Development of Microsatellite Loci of Pod Mahogany, Afzelia quanzensis (Fabaceae), by Illumina Shotgun Sequencing, and Cross-Amplification in A. africana

Authors: Jinga, Percy, Palagi, Jason, and Ashley, Mary V.

Source: Applications in Plant Sciences, 4(6)

Published By: Botanical Society of America

URL: https://doi.org/10.3732/apps.1600010

Applications in Plant Sciences 2016 4(6): 1600010

PRIMER NOTE

DEVELOPMENT OF MICROSATELLITE LOCI OF POD MAHOGANY, *AFZELIA QUANZENSIS* (FABACEAE), BY ILLUMINA SHOTGUN SEQUENCING, AND CROSS-AMPLIFICATION IN *A. AFRICANA*[1]

PERCY JINGA[2,3,4], JASON PALAGI[2], AND MARY V. ASHLEY[2]

[2]Department of Biological Sciences, University of Illinois at Chicago, 845 West Taylor Street, Chicago, Illinois 60607 USA; and
[3]Department of Biological Sciences, Bindura University of Science Education, Private Bag 1020, Bindura, Zimbabwe

- *Premise of the study:* Microsatellite loci were developed for *Afzelia quanzensis* (Fabaceae) as a first step toward investigating genetic diversity and population structure of the species in its native range.
- *Methods and Results:* Illumina shotgun sequencing was used to generate raw sequence reads, which were searched for potential microsatellite loci. A total of 70 potential microsatellite loci were tested for amplification and polymorphism, and 39 successfully amplified. Of the 39 loci that amplified, 12 were polymorphic while 27 were monomorphic. The 12 polymorphic loci were cross-amplified in *A. africana*, and eight successfully amplified.
- *Conclusions:* The 12 polymorphic microsatellite loci can be used for genetic studies of *A. quanzensis*, which can help determine its conservation status. Eight loci can also be used for genotyping in *A. africana*.

Key words: *Afzelia africana*; *Afzelia quanzensis*; Fabaceae; Illumina; microsatellite; PAL_FINDER.

Afzelia quanzensis Welw. (Fabaceae) is a deciduous, medium to large tree that naturally occurs in eastern and southern Africa. It is a lowland species that grows well in hot temperatures and sandy soils. Its wood possesses an ornamental grain, which is very strong and flexible. It glues firmly and takes a good varnish, properties that make it eagerly sought after by woodcarvers. Apart from woodcarving, *A. quanzensis* is also used for railway sleeper and door construction, and as timber for roofing and fencing. As a result, it has been heavily logged in its native range (Gerhardt and Todd, 2009). The International Union for Conservation of Nature (IUCN) has regionally listed *A. quanzensis* as vulnerable in Malawi (Golding, 2002), while in South Africa, it is now a protected species. No microsatellite loci have been developed specifically for the species. Here, we describe the development of microsatellite loci that will be used in genetic studies.

METHODS AND RESULTS

Genomic DNA was extracted from a leaf of one *A. quanzensis* individual (population geographic coordinates: 19°36.056′S, 32°30.084′E; representative voucher deposited at the National Herbarium and Botanic Garden, Harare, Zimbabwe [SRGH], voucher number 1) using the DNeasy Plant Mini Kit (QIAGEN, Valencia, California, USA) following the manufacturer's instructions. The DNA was used to prepare a sequencing library using the KAPA DNA Library Preparation Kit for Illumina Sequencing (Kapa Biosystems, Wilmington, Massachusetts, USA) following the manufacturer's instructions. The final library was quantified using the KAPA Library Quantification Kit for Illumina. The DNA library was sequenced by an Illumina MiSeq Benchtop Sequencer (Illumina, San Diego, California, USA).

The resulting raw Illumina paired-end sequencing reads were analyzed with a Perl script, PAL_FINDER_v0.02.04 (available at http://sourceforge.net/projects/palfinder), which identifies microsatellite loci without the need for prior sequence trimming and assembly (Castoe et al., 2012). The Perl script was run with Primer3 version 2.0.0 (Rozen and Skaletsky, 1999) for simultaneous primer design. Default settings were used except for the following adjustments: primer minimum annealing temperature (T_a) 50°C, primer maximum T_a 60°C, and primer optimum T_a 55°C. A total of 961,804 potential loci were identified, of which 7789 had primer pairs. We tested 70 potentially amplifiable loci with amplifiable primer pairs that occurred only once.

