Bairu et al. *New Zealand Journal of Forestry Science* (2020) 50:6
https://doi.org/10.33494/nzjfs502020x56x
E-ISSN: 1179-5395
Published on-line: 14/08/2020

RESEARCH ARTICLE Open Access

New Zealand Journal of Forestry Science

Development of an SSR-based DNA fingerprinting method for black wattle (*Acacia mearnsii* De Wild)

Michael W. Bairu[1,2], Willem G. Coetzer[1,3], Assefa B. Amelework[2*]

[1]*Institute for Commercial Forestry Research, P. O. Box 100281, Scottsville 3209, South Africa*
[2]*Agricultural Research Council, Vegetable and Ornamental Plants, Private Bag X293 Pretoria 0001, South Africa (current address)*
[3] *Department of Genetics, University of the Free State, P. O. Box 339, Bloemfontein 9300, South Africa (current address)*

*Corresponding author: assefaa@arc.agric.za

(Received for publication 28 May 2019; accepted in revised form 20 July 2020)

Abstract

Background: The most commonly used method for extracting DNA from plant leaf tissue involves cetyl trimethylammonium bromide but some species, such as *Acacia mearnsii,* contain high levels of secondary metabolites and polysaccharides that interfere with this process. Various modifications have been proposed for effective removal of these biomolecules but these methods can be time consuming. Therefore, this study was initiated to optimise the cetyl-trimethylammonium bromide protocol for the extraction of high-quality genomic DNA and to develop a fingerprinting tool using cross species transferable simple sequence repeat markers for genetic diversity studies in *A. mearnsii*.

Methods: Five CTAB-based modification were examined and 49 cross-species microsatellite markers, developed for several *Acacia* species, were tested in four multiplex panels of *A. mearnsii* populations.

Results: The modified protocol yields high quantity and quality DNA from *A. mearnsii* leaves using high concentration of NaCl to remove polysaccharides and polyvinylpolypyrrolidone (PVPP) to eliminate polyphenols during DNA purification. In addition, omitting the selective precipitation and NaCl gradient steps in the extraction protocol, enabled us to extract DNA 10–20 min faster than the normal protocol. Of the tested microsatellite loci, 11 were successful in amplifying sharp and high-intensity bands in all the four multiplex panels and were polymorphic. The level of polymorphism ranged from 0.115 to 0.794, with a mean 0.50 and mean number of alleles varied from 2 to 10, with overall mean of 6 alleles per locus. The mean observed and expected heterozygosity ranged from 0.058 to 0.970 and 0.102 to 0.796, respectively. The 11 microsatellite loci that were effectively amplified from *A. mearnsii* DNA were adequate in detecting genetic variation among the tested populations.

Conclusions: These PCR-based, multi-allelic, co-dominant microsatellite markers provide a powerful tool for genetic, breeding and conservation studies in *A. mearnsii*.

Keywords: *Acacia mearnsii*, cross-species microsatellite markers, DNA extraction, PCR amplification

Introduction

Acacia mearnsii is a well-known commercially grown tree species in South Africa. This species is cultivated mainly for its timber, timber products, pulp and its tannin-rich bark for the leather tanning industry. *Acacia mearnsii* is indigenous to Australia but is currently grown intensively in India, Japan, South Africa, Kenya, Tanzania, Uganda, Brazil, Uruguay and Argentina (ILDIS 2015). *Acacia mearnsii* was first introduced to South Africa in 1864 (De Beer 1986). South Africa is the third largest, and one of the oldest, plantation resource areas in the southern hemisphere (Owen & van der Zel 2000). *Acacia mearnsii* is grown in approximately 130 000 hectares, that stretches from south-eastern Mpumalanga (27.0245° S, 30.7925° E) to KwaZulu-Natal (29.7285° S, 30.5319° E) in the north (The Department of Forestry

and Water Affair 2003).

Acacia mearnsii was first introduced into South Africa via a bag of seeds with no prior information on the genetic variability. Consequently, *A. mearnsii* plantations were presumed to have a narrow genetic base. The high out crossing nature and the low self-incompatibility of *Acacia* species and their broad geographic adaptability enabled the species to maintain high genetic diversity (Duminil et al. 2009). However, the out crossing rate is highly dependent on flower fecundity, pattern and synchronization (Butcher et al. 2000). These factors are highly influenced by the environment. Intensive breeding was carried out over the years and populations were established in geographically isolated regions depending on the objective of breeding (frost tolerance, disease resistance and timber and bark quality). These populations were exposed to specific environmental factors that led to populations with distinct phenotypes. It was suggested that fragmentation of populations in a given region might result in an increase in the level of inbreeding and a decrease in allelic richness (Butcher et al. 2000). Natural selection, mutation, genetic drift and mating systems all affect the patterns of genetic variation among populations (Millar et al. 2008). To investigate the influence of the above genetic forces on the genetic diversity and patterns of variation, it is helpful to develop appropriate genetic analysis tools.

Population genetic diversity and conservation studies using molecular techniques are important for population or species survival; and the extraction of high quality genomic DNA and polymerase chain reaction (PCR) amplification protocols are essential pre-requisites (Bonin et al. 2004; Tan & Yiap 2009). Extraction of high-quality DNA from plant tissue is challenging because plants have variable levels of metabolites and structural biomolecules that interfere with existing DNA extraction protocols (Salblok et al. 2009; Sahu et al. 2012). Secondary metabolites, polysaccharides and polyphenols, are plant biomolecules that often interfere with DNA isolation, enzymatic digestion and PCR (Weising et al. 2005). In the case where species are characterized by their high secondary metabolite content, such as *A. mearnsii*, DNA extraction procedures require intensive testing and adaptation of different protocols.

The most commonly used basic DNA extraction methods from plant leaf tissue are derived from the original cetyl trimethylammonium bromide (CTAB) based protocol (Doyle & Doyle 1990). For species like *A. mearnsii,* which contain high levels of secondary metabolites and polysaccharides, ensuring extraction of high-quality DNA is important for downstream genetic analysis (Sahu et al. 2012; Healey et al. 2014). To contend with the problems associated with secondary metabolites various modifications have been proposed for effective removal of these biomolecules (Weising et al. 2005). The modified protocols include the addition of polyvinylpyrrolidone (PVP), insoluble PVPP, bovine serum albumin (BSA), reducing agents like β-mercaptoethanol or high salt solutions to the extraction buffer (Allen et al. 2006; Varma et al. 2007; Lade et al. 2014; Arruda et al. 2017). However, these methods can be time consuming, relying on long incubation steps, or requiring multiple DNA washes and precipitations that decrease overall yield (Healey et al. 2014). Therefore, this study was initiated to optimise the CTAB protocol for the extraction of high-quality genomic DNA and to develop a fingerprinting tool using cross species SSR markers for genetic diversity studies in *A. mearnsii*.

Methods

Sample preparation

The development of the DNA extraction and SSR-PCR amplification protocols were part of the genetic diversity study to support the Institute for Commercial Forestry Research's *A. mearnsii* breeding and seed production programmes (https://www.icfr.ukzn.ac.za). The four breeding populations used in this study were derived from a base population constituted with 1081 genotypes collected from different progeny trials resulting from controlled crosses and targeted selfing (Dunlop et al. 2003). The four breeding population were reconstituted based on similarity in disease resistance, bark yield and quality, stem form and vigour. Two hundred and twenty-eight individuals were selected from the four breeding populations (TP1 = 78, TP2 = 34, TP3 = 64, and TP4 = 52). At sampling time, three of these breeding populations were seven years old and one was a 15-year-old population (TP1). Young, fresh, and healthy leaves free from visible contamination were sampled, wrapped in moist paper and placed in a cooler box. Upon arrival to the laboratory, each sample was ground into fine powder with mortar and pestle following the addition of liquid nitrogen. The powder was transferred to duplicate 2 ml reaction tubes and immediately stored in a -80 °C freezer. One set of samples were used for DNA extraction and the duplicates were stored for future reference.

DNA Extraction

Various CTAB DNA extraction protocols were tested for their suitability with *A. mearnsii* leaf material. The initial protocol used here was developed by Centro Internacional de Mejoramiento de Maíz y Trigo (CIMMYT 2005) and modified by Bairu et al. (2006) and Moyo et al. (2008). The protocol was amended to increase DNA quantity and quality and to shorten the time spent on extractions. Briefly, some of the modifications included an altered NaCl concentration, to adding PVPP to the CTAB extraction buffer, and the collapse of the chloroform and CTAB/NaCl steps into a single step as described below and summarised in Table 1.

