5–30 ng DNA template, 6.25 μL QIAGEN Multiplex PCR Master Mix, 1.25 μL Q solution, 0.1 μM of each forward primer (labeled; unique to primer), 0.1 μM of each reverse primer (unlabeled), and sterile H_2O to 12.5 μL. The multiplex PCR was conducted in a Veriti 96-Well Thermocycler (Applied Biosystems) with an initial activation step at 95°C for 5 min; followed by nine cycles of 95°C for 30 s, a 1°C touchdown starting at 65°C for 180 s, and 72°C for 15 s; followed by 25 cycles at 95°C for 30 s, 56°C for 180 s, and 72°C for 15 s; and then a final extension at 60°C for 30 min. Capillary electrophoresis and fragment scoring was carried out using the same method as for cloned microsatellites, mentioned above. Eight individuals of *P. elliptica* were also genotyped using the above protocol, to test for cross-species amplification for these 12 loci.

One of the nine loci isolated from microsatellite-enriched libraries for *P. elliptica* did not amplify reliably in the target species (Table 1). Six out of 12 loci isolated from 454 pyrosequencing for *P. longifolia* were monomorphic, and three did not amplify reliably (Table 1). Cross-species amplification from *P. elliptica* to *P. longifolia* was successful for three loci (Table 1). Cross-species amplification from *P. longifolia* to *P. elliptica* was successful at one locus (Table 1).

More than 15 individuals from the two populations of both *P. elliptica* and *P. longifolia* were then genotyped using nine and six loci, respectively; all loci were polymorphic and amplified reliably from the previous screening. Genetic diversity parameters and deviation from Hardy–Weinberg equilibrium (HWE) were calculated using GenAlEx version 6.4 (Peakall and Smouse, 2006) (Table 2). We used MICRO-CHECKER 2.2.3 (van Oosterhout et al., 2004) to check each locus for evidence of null alleles, scoring error due to stuttering, and large allele dropout.

Table 1. Characteristics of microsatellite primers developed in *Persoonia elliptica* and *P. longifolia*.

Locus	Method[a]	Primer sequences (5′–3′)	Repeat motif	Allele size range (bp)[b]	WellRED dye label (Multiplex marker set)	GenBank accession no.
PeA102	GIS	F: ACCCACCAGCTAACTCTTTTG R: CCATCTGATCGTACCAACGTA	$(CA)_{14}$	94–138	D3 (3)	JX987704
PeA103	GIS	F: CCTAATAACCCCACCCTACTAC R: AGGTCCTTTCACACTTTTGAT	$(GT)_{15}$	243–299; 243–289	D3 (2)	JX987705
PeA106	GIS	F: TCCAACTGGATGCCAACTC R: TCTTGGGATCAGCGTATGC	$(GC)_5(AC)_{15}$	179–187; 167–169	D2 (1)	JX987706
PeB103	GIS	F: AATGGGGTTGCTGTGTCT R: TCCCTTCTCTTCCACCAC	$(GA)_{19}$	272–314; 270–282	D4 (2)	JX987707
PeB106	GIS	F: TGCTTCTTGTATGCCAAAGAG R: CCTTCTAAGGGAGATGCAGAG	$(GTT)_9$	164–226	D2 (3)	JX987708
PeB2	GIS	F: TTCGCTTCCTTCTTTGTTAAAG R: AGGGAGGTATCACATGAGGAC	$(CT)_{19}$	285–313	D4 (1)	JX987709
PeC104	GIS	F: CTTGAGCATGAGCTTGAGATC R: ATCAACCATGAGGAGAAGAGAG	$(GTT)_9$	146–168	D4 (1)	JX987710
PeC2	GIS	F: GATCCTTAGGTTTGGTTTCTTG R: TGCCCTATCAAGCATAGTTATG	$(GTT)_{10}$	235–267	D3 (1)	JX987711
Pl13	454	F: GAGGCCATATAGGGTGGTGA R: CACAGGATCATAGTTGCCTCG	$(AT)_8$	na; 110–132	D2 (5)	JX987701
Pl15	454	F: ATAATGGGCCTAATGATGCG R: ACTCCCTAGGCTATGCCTCC	$(GA)_9$	145–163; 135–181	D3 (3)	JX987702
Pl18	454	F: TGGACACTTCAGGAGCTAACA R: GGGAGGTGTATCCACTCCTTG	$(GA)_{15}$	na; 188–206	D2 (5)	JX987703

[a]The two development methods were (a) microsatellite-enriched library construction by Genetic Identification Services (GIS; La Cañada Flintridge, California, USA; http://www.genetic-id-services.com/) for *P. elliptica* and (b) 454 GS-FLX shotgun sequencing for *P. longifolia*.

[b]"na" indicates that the primer was not used for *P. elliptica*; for primers that are used in both species, values for *P. elliptica* are listed first.