Of the 70 loci tested, 39 amplified successfully and these were checked for polymorphisms in 40 individuals randomly collected from a population near Chaseyama, southeastern Zimbabwe. Forward primers were tagged with a labeled M13 primer tail (TGTAAAACGACGGCCAGT). All PCR reactions were performed in a total volume of 10 µL, with 10 ng of template DNA, 0.6 µM of the reverse primer, 0.15 µM of the forward primer, 0.25 mM each dNTP, 0.6 µL bovine serum albumin (BSA; 10% w/v), 1 µL 10× reaction buffer with 15 mM $MgCl_2$, and 0.25 units of *Taq* DNA polymerase (Bulldog Bio, Rochester, New York, USA). Loci Afq45, Afq51, Afq62, Afq68, and Afq69 had an additional 0.1 mM $MgCl_2$. The thermocycling profile consisted of an initial denaturation at 94°C for 5 min; then 35 cycles of 94°C for 30 s, 55.0°C or 59.4°C (Table 1) for 30 s, 72°C for 30 s; and a final extension at 72°C for 7 min. The PCR amplicons were electrophoresed on an ABI 3730 DNA analyzer with GeneScan 500 LIZ (Applied Biosystems, Foster City, California, USA) as the size standard. The genotypes were scored using GeneMapper version 3.7 (Applied Biosystems).

Table 1 shows the 39 loci that amplified, their repeat motifs, number of alleles per locus, allele size range, and T_a. Twenty-seven loci were monomorphic while 12 were polymorphic. For the 12 polymorphic loci, number of alleles per locus (A), observed heterozygosity (H_o), and expected heterozygosity (H_e) were calculated using GenAlEx version 6.5 (Peakall and Smouse, 2006, 2012), and are shown in Table 2. The program Arlequin version 3.5 (Excoffier and Lischer,

[1]Manuscript received 26 January 2016; revision accepted 16 February 2016.

The authors thank K. Feldheim for assistance with the sequencing run and K. Ramanauskas for providing the platform to run the Perl script. Appreciation also goes to A. S. L. Donkpegan for providing *Afzelia africana* samples. Funding was provided by the University of Illinois at Chicago (UIC) Biological Sciences Department's Hadley Award, the UIC Office of International Affairs' Chicago Consular Corps Award, and the Institute of International Education's Dr. Benjamin L. Van Duuren Grant.

[4]Author for correspondence: pjinga2@uic.edu

doi:10.3732/apps.1600010

TABLE 1. Characteristics of 39 microsatellite loci developed for *Afzelia quanzensis*.