The final protocol contained the following steps: about 150 mg ground leaf material was added to 800 µL preheated at 65 °C CTAB extraction buffer (2% w/v CTAB, 100 mM Tris-HCl (PH 8.0) 20 mM EDTA, 1.4 M NaCl), with 8 µL 2-mercaptoethanol and 2% PVPP in a clean 2 mL tube. The mixture was incubated in a water bath at 65 °C for 1 hour, while mixing every 10 min by inversion. Immediately following incubation 800 µL chloroform: isoamyl alcohol (24:1) and 160 µL preheated at 65 °C CTAB/NaCl (10% CTAB,

TABLE 1: Comparison between the different modified CTAB protocols tested for DNA extraction from *Acacia mearnsii* leaf material.

CTAB Protocol	Mean DNA concentration (ng/µL)	Mean 260/280 ratio	Mean 260/230 ratio
1 (0.7 M NaCl in CTAB buffer)	44.217[#]	2.089[#]	-
2 (with CTAB/NaCl step)	15.981	1.787	-
3 (1.4 M NaCl in CTAB buffer)	113.222	1.908	0.966
4 (combined CTAB/NaCl step)	131.214	1.915	1.309
5 (final protocol) (PVPP added)	159.515	1.887	1.754

[#]The mean taken from four samples

0.7 M NaCl) was added and mixed for 10 min by inversion. The mixture was centrifuged for 10 min at 10000 rpm followed by recovery of the top aqueous layer into a 1.5 ml clean tube. Precipitation of DNA was achieved by adding 800 µL ice-cold isopropanol to the supernatant and mixed gently by inversion. The mixture was then centrifuged for 15 min at 10 000 rpm followed by careful decanting of the isopropanol, leaving the pellet in the tube. The pellet was washed in two steps, first with 70% ice cold ethanol and then 95% ethanol by inversion. After removing the ethanol by decanting, the pellet was allowed to air dry in a laminar flow bench. Finally, the dried pellet was re-suspended in 100 µL TE buffer (10 mM Tris-HCl, 1 mM EDTA, PH 8.0) and stored at -20 °C. Following DNA extraction, the concentration and purity of DNA was estimated using the Jenway Genova Nano spectrophotometer (Bibby Scientific Ltd., Staffordshire, UK).

SSR marker screening

Due to the lack of species specific SSR markers for *A. mearnsii*, 49 simple sequence repeat (SSR) markers were used from other *Acacia* species (Additonal File). The markers were chosen based on their ability to amplify across species (Adamski et al. 2013; Aggarwal et al. 2011; Miller 2009; Ng et al. 2005; Butcher et al. 2000). Initially, unlabelled primer sets were tested using PCR amplification and the PCR products were visualization on a 2% MetaPhor™ agarose gel (Lonza Rockland Inc., Rockland, USA) with TAE buffer. Only those markers with visible bands were selected for further evaluation. The forward primer of each selected primer set was then labelled with a fluorescent dye and used for downstream amplification. The PCR products were sent to the Central Analytical Facility at Stellenbosch University, South Africa for fragment analysis. The software program, Gene Marker® v2.4.0 (Soft Genetics) was used to score the electropherograms. After this step, only polymorphic markers that amplified more than two alleles across eight *A. mearnsii* samples of known genetic variation in the breeding programme were selected for further analyses.

PCR amplification

The initial PCR reactions were performed using the DreamTaq master mix, ready-to-use solution containing DreamTaq DNA Polymerase, optimized DreamTaq buffer, $MgCl_2$, and dNTPs, from Thermo Scientific. Each reaction mixture contained; ~150 ng DNA, 5 µL DreamTaq PCR master mix (Thermo Scientific), 0.4 µM of each primer, 0.8 µL BSA (1mg/ml) and dH_2O to make up the final volume of 10 µL. A gradient PCR was performed to determine the optimal annealing temperature (Ta) for each primer set tested. The temperature range for the gradient PCR was set from 48–60 °C, while the remaining parameters set were according to the manufacturer's protocols. The PCR products were viewed on a 2% MetaPhor™ agarose gel. The optimum Ta was established for each primer set and the touchdown PCR method was used for all subsequent PCR amplifications using the DreamTaq master mix. The touchdown protocol had the following conditions: initial denaturation step at 95 °C for 3 min, 10 cycles of 95 °C for 20 seconds, Ta 10 °C higher than the optimum and decreasing every cycle by 1 °C for 20 seconds and an elongation step of 72 °C for 30 seconds, followed by 25 cycles of 95 °C for 20 seconds, optimum Ta for 20 seconds, 72 °C for 30 seconds, with a final elongation step of 72 °C for 5 minutes. Additional magnesium chloride ($MgCl_2$; 1 mM) was added to PCR mixtures for primer sets with weak amplicons. The various annealing temperatures for each primer pair (Ta = 48–52 °C) and the different levels of amplicons across the primer sets, made multiplexing difficult. Each primer was, therefore, amplified in single-plex and added to a single tube for fragment analysis as a 'multiplex'. Four multiplex sets (Multiplex A, B, C and D) were established (Table 2). All amplified products were sent to the Central Analytical Facility at Stellenbosch University, South Africa for fragment analysis. Gene Marker® v2.4.0 was used for scoring all genotypes. This process was time consuming and an alternative was needed.

The KAPA2G Fast Multiplex kit (KAPA Biosystems, Cape Town, South Africa) was tested for its multiplexing utility, using the selected *Acacia* SSR primer sets. The KAPA2G Fast Multiplex kit contains KAPA2G Fast HotStart DNA Polymerase, a buffer optimised for multiplex PCR, with 0.2 mM of each dNTP and 3 mM $MgCl_2$ (at 1X). This kit is pre-optimized for multiplex PCR reactions to use with primer interactions, primer concentrations, DNA quality and quantity, with marginal changes to annealing temperatures. Each PCR reaction mixture contained: ~150 µL DNA, 5 µL KAPA2G Fast Multiplex mix, 0.6 µL BSA (1mg/mL), 0.1–0.4 µM of each primer and dH_2O to make up the final volume of 10 µL. The PCR cycle

TABLE 2: Detail of the eleven selected SSR markers with bands amplified in *A. mearnsii*.

Multiplex	Dye	Primer sequence (5'-3')	Repeat motifs	Type repeat	T_a	Size range (bp)	N_a	Source
A								
AH2-1	FAM	F: GACAGAGGGAGCATTTTGTA	$(CT)_{12}$	Di-	60	146-160	10	Aggarwal et al. 2011
		R: CAGACAAGACCAGAGAATGAC						
AH3-18	FAM	F: TGAGACAATTAATGGTGGTG	$(TAA)_5$	Tri	60	209-221	4	Aggarwal et al. 2011
		R: TTTACAAGGGAAAAGCTGAG						
Am465	Cy 3.5	F: TGGGTATCACTTCCACCATT	$(AC)_{23}$	Di-	60	113-131	5	Butcher et al. 2000
		R: AGGCTGCTTCTTTGTGCAGG						
B								
AH3-1	PET	F: CTAAGGCACTTGGATCATTC	$(TCT)_5$	Tri-	60	214-217	2	Aggarwal et al. 2011
		R: AGAGAGAGAGAGGCACACTG						
AH3-10	FAM	F: AGGGATATCGGATGCTTACT	$(GAT)_7$	Tri-	60	178-202	10	Aggarwal et al. 2011
		R: AAAGATGCAGCAGACCTATC						
Ak15	VIC	F: CACCCCCACGTTATCTTACA	$(TAT)_5$	Tri-	60	297-309	4	Adamski et al. 2013
		R: GACTGGCGAAAGAGTCGAA						
C								
AH16	HEX	F: GAGGGTAATGCTTCAAGTAGAC	$(GA)_{16}$	Di-	60	86-88	2	Ng et al. 2005
		R: TGCGTGTCTCCCCACTACTC						
AH56	Cy3.5	F: GATAGCTCATAGAAACACCATACC	$(GA)_9$	Di-	60	123-129	4	Ng et al. 2005
		R: GGCGAAGCTCTCTCTCTCTCTCTCTCT						
Ak89	FAM	F: AGGGGAAGGACGAAAGTTGT	$(AC)_7$	Di-	60	160-174	5	Adamski et al. 2013
		R: GCAAGAGGAGCTTCAAGTGG						
D								
AH01	FAM	F: TTGAGGTTGAGGGTGATGAA	$(GA)_6$	Di-	60[#]	106-116	5	Ng et al. 2005
		R: GGCAAGCCTCTCTCTCTCT						
AH2-13	NED	F: GAAGAAGCAGGAGGAGGTAG	$(AG)_7$	Di-	60[#]	143-151	7	Aggarwal et al. 2011
		R: TGTTTTCCACTTCTCACACA						

[#] 0.5 mM additional $MgCl_2$ added; T_a= annealing temperature; N_a= number of observed alleles. Only the annealing temperature used for the KAPA2G Fast Multiplex kit is reported.

parameters were set up following the manufacturer's protocols. The PCR cycle parameters were as follows: initial denaturation at 95 °C for 3 minutes, 30 cycles of 95 °C for 15 seconds, 60 °C for 30 seconds, and 72 °C for 20 seconds, with a final elongation step of 72 °C for 1 minute. The above mentioned multiplex sets were used for amplification and the amplified products were sent to the Central Analytical Facility at Stellenbosch University, South Africa for fragment analysis. All loci were successfully amplified using the KAPA2G Fast Multiplex kit. All subsequent amplifications were done using this kit.