Table 2. Results of primer screening in two populations each of *Persoonia elliptica* and *P. longifolia*.[a]

	P. elliptica								*P. longifolia*							
	Pe-AW (n = 15)				Pe-AE (n = 15)				Pl-MD (n = 30)				Pl-CE (n = 25)			
Locus	A	H_o	H_e	HWE[b]	A	H_o	H_e	HWE[b]	A	H_o	H_e	HWE[b]	A	H_o	H_e	HWE[b]
PeA102	10	0.46	0.84	*	10	0.73	0.82	ns	—	—	—	—	—	—	—	—
PeA103	7	0.80	0.78	*	5	0.80	0.65	ns	6	0.70	0.68	ns	6	0.76	0.68	ns
PeA106	3	0.47	0.42	ns	4	0.67	0.61	ns	3	0.07	0.07	ns	2	0.04	0.04	ns
PeB103	14	0.87	0.86	ns	10	0.67	0.80	ns	6	0.40	0.46	***	5	0.60	0.68	ns
PeB106	13	0.69	0.84	**	11	0.87	0.86	ns	—	—	—	—	—	—	—	—
PeB2	8	0.93	0.84	ns	10	0.93	0.88	ns	—	—	—	—	—	—	—	—
PeC104	4	0.73	0.66	ns	4	0.87	0.68	ns	—	—	—	—	—	—	—	—
PeC2	6	0.93	0.78	*	8	0.80	0.81	ns	—	—	—	—	—	—	—	—
Pl13	—	—	—	—	—	—	—	—	3	0.30	0.32	ns	5	0.56	0.68	***
Pl15	10	0.86	0.85	ns	8	0.93	0.82	ns	13	0.73	0.84	ns	10	0.76	0.79	***
Pl18	—	—	—	—	—	—	—	—	6	0.57	0.68	ns	6	0.88	0.68	ns

Note: — = primer was not genotyped; A = number of alleles; H_e = expected heterozygosity; H_o = observed heterozygosity; HWE = Hardy–Weinberg equilibrium; n = number of individuals sampled.

[a]All values are based on samples from populations in southwestern Australia (Pe-AE: 31.572°S 116.181°E, Pe-AW: 31.613°S 116.151°E, Pl-MD: 32.176°S 116.245°E, Pl-CE: 33.356°S 116.345°E).

[b]Significant departure from HWE is shown as: *$P < 0.05$, **$P < 0.01$, *** $P < 0.001$; ns = not significant.

In *P. elliptica,* between three and 14 alleles per locus were found. Observed (H_o) and expected (H_e) heterozygosities ranged from 0.46 to 0.93 and 0.42 to 0.88, respectively (Table 2). For *P. longifolia,* between two and 13 alleles per locus were found. H_o and H_e ranged from 0.04 to 0.88 and 0.04 to 0.84, respectively (Table 2). Significant departure from HWE was detected in four of the nine loci for *P. elliptica* and three of the six loci for *P. longifolia* (Table 2). No loci showed significant null allele frequencies, large allele dropout, or evidence of scoring error due to stuttering.

CONCLUSIONS

The microsatellite markers developed here will enable an examination of population genetic patterns and realized dispersal of pollen and seed in *P. elliptica* and *P. longifolia*. The effort required in developing microsatellite markers has long been an important consideration, especially for low-budget studies on nonmodel organisms (Squirrell et al., 2003). Next-generation sequencing techniques such as 454 pyrosequencing have now been widely used for the development of microsatellite markers more rapidly and cheaply than traditional cloning approaches (Gardner et al., 2011; Malausa et al., 2011). Our experience suggests that, although there were efficiencies to be gained from employing next-generation sequencing for microsatellite marker development, particularly in the identification of microsatellite loci and primer sequences, it is not a panacea. A significant investment of time and effort in the screening, as well as optimization of polymorphic microsatellite markers, is still required with this relatively new approach to marker development.

LITERATURE CITED

Gardner, M. G., A. Fitch, T. Bertozzi, and A. J. Lowe. 2011. Rise of the machines—Recommendations for ecologists when using second generation sequencing for microsatellite development. *Molecular Ecology Resources* 11: 1093–1101.

Jobes, D., D. Hurley, and L. Thien. 1995. Plant DNA isolation: A method to efficiently remove polyphenolics, polysaccharides, and RNA. *Taxon* 44: 379–386.

Li, Y. C., A. B. Korol, T. Fahima, A. Beiles, and E. Nevo. 2002. Microsatellites: Genomic distribution, putative functions and mutational mechanisms: A review. *Molecular Ecology* 11: 2453–2465.

Malausa, T., A. Gilles, E. Meglécz, H. Blanquart, S. Duthoy, C. Costedoat, V. Dubut, et al. 2011. High-throughput microsatellite isolation through 454 GS-FLX Titanium pyrosequencing of enriched DNA libraries. *Molecular Ecology Resources* 11: 638–644.

Meglécz, E., C. Costedoat, V. Dubut, A. Gilles, T. Malausa, N. Pech, and J. F. Martin. 2010. QDD: A user-friendly program to select microsatellite markers and design primers from large sequencing projects. *Bioinformatics (Oxford, England)* 26: 403–404.

Meglécz, E., G. Nève, E. Biffin, and M. G. Gardner. 2012. Breakdown of phylogenetic signal: A survey of microsatellite densities in 454 shotgun sequences from 154 non model eukaryote species. *PLoS ONE* 7: e40861.

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

Rozen, S., and H. Skaletsky. 2000. Primer3 on the WWW for general users and for biologist programmers. *Methods in Molecular Biology (Clifton, N.J.)* 132: 365–386.

Squirrell, J., P. M. Hollingsworth, M. Woodhead, J. Russell, A. J. Lowe, M. Gibby, and W. Powell. 2003. How much effort is required to isolate nuclear microsatellites from plants? *Molecular Ecology* 12: 1339–1348.

van Oosterhout, C., W. F. Hutchinson, D. P. M. Wills, and P. Shipley. 2004. MICRO-CHECKER: Software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* 4: 535–538.

Weston, P. H. 1995. *Persoonia. In* A. S. George [ed.], Flora of Australia, Vol. 16, Elaeagnaeae, Proteaceae 1. CSIRO, Melbourne, Australia.