Genetic diversity and population structure analysis reveals the unique genetic composition of South African selected macadamia accessions

Mary Ranketse[1,2], Charles A. Hefer[2,3], Rian Pierneef[2], Gerda Fourie[1], Alexander A. Myburg[1]

[1] Department of Biochemistry, Genetics and Microbiology, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Pretoria, South Africa

[2] Biotechnology Platform, Agricultural Research Council – Onderstepoort Veterinary Institute, Pretoria, South Africa

[3] AgResearch Ltd, Lincoln, New Zealand

Corresponding author:

Prof Alexander A. Myburg

Department of Biochemistry, Genetics and Microbiology

Forestry and Agricultural Biotechnology Institute (FABI)

University of Pretoria,

Private bag X20

Pretoria, 0028

South Africa

Email: zander.myburg@fabi.up.ac.za

Tel: +27-012-420-4945

Abstract

Macadamia nuts are known globally for their high quality and economic value. Global macadamia commercial nut production amounts to 60 000 metric tonnes and is increasing steadily. South Africa is the leading producer with 29% of world-wide kernel production. Commercial macadamia germplasm was originally selected from a small genepool (mainly *Macadamia integrifolia* species) from a limited geographic distribution in Australia. These accessions were subsequently bred, cloned and exported across the world to start local macadamia industries. The South African macadamia industry was established with pre-commercial and commercial macadamia from different parts of the world, and local selections were also performed. Many of these accessions have unique genetic compositions that have not been characterized yet. We used 13 nuclear microsatellite markers to study the genetic diversity and structure of macadamia germplasm cultivated in South Africa. We compared four groups of accessions including 31 originating from the Hawaiian Agricultural Experimental Station (HAES), 19 from Australia (AUS), two from California and one from Israel (OTH), 31 from South Africa's locally selected accessions (SA) and 26 from two local Farmers (FARM). We used STRUCTURE, PCoA and Neighbour-Joining phylogenetic analyses to show that the South African selected accessions include diverse hybrid genotypes with strong *Macadamia tetraphylla* composition, unlike the Hawaiian commercially released and Australian representative collections that mostly have *M. integrifolia* or hybrid composition. Our results suggest that the South African selections represent a unique and diverse set of germplasm for future macadamia improvement efforts that will benefit from genomic breeding technologies.

Keywords: Macadamia, microsatellite, genetic diversity, population structure, molecular breeding

Introduction

South Africa is the world's largest exporter of Macadamia nuts, surpassing Australia and Hawaii (INC 2019-2020), despite not having any formal cultivar breeding programs. The application of molecular genetics technologies has the potential to guide and accelerate breeding of macadamia cultivars. DNA marker technologies such as microsatellite or simple sequence repeat (SSR) markers have proven to be invaluable in the management of genetic resources and breeding programs, allowing for accurate cultivar identification; parentage analysis and population genetic analysis amongst other uses (Moose and Mumm 2008).

Furthermore, breeding of tree species is challenging due to their long breeding cycles and the high cost of field trials. The application of molecular genetics to nut trees has proven to be useful for genetic resource management and routine breeding operations in many species including brazil nuts (Cabral et al. 2017; Giustina et al. 2017; Baldoni et al. 2020), hazelnuts (Helmstetter et al. 2019), cashews (dos Santos et al. 2019) and walnuts (Bernard et al. 2018).

Due to the extensive and laborious production process and the resulting high-quality product (Weinert 1993; Coverdale et al. 1999), macadamia is the most expensive nut in the world. Kernels are valued at approximately $18-$20 per kg, and Nuts-in-shell (NIS) are approximately $5-$6.50 per kg (SAMAC 2020) making the genus *Macadamia* one of the most economically important in the Proteaceae family (O'Connor et al. 2015). Commercial macadamia accessions originate mainly from two species namely *Macadamia integrifolia* (Maiden & Betche) and *Macadamia tetraphylla* (L.A.S. Johnson), and hybrids thereof (Peace et al. 2008; Hardner et al. 2009). Hybridization also occurs naturally in overlapping zones in the wild (Peace 2004; O'Connor et al. 2015; Mai et al. 2020). The two species are morphologically different, but do exhibit overlapping characteristics. In terms of tree height *M. integrifolia* grows taller and has only white/cream flowers, whereas *M. tetraphylla* has pink flowers. Another differentiating feature is the texture of the outer nut shell, with *M. integrifolia* having a smooth shell and *M. tetraphylla* a rough shell, and hybrids displaying a range of characteristics (Simpson and Allan 1998; Hardner et al. 2009).

Global commercial macadamia cultivation started at the Hawaiian Agricultural Experimental Station (HAES) following its introduction from Australia in the late 19th century (Hardner et al. 2009; Hardner 2016). The success of grafting during the 1930's accelerated clonal breeding programs (Hardner 2016). Selections at HAES were based on traits such as kernel weight, kernel size and high nut yields across Hawaii's various climatic zones (Hardner et al. 2009; Hardner 2016), and roasting ability of the nut which at the time was based on oil content (Hardner et al. 2009; Hobson 2013; Allan et al. 2016). Many of these selections had *M. integrifolia* ancestry (Nock et al. 2019) as *M. tetraphylla* was observed to have variable yield and the kernels were generally below standard (Hardner et al. 2009; Hobson 2013; Allan et al. 2016; Hardner 2016). Official macadamia selections in Australia were first

performed by the Queensland Department of Agriculture and Stock during the mid-1950s, together with a few local farmers (Hardner et al. 2009). Commercial macadamia selections from Hawaii were introduced to Australia mainly from the 1970s onwards. The early Australian selections together with the commercial accessions from Hawaii served as the foundation of macadamia cultivation and breeding projects in Australia, with some accessions used as parents for new local crosses (Hardner et al. 2009). These accessions have been recorded mainly as *M. integrifolia* species and *M. integrifolia* x *M. tetraphylla* hybrids (Hardner et al. 2009; Hardner 2016; O'Connor et al. 2019). Superior accessions from HAES, and Australia's early selections and accessions from breeding programs have been exported worldwide to start local industries (Allan 2016; Hardner 2016). The nut crop is now grown extensively in Australia; South Africa; USA (Hawaii and California); Kenya; Guatemala; Malawi; Brazil and China (Peace et al. 2008; SAMAC 2020). Current commercial macadamia accessions in Hawaii and elsewhere are only two to four generations removed from their wild progenitors in Australia (Peace et al. 2005; Hardner 2016) and should contain ample genetic variation for selective breeding.

Due to multiple introductions from around the globe as well as the prevalence of local selections from *M. tetraphylla*, South Africa is thought to possess a unique diversity of genetic material (Peace et al. 2005). South Africa initially acquired macadamia seed from Australia and Hawaii during the 1930s (Peace et al. 2005). These imports included superior grafted selections from HAES; grafted and seed-derived accessions from Australia and a few accessions from California imported during and after the late 1960s (Hardner et al. 2009; Allan 2016; Penter 2016). Macadamia adapted well to the South African subtropical climate, which is similar to its provenance on the east coast of Australia (Köppen-Geiger Climate Classification [Source: Köppen-Geiger.vu-wien.ac.at]). Although South Africa is a major role player in the global macadamia market, the country does not have any national macadamia breeding programs. Breeding attempts were made when macadamia was first introduced into South Africa. This local breeding was mainly with *M. tetraphylla* genotypes as the species was preferred due to it's perceived sweeter taste (Hardner et al. 2009; Allan 2016), however only the Nelmak (Nel-) selections were absorbed into the local industry (Peace et al. 2005; Penter 2016).

Breeding programs have evolved and improved over time with technological advances in the field of genetics. Traditional identification and breeding of macadamia was based on morphological characters. Morphological data is however not robust to differentiate between closely related accessions or hybrids (Peace et al. 2002a; Hardner 2016). The knowledge of cultivar relatedness, genetic structure and genetic diversity has been shown to improve breeding programs in terms of avoiding inbreeding and maintaining diverse breeding material (Collard and Mackill 2008). DNA marker systems have allowed for more accurate genetic identification and diversity studies in macadamia (Vithanage and Winks 1992; Peace et al. 2002a/b; Steiger et al. 2003; Peace et al. 2004; Peace et al. 2005; Schmidt et al. 2006; Nock et al. 2014; Alam et al. 2018; Langdon et al. 2019; O'Connor et al. 2019; also reviewed in Hardner et al. 2009). Microsatellite DNA markers, being highly informative; multi-allelic; co-dominantly inherited; polymorphic; and time and cost efficient, have been the preferred DNA marker system for genetic resource management (Varshney et al. 2005; Kalia et al. 2011; Guichoux et al. 2011; Vieira et al. 2016). These molecular tools have not been used for comprehensive genetic analysis of macadamia planted in South Africa.