Statistical analyses

Estimates of null allele frequencies were performed using the software program FreeNA (Chapuis and Estoup 2007) using the Expectation Maximization Algorithm (EM) (Dempster et al. 1977). The influence of null alleles on genetic diversity estimates was assessed with a Wilcoxon signed rank test using R (R Core Team 2015) and F_{st} per locus values for corrected uncorrected null alleles were made using the excluding null alleles (ENA) method (Chapuis & Estoup 2007). Genetic diversity per locus was assessed using the mean number of alleles (N_a), observed (H_o) and unbiased expected (uH_e) heterozygosity using GenAlEx v6.5 (Peakall & Smouse 2012), and the inbreeding coefficient (F_{is}) was determined using Genepop v4.3 (Rousset 2008). Polymorphic information content (PIC) was estimated using Cervus v3.0 (Kalinowski et al. 2007), probability of identity (P_{ID}) for unrelated individuals and probability of identity for full siblings (P_{IDSibs}) were estimated using GeneAlEx V6.5, and allelic richness (Ar), using the rarefaction method, as implemented in FSTAT (Goudet 2001). Deviations from Hardy-Weinberg equilibrium were calculated using Genepop v4.3. P_{ID} and P_{IDSibs} estimate the probability that two randomly chosen full-sibs within a given population that have the same genotype on a set of makers. The P_{ID} assumes that there is no linkage disequilibrium and population substructure. When such assumptions do not hold, the P_{IDSibs} is often used as a conservative upper bound of the "real" probability.

Results

DNA extraction, quality and quantity

In this study, the CTAB DNA isolation technique was employed to extract DNA from *Acacia mearnsii* leaf samples with slight modifications. The mean concentration and purity of DNA samples extracted from the leaves of *Acacia mearnsii* based on the five protocols tested are presented in Table 1. The mean DNA concentration varied significantly ranging from 44 to 159 ng/µL, at low concentration of NaCl and at high concentration of polyvinyl polypyrrolidone (PVPP), respectively. The addition of a high concentration NaCl and PVPP provided the best results, with high DNA concentrations (>100 ng/µL) and near optimum wavelength ratios (mean 260/280 = 1.887; mean 260/230 = 1.754). Efficient DNA extraction was achieved with the optimised CTAB protocol, with less than 5% re-extraction. However, the 260/230 ratio values from the spectrophotometer assessments indicated that some residual phenols and/or carbohydrates might still be present in the DNA extracts. Subsequently, Bovine Serum Albumin (BSA) was used during PCR amplification to eliminate the effect of the co-extracted compounds.

SSR screening and summary statistics

Of the 49 assessed microsatellite makers developed for other Acacia species, only 11 (22.5%) loci amplified clear reproducible high-quality DNA bands in *A. mearnsii* (Table 2). These markers were polymorphic and consistent amplification was achieved among the four populations of *A. mearnsii* tested. The proportion of missing data ranged from 0% to 3.5% per locus. Missing data of up to ~4% of the scored genotypes is, however, acceptable for population genetic studies (Putman and Carbone 2014) and no genotype was eliminated from the analysis. The proportion of null alleles calculated per locus over all samples ranged from 0.00 (*Ak89*) to 25.8 (*AH3-10*), with a mean of 10.8 over all loci. The null allele frequencies (NAF) for locus *AH2-13* and *AH3-10* were higher than 20% (Table 3). The observed null allele frequencies per population for *AH2-13 and AH3-10* ranged from 15.5–20.6% and from 19.4–32.1% per population, respectively (data not shown). Although the values are high, they are still within the range of values often reported in other studies using SSR loci develop in other species (Dakin and Avise 2004). There was no significant difference detected between Excluding Null Allele (ENA) corrected and uncorrected F_{ST} values (*p*-value > 0.05; Bonferroni corrected) and therefore, it was decided to keep all loci for the subsequent analyses. The genotypes showed a wide range of allelic diversity from 2 to 10 alleles per locus. The highest allele number (N_a = 10) was observed at markers *AH2-1* and *AH3-10* and the lowest was for *AH16*, with an overall mean N_a of 5.3 (Table 2). Based on alleles detected among four multiplex panels in all four *A. mearnsii* populations, di-nucleotide repeat SSRs were relatively more polymorphic than those with tri-nucleotide repeats.

The unbiased expected heterozygosity (uH_e) ranged from 0.103 to 0.796 per locus, while the observed heterozygosity (H_O) varied from 0.058 to 0.97 (Table 3). Only three loci (*AH3-1, AH3-18* and *Am465)* did not show significant deviation from Hardy-Weinberg equilibrium (Table 3). A high level of variation was observed for the inbreeding coefficient (F_{IS}) estimates among loci (F_{IS} = -0.567 to 0.653). Negative F_{IS} values were only observed at two loci (*AH3-18*, *Ak89*), indicating a heterozygote excess for *Ak89*. Two loci were found to be highly polymorphic (*AH2-1* and *AH3-10*). Both loci had PIC values greater than 0.7, with high allelic richness (Ar) values. More than 50% of the markers had PIC values greater than 0.50. The combined P_{ID} value for the SSR panel was 7.2×10^{-8}. Therefore 1 in about 14 million trees will have the same genotype. The P_{IDSibs} value of 1×10^{-3} provides a lower bound for the number of loci required for the successful identification of individuals, with 1 in about 1000 individuals sharing the same genotype if all individuals are full siblings. This indicates that these markers had a high discriminatory power and were found to be highly suitable for genetic diversity analysis. The allelic richness (A_r) estimates were based on a minimum sample size of 30 individuals, the values ranged from 2 (*AH16 and AH3-1*) to 9 (*AH3-1*), with a mean of 4.8 alleles per locus.

Discussion

Exploiting the differences in solubility of polysaccharides and DNA in the CTAB buffer by adjusting the concentration of sodium chloride can aid the removal of polysaccharides (Weising et al. 2005). It has been

TABLE 3: Genetic diversity parameter values measured per locus over all samples

Locus	N_A	Ar	H_o	uH_e	F_{IS}	P_{ID}	P_{IDSibs}	PIC	NAF	HWD *p*-value
AH2-1	8.25	9.96	0.514	0.779	0.393	0.057	0.36	0.792	0.18	0.00
AH2-13	4.00	7.96	0.216	0.446	0.608	0.340	0.60	0.410	0.21	0.00
AH3-18	3.75	4.97	0.407	0.431	0.001	0.360	0.62	0.395	0.00	0.19
Am465	4.25	5	0.634	0.663	0.042	0.170	0.46	0.608	0.02	0.37
AH3-10	9.00	10	0.336	0.796	0.575	0.054	0.36	0.794	0.26	0.00
AH3-1	2.00	2	0.095	0.102	0.080	0.780	0.88	0.115	0.02	0.23
Ak15	3.25	4	0.361	0.563	0.403	0.270	0.53	0.485	0.16	0.00
AH16	2.00	2	0.058	0.164	0.653	0.710	0.85	0.151	0.14	0.00
AH56	4.50	6	0.337	0.598	0.445	0.210	0.49	0.557	0.17	0.00
Ak89	5.50	6.97	0.970	0.608	-0.567	0.210	0.50	0.551	0.00	0.00
AH01	6.25	7	0.622	0.639	0.005	0.170	0.48	0.594	0.04	0.00
Mean	4.80	5.99	0.416	0.524	0.240	-	-	0.496	0.11	-
SE	0.36	0.83	0.041	0.035	0.11	-	-	0.067	0.03	-

Ar= allelic richness; H_o= observed heterozygosity; H_e= unbiased expected heterozygosity; F_{IS}= Inbreeding coefficient; P_{ID}= probability of identity per locus; P_{IDSibs}= probability of identity for full siblings per locus; PIC= Polymorphic information content; NAF(%)= null allele frequency as percentage; HWD= deviation from Hardy-Weinberg; SE= Standard error

described that a high salt concentration in the extraction buffer assists in eliminating polysaccharides by increasing their solubility in ethanol (Fang et al. 1992; Lodhi et al. 1994; Varma et al. 2007). The addition of NaCl at concentrations higher than 0.5 M, along with CTAB, successfully removes polysaccharides during DNA extraction (Moreira & Oliveira 2011; Lucas et al. 2019). In the present study, a concentration of 1.4 M NaCl was used in the extraction buffer that further improved the quality of the extracted DNA. Endogenous DNases can degrade the extracted DNA unless EDTA is added (Weising et al. 2005). EDTA has an inhibitory effect on magnesium-dependant DNases by binding to magnesium ions through chelation (Weising et al. 2005). In this study, the addition of PVPP, which have strong H-receptor for binding and removal of polyphenolics into the CTAB buffer was helpful in removing the polyphenols and polysaccharides from leaf samples of *Acacia mearnsii* (Kolosova et al. 2004).