Locus	Primer sequences (5′–3′)	Repeat motif	*A*	Allele size range (bp)	T_a (°C)	NCBI Probe Database accession no.
Afq1	F: CCTATACCAGAAATTGATAAATTAGAGAGC R: GCTTAGCCAAGGGACATTGC	$(ATT)_{78}$	1	417	55.0	Pr032805619
Afq5	F: GACTCACAAGTGGCAAGTGAGG R: GTCCAACGATTGAAAGATTTAAGG	$(AAAT)_{20}$	1	341	55.0	Pr032805635
Afq6	F: CATGACCCAAGCATGACTCC R: CCTACAGTTGTTGAGAAGTCCGC	$(ACCCC)_{25}$	1	252	55.0	Pr032805642
Afq8	F: TTAATAATGCAAAGATGATTGGC R: GGGGCAATAAGTCAAAATGG	$(AAAT)_{24}$	1	410	55.0	Pr032805644
Afq9	F: CATTGACAAAGATGCATGATAGC R: TCATTGTAGTTTTCATTCACAACCC	$(AAAG)_{20}$	1	205	55.0	Pr032805645
Afq10	F: CAGGCAAGGGGTAAAATTGG R: CTGCTCCAAATTCCAAAGCC	$(TTC)_{33}$	1	154	55.0	Pr032805620
Afq12	F: CTCCTCTGCGCCACTATTCC R: CACTCCTCTCTCAGGCAGGG	$(AAC)_{15}$	3	265–271	55.0	Pr032754338
Afq13	F: AAATATTTTCGAGACCACAAACG R: AACTCGATTTCTTCATGTACGG	$(ATT)_{18}$	1	170	55.0	Pr032805621
Afq15	F: AGAAAACCAGCGGTACGAGC R: CATTATCGCCGGTAAGCTGC	$(CGG)_{18}$	1	212	55.0	Pr032805622
Afq20	F: AGAAAACCAGCGGTACGAGC R: CATTATCGCCGGTAAGCTGC	$(CGG)_{18}$	1	326	55.0	Pr032805623
Afq24	F: GGAAAGACTCCAGATCACTTCCC R: ACAAAACTGACCTGAACAAGGC	$(ATT)_{15}$	1	350	55.0	Pr032805624
Afq31	F: TGCACGAATGCAAATAGTGG R: ATTCTAAGGCATTAACATGGAGC	$(AAAT)_{20}$	1	249	55.0	Pr032805625
Afq33	F: GGATTCCATTCTAACCAGAGACC R: AAAGTTAGCTTTGCACCCTCC	$(ATTTT)_{20}$	3	210–220	55.0	Pr032754339
Afq34	F: AAACTGATGCAAATAAGATGGG R: TCCTAGTTTGATACCAATTAATGTAACG	$(ATTTT)_{20}$	1	349	55.0	Pr032805626
Afq35	F: TGATATCGGTTATGTGCAGGG R: TGCTGGGTCATATTTACTAGTGCC	$(ATT)_{21}$	6	364–382	55.0	Pr032754340
Afq38	F: ACACCATGGGTGAACTTGAGG R: CCCAGAGATTCAGCTTAGGCG	$(TC)_{32}$	1	150	55.0	Pr032805627
Afq39	F: AGGTGGTCATCCACAGTCCC R: GCTCACATTTAGACGGTGACG	$(AT)_{30}$	1	152	55.0	Pr032805628
Afq40	F: CATGCATATATGACGATTTTGTCC R: TGACTGTTCATTTATATACACACATTCACC	$(AT)_{32}$	1	265	55.0	Pr032805629
Afq41	F: TGCATAACCACCCAAAAGGG R: TCCTAATGGTTGATAGGTCCCC	$(ATT)_{21}$	1	290	55.0	Pr032805630
Afq42	F: AATGGCATGTTGCGTACACC R: AAAGCATTTGAAGATTTGGTAGGG	$(ATT)_{33}$	1	343	55.0	Pr032805631
Afq43	F: GAAGAAGGAAGCTTGTCGGC R: ATCACATTACCCGCATTGGG	$(TCC)_{21}$	3	227–239	55.0	Pr032754341
Afq44	F: AATTTACATTTGCTTCAACAGGG R: AAACACTCTTATTAGTTTATTCACCTGG	$(ATCT)_{20}$	4	147–163	55.0	Pr032754342
Afq45	F: CAAAACTAAACGACATCTCCTGC R: TTCCCTTCTTGCTTAGGGAGC	$(TTGGGC)_{24}$	4	297–315	55.0	Pr032754343
Afq46	F: CCATGTGTGAATATATCCCTTTGC R: GGAGGATGTTGTTCCTGTCG	$(AAAAC)_{20}$	1	230	55.0	Pr032805632
Afq47	F: TGACATCAGTTTCCTTGTGCC R: TTTTGCCTAAAGAAAATAGGTTTGG	$(AAAAG)_{20}$	1	195	55.0	Pr032805633
Afq48	F: TTGACCCACGTTCCTTCC R: TCACATGACTTCACAATATTTCCG	$(AAATT)_{20}$	1	152	55.0	Pr032805634
Afq49	F: ATCCTTTTGCCCATTCCTGC R: ATGGCACCCAAAGAAGAAGC	$(TC)_{26}$	7	273–291	55.0	Pr032754344
Afq50	F: CCAAAGGAATAGTTGGGTTTGC R: TATCGCCTTGTTCAACTGCC	$(AC)_{26}$	1	223	55.0	Pr032805636
Afq51	F: CATGGCTTCAACCTATCCTGG R: CCTTTCTCTGGTCCTTCCCC	$(ATT)_{39}(ATT)_{24}$	7	247–264	55.0	Pr032754345
Afq52	F: GGCAGGATTCATAGTTTACTTTCG R: ACAGGTGACATCGGAGTTGC	$(ATT)_{18}(ATT)_{15}$	1	319	55.0	Pr032805637
Afq54	F: CAAAGAGTAACAAAATCCCTGCG R: CATCGCTGGTTAGATGTTTTAGC	$(AAAT)_{20}(AAAT)_{20}$	1	352	55.0	Pr032805638
Afq56	F: TGCGAACAAGGTTCCTAACG R: TTTGGCATATGACAGTTGATGG	$(ATT)_{36}(ATT)_{15}$	1	408	55.0	Pr032805639
Afq57	F: CCTATTTGAAAGGTAATTTCTAAGACCC R: TCCCACACTTCATAAAACGGG	$(ATT)_{18}(ATT)_{18}$	1	271	55.0	Pr032805640
Afq58	F: TGTTAGCAGCATTGTTGAGGG R: CACTAATGGATTGCCTTTTCCC	$(ATT)_{30}(ATT)_{30}$	1	362	55.0	Pr032805641