This study aimed to characterize the genetic relationships and structure of macadamia accessions planted commercially in South Africa, in the context of international macadamia genetic diversity. First, we report DNA fingerprinting for South African accessions that will, in future, allow identification of accessions in commercial deployment and breeding programs. Second, we describe the genetic relationships and structure of macadamia accessions originating from Australia; Hawaii and South Africa. The results of this study will inform cultivar identification; orchard design; breeding and genetic resource management for macadamia improvement.

Materials and Methods

Plant material and DNA extraction

Leaf samples were obtained from macadamia collections originating from the Hawaiian Experimental Research Station (HAES); Australian Hidden Valley Plantation and pre-

commercial selections (AUS); California and Israel (OTH); South Africa's commercial and non-commercial local selections (S.A) and two local Farmers (FARM) (Table S1). The South African commercial accessions and the international representative accessions were collected from the Agricultural Research Council – Institute for Tropical and Subtropical Crops (ARC-ITSC) in Nelspruit, Mpumalanga province (n = 53). Leaf samples representing South African non-commercial local selections were obtained from the University of KwaZulu-Natal, Ukulinga Research Farm in Pietermaritzburg, KwaZulu-Natal province (n = 31). A further 25 samples were supplied by a local farmer from his breeding population, and an additional sample from a second farmer (grouped as one in the current study). Altogether, 110 accessions were used in this study.

Genomic DNA was extracted using the DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) following the manufacturer's protocol with minor modifications: a. leaf samples were ground in liquid nitrogen and incubated for 1 hr at 65°C with lysis buffer API and RNase A; b. DNA was eluted in 50 µl buffer AE to maximise DNA concentration. The DNA was quantified using the Qubit dsDNA BR Assay (Life Technologies, Carlsbad, USA). DNA quality was assessed by electrophoresis on 1% TAE agarose gels stained with ethidium bromide. The extracted DNA was stored at -20°C until further analysis.

SSR genotyping

A selection of 13 microsatellite markers from literature (Nock et al. 2014; Langdon et al. 2019) were used, and multiplexed into two panels for rapid and economical analysis. The 25 µl reaction contained 20 ng DNA; 12.5 µl Qiagen Multiplex PCR Master Mix (Qiagen, Hilden, Germany); primer concentration ranging from 0.1 – 0.4 µM (Table S2) (primers were manufactured at Inqaba Biotechnical Industries (Pty) Ltd, Pretoria, South Africa) and nuclease free water (Invitrogen, USA). PCR amplification was performed using a Bio-Rad Thermo Cycler (Bio-Rad Laboratories Inc. Hercules, California, USA) with the following conditions: initial denaturation at 95°C for 15 min; 35 cycles of 94°C for 20 s, 57°C for 90 s, 72°C for 60 s; followed by final extension at 60°C for 30 min. PCR products were sent to Inqaba Biotec™ for capillary analysis on a ABI 3500 XL Genetic Analyzer (Applied Biosystems, California, USA), with the DS-33 Matrix Standard (G5 dye set) and LIZ500

(Applied Biosystems, California, USA). Allele scoring was done using GeneMarker HID software v2.9.5 (SoftGenetics, State College, California, USA).

Genetic diversity

Genetic diversity parameters included number of alleles per locus (*Na*); number of effective alleles (*Ne*); observed heterozygosity (*Ho*); expected heterozygosity (*He*); Wrights allelic fixation index (F_{IS}); probability of identity (*PI*); cumulative probability of identity (*PI cum*); probability of identity for siblings (*PIsib*), an upper bound estimator for loci required for identify (Waits et al. 2001); cumulative probability of *PIsib* (*PIsib cum*); number of private alleles; Shannon's information index (*I*) and unbiased heterozygosity (*uHe*) were estimated using GenAlEx 6.502 software (Peakall and Smouse 2012). Marker polymorphism information content (*PIC*) was calculated in Cervus 3.0.7 (Kalinowski et al. 2007).

Population structure

Accessions were assigned according to the Hawaiian collection (HAES); Australian collection (AUS); California and Israel collections (OTH); South African locally selected collection (S.A.); and the local Farmers' accessions (FARM). Genetic differentiation among the different representative collections was analysed with principal coordinate analysis (PCoA) using the genetic dissimilarity matrix in GenAlEx 6.502 (Peakall and Smouse 2012) and a complimentary 3D scatter plot was generated using the Plotly R package (Sievert 2020) (Script S2). Genetic relationships among the accessions was analysed using a neighbour-joining (NJ) tree, constructed from Nei's standard genetic distance (Nei 1983) using Populations v.1.2.34 software (Langella 1999) with 1,000 bootstrap replicates specified. The resulting tree was visualised in the iTOL online tool (Letunic and Bork 2019). Population structure was further assessed with a Bayesian clustering approach using the LEA (Frichot and Francois 2014) and plyr (Wickham 2011) R packages for *K*1 to *K*4 (Script S3). Accessions were sorted according to the membership coefficient (Q) for *K*4 using the STRUCTURE "Sort by Q" method (Francis 2015), the *K*2 and *K*3 accessions were ordered according to *K*4. Population structure was confirmed with the STRUCTURE v2.3.4 software (Pritchard et al. 2000) using a burn-in period of 10,000 and a Markov chain Monte Carlo (MCMC) of 50,000 iterations. Each *K* value between $K = 1$ and $K = 4$ was run independently twenty times with

correlated alleles frequencies and population information was supplied. The best value of *K* was determined by Δ*K* (Evanno et al. 2005) statistics using the Structure Harvester web-based tool (Earl and von Holdt 2012). CLUMPAK (Kopelman et al. 2015) was used to process and visualise the STRUCTURE results.

Results

SSR marker analysis

We used 13 SSR markers to analyse genetic diversity in 110 macadamia accessions represented in South Africa (some of which have been imported from Hawaii; Australia; California and Israel). To determine the applicability and performance of the previously published microsatellite DNA markers, individual marker statistics were calculated (Table 1). The number of alleles per marker ranged from 8 (Mac001; Mac005 and Mac008) to 24 (Mac019), with a mean of 13 alleles. Observed heterozygosity ranged from 0.40 to 0.91, and expected heterozygosity ranged from 0.60 to 0.91. The mean observed heterozygosity (0.66) was lower than the expected heterozygosity (0.80). Polymorphic information content of the marker ranged from 0.56 to 0.90, with a mean of 0.80. Wright's fixation index (F_{IS}) ranged from -0.04 to 0.52 and negative F_{IS} values were observed in two of the markers (-0.04). Probability of identity of siblings (*PIsib*) ranged between 0.30 and 0.50, and the 13 loci have a cumulative *PIsib* of 1.53E-06.