This protocol resulted in a mean total DNA amount of 15.95 µg extracted from 150 mg leaf tissue. Similar results were obtained with modified CTAB methods in cotton (15–30 µg from 100 mg plant tissue; Ali et al. 2019), as well as *Arabidopsis thaliana*, *Zea mays* and *Nicotiana sp.* (5–30 µg from 200 mg plant tissue; Allen et al. 2006). However, the values obtained in this study are much lower than those obtained from other species characterized by high levels of secondary metabolites (Sahu et al. 2012; Arruda et al. 2017). Tiwari et al. (2012), obtained a DNA concentration ranged from 179 to 833 ng/µL using the modified CTAB protocol in selected medicinal plants. Similarly, Sahu et al. (2012) achieved DNA concentration ranged from 8.8 to 9.9 µg/µL that was amenable of RAPD markers analysis. The difference in DNA concentration could be attributed to the various modifications such as high salt concentration, PVP, PVPP and avoiding the use of liquid nitrogen and selective precipitation and washing steps in addition to species differences.

The 260/280nm and 260/230nm wavelength ratios are well-known measures of nucleic acid quality. The recommended values for the 260/280 ratio ranged from 1.8 to 2.0 and the optimised CTAB protocol in this study resulted in an absorbance value of 1.89, which is within the accepted range indicating the protocol is efficient in obtaining high-quality DNA samples. The quality of the DNA obtained using the optimized protocol is sufficient since ratios in the range of 1.6–1.8 are acceptable for PCR reactions. Similarly, in a pure DNA sample, the reference interval for 260/230 ratio is 2.0–2.2, however, the ratio obtained in this study was 1.75 which falls below the recommended level. This value was similar to those reported by Murray and Thompson (1980) but lower than values reported by Arruda et al. (2017). This indicates that the quality of the DNA is low due to possible contamination by residual polyphenolic compounds and/or carbohydrates in the extract (Moncada et al. 2013). The high quality DNA obtained by Arruda et al. (2017) might be attributed to the higher concentration of CTAB (3%) and NaCl (2.5 M) compared to CTAB (2%) and NaCl (1.4 M) used in this study. This could aid to more efficient elimination of polysaccharides since the composition and concentration of reagents can interfere with the quality and quantity of extracted DNA (Borges et al. 2012). In addition, the low 260/280 and 260/230 ratios obtained in this study might be attributed to high quantities of tannin in *A. mearnsii* leaves (Elgailani & Ishak 2014).

The systematic exploration of microsatellite markers across species should be the first step in developing SSRs

in under-studied species. Developing new SSR primers specific to the species of interest is both time consuming and costly (Yosodha et al. 2005; Ravishankar et al. 2015). It is, therefore, more feasible to source SSR primers developed for other related species. The use of SSR markers across species from the same genus (Omondi et al. 2010; Aggarwal et al. 2011; Adamski et al. 2013; Le Roux et al. 2013; Roberts et al. 2013) and across different genera (Peakall et al. 1998) have been reported for Leguminosae. In this study, the possibility of cross-species transferability of 49 microsatellites, derived from *A. mangium*, *A. koa* and *A. auriculiformis,* into *A. mearnsii* was investigated. Of the tested 49 cross-species markers, only 11 were effective in amplifying sharp and high intensity bands. The transfer rate was 22.5%, which is moderately low. This result was in agreement with Butcher et al. (2000), who reported low levels of cross-species SSR amplification within *Acacia* species. The low transferability might be attributed to the genetic divergence among the *Acacia* species due to exposure to different bottlenecks and geographic isolation (Varshney et al. 2005; Barbará et al. 2007). Similarly, several studies indicated the low transferability rate of SSRs among plant species (Luro et al. 2008; Koppolu et al. 2010; Lee et al. 2011). Although transferability was in general very low, the mean number of alleles amplified per locus was relatively higher (6.0) in this study compared to the 2.9 allele per locus reported by (Adamski et al. 2013) using 16 *Acacia koa* SSR markers on seven *Acacia* species. The high levels of polymorphisms of SSR loci indicated that the application of these markers for genetic diversity studies in *A. mearnsii* was useful and cost-effective compared to developing new SSR markers specific to the species. This study provided an important insight in the development of microsatellite markers suitable for genetic studies in *A. mearnsii*. In the current study, we only used SSRs to study their cross-species transferability and on the genetic diversity of four *A. mearnsii* breeding population. However, their genomic distribution, biological function and other uses should be further investigated (Vieira et al. 2016).

Various factors such as genotyping errors of loci should be considered in choosing a sufficient SSR panel for population genetic and parentage studies. A well-known problem with SSR is the occurrence of null alleles and over estimation of alleles due to polymerase strand-slippage in DNA replication (Vieira et al. 2016). The analysis of null allele frequencies per genotype conducted using the maximum likelihood estimates in this study confirmed the presence of null alleles in all loci but *Ak89*. For these loci, null-allele frequencies estimates ranged between 0% and 25.8%, with 64% of the loci showing null-allele frequencies above 5%. In this case, the null alleles occur when an allele fails to amplify via PCR and heterozygotes might falsely be scored as homozygote. Null alleles can possible be either caused by mutations in the flanking regions of a microsatellite sequence (Chapuis & Estoup 2007; Dakin & Avise 2004), or associated with shorter alleles outperforming the longer alleles, usually due to lower DNA quality or quantity (Gagneux et al. 1997; Wattier et al. 1998). Consequently, null alleles may inflate genetic diversity parameter estimates (Chapuis & Estoup 2007; Carlsson 2008) and result in inaccurate parentage assessments (Dakin & Avise 2004). Carlsson (2008) reported that in datasets containing null alleles, the genetic differentiation estimate F_{ST} were slightly overestimated. If the frequency of a null allele is <0.1, some estimators can be used directly without adjustment; if it is >0.5, the potency of estimation is too low and such a locus should be excluded (Huang et al. 2016). In this study, the estimated null allele frequencies were generally below the threshold levels.

All SSR loci were used for the downstream diversity analyses. It was reported by Carlsson (2008) that loci affected by null alleles can still be useful for population level studies. However, breeders should interpret the results with caution as the effect of null alleles on parentage analysis is more substantial, as it could lead to false parentage assignments (Pemberton et al. 1995). Since these breeding populations were derived from the same base population, it is therefore, advisable to use parentage assignment methods that consider the presence of null alleles in the dataset (Dakin & Avise 2004) during further selection and breeding planning.

The mean genetic diversity value observed in *A. mearnsii* (H_e = 0.524) in this study was much higher than the study conducted on *A. longifolia* in Portugal (H_e = 0.190) (Vicente et al. 2018) using SSRs. However, the value is more comparable with a study of *A. senegal* (H_e = 0.519) (Djibo et al. 2017), *A. auriculiformis* (H_e = 0.60) (Son et al. 2016) and *A. mangium* (H_e = 0.56) (Son et al. 2016) using SSRs. The inbreeding coefficient (F_{IS}) measures the deviation of heterozygosity from expected values under the assumptions of Hardy Weinberg. A negative and positive F_{IS} value indicates an excess or deficiency of heterozygosity in a population relative to the ratio under HWE. All the markers except Ak89 revealed a positive F_{IS} value suggesting a deficiency of heterozygotes. The deficiency of heterozygotes can be attributed to either inbreeding or existence of null alleles or the presence of subpopulations within populations (Wahlund effect) (Jordana 2003). Of the 11 microsatellite loci analysed, 9 loci deviated from HWE in *A. mearnsii* ($P < 0.001$). This might be due to the small sample size used or the existence of null alleles (Wang 2008). Linkage disequilibrium observed among pairs of loci in this study appeared to be associated with the presence of null alleles in some of the analysed loci. All the microsatellite loci, except *Ak89* and *AH01*, which were found to depart from the Hardy-Weinberg equilibrium (HWE), had null allele frequencies exceeding 10%.