TABLE 1. Continued.

Locus	Primer sequences (5′–3′)	Repeat motif	*A*	Allele size range (bp)	T_a (°C)	NCBI Probe Database accession no.
Afq62	F: TGTATACAAAACGATTTGACGGC R: TTTCCAATCAAGCAAATCTCG	$(ATT)_{21}(ATT)_{18}$	4	219–234	59.4	Pr032754346
Afq66	F: TGAACAGATCAATCAAAGTGCG R: CCATATTCATCCCACTCCCG	$(TC)_{18}$	1	276	55.0	Pr032805643
Afq67	F: CTTCATCATATAGCATAAGATAATCGG R: TTTAAGATAGGCTCAAGGACGG	$(AC)_{26}$	8	330–354	55.0	Pr032754347
Afq68	F: AGGCACACGAGCACACTAGG R: CAGGACCCTCCAGTGTTTCC	$(TC)_{20}$	8	215–235	55.0	Pr032754348
Afq69	F: TGACCGTTTTAAGAAAAGTCAAGC R: TCGATGATCCAGGAAAGTTGG	$(TC)_{16}$	10	294–318	55.0	Pr032754349

Note: *A* = number of alleles per locus; NCBI = National Center for Biotechnology Information; T_a = annealing temperature.

2010) was used to perform an exact test (Guo and Thompson, 1992) with a Markov chain for Hardy–Weinberg equilibrium (HWE), while the program MICRO-CHECKER version 2.2.3 (van Oosterhout et al., 2004) was used to estimate null allele frequencies (F_{NULL}) with Bonferroni correction.

The number of alleles for polymorphic loci ranged from three to 10 with an average of 5.583. H_o ranged from 0.138 to 0.737, while H_e ranged from 0.313 to 0.832. Three loci, Afq12, Afq51, and Afq67, showed evidence of null alleles. Loci Afq35, Afq51, and Afq67 showed departure from HWE (Table 2). It is suspected that a relatively small sample size may have contributed to the departure from HWE, and the result does not invalidate the utility of the markers. The alleles that show potential null alleles can be used with adjusted genotypes. Cross-amplification of the 12 polymorphic loci was tested in 24 individuals of a congeneric species, *A. africana* Sm., another African tree species prized for its high-quality wood. Eight loci amplified in *A. africana*, with the number of alleles ranging from one to five, as shown in Table 3.

CONCLUSIONS

Thirty-nine microsatellite loci were developed, of which 27 were monomorphic and 12 showed polymorphisms. The microsatellite loci are the first developed specifically for *A. quanzensis*. The loci provide a set of markers to study the genetic diversity, gene flow patterns, and population structure of the species. The species is being increasingly harvested for its wood, so the effects of overharvesting on long-term genetic viability need to be understood for conservation purposes. Eight loci can also be used for genetic studies of *A. africana*, another species that is being logged for its valuable wood.