Table 1 Marker statistics for 13 SSR markers assessed in 110 macadamia accessions

Locus	*Na*	*Ne*	*Ho*	*He*	*PIC*	F_{IS}	*PI*	*PI cum*	*PIsib*	***PIsib cum***
MAC001	8	4.04	0.72	0.75	0.72	0.05	0.09	9.44E-02	0.40	3.97E-01
MAC002	9	3.73	0.43	0.73	0.71	0.42	0.10	9.06E-02	0.41	1.62E-01
MAC005	8	4.17	0.79	0.76	0.72	-0.04	0.02	1.52E-04	0.30	4.93E-02
MAC007	19	8.56	0.85	0.88	0.87	0.04	0.03	4.22E-06	0.32	1.57E-02
MAC008	8	2.51	0.52	0.60	0.56	0.14	0.21	8.67E-07	0.50	7.87E-03
MAC009	10	4.31	0.40	0.77	0.74	0.48	0.08	7.34E-08	0.39	3.05E-03
MAC010	16	8.47	0.83	0.88	0.87	0.06	0.02	1.80E-09	0.32	9.61E-04
MAC011	12	7.94	0.91	0.87	0.86	-0.04	0.03	5.12E-11	0.32	3.08E-04
MAC013	11	2.85	0.50	0.65	0.61	0.24	0.16	8.44E-12	0.47	1.43E-04
MAC014	15	6.35	0.63	0.84	0.83	0.25	0.04	3.42E-13	0.34	4.86E-05
MAC017	11	7.52	0.42	0.87	0.85	0.52	0.03	1.08E-14	0.32	1.58E-05
MAC018	18	6.18	0.77	0.84	0.83	0.08	0.04	4.15E-16	0.34	5.37E-06
MAC019	24	10.52	0.82	0.90	0.90	0.10	0.01	3.46E-18	0.29	1.53E-06
Mean	*13*	*5.93*	*0.66*	*0.80*	*0.78*	*0.18*				

Na number of alleles per locus, *Ne* number of effective alleles, *Ho* observed heterozygosity, *He* expected heterozygosity, *PIC* polymorphic information content, F_{IS} Wrights allelic fixation index, *PI* probability of identity that two individuals share the same genotype at a locus, *PI cum* cumulative probability of identity achieved by adding markers from top to bottom, *PIsib* probability of identity for siblings sharing the same genotype at a locus, *PIsib cum* cumulative probability of *PIsib* from top to bottom

Population diversity and structure analysis

For the analysis of population diversity we excluded the California and Israel collections as they were represented by fewer than five samples (Table 2). Heterozygosity was relatively high in all four remaining collections, with *He* = 0.62 in HAES; *He* = 0.74 in both the Australian and Farmers' collections; and the highest *He* = 0.78 in the South African collection. Private alleles were observed at a mean of 0.38 in the HAES collection; 0.85 in

the Australian collection; 1.62 in the South African collection and 0.77 in the Farmers' collection across all 13 markers.

Table 2 Mean genetic diversity parameters across collections

Pop	*N*	*Na*	*Ne*	*He*	*No. of private alleles*	*I*	*uHe*
HAES	31	6.69	3.33	0.62	0.38	1.29	0.63
AUS	19	8.46	4.89	0.74	0.85	1.69	0.76
S.A.	31	9.31	5.28	0.78	1.62	1.82	0.80
FARM	26	7.92	4.43	0.74	0.77	1.63	0.76
Mean	*27*	*8.10*	*4.48*	*0.72*	*0.91*	*1.61*	*0.74*

N Sample number, *Na* mean number of alleles per collection, *Ne* mean number of effective alleles per collection, *He* expected heterozygosity, mean number of private alleles, *I* Shannon's information index, *uHe* unbiased heterozygosity. Unbiased heterozygosity accounts directly for related and inbred individuals.

Population differentiation analysed using principal coordinate analysis (PCoA Fig. 1, Fig. S4) identified three major groups corresponding to the known or suspected species composition of individual accessions within each group. One group is composed mainly of HAES and some AUS accessions, most of which have been recorded as *M. integrifolia* derived (green group). The second group is mainly South African accessions, some of which are recorded as *M. tetraphylla* derived (pink group). The third group includes the majority of the Farmer's breeding accessions (blue group). The first three coordinates of the PCoA analysis explained 22.31%, 9.30% and 7.90% of the total variance, respectively, and jointly explained 39.51% of the total genetic variation. The phylogenetic relationships of the accessions visualised using a NJ tree was consistent with the results of the PCoA, with distinct clustering of the two species derived groups (Fig. 2). Accessions recorded with either a strong *M. integrifolia* or *M. tetraphylla* genetic background formed separate clusters with known hybrids (A4; A16; A38) positioned intermediately and many of the Farmer's breeding population forming a separate cluster. The species-associated clustering coincided with the representative HAES and some Australian accessions in the putative *M. integrifolia*-derived cluster and the South African accessions in the putative *M. tetraphylla*-derived cluster on the opposite end, with both clusters likely including interspecific hybrids of these species.

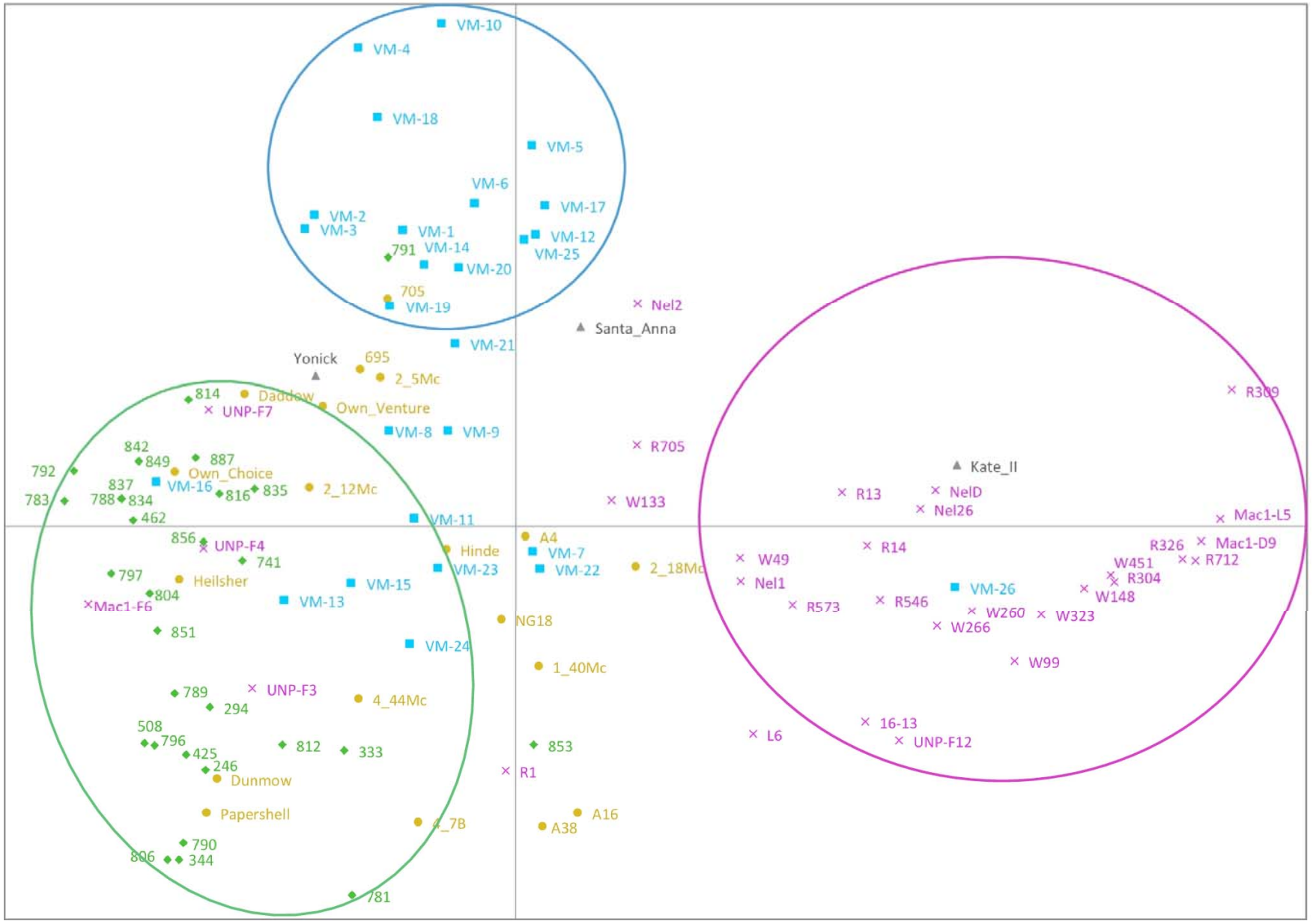
VM-10
VM-4
VM-18
VM-5
VM-6
VM-17
VM-2
VM-3
VM-1
VM-12
VM-25
791
VM-14
VM-20
705
VM-19
Nel2
Santa_Anna
VM-21
Yonick
695
2_5Mc
814
Daddow
UNP-F7
Own_Venture
VM-8
VM-9
R705
842
849
887
792
Own_Choice
837
VM-16
835
2_12Mc
783
788
834
816
462
VM-11
W133
R309
Kate_II
R13
NelD
Nel26
Mac1-L5
856
UNP-F4
741
Hinde
VM-23
797
Heilsher
804
VM-15
Mac1-F6
VM-13
851
VM-24
A4
VM-7
VM-22
2_18Mc
NG18
1_40Mc
R14
W49
Nel1
R573
R546
VM-26
W148
W260
W323
W266
W99
R326
Mac1-D9
R712
W451
R304
16-13
L6
UNP-F12
789
UNP-F3
294
4_44Mc
508
796
425
812
333
246
Dunmow
853
R1
Papershell
4_7B
A16
A38
790
806
344
781