The probability of identity (P_{ID}) is also a widely used individual identification estimator (Peakall & Sydes 1996; Reed et al. 1997; Waits et al. 2001; Ferrie et al. 2013), which provides the probability that two randomly drawn individuals within a given population will have identical genotypes at multiple loci. For P_{ID}, it is assumed that there is neither linkage disequilibrium nor population substructure. In the case of such assumptions not holding true, the P_{ID} might drastically underestimate the above-mentioned probability. Hence,

probability of identity for siblings ($P_{IDsibling}$) is often used as a conservative upper bound of the "real" probability (Waits et al. 2001). Low P_{ID} values are required to accurately estimate individual identity and not falsely classify individuals as identical (Waits et al. 2001; Vidya & Sukumar 2005). Locus P_{ID} values in the range of 0.01 – 0.0001 are generally used in wildlife forensics (Waits et al. 2001), suggesting the values in the current study are slightly higher than ideal. It is important to take the population size and genetic history of the species under study into account in choosing an appropriate marker panel using P_{ID} estimates (Schwartz & Monfort 2008). For example, if the study population comprise of 10000 individuals, a $P_{ID} < 0.0001$ (1 in about 10000 individuals) should be sufficient for individual identification.

Conclusions

In this study, we managed to get high quantity and quality DNA from *A. mearnsii* leaves by modifying the CTAB-based DNA extraction procedure using high concentration of NaCl to remove polysaccharides and PVPP to eliminate polyphenols during DNA purification. In addition, by combining the CTAB and NaCl steps in the extraction protocol, we succeeded in the extraction of DNA 10–20 min faster than the normal protocol. According to our knowledge, this is the first assessment of cross-species microsatellite marker transferability and the use of these markers to study the genetic variation in *A. mearnsi*. However, it was also demonstrated that the majority of cross-species microsatellite markers were not transferable across species in the genus *Acacia*. Considering the high polymorphism demonstrated by the studied microsatellite markers, the cross-species amplification is an interesting alternative to the development of new microsatellites in *A. mearnsii*. These multi-allelic, PCR-based, co-dominant microsatellite loci provides a powerful tool for genetic, breeding and conservation genetic studies of *A. mearnsii*.

Competing interests

The authors declare no competing interest. The institution where the research was done, and the funding body are duly acknowledged.

Authors' contributions

MB designed, initiated and managed the project, and contributed to laboratory work and manuscript preparation. WC undertook most of the laboratory work, and contributed to data analysis and manuscript preparation. AA undertook data analysis and manuscript preparation.

Additional Files

Additional File 1: List of 49 cross-species SSR markers used in this study

Acknowledgements

The Institute for Commercial Forestry Research (ICFR), South Africa for affording the authors the opportunity to do this research and the South African Wattle Growers Union (through ICFR) for making the funding available.

References

Adamski, D.J., Dudley, N.S., Morden, C.W. & Borthakur, D. (2013). Cross-amplification of nonnative *Acacia* species in the Hawaiian Islands using microsatellite markers from *Acacia koa*. *Plant Biosystems 147*, 1088-1091. https://doi.org/10.1080/11263504.2012.749958

Aggarwal, R.K., Allainguillaume, J., Bajay, M.M., Barthwal, S., Bertolino, P., Chauhan, P., et al. (2011). Permanent Genetic Resources added to Molecular Ecology Resources Database 1 August 2010 – 30 September 2010. *Molecular Ecology Resources 11*, 219-222. https://doi.org/10.1111/j.1755-0998.2010.02944.x

Ali, Q. Salisu, I.B., Raza, A., Shahid, A.A., Rao, A.Q., & Husnain, T. (2019). A modified protocol for rapid DNA isolation from cotton (*Gossypium* spp.). *MethodsX 6*, 259-264. https://doi.org/10.1016/j.mex.2019.01.010

Allen, G.C., Flores-Vergara, M.A., Krasynanski, S., Kumar, S., & Thompson. W.F. (2006). A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. *Nature Protocols 1*, 2320-2325. https://doi.org/10.1038/nprot.2006.384

Arruda, S.R., Pereira, D.G., Silva-Castro, M.M., Brito, M.G., & Waldschmidt, A.M. (2017). An optimized protocol for DNA extraction in plants with a high content of secondary metabolites, based on leaves of *Mimosa tenuiflora* (Willd.) Poir. (Leguminosae). *Genetics and Molecular Research 16*, 1-9. https://doi.org/10.4238/gmr16039063

Bairu, M.W., Fennell, C.W., & van Staden, J. (2006). The effect of plant growth regulators on somaclonal variation in Cavendish banana (Musa AAA cv. 'Zelig'). *Scientia Horticulturae 108*, 347–351. https://doi.org/10.1016/j.scienta.2006.01.039

Barbará, T., Palma-Silva, C., Paggi, G.M., Bered, F., Fay, M.F., & Lexer, C. (2007). Cross-species transfer of nuclear microsatellite markers: Potential and limitations. *Molecular Ecology 16*, 3759-3767. https://doi.org/10.1111/j.1365-294X.2007.03439.x

Bonin, A., Bellemain, E., Bronken, E.P., Pompanon, F., Brochmann, C., & Taberlet, P. (2004). How to track and assess genotyping errors in population genetics studies. *Molecular Ecology 13*, 3261-3273. https://doi.org/10.1111/j.1365-294X.2004.02346.x

Borges, D.B., Amorim, M.B., Waldschmidt, A.M., Mariano-Neto, E., Vivas, C.V., & Pereira, D.G. (2012). Optimization of DNA extraction from fresh leaf tissues of *Melanoxylon brauna* (Fabaceae). *Genetics and Molecular Research 11*, 1586-1591. https://doi.org/10.4238/2012.May.22.8

Butcher, P.A., Decroocq, S., Gray, Y., & Moran, G.F. (2000). Development, inheritance and cross-species amplification of microsatellite markers from *Acacia*

mangium. *Theoretical and Applied Genetics 101*, 1282-1290. https://doi.org/10.1007/s001220051608

Carlsson, J. (2008). Effects of microsatellite null alleles on assignment testing. *Journal of Heredity 99*, 616-623. https://doi.org/10.1093/jhered/esn048

Chapuis, M.-P., & Estoup, A. (2007). Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution 24*, 621-631. https://doi.org/10.1093/molbev/msl191

CIMMYT (Centro Internacional de Mejoramiento de Maíz y Trigo). (2005). *Laboratory Protocols*: *CIMMYT Applied Molecular Genetics Laboratory*. 3rd ed. El Batán, Mexico: CIMMYT https://repository.cimmyt.org/xmlui/handle/10883/1333

Dakin, E.E., & Avise, J.C. (2004). Microsatellite null alleles in parentage analysis. *Heredity 93*, 504-509. https://doi.org/10.1038/sj.hdy.6800545

De Beer, H. (1986). *A. mearnsii*. [Weeds A. 24/1986]. Pretoria, South Africa: Department of Agriculture and Water Supply.

Dempster, A.P., Laird, N.M., & Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society. Series B (methodological) 39*, 1-38. https://doi.org/10.1111/j.2517-6161.1977.tb01600.x

Department of Water Affairs and Forestry (2003). *South African Forestry Facts for the Year 2001/2002*. [Newsletter].

Djibo, E.S., Aichatou, A., Manssour, A.M., Bilassanou, H.I., Djibo, M., Moussa, T., & Zouberou, A.M. (2017). Exploring genetic diversity and structure of *Acacia senegal* (L.) Willd. to improve its conservation in Niger. *African Journal of Biotechnology 16*,1650-1659. https://doi.org/10.5897/AJB2016.15846

Doyle J.J. and Doyle J.L. (1990). Isolation of plant DNA from fresh tissue. *Focus 12*, 13-15.

Duminil, J., Brown, R.P., Ewédjè, E-E.B.K., Mardulyn, P., Doucet, J-L., & Hardy O.J. (2013). Large-scale pattern of genetic differentiation within African rainforest trees: Insights on the roles of ecological gradients and past climate changes on the evolution of *Erythrophleum* spp. (Fabaceae). *BMC Evolutionary Biology 13:* 195.

Dunlop, R.W., Barnes, R.D. Nixon, K.M. & Hagedorn, S.F. (2003). A breeding strategy for *Acacia measrnsii* (black wattle) In South Africa. *Southern African Forestry Journal 199*, 55-63. https://doi.org/10.1080/20702620.2003.10431749

Elgailani, I.E.H., & Ishak, C.Y. (2014). Determination of tannins of three common *Acacia* species of Sudan. *Advances in Chemistry* 2014, 1-5. https://doi.org/10.1155/2014/192708

Fang, G., Hammar, S., & Grumet, R. (1992). A quick and inexpensive method for removing polysaccharides from plant genomic DNA. *Biotechniques 13*, 52-56.