TABLE 2. Polymorphic microsatellite locus-specific measures of genetic diversity of a population of 40 individuals of *Afzelia quanzensis*.[a]

Locus	*A*	H_o	H_e	F_{NULL}
Afq12	3	0.487	0.605	0.1075*
Afq33	3	0.410	0.347	–0.0833
Afq35†	6	0.730	0.728	–0.0010
Afq43	3	0.550	0.546	–0.0037
Afq44	4	0.350	0.343	–0.0095
Afq45	4	0.556	0.564	0.0079
Afq49	7	0.641	0.612	–0.0231
Afq51†	7	0.138	0.832	0.8119*
Afq62	4	0.359	0.313	–0.0690
Afq67†	8	0.325	0.423	0.1304*
Afq68	8	0.583	0.606	0.0195
Afq69	10	0.737	0.748	0.0077

Note: *A* = number of alleles per locus; F_{NULL} = null allele estimates (Chakraborty et al., 1992); H_e = expected heterozygosity; H_o = observed heterozygosity; † = loci not in Hardy–Weinberg equilibrium; * = loci showing evidence of null alleles.

[a]Geographic coordinates for the population are 19°36.056′S, 32°30.084′E. A voucher is deposited at the National Herbarium and Botanic Garden, Harare, Zimbabwe (SRGH), with voucher number 1. The specimen was collected by Percy Jinga.

LITERATURE CITED

CASTOE, T. A., A. W. POOLE, A. P. J. DE KONING, K. L. JONES, D. F. TOMBACK, S. J. OYLER-MCCANCE, J. A. FIKE, ET AL. 2012. Rapid microsatellite identification from Illumina paired-end genomic sequencing in two birds and a snake. *PLoS ONE* 7: e30953.

CHAKRABORTY, R., M. DE ANDRADE, S. P. DAIGER, AND B. BUDOWLE. 1992. Apparent heterozygote deficiencies observed in DNA typing data and their implications in forensic applications. *Annals of Human Genetics* 56: 45–57.

EXCOFFIER, L., AND H. E. L. LISCHER. 2010. Arlequin suite ver3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* 10: 564–567.

GERHARDT, K., AND C. TODD. 2009. Natural regeneration and population dynamics of the tree *Afzelia quanzensis* in woodlands in Southern Africa. *African Journal of Ecology* 47: 583–591.

GOLDING, J. S. 2002. Southern African plant red data lists. Southern African Botanical Diversity Network Report No. 14. SABONET, Pretoria, South Africa.

TABLE 3. Genetic properties of eight *Afzelia quanzensis* microsatellite loci tested in 24 *A. africana* individuals.[a]

Locus	*A*	Allele size range (bp)	H_o	H_e
Afq12	2	265–268	0.105	0.100
Afq33	2	210–215	0.348	0.386
Afq35	2	379–382	0.333	0.420
Afq43	3	230–236	0.087	0.084
Afq44	2	159–163	1.000	0.500
Afq49	5	283–291	0.750	0.622
Afq67	1	348	0.000	0.000
Afq69	1	289	0.000	0.000

Note: *A* = number of alleles per locus; H_e = expected heterozygosity; H_o = observed heterozygosity.

[a]Geographic coordinates for the population are 1.55586°N, 9.26674°E. Vouchers are deposited at the Université Libre de Bruxelles, Belgium (BRLU), with voucher numbers AD61–AD85. The population is located in Bassila, central Benin. Specimens were collected by Dr. Olivier Hardy.

Guo, S. W., and E. A. Thompson. 1992. Performing the exact test of Hardy–Weinberg proportion for multiple alleles. *Biometrics* 48: 361–372.

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.

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.

Rozen, S., and H. Skaletsky. 1999. Primer3 on the WWW for general users and for biologist programmers. *In* S. Misener and S. A. Krawetz [eds.], Methods in molecular biology, vol. 132: Bioinformatics methods and protocols, 365–386. Humana Press, Totowa, New Jersey, USA.

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.