Fig. 1 Principal coordinate analysis of macadamia analysed based on genetic distance using GenAlex v6.502. Accessions differentiate into three genetic clusters. The font colour represents cultivar origin with the pink font representing the S.A. (South African), the green font representing the HAES (Hawaiian Agricultural Experimental Station) representative accessions, gold font representing the AUS (Australian) representative accessions and the grey font representing the OTH (Californian and Israel) representative accessions. The pink cluster contains mainly South African accessions forming a *M. tetraphylla*-derived group, the green cluster contains mainly HAES and Australian accessions forming the *M. integrifolia*-derived group and the blue cluster contains the local FARM (Farmer's breeding population). A 3D interactive version is available here: https://chart-studio.plotly.com/~mranketse1/17/#/

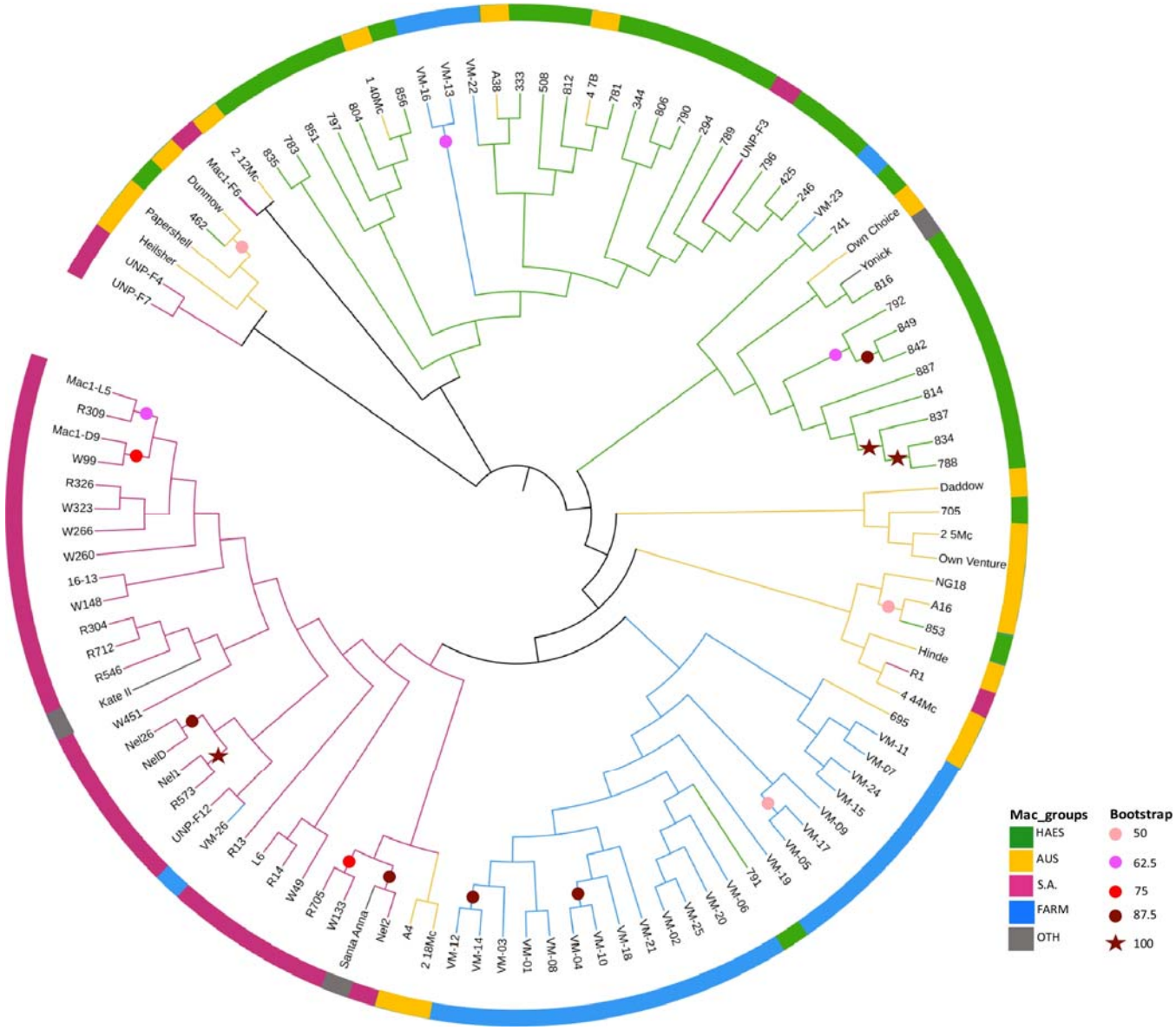
Mac1-L5
R309
Mac1-D9
W99
R326
W323
W266
W260
16-13
W148
R304
R712
R546
Kate II
W451
Nel26
NelD
Nel1
R573
UNP-F12
VM-26
R13
L6
R14
W49
R705
W133
Santa Anna
Nel2
A4
2 18Mc
VM-12
VM-14
VM-03
VM-01
VM-08
VM-04
VM-10
VM-18
VM-21
VM-02
VM-25
VM-20
VM-06
791
VM-19
VM-05
VM-17
VM-09
VM-15
VM-24
VM-07
VM-11
695
4 44Mc
R1
Hinde
853
A16
NG18
Own Venture
2 5Mc
705
Daddow
788
834
837
814
887
842
849
792
816
Yonick
Own Choice
741
VM-23
246
425
796
UNP-F3
789
294
790
806
344
781
4 78
812
508
333
A38
VM-22
VM-13
VM-16
856
1 40Mc
804
797
851
783
835
2 12Mc
Mac1-F6
Dunmow
462
Papershell
Heilsher
UNP-F4
UNP-F7
Mac_groups
HAES
AUS
S.A.
FARM
OTH
Bootstrap
50
62.5
75
87.5
100

Fig. 2 Neighbour-joining tree using Nei's genetic distance of macadamia displaying phylogenetic relationship amongst representative national collections visualised in iTOL. Accessions separated into three major genetic clades. One major clade consists mostly of accessions from HAES, many of which are *M. integrifolia*-derived. The second major clade consists of accessions from the local Farmer's population. The final major clade consists mainly of accessions from S.A., many of which are *M. tetraphylla*-derived. The colors are co-ordinated according the cultivar origin, green: HAES (Hawaiian Agricultural Experimental Station) representative accessions, gold AUS (Australian) representative accessions), blue: FARM (local Farmers samples), pink: S.A. (South African) representative accessions.