Ferrie, G.M., Cohen, O.R., Schutz, P., Leighty, K.A., Plasse, C., Bettinger, T.L., & Hoffman, E.A. (2013). Identifying parentage using molecular markers: Improving accuracy of studbook records for a captive flock of marabou storks (*Leptoptilus crumeniferus*). *Zoo Biology 32*, 556-564. https://doi.org/10.1002/zoo.21090

Gagneux, P., Boesch, C., & Woodruff, D.S. (1997). Microsatellite scoring errors associated with non-invasive genotyping based on nuclear DNA amplified from shed hair. *Molecular Ecology 6*, 861–868. https://doi.org/10.1111/j.1365-294X.1997.tb00140.x

Goudet, J. (2001). *FSTAT: A program to estimate and test gene diversities and fixation indices* (version 2.9.3). Available from http://www2.unil.ch/popgen/softwares/fstat.htm [Accessed 22 September 2018].

Healey, A., Furtado, A., Cooper, T., & Henry, R.J. (2014). Protocol: a simple method for extracting next-generation sequencing quality genomic DNA from recalcitrant plant species. *Plant Methods 10*: 21. https://doi.org/10.1186/1746-4811-10-21

Huang, K., Ritland, K., Dunn, D.W., Qi, X., Guo, S., & Li, B. 2016. Estimating relatedness in the presence of null alleles. *Genetics 202*, 247-260. https://doi.org/10.1534/genetics.114.163956

ILDIS. (2015). *International Legume Database and Information Service*. Reading, UK:, University of Reading School of Plant Sciences. http://www.ildis.org

Jordana, J. (2003). Genetic structure of eighteen local South European beef cattle breeds by comparative F-statistics analysis. *Journal Animal Breeding and Genetics 120*, 73-87. https://doi.org/10.1046/j.1439-0388.2003.00384.x

Kalinowski, S.T., Taper, M.L., & Marshall, T.C. (2007). Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology 16*, 1099-1106. https://doi.org/10.1111/j.1365-294X.2007.03089.x

Kolosova N., Miller B., Ralph S., Ellis B.E., Douglas C., Ritland K. and Bohlmann J. (2004). Isolation of high-quality RNA from gymnosperm and angiosperm trees. *Biotechniques 35*, 821-824. https://doi.org/10.2144/04365ST06

Koppolu, R., Upadhyaya, H., Dwivedi, S., Hoisington, D., & Varshney, R. (2010). Genetic relationships among seven sections of genus *Arachis* studied by using SSR markers. *BMC Plant Biology 10*:15. https://doi.org/10.1186/1471-2229-10-15

Lade, B.D., Patil, A.S., & Paikrao, H.M. (2014). Efficient genomic DNA extraction protocol from medicinal rich *Passiflora foetida* containing high level of polysaccharide and polyphenol. *SpringerPlus* 3: 457. https://doi.org/10.1186/2193-1801-3-457

Le Roux, J.J., Richardson, D.M., Wilson, J.R.U., & Ndlovu, J. (2013). Human usage in the native range may determine future genetic structure of an invasion: insights from *Acacia pycnantha*. *BMC Ecology* 13: 37. https://doi.org/10.1186/1472-6785-13-37

Lee, G-A., Kwon, S-J., Park, Y-J., Lee, M.C., Kim, H.H., Lee, J.S., Lee, S-Y., Gwag, J-G., Kim, C-K., & Ma, K-H. (2011). Cross-amplification of SSR markers developed from *Allium sativum* to other *Allium* species. *Scientia Horticulturae 128*, 401-407. https://doi.org/10.1016/j.scienta.2011.02.014

Leberg, P.L. (2002). Estimating allelic richness: Effects of sample size and bottlenecks. *Molecular Ecology 11*, 2445–2449. https://doi.org/10.1046/j.1365-294X.2002.01612.x

Lodhi, M.A., Ye, G.N., Weeden, N.F., & Reisch, B.I. (1994). A simple and efficient method for DNA extraction from grapevine cultivars and *Vitis* species. *Plant Molecular Biology Reporter 12*, 6-13. https://doi.org/10.1007/BF02668658

Lucas, M.S., Carvalho, C.S. Hypolito, G.B., & Côrtes, M.C. (2019). Optimized protocol to isolate high quality genomic DNA from different tissues of a palm species. *Hoehnea* 46: e942018. https://doi.org/10.1590/2236-8906-94/2018

Luro, F.L., Costantino, G., Terol, J., Argout, X., Allario, T., Wincker, P., Talon, M., Ollitrault, P., & Morillon, R. (2008). Transferability of the EST-SSRs developed on Nules clementine (*Citrus clementina* Hort ex Tan) to other *Citrus* species and their effectiveness for genetic mapping. *BMC Genomics 9*: 287. https://doi.org/10.1186/1471-2164-9-287

Millar, M.A., Byrne, M., Nuberg, I., & Sedgley, M. (2008). High outcrossing and random pollen dispersal in a planted stand of *Acacia saligna* subsp. *saligna* revealed by paternity analysis using microsatellites. *Tree Genetics & Genomes 4*, 367-377. https://doi.org/10.1007/s11295-007-0115-z

Millar, M.A. 2009. Characterisation of microsatellite DNA markers for the rare *Acacia woodmaniorum* (*Leguminosae: Mimosaceae*). *Conservation Genetics Resources 1*, 441-445. https://doi.org/10.1007/s12686-009-9102-z

Moncada X., Payacán C., Arriaza F., Lobos S., Seelenfreund D. and Seelenfreund A. (2013). DNA extraction and amplification from contemporary Polynesian bark-cloth. *PLoS One 8*: e56549. https://doi.org/10.1371/journal.pone.0056549

Moreira, P.A., & Oliveira, D.A. (2011). Leaf age affects the quality of DNA extracted from *Dimorphandra mollis* (Fabaceae), a tropical tree species from the Cerrado region of Brazil. *Genetics and Molecular Research* 10, 353-358. https://doi.org/10.4238/vol10-1gmr1030

Moyo, M., Amoo, S.O., Bairu, M.W., Finnie, J.F., & Van Staden, J. (2008). Optimising DNA isolation for medicinal plants. *South African Journal of Botany 74*, 771–775. https://doi.org/10.1016/j.sajb.2008.07.001

Murray, M.G., & Thompson, W.F. (1980). Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Research 8*, 4321-4326. https://doi.org/10.1093/nar/8.19.4321

Ng, C.H., Koh, S., Lee, S., Ng, K., Mark, A., Norwati, M., & Wickneswari, R. (2005). Isolation of 15 polymorphic microsatellite loci in *Acacia* hybrid (*Acacia mangium* × *Acacia auriculiformis*). *Molecular Ecology Notes 5*, 572-575. https://doi.org/10.1111/j.1471-8286.2005.00994.x

Omondi, S.F., Kireger, E., Dangasuk, O.G., Chikamai, B., Odee, D.W., Cavers, S., & Khasa, D.P. (2010). Genetic diversity and population structure of *Acacia senegal* (L) Willd. in Kenya. *Tropical Plant Biology 3*, 59-70. https://doi.org/10.1007/s12042-009-9037-2

Owen, D.L., & van der Zel, D.W. (2000). Chapter 1: Trees, forests and plantations in Southern Africa. In D.L. Owen (Ed.) *South African Forestry Handbook*. Pretoria, South Africa: South African Institute of Forestry. pp. 3-8.

Peakall, R., & Smouse, P.E. (2012). GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics 28*, 2537-2539. https://doi.org/10.1093/bioinformatics/bts460

Peakall, R., & Sydes, M.A. (1996). Defining priorities for achieving practical outcomes from the genetic studies of rare plants. In S. Stephens, & S. Maxwell (Eds.), *Back from the brink: refining the threatened species recovery process*. (pp. 119-129). Sydney: Surrey Beatty and Sons.

Peakall, R., Gilmore, S., Keys, W., Morgante, M., & Rafalski, A. (1998). Cross-species amplification of soybean (*Glycine max*) simple sequence repeats (SSRs) within the genus and other legume genera: Implications for the transferability of SSRs in plants. *Molecular Biology and Evolution 15*, 1275-1287. https://doi.org/10.1093/oxfordjournals.molbev.a025856

Pemberton, J.M., Slate, J., Bancroft, D.R., & Barrett, J.A. (1995). Non-amplifying alleles at microsatellite loci: A caution for parentage and population studies. *Molecular Ecology 4*, 249–252. https://doi.org/10.1111/j.1365-294X.1995.tb00214.x

Putman, A.I., & Carbone, I. (2014). Challenges in analysis and interpretation of microsatellite data for population genetic studies. *Ecology and Evolution 4*, 4399-4428. https://doi.org/10.1002/ece3.1305

R Core Team (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available from: http://www.R-project.org.