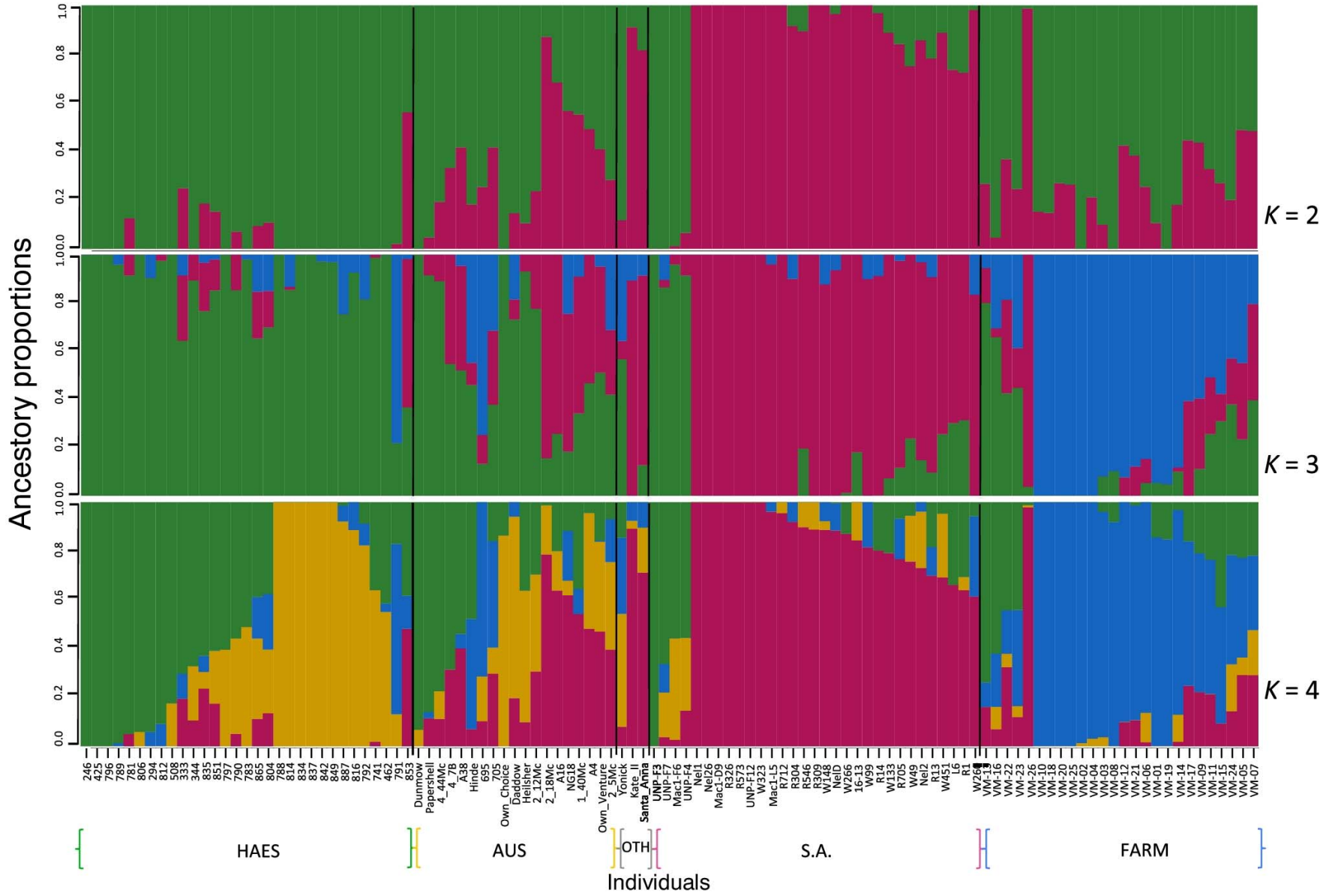
Ancestory proportions
K = 2
K = 3
K = 4
HAES
AUS
OTH
S.A.
FARM
Individuals

Fig. 3 STRUCTURE analysis of macadamia accessions from *K* = 2 to *K* = 4. Accessions from national country representative collections are separated by black vertical lines. Commercial macadamia originate from two ancestral species, *M. integrifolia* in green, and *M. tetraphylla* in pink. The HAES (Hawaiian Agricultural Experimental Station) representative collection has a higher proportion of *M. integrifolia,* the AUS (Australian) representative collection consists mostly of hybrids, the S.A. (South African) representative collection has a higher proportion of *M. tetraphylla*, and the Farmer's (FARM) breeding population is genetically unique from the other three collections.

STRUCTURE analysis revealed the genetic structure and likely admixture of the macadamia germplasm (Fig. 3). The analysis confirmed that most commercial macadamia were derived from two ancestral populations, likely *M. integrifolia* and *M. tetraphylla.* The most likely number of genetic structures defined by Δ*K* identified *K* = 3 (Fig. S5) as the number of clusters beyond which there is no further increase in likelihood. All country origin groups have unique genetic backgrounds across all *K*s. Many of the AUS representative accessions are observed to be admixed. The local Farmer's beeding population also has a unique genetic composition.

Discussion

Genetic parameters and DNA fingerprinting

South Africa is an important player in the global macadamia production industry and there is therefore a need to determine the genetic parameters of locally selected accessions in the context of global genetic diversity. Comparison of different population representatives is possible due to the availability of published microsatellite DNA markers (Nock et al. 2014; Langdon et al. 2019). The mean heterozygosity for markers (*Ho* = 0.66 and *He* = 0.80) was higher in the present study compared to previous studies using similar microsatellite markers to assess HAES, AUS and some wild accessions (Nock et al. 2014) and for paternity testing in a varietal plot containing HAES and AUS accessions (Langdon et al. 2019). This suggests greater allelic richness for some markers in the present study (total number of alleles = 170). High heterozygosity is not unusual for studies with many hybrid accessions and was observed for some loci in the study by Nock et al. (2014), as well as studies on walnuts (Schneider et al. 2019) and cocoa (Everaert et al. 2017). The *PIC* values for all 13 markers were above 0.5, and in 11 of the markers *PIC* was above 0.7, confirming the informativeness of the markers used (Hildebrand et al. 1992). The lowest diversity and *PIC* values were observed in two microsatellite DNA markers, Mac008 and Mac013. These markers were also observed to have low diversity in the study by Langdon et al. (2019), but are still useful in combination with other markers for DNA fingerprinting or identification purposes. We found that *PIsib* for markers ranged from 0.30 to 0.47 (Table 1), Langdon et al.

(2019) found *PIsib* ranged between 0.46 and 0.88 within a breeding population that included a high number of HAES accessions. All but five accessions (824/849; 788/834/837) had unique microsatellite profiles in the present study, these may need to be resolved using additional markers. Together, these markers can be used to successfully differentiate macadamia grown in South Africa.

DNA markers used for fingerprinting and identification purposes are important for applications such as genetic resource management and routine breeding (Govindraj et al. 2015). Previous studies that have demonstrated the usefulness of genetic identification methods in macadamia include the analysis of an Australian macadamia breeding population that identified genetically identical accessions using microsatellite DNA markers (Langdon et al. 2019) and cultivar mislabelling using SNP and silicoDArT markers (Alam et al. 2018). Different labelling errors, such as homonymous/intraplant and synonymous/duplicate are common (Everaert et al. 2017). These errors highlight the importance of genetic identification and fingerprinting, as phenotypic identification is not always reliable, especially with a genus such as *Macadamia* which readily hybridizes (Hardner 2016).

Population diversity and structure

We analysed the diversity and genetic structure of macadamia germplasm grown in South Africa, including sourced representative collections from Hawaii and Australia. Commercial macadamia from HAES were clonally propagated from a relatively small genepool of the *M. integrifolia* species and were used for subsequent crosses and breeding programmes (Hardner et al. 2009; Hardner 2016; Nock et al. 2019). Previous molecular studies found HAES accessions had lower heterozygosity and genetic diversity, confirming that these accessions come from a smaller gene pool, with a strong *M. integrifolia* background (Steiger et al. 2003; Peace et al. 2002a/b; Peace et al. 2005; Schmidt et al. 2006; Nock et al. 2014; O'Connor et al. 2015; Alam et al. 2018; Nock et al. 2019; Langdon et al. 2019). Some accessions from the current study were initially selected in Australia but recorded as later selections released from HAES and given an HAES number (Hardner et al. 2009; Hardner 2016), e.g. 705 (Peace et al. 2002a; O'Connor et al. 2019). Australian accession Own Choice is another example that is recorded as HAES 772 (Peace et al. 2005; Hardner et al. 2009;

Hardner 2016; O'Connor et al. 2019). Accession 695 is also a well known example, as it was originally selected in Australia with the name NSW44; thereafter taken to HAES and named as 695, and finally introduced to California and named Beaumont. These accessions that were introduced to Hawaii from Australia and were renamed with HAES numbers may have given rise to some of the apparent discrepancy between suspected origin and observed genetic grouping.

Understanding population structure is important for genetic resource management in macadamia. The South African macadamia collection is sourced from Hawaii; Australia; California and Israel with selections from local crosses and breeding material also contributing to the total genetic diversity in the country (Penter 2016). Our PcoA analysis differentiated accessions into groups with suspected species compositions (Fig. 1, Fig. S4). Most of the HAES accessions were included in one group together with accessions recorded as having *M. integrifolia* ancestry, such as 741; 246 and 425 (Hardner et al. 2009; Hardner 2016; Nock et al. 2019). Many of the South African accessions were included in a separate group together with accessions recorded as having putative *M. tetraphylla* ancestry or hybrids with strong *M. tetraphylla* genetic background such as Nel1; Nel2; R14 and W148 (Peace et al. 2005; Hardner et al. 2009). The Californian accession Cate_II (Cate, Kate_II in the present study) and Santa Ana (Santa Anna in the present study) were recorded as a selection from *M. tetraphylla* (Hardner at al. 2009). Our results show that the Kate_II accession groups/clusters with the majority of the South African accessions and the Santa Anna and Nel2 accessions are intermediary between the South African and Farmer's breeding groups.

Some Australian accessions included in this study (eg. Own Choice; Heilsher; Dunmow) grouped with the HAES accessions in the putative *M. integrifolia* group, whereas other accessions such as A4; A16 and A38 which are recorded as hybrids (Hardner et al. 2009; Hardner 2016; Nock et al. 2019) are intermediate to the putative *M. integrifolia* and *M. tetraphylla* groups. The local Farmer's private breeding population formed a third group that is composed of highly related breeding material with sibling and progeny relationships. This group includes the accession 791 which is in itself a distinctive tri-hybrid thought to contain *Macadamia ternifolia* in addition to *M. integrifoilia* and *M. tetraphylla* in its genome

(Peace et al. 2005; Hardner et al. 2009). The groupings in the PCoA plot are also reflected in the NJ tree (Fig. 2). Studies performed by Nock et al. (2014) and Langdon et al. (2019) using similar SSR markers; Peace et al. (2002a/b) and Peace et al. (2005) using RAF markers; and Alam et al. (2018) and Mai et al. (2020) using silico-DArT and SNP markers observed similar species associations.

The STRUCTURE analysis gives a clear indication of the genetic composition of the different country representative collections (Fig. 3). The most likely number of genetic clusters for the present study was $K = 3$ for ΔK, however higher K values may identify unique genetic structures as hybridization readily occurs in macadamia. It has been recorded that the Hawaiian breeding program focused on making selections from *M. integrifolia* as this species consistently produced good quality nuts that roasted well, which was the main consumer product (Hardner et al. 2009; Hobson 2013; Harder 2016; Allan 2016). In this study, two genetically distinct groupings were observed for the HAES collection in the STRUCTURE plot at $K = 4$, which coincides with two groups in the PCoA and the two sub-clades observed in the NJ tree. This close but distinct grouping was also observed in previous studies with various HAES accessions (Steiger et al. 2003; Peace et al. 2002b, 2005; Nock et al. 2019; O'Connor et al. 2019). One subgroup includes the first generation HAES accessions 246; 294; 333; 344; 425 and 508 from the present study (Hardner et al. 2009; Hardner 2016; Nock et al. 2019) together with some progeny of these accessions and advanced generation accessions. The second HAES subgroup includes mostly advanced generation accessions (Hardner et al. 2009; Hardner 2016). Hardner (2016) recorded that the advanced generation accessions have HAES numbers between 747-865, some of these are open-pollinated progeny of earlier selections. The Australian released accessions consist of a mixture of recorded *M. integrifolia*, and *M. integrifolia* x *M. tetraphylla* hybrids (Hardner et al. 2009). This may suggest that while earlier breeding projects in Australia tended to select for *M. integrifolia* species and *M. integrifolia* x *M. tetraphylla* hybrids, newer breeding projects may have involved different genotypes. The genetic composition of the Hawaiian and Australian accessions is further evident in the STRUCTURE plot, agreeing with previous studies (Peace et al. 2002a; Peace et al. 2005; Alam et al. 2018; O'Connor et al. 2019; Nock et al. 2019; Mai et al. 2020).

The South African locally selected accessions have a different STRUCTURE profile compared to the Hawaiian and Australian groups. According to literature, the South African farmers who imported macadamia in the early 1930s preferred the perceived sweeter taste of *M. tetraphylla* and thus initially made selections from this species (Peace et al. 2005; Allan 2016). Much earlier studies, summerised by Hardner et al. (2009), recorded that *M. tetraphylla* kernels may taste sweeter and have a higher sugar content compared to *M. integrifolia.* The Nelmak (Nel-), Reim's Nursery (R-) Windy Hill (W-) and Wyatts (Mac1-) series accessions are thought to have been selected from *M. tetraphylla* seedlings (Allan 2016). The study by Peace et al. (2005) showed that the South African selections were hybrids with a strong *M. tetraphylla* background. This strong *M. tetraphylla* genetic background is evident in the STRUCTURE plot, confirming the results from previous studies using RAF markers (Peace et al. 2002b; Peace et al. 2005). The Farmer's breeding population contains a unique genetic structure compared to the country-representative collections. This could be the result of distinct founder material used in the Farmer's breeding population and sibling relationships in the group of accessions.

In our study, most of the South African selections formed a distinct group/cluster, with some accessions intermediate to the major groups. Nelmak1, 2 and 26 were previously reported to be hybrids of *M. integrifolia* and *M. tetraphylla* based on RAF marker analysis, with Nel1 and Nel26 having 60-70% *M. tetraphylla* in their genome (Peace et al. 2005). The Nelmak selections have favourable phenotypic characteristics that may be attributed to their hybrid genetic background (Allan 2016). The Nelmak accessions produce larger, oval shaped nuts that need to be roasted at lower temperatures for longer periods compared to nuts from *M. integrifolia* (Lee 1995). A further five accessions were selected from open-pollinated Faulkner (Californian-bred *M. integrifolia*) accession, known as the UNP-F- series in South Africa (Allan 1992; Allan and Hauff 1995; Peace et al. 2005; Allan 2016). Allan (1992) reported on the trial progress of the UNP-F- accessions, as well as *M. tetraphylla* selections Mac1-F5 and R573 included in the present study. The trial found that these accessions produced nuts that met the HAES macadamia selection standards (Allan 1992). Other trials found that UNP-F- accessions performed satisfactorily in comparison to the HAES accessions (Allan and Hauff 1995; Allan et al. 1997). Unfortunately, not many trials included the W-, R-, Mac1- and UNP-F- lines and as such no further analysis of these

accessions was performed. Some of these accessions showed promise, but only the Nelmak germplasm was selected for commercial planting (Penter 2016; SAMAC 2020). These local accessions do however represent unique germplasm that can be included in breeding programs as they may have favourable traits, or may open up new markets (Allan 2016).

First generation macadamia breeding initiated by the Australian macadamia breeding program focused on increasing genetic diversity of the population, but a recent DNA marker study of one subset population (including parents and progeny) showed it still has relatively low genetic diversity (O'Connor et al. 2019), which may be due to the HAES accessions in the population that are known to come from a narrow gene pool (Langdon et al. 2019; Nock et al. 2019). Current molecular studies being performed on these populations indicate the importance of DNA marker technologies in establishing such breeding programs. A successful breeding program initiated to increase genetic diversity is the 1992 Australian project that included accessions from HAES; Hidden Valley Plantation (HVP) and pre-commercial (non-grafted) Australian selections (Alam et al. 2018). O'Connor et al. (2019) observed that increasing genetic diversity of progeny in a breeding program can also be achieved by selecting hybrid parents, instead of pure-species, as this will provide a wide range of genotypes in the progeny to select from in subsequent breeding. Interspecific hybridization serves as a source of new genetic diversity for breeding programs and also assists in breeding genotypes adapted to environmental changes. Diversity studies done on wild macadamia populations of the four species (Mai et al. 2020) and wild *M. tetraphylla* accessions (Spain and Louw 2011; O'Connor et al. 2015) found relatively high genetic diversity, even in local populations. These studies demonstrated the importance of DNA marker technologies for breeding and management and shed light on the diversity of macadamia present in the wild. In our study, the results reflect that breeding efforts of different countries focused on different criteria, which is evidenced by the Australian accessions being more mixed, the Hawaiian accessions being predominantly *M. integrifolia* and the South African accessions having a more *M. tetraphylla* genetic background.