Ravishankar, K., Dinesh, M., Nischita, P., & Sandya, B. (2015). Development and characterization of microsatellite markers in mango (*Mangifera indica*) using next-generation sequencing technology and their transferability across species. *Molecular Breeding 35*, 1–13. https://doi.org/10.1007/s11032-015-0289-2

Reed, J.Z., Tollit, D.J., Thompson, P.M., & Amos, W. (1997). Molecular scatology: The use of molecular genetic analysis to assign species, sex and individual identity to seal faeces. *Molecular Ecology 6*, 225-234. https://doi.org/10.1046/j.1365-294X.1997.00175.x

Roberts, D.G., Forrest, C.N., Denham, A.J., & Ayre, D.J. (2013). Microsatellite markers for vulnerable Australian arid zone Acacias. *Conservation Genetics Resources 5*, 199-201. https://doi.org/10.1007/s12686-012-9767-6

Rousset, F, (2008). Genepop'007: A complete reimplementation of the Genepop software for Windows and Linux. *Molecular and Ecology Resources 8*, 103-106. https://doi.org/10.1111/j.1471-8286.2007.01931.x

Sahu, S.K., Thangaraj, M., & Kathiresan, K. (2012). DNA extraction protocol for plants with high levels of secondary metabolites and polysaccharides without using liquid nitrogen and phenol. *International Scholarly Research Network Molecular Biology 2012: 205049.* https://doi.org/10.5402/2012/205049

Sablok, G., Gahlot, P., Gupta, A.K., Pareek, K., & Shekhawat, N.S. (2009). Extraction of PCR-usable DNA from trees adapted to arid environment. *Plant Omics 2*, 103-109.

Schwartz, M.K., & Monfort, S.L. (2008). Genetic and endocrine tools for carnivore surveys. In: R.A. Long, P. Mackay, W.J. Zielinksi, & J.C. Ray (Eds.), *Non-invasive Survey Methods for Carnivores*. (pp. 238–262). Washington DC, USA: Island Press.

Son L., Wickneswari R., Harwood C.E., Larcombe M.J., Griffin R.A., Koutoulis A., Harbard J.L., Cyer K.S., Yee L.W., Thinh H.H. and Vaillancourt R.E. (2016). A multiplexed set of microsatellite markers for discriminating *Acacia mangium, A. auriculiformis*, and their hybrid. *Tree Genetics & Genomes 12*, 1-10. https://doi.org/10.1007/s11295-016-0990-2

Tan, S.C., & Yiap, B.C. (2009). DNA, RNA, and protein extraction: The past and the present. *Journal of Biomedicine and Biotechnology* 2009: 574398. https://doi.org/10.1155/2009/574398

Thermo Scientific (2013). *Assessment of nucleic acid purity*. [T042 Technical Bulletin, NanoDrop Spectrophotometers].

Tiwari, K.L., Jadhav, S.K., & Gupta, S. (2012). Modified CTAB technique for isolation of DNA from some medicinal plants. *Research Journal of Medicinal Plants 6*, 5-73. https://doi.org/10.3923/rjmp.2012.65.73

Varma, A., Padh, H., & Shrivastava, N. (2007). Plant genomic DNA isolation: An art or a science. *Biotechnology Journal: Healthcare Nutrition Technology 2*, 386-392. https://doi.org/10.1002/biot.200600195

Varshney, R.K., Graner, A., & Sorrells, M.E. (2005). Genetic microsatellite markers in plants: Their features and applications. *Trends in Biotechnology 23*, 48-55. https://doi.org/10.1016/j.tibtech.2004.11.005

Vicente, S., Máguas, C., & Trindade, H. (2018). Genetic diversity and differentiation of invasive *Acacia longifolia* in Portugal. *Web Ecology 18*, 91-103. https://doi.org/10.5194/we-18-91-2018

Vidya, T.N.C., & Sukumar, R. (2005). Amplification success and feasibility of using microsatellite loci amplified from dung to population genetic studies of the Asian elephant (*Elephas maximus*). *Current Science 88*, 489-492.

Vieira, M.L.C., Santini, L., Diniz, A.L., & Munhoz, C.F. (2016). Microsatellite markers: What they mean and why they are so useful. *Genetics and Molecular Biology 39*, 312-328. https://doi.org/10.1590/1678-4685-GMB-2016-0027

Waits, L.P., Luikart, G., & Taberlet, P. (2001). Estimating the probability of identity among genotypes in natural populations: Cautions and guidelines. *Molecular Ecology 10*, 249-256. https://doi.org/10.1046/j.1365-294X.2001.01185.x

Wang, Y. (2008). Microsatellite loci transferability in chestnut. *Journal of American Society of Horticultural Science 133*, 692-700. https://doi.org/10.21273/JASHS.133.5.692

Wattier, R., Engel, C.R., Saumitou-Laprade, P., & Valero, M. (1998). Short allele dominance as a source of heterozygote deficiency at microsatellite loci: Experimental evidence at the dinucleotide locus Gv1CT in *Gracilaria gracilis* (Rhodophyta). *Molecular Ecology 7*, 1569–1573. https://doi.org/10.1046/j.1365-294x.1998.00477.x

Weising, K., Nybom, H., Pfenninger, M., Wolff, K., & Kahl, G. (2005). *DNA Fingerprinting in Plants: Principles, Methods, and Applications: Chapter 4 – Methodology*. (pp. 81-106). New York: CRC Press. https://doi.org/10.1201/9781420040043

Yasodha, R., Ghosh, M., Sumathi, R., & Gurumurthi, K. (2005). Cross-species amplification of *Eucalyptus* SSR markers in *Casuarinaceae*. *Acta Botanica Croatica 64*, 115-120.

 https://doi.org/10.33494/nzjfs492020x56x

Additional File 1: List of 49 cross-species SSR markers used in this study

Development of an SSR-based DNA fingerprinting method for black wattle (*Acacia mearnsii* De Wild)

Michael W. Bairu, Willem G. Coetzer, Assefa B. Amelework

No.	Name	Dye	Primer sequence (5'-3')	Selected (Y/N)	Repeat motifs	Nucleotide repeat	Annealing Temp.	Size range-bp	No. of alleles	Reference
1	AH2-1	FAM	F: GACAGAGGGAGCATTTTGTA R: CAGACAAGACCAGAGAATGAC	Y	$(CT)_{12}$	Di-	60	146-160	10	Aggarwal et al. 2011
2	AH2-2	VIC	F: CGGTTTAGCAGTCACAGAAG R: TACAAGCATCATCATGGAAG	N	$(CA)_{7}$	Di-	52	170	1	Aggarwal et al. 2011
3	AH2-4	FAM	F: GGATTATAAATGGCTGATCG R: TGGTCCCCTAACTACAAATG	N	$(TA)_{18}$	Di-	-	-	-	Aggarwal et al. 2011
4	AH2-5	None	F: TGAGTCAACCAACTTCCTTC R: CGATTCGCAACTAAAAGTG	N	$(TG)_{11}$	Di-	-	-	-	Aggarwal et al. 2011
5	AH2-6	None	F: ACAGTGGTATGATGGGAGTG R: CACTAACGTCACAATGATCG	N	$(TA)_{8}$	Di-	-	-	-	Aggarwal et al. 2011
6	AH2-9	HEX	F: CGTCTCATCGATCTTCTTTC R: GGAGGCATAACATCAAACAT	N	$(AT)_{7}$	Di-	60	152	1	Aggarwal et al. 2011
7	AH2-10	None	F: CAGTGTTGTGGTCCTTTTTC R: TTTACTCTCGAGCAAACACC	N	$(TA)_{10}$	Di-	-	-	-	Aggarwal et al. 2011
8	AH2-11	Cy3	F: CAAGCGTTCAGTAGAGGTTT R: CCCTGTAGATATAAGCACTGAAC	N	$(AT)_{7}$	Di-	60	222	1	Aggarwal et al. 2011
9	AH2-12	Cy3.5	F: CTTGTATTTCCATGGTGAGTC R: AGCGATTGATATCCTTGAGA	N	$(AT)_{10}$	Di-	-	-	-	Aggarwal et al. 2011
10	AH2-13	NED	F: GAAGAAGCAGGAGGAGGTAG R: TGTTTTCCACTTCTCACACA	Y	$(AG)_{7}$	Di-	60	143-151	7	Aggarwal et al. 2011
11	AH2-14	HEX	F: CGGAAGAAGAAGAAGAAGAA R: AATACAGCACTTGGCAACA	N	$(AT)_{25}$	Di-	-	-	-	Aggarwal et al. 2011
12	AH2-15	None	F: TCCGAAATGTTGAACTAAGG R: TATGAAAGCCAACCAGAAAC	N	$(CA)_{10}$	Di-	-	-	-	Aggarwal et al. 2011