Genetic diversity is also important when it comes to nut production and orchard design for macadamia. Macadamia is partially self-incompatible (Heard et al. 1993; Trueman et al. 2013) with some accessions being fully self-incompatible, and others accessions have higher

levels of self-pollination (Sedgley et al. 1990; Langdon et al. 2019). Cross pollination is favoured as it has been found to improve nut set and kernel yield (Wallace et al. 1996; Penter et al. 2007; Howlett et al. 2015; Langdon et al. 2019). Macadamia are long-lived evergreens that can produce nuts for up to 100 years. Therefore accurate identification of cultivars with DNA markers can be used to design orchards which include compatible accessions to maximise cross-pollination, allowing for improved nut production and successful yields through the years (Trueman et al. 2013; Howlett et al. 2015; Langdon et al. 2019).

Conclusion

Our study highlights the uniqueness of the South African selected macadamia germplasm compared to the HAES and Australian accessions used in this study, and demonstrates the utility of microsatellite DNA markers for breeders and researchers to timeously profile and resolve the genetic structure of macadamia accessions. The 13 SSR markers profiled in 110 macadamia accessions provide insight into the diversity present in macadamia grown in South Africa, comprising local selections and imported germplasm. Understanding the genetic structure of macadamia in the country aids in identifying unique genotypes present compared to international selections, such as the Nelmak accessions that can be capitalized on in breeding programs. Microsatellite DNA markers are useful for routine and affordable analysis, however they do have limitations such as the limited number of loci sampled and the presence of null alleles in some species which may skew the results. Future studies can make use of recently published macadamia genome assemblies (Niu et al. 2020; Nock et al. 2020) to mine more SSR markers across the genome thereby improving discriminating power of the existing marker panels. A further limitation of our study was not having wild representatives of *M. integrifolia* and *M. tetraphylla* to serve as species references, hence the species composition of the groupings observed are putative and based on records of the accessions they contain. Future studies aiming to understand genome-wide genetic structure and diversity would benefit from high throughput single nucleotide polymorphism (SNP) genotyping. Whole genome resequencing or technologies such as Genotyping-by-sequencing (GBS) and Diversity Arrays Technology (DArTseq) generate thousands of SNP

markers covering the entire genome for improved analysis. Also, chromosome level ancestry or admixture mapping can be performed using genome-wide SNP markers to unravel the suspected hybrid composition of many Macadamia accessions and thereby add to the genomic tools available in Macadamia.

Declarations

Electronic Supplementary Information The online version contains supplementary material.

Acknowledgements The authors thank Mark Penter (Agricultural Research Council – Institute for Tropical and Subtropical Crops (ARC-ITSC) of South Africa) for assistance with sample collection from Mpumalanga, constructive comments on the manuscript and valuable information and discussions. The authors also thank Matthew Erasmus for the sample collection from University of KwaZulu-Natal – Ukulinga Research Farm. We would like to thank the local Farmer for allowing us to use the breeding population and the second farmer for his sample. Further thanks to Dr Catherine Nock (Southern Cross University) for assistance with the microsatellite markers. We would like to thank Dr Nanette Christie (University of Pretoria) and Annie Chan (Agricultural Research Council – Biotechnology Platform) for assistance with RStudio, and the Forest Molecular Genetics Team (University of Pretoria) for their guidance with the GeneMarker. This work was funded by the National Research Foundation (NRF) of South Africa (UID 112133) and Macadamias South Africa (SAMAC). MR acknowledges PhD scholarship support from the Agricultural Research Council – Professional Development Programme (ARC - PDP) of South Africa.

Data Archiving Statement All of the data are included in the electronic supplementary material (Table S3).

Author's contributions The experimental design was contributed by CH, AAM and MR. MR did data collection, data analysis and wrote the first draft of the paper as part of her PhD thesis. The study was supervised by AAM, CH, GF and RP. The final manuscript was edited and approved by all the authors.

Funding information This work was funded by the National Research Foundation (NRF) of South Africa (Grant to CH, UID 112133). Further funding was received from Macadamias South Africa (SAMAC, strategic grant to GF).

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Alam M, Neal J, O'Connor K, Kilian A, Topp B (2018) Ultra-high-throughput DArTseq-based silicoDArT and SNP markers for genomic studies in macadamia. PLOS ONE **13**(8):e0203465

Allan P Quality of macadamia cultivars and selections in subtropical areas. In: Proceedings of the First International Macadamia Research Conference, Hawaii, 1992. pp 28-30

Allan P (1997) Future macadamia cultivar research. Macadamia Yearbook **1997**:53-55

Allan P, Apostolides K, Blore N, Williams M (1997) Cultivar evaluation and rooting at Pietermaritzburg. Macadamia Yearbook **1997**:49-55

Allan P (2016) Development of the macadamia nut industry in South Africa. Subtop Quarterly Journal **13** (January-March):24-30

Allan P, Hauff N (1995) Evaluation of macadamia cultivars and selections at Pietermaritzburg. South African Macadamia Growers Association Yearbook **1995**:24-28

Baldoni AB et al. (2020) Genetic diversity of Brazil nut tree (*Bertholletia excelsa* Bonpl.) in southern Brazilian Amazon. Forest Ecology and Management **458**:117795

Beck HE, Zimmermann NE, McVicar TR, Vergopolan N, Berg A, Wood EF (2018) Present and future Köppen-Geiger climate classification maps at 1-km resolution. Scientific Data **5:**180214

Bernard A, Barreneche T, Lheureux F, Dirlewanger E (2018) Analysis of genetic diversity and structure in a worldwide walnut (*Juglans regia* L.) germplasm using SSR markers. PLOS ONE **13**(11):e0208021

Bolger M et al. (2017) From plant genomes to phenotypes. Journal of Biotechnology **261**:46-52

Cabral J, Baldoni A, Tonini H, Azevedo V, Giustina L, Tiago A, Rossi A (2017) Diversity and genetic structure of the native Brazil nut tree (*Bertholletia excelsa* Bonpl.) population. Genetics and Molecular Research **16**(3):gmr16039702

Collard BCY, Mackill DJ (2008) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philosophical Transactions of the Royal Society B: Biological Sciences **363**(1491):557-572

Coverdale C, Hardner C Wegener, Malcolm K (1999) Evaluation of economic weights for selection and breeding in macadamia. Australian Agricultural and Resource Economics Society Conference (43th), January 20-22, Christchurch, New Zealand 123796

dos Santos JO, Mayo SJ, Bittencourt CB, de Andrade IM (2019) Genetic diversity in wild populations of the restinga ecotype of the cashew (*Anacardium occidentale*) in coastal Piauí, Brazil. Plant Systematics and Evolution **305**(10):913-924

Earl DA (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources **4**(2):359-361

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Molecular Ecology **14**(8):2611-2620

Everaert H et al. (2017) Molecular characterization of Vietnamese cocoa genotypes (*Theobroma cacao* L.) using microsatellite markers. Tree Genetics & Genomes **13**(5):1-11

Francis R (2015) Structure 'Sort by Q' explained. https://www.royfrancis.com/structure-sort-by-q-explained/. Accessed 20 October 2020

Frichot E, François O (2015) LEA: An R package for landscape and ecological association studies. Methods in Ecology and Evolution **6**:925-929

Govinadraj M, Vetriventhan M, Srinivasan M (2015) Importance of genetic diversity assessment in crop plants and its recent advances: An overview of its analytical perspectives. Genetics Research International **2015**

Giustina L et al. (2017) Genetic diversity between and within half-sib families of Brazil nut tree (*Bertholletia excelsa* Bonpl.) originating from native forest of the Brazilian Amazon. Genetics and Molecular Research **16**(4):gmr16039839

Guichoux E et al. (2011) Current trends in microsatellite genotyping. Molecular Ecology Resources **11**(4):591-611

Hardner C (2016) Macadamia domestication in Hawai'i. Genetic Resources for Crop Evolution **63**(8):1411-1430