 https://doi.org/10.33494/nzjfs492020x56x

No.	Name	Dye	Primer sequence (5'-3')	Selected (Y/N)	Repeat motifs	Nucleotide repeat	Annealing Temp.	Size range-bp	No. of alleles	Reference
13	AH2-17	None	F: AAATTCTCTTCGCAACCAC R: TCTGAGGTATTCCATTAGGC	N	$(CT)_7$	Di-	-	-	-	Aggarwal et al. 2011
14	AH01	FAM	F: TTGAGGTTGAGGGTGATGAA R: GGCAAGCCTCTCTCTCTCT	Y	$(GA)_6$	Di-	58#	106-116	5	Ng et al. 2005
15	AH16	HEX	F: GAGGGTAATGCTTCAAGTAGAC R: TGCGTGTCTCCCCACTACTC	Y	$(GA)_{16}$	Di-	60	86-88	2	Ng et al. 2005
16	AH18	Cy3.5	F: GGCGCAACTCTCTCTCTCT R: TTGGTCACTTAGCGCATGCC	N	$(CT)_6(CA)_6$	Di-	60	38	1	Ng et al. 2005
17	AH29	Cy3	F: GGCCATCTCTATCCATCTCC R: CCTTCCCCAATTCCTTTGCTT	N	$(GA)_{10}$	Di-	58#	122-134	4	Ng et al. 2005
18	AH56	Cy3.5	F: GATAGCTCATAGAAACACCATACC R: GGCGAAGCTCTCTCTCTCTCTCTCTCT	Y	$(GA)_9$	Di-	60	123-129	4	Ng et al. 2005
19	AH59	HEX	F: CTACGAGGCACAAAAGTT R: GTGAGGGCTCTCTCTCTCTCTCTCTCT	N	$(GA)_9$	Di-	58#	119-139	3	Ng et al. 2005
20	AH71	FAM	F: GGGGGAGCTCTCTCTCTCTCTCTCT R: GCTACTAAGGTTTCTTTCACGG	N	$(CT)_{11}$	Di-	-	-	-	Ng et al. 2005
21	AH76	HEX	F: GGGGAGGCTCTCTCTCTCTCTCTCT R: GTGACCTGAGTTAGGAAGGAGC	N	$(CT)_{10}T(CT)_4$	Di-	60	135-145	2	Ng et al. 2005
22	AH3-1	PET	F: CTAAGGCACTTGGATCATTC R: AGAGAGAGAGAGGCACACTG	Y	$(TCT)_5$	Tri-	60	214-217	2	Aggarwal et al. 2011
23	AH3-4	None	F: GATCTCAGCAGCAGCAAC R: CTGGTGGATGTGGTTTGG	N	$(AAG)_9$	Tri-	-	-	-	Aggarwal et al. 2011
24	AH3-6	HEX	F: AGCCAAGGTTGAGACTGTAA R: TCCTTTTTCTCAGCTTTGTC	N	$(AAG)_5$	Tri-	-	-	-	Aggarwal et al. 2011
25	AH3-10	FAM	F: AGGGATATCGGATGCTTACT R: AAAGATGCAGCAGACCTATC	Y	$(GAT)_7$	Tri-	60	178-202	10	Aggarwal et al. 2011
26	AH3-13	Cy3	F: GAGGGATGAGATCTGTTTGA R: CACCGAAATCATCAGGATAG	N	$(CTT)_6$	Tri-	60	218-227	2	Aggarwal et al. 2011
27	AH3-17	Cy3.5	F: AATACTGGCATTTCGTGTCT R: AACAAAACATCACCAAGGTC	N	$(TTC)_7$	Tri-	60	155	1	Aggarwal et al. 2011

 https://doi.org/10.33494/nzjfs492020x56x

No.	Name	Dye	Primer sequence (5'-3')	Selected (Y/N)	Repeat motifs	Nucleotide repeat	Annealing Temp.	Size range-bp	No. of alleles	Reference
28	AH3-18	FAM	F: TGAGACAATTAATGGTGGTG R: TTTACAAGGGAAAAGCTGAG	Y	$(TAA)_5$	Tri	60	209-221	4	Aggarwal et al. 2011
29	Am30	HEX	F: GAGGTAATATTTTGAATTCCTTGAAC R: GGTGTATACCTCTTTCCTGTGG	N	$(AT)_9(GT)_{15}$	Di-	60	81	1	Butcher et al. 2000
30	Am41	FAM	F: TAGGCTAATGGTCATATTCCTAG R: AGAGATAGGGGTACACACTAAAAAAC		$(GT)_{36}$	Di-	-	-	-	Butcher et al. 2000
31	Am136	HEX	F: CCCATTGCCGTTTCTTTG R: GCATTTCCCTTGGAACAGTC	N	$(CT)_{20}$	Di-	-	-	-	Butcher et al. 2000
32	Am164	HEX	F: ACCCGGACGTATAGAAATAAATACA R: CGTGGAGGCAAGCAATATC	N	$(TG)_{93}$	Di-	60	46	1	Butcher et al. 2000
33	Am387	FAM	F: TGATACAAGGGAAGACAGAGTGG R: CCAACTCAAAACCTGACAACG	N	$(AT)_2(GT)_2(AT)_2(GT)_{17}$-$(TA)_8$	Di-	-	-	-	Butcher et al. 2000
34	Am424	NED	F: AATACATGGAAGAGGATGAGATG R: ATTGCATTTCATTTGTTGCC	N	$(GT)2\ (TG)_2$,$A(GT)_4$,$GA(GT)_9$	Di-				Butcher et al. 2000
35	Am465	Cy 3.5	F: TGGGTATCACTTCCACCATT R: AGGCTGCTTCTTTGTGCAGG	Y	$(AC)_{23}$	Di-	60	113-131	5	Butcher et al. 2000
36	Am502	VIC	F: CAAATGGCCAAGTTACGACTG R: TTCTGGTAATCCAAACTTATGTGG	N	$(TTC)_3$-$(GGA)_8$,$AGA(GGA)_2$	Tri-	-	-	-	Butcher et al. 2000
37	Am770	FAM	F: CAGAGGTGGCAGATGATGTC R: AAGCCTTTAGTTGGGCGTTC	N	$CTC(CAC)_5CGC(CAC)_3$	Tri-	60	84	1	Butcher et al. 2000
38	Ak06	FAM	F: AGGTTGATGAAAAGGCATGG R: TCTCAGGTTTGGTGGGTTTT	N	$(AAG)_6$	Tri-	-	-	-	Adamski et al. 2013
39	Ak08	VIC	F: ACAGTTCCACCTCACCGTTC R: CGACCCTATCACCTTCTTGC	N	$(TACA)_8$	Tetra-	-	-	-	Adamski et al. 2013
40	Ak15	VIC	F: CACCCCCACGTTATCTTACA R: GACTGGCGAAAGAGTCGAA	Y	$(TAT)_5$	Tri-	60	297-309	4	Adamski et al. 2013
41	Ak16	PET	F: GCTCCTTGTCATGCTCTTTCA GCTGGCAGCTGCTGTAGTTT	N	$(A4G)_2AA(A4G)_2$	Di-	TD: 54	220	1	Adamski et al. 2013
42	Ak28	NED	F: ACTGGTGCAGTGTCTGTTGC R: ACGCAGGTCAATTGTGTTGA	N	$(ATTA)_4$	Tetra-	-	-	-	Adamski et al. 2013
43	Ak39	VIC	F: AGCAAACTTGGCCTTCAAGA R: CAACTGCTCCTGTTGGTGAA	N	$(GTGC)_3$	Tetra-	-	-	-	Adamski et al. 2013
										Adamski et al. 2013

 https://doi.org/10.33494/nzjfs492020x56x

No.	Name	Dye	Primer sequence (5'-3')	Selected (Y/N)	Repeat motifs	Nucleotide repeat	Annealing Temp.	Size range-bp	No. of alleles	Reference
44	Ak50	PET	F: AGGTTGATGAAAAGGCATGG R: TCTCAGGTTTGGTGGGTTTT	N	$(AAG)_8$	Tri-	TD: 54	234	1	
45	Ak89	FAM	F: AGGGGAAGGACGAAAGTTGT R: GCAAGAGGAGCTTCAAGTGG	Y	$(AC)_7$	Di-	60	160-174	5	Adamski et al. 2013
46	Ak219	FAM	F: AACAAATACCAAGGCGACAAA R: GGGTACTGGAAGAGCAGGTG	N	$(GAA)_4$	Tri-	60	190	1	Adamski et al. 2013
47	As2.17	HEX	F: TCCTCGCTTCTCGACATTTT R: GCTCGAACCTTTCAAACGAA	N	$(AC)7(TC)_7$	Di-	60	111	1	Millar 2009
48	As2.46	HEX	F: GTTCTCTTGCCCTGTTTGCT R: AGGCTGGAAATAAATGGAGGA	N	$(TC)_7$	Di-	-	-	-	Millar 2009
49	As2.61	FAM	F: CTGAATGTGCTTCTTCTCTCTTGG R: GGGAATCTGCCTTTAGTTTGC	N	$(TC)_{12}$	Di-	-	-	-	Millar 2009

[#] 0.5 mM additional $MgCl_2$ added