Hardner CM, Peace C, Lowe AJ, Neal J, Pisanu P (2009) Genetic resources and domestication of macadamia vol 35. Horticultural Reviews. John Wiley & Sons, Inc, USA

Heard TA (1993) Pollinator requirements and flowering patterns of *Macadamia integrifolia.* Australian Journal of Botany **41**(5):491-497

Helmstetter AJ, Oztolan-Erol N, Lucas SJ, Buggs RJA (2020) Genetic diversity and domestication of hazelnut (*Corylus avellana* L.) in Turkey. Plants, People, Planet **2**(4):326-339

Hildebrand CE, Torney DC, Wagner RP (1992) Informativeness of polymorphic DNA markers. Los Alamos Science **November 20**

Hobson L (2013) A few thoughts on macadamia varieties. SUBTROP Quarterly Journal **1**(Jan - Mar):24-25

Howlett BG, Nelson WR, Pattemore DE, Gee M (2015) Pollination of macadamia: Review and opportunities for improving yields. Scientia Horticulturae **197**:411-419

INC (2019-2020) Nut and dried fruits statistical yearbook 2019/2020 vol 2019-2020. International Nut and Dried Fruit Council Foundation, Spain

Kalia RK, Rai MK, Kalia S, Singh R, Dhawan AK (2011) Microsatellite markers: an overview of the recent progress in plants. Euphytica **177**(3):309-334

Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Molecular Ecology **16**(5):1099-1106

Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I (2015) Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Molecular Ecology Resources **15**(5):1179-1191

Langdon K, Graham JK, Nock CJ (2019) DNA paternity testing indicates unexpectedly high levels of self-fertilisation in macadamia. Tree Genetics & Genomes **15**:29

Langella O (1999) Populations 1.2.30: Population genetic software (individuals or populations distances, plylogenetic tree). http://www.bioinformatics.org/populations/

Lee P (1995) Macadamia cultivars for Southern Africa. Macadamia Yearbook **1995**:42-46

Letunic I, Bork P (2019) Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Research **47**(W1):W256-W259

Mai T, Alam M, Hardner C, Henry R, Topp B (2020) Genetic structure of wild germplasm of macadamia: Species assignment, diversity and phylogeographic relationships. Plants **9**:714

Moose SP, Mumm RH (2008) Molecular plant breeding as the foundation for 21st century crop improvement. Plant Physiology **147**(3):969-977

Nei M, Tajima F, Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular data. Journal of Molecular Evolution **19**(2):153-170

Nock CJ, Elphinstone MS, Ablett G, Kawamata A, Hancock W, Hardner CM, King GJ (2014) Whole genome shotgun sequences for microsatellite discovery and application in cultivated and wild *Macadamia* (Proteaceae). Applications in Plant Sciences **2**(4):apps.1300089

Nock CJ et al. (2019) Wild origins of macadamia domestication identified through intraspecific chloroplast genome sequencing. Frontiers in Plant Science **10**:334

Niu Y-F et al. (2020) Genome assembly and annotation of *Macadamia tetraphylla*. bioRxiv:2020.2003.2011.987057

Nock CJ et al. (2020) Chromosome-scale assembly and annotation of the macadamia genome (*Macadamia integrifolia* HAES 741). G3: Genes, Genomes, Genetics 10(10):3497-3504

O'Connor K, Powell M, Nock C, Shapcott A (2015) Crop to wild gene flow and genetic diversity in a vulnerable *Macadamia* (Proteaceae) species in New South Wales, Australia. Biological Conservation **191**:504-511

O'Connor K et al. (2019) Population structure, genetic diversity and linkage disequilibrium in a macadamia breeding population using SNP and silicoDArT markers. Tree Genetics & Genomes **15**(2):24

Peace C, Ming R, Schmidt A, Manners J, Vithanage V (2008) Genomics of macadamia, a recently domesticated tree nut crop. In: Moore PH, Ming R (eds) Genomics of Tropical Crop Plants, vol 1. Plant Genetics and Genomics:Crops and Models. Springer, New York, pp 313-332

Peace CP, Allan P, Vithanage V, Turnbull CN, Carroll BJ (2005) Genetic relationships amongst macadamia varieties grown in South Africa as assessed by RAF markers. South African Journal of Plant and Soil **22**(2):71-75

Peace CP, Vithanage V, Neal J, Turnbull CGN, Carroll BJ (2004) A comparison of molecular markers for genetic analysis of macadamia. The Journal of Horticultural Science and Biotechnology **79**(6):965-970

Peace C, Vithanage V, Turnbull C, Carroll BJ (2002a) Characterising macadamia germplasm with codominant radiolabelled DNA amplification fingerprinting (RAF) markers In: International Symposium on Tropical and Subtropical Fruits, 2002. International Society for Horticultural Science (ISHS), Leuven, Belgium, pp 371-380. doi:10.17660/ActaHortic.2002.575.42

Peace C, Hardner C, Brown A, O'Connor K, Vithanage V, Turnbull C, Carroll B (2002b) The diversity and origins of macadamia cultivars. WANATCA Yearbook **26**:19-25

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. Bioinformatics (Oxford, England) **28**(19):2537-2539

Penter M (2016) Progress of new cultivar trials for the South African macadamia industry. SUBTROP Quarterly Journal **15**(July - September 2016):24-27

Penter M, Schoeman S, Nkwana E (2007) Cross-pollination for improved nut set in 'Beaumont' macadamias. South African Macadamia Growers Association Yearbook **15**:13-16

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics **155**(2):945-959

RStudio_Team (2015) RStudio: Integrated development for R. RStudio, Inc., Boston, MA. http://www.rstudio.com

SAMAC (2020) Statistics of the Southern African Macadamia Industry. SAMAC, Macadamias South Africa. https://www.samac.org.za/industry-statistics/. Accessed 11/05/2020

Schmidt AL, Scott L, Lowe AJ (2006) Isolation and characterization of microsatellite loci from *Macadamia*. Molecular Ecology Notes **6**(4):1060-1063

Schneider SJ, Hwang AY, Land SD, Chen L-L, Thomas AL, Hwang C-F (2019) Genetic diversity of ten black walnut (*Juglans nigra* L.) cultivars and construction of a mapping population. Tree Genetics & Genomes **15**(4):1-10

Sedgley M, Bell F, Bell D, Winks C, Pattison S, Hancock T (1990) Self-and cross-compatibility of macadamia cultivars. Journal of Horticultural Science **65**(2):205-213

Sievert C (2020) Interactive web-based data visualization with R, plotly, and shiny. CRC Press, Florida

Simpson WA, Allan P (1998) Revised key for identifying macadamia cultivars. South African Macadamia Growers' Association Yearbook **6**:47-56

Spain CS, Lowe AJ (2011) Genetic consequences of subtropical rainforest fragmentation on *Macadamia tetraphylla* (Proteaceae). Silvae Genetica **60**(1-6):241-249

Steiger DL, Moore PH, Zee F, Liu Z, Ming R (2003) Genetic relationships of macadamia cultivars and species revealed by AFLP markers. Euphytica **132**:269-277

Trueman SJ (2013) The reproductive biology of macadamia. Scientia Horticulturae **150**:354-359

Varshney RK, Graner A, Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. Trends in Biotechnology **23**(1):48-55

Vieira MLC, Santini L, Diniz AL, Munhoz CdF (2016) Microsatellite markers: what they mean and why they are so useful. Genetics and Molecular Biology **39**(3):312-328

Vithanage V, Winks C (1992) Isozymes as genetic markers for *Macadamia.* Scientia Horticulturae **49**(1-2):103-115

Waits LP, Luikart G, Taberlet P (2001) Estimating the probability of identity among genotypes in natural populations: Cautions and guidelines. Molecular Ecology **10**(1):249-256

Wallace HM, Vithanage V, Exley E (1996) The effect of supplementary pollination on nut set of *Macadamia* (Proteaceae). Annals of Botany **78**(6):765-773

Weinert IAG (1993) Macadamia nut processing. The Southern African Macadamia Growers Association Yearbook **1993**:54-75

Wickham H (2011) The split-apply-combine strategy for data analysis. Journal of Statistical Software **40**(1):1-29