Tree Genetics & Genomes
DOI 10.1007/s11295-009-0205-1

ORIGINAL PAPER

AFLP and SSR diversity in the African fruit tree *Allanblackia*: implications for management of a genus newly subject to domestication for the edible oil industry

J. R. Russell · C. A. C. Kadu · R. Jamnadass · A. Booth · N. J. Cordeiro · M. Woodhead · I. K. Dawson

Received: 22 May 2008 / Revised: 2 December 2008 / Accepted: 1 March 2009

Abstract *Allanblackia*, a dioecious fruit tree native to sub-Saharan Africa, is the subject of increased international interest for oil production for the global food market. Until recently, however, *Allanblackia* has been an overlooked wild tree, with very little known about its biology that could guide domestication. Here, we applied amplified fragment length polymorphisms (AFLPs) to assess the genetic composition of populations of five important *Allanblackia* species. Data indicated significant differentiation between certain species and occasional misidentification of taxa during collection. Misclassification suggested that care is required when sampling germplasm, especially when domesticating single species in areas where related taxa are sympatric. Genetic relatedness between species and the geographic proximity of distributions sometimes but did not always correspond. This likely reflects complex evolutionary processes related to migration and dispersal in the genus and indicated that a simple 'sampling-by-distance' model for assessing variation is not always appropriate. High AFLP variation suggested that Cameroon presents particular opportunities for domestication. In a comparison with AFLPs, we tested the value of *de novo* simple sequence repeats (SSRs) for detecting genetic variation in the same populations of *Allanblackia*, with a view to later applying these markers to determine optimal tree-planting ratios (female to male trees) and configurations during on-farm planting. Four primer pairs from a genomic library of one member of the genus—*Allanblackia stuhlmannii*—appeared suitable for research in this taxon (revealing 4.75 alleles per locus on average). However, cross-species application appeared limited (occurrence of null alleles?), suggesting either generic expressed sequence tag SSRs or specific suites of primers for each taxon are required.

Keywords Agroforestry · Domestication · Fruit tree · Poverty reduction

Communicated by A. Kremer

J. R. Russell · A. Booth · M. Woodhead
SCRI, Invergowrie,
Dundee DD2 5DA, Scotland, UK

C. A. C. Kadu · R. Jamnadass · I. K. Dawson (✉)
The World Agroforestry Centre (ICRAF),
P.O. Box 30677, Nairobi, Kenya
e-mail: iankdawson@aol.com

N. J. Cordeiro
Departments of Botany and Zoology, The Field Museum,
1400 S Lake Shore Drive,
Chicago, IL 60605-2496, USA

Present Address:
C. A. C. Kadu
Department of Genetics, Federal Research and Training Centre for Forests, Natural Hazards and Landscape,
Hauptstraße 7,
1140 Vienna, Austria

Introduction

Allanblackia (Clusiaceae), a dioecious fruit tree restricted to the humid forests of west, central and east Africa, has been the subject of increased international interest over the last 5 years because the seed yields an edible oil that has the potential for the production of healthy spreads (low in trans-fats) in the global food market. The oil is attractive to the food industry since it requires little chemical processing and refraction, whilst the *Allanblackia* tree appears easier to cultivate in more environmentally friendly ways than current alternatives for fat production such as oil palm. Market value chains for *Allanblackia* seed harvested from

Published online: 14 April 2009

natural stands are currently under development by Unilever (who provide a guaranteed price for product) and other commercial parties in Cameroon, Ghana, Nigeria and Tanzania (Attipoe et al. 2006). In agreement with a key development concern, current experience shows that the *Allanblackia* business compares favourably with other enterprises in terms of the gender ratio of received benefits, with 47% of 6,000 seed sellers in Tanzania being women in 2006 (most recent compiled figures; Harrie Hendrickx, Unilever, personal communication).

At the same time as market value chain development, a domestication programme to bring *Allanblackia* into cultivation has commenced in the above countries. This initiative, which is led by the World Agroforestry Centre (ICRAF) in collaboration with smallholder farmers, will rely on bringing into cultivation selected vegetatively propagated high-value clones for oil production (Atangana et al. 2006). Smallholder planting of the genus has significant potential for increasing and diversifying farmers' incomes in the relevant countries, with 100,000 farmers predicted to be earning an extra 200 Euros per annum on average from the oil business by 2015, comparing well in returns with the cultivation of other perennial species such as cacao, oil palm and tea (www.worldagroforestrycentre.org). In addition, smallholder cultivation has the potential to take harvesting pressure off natural stands of the genus, if value chains are developed in appropriate ways. However, as until recently *Allanblackia* has been an overlooked wild tree, very little is known about the biology of the genus that would help guide the process of transition from wild harvest to cultivation.

Critical to the development of a sound domestication programme is the need for adequate biological information that will help balance environmental concerns with the needs of local communities (Kremen 2005). Biological knowledge is particularly important when the negative consequences of inappropriate management are considered for associated flora and fauna, such as rare mammals that feed on the fruit or pollinators that forage for nectar, in natural and neighbouring farm landscapes (Amanor et al. 2003). This is especially so because the regions in which *Allanblackia* is found are high in biodiversity, rich in endemic species and subject to deforestation and habitat fragmentation and therefore represent areas of high conservation significance on the continent as well as globally (Mittermeier et al. 2004; Bergl et al. 2007; Burgess et al. 2007). On the other hand, proper handling of *Allanblackia* could contribute greatly to driving the participatory forest management practises that are sorely needed in biodiversity hot spots (Wily 2003). Furthermore, proper management may enhance 'biodiversity-friendly' eco-agricultural methods that contribute to habitat and species conservation in farmland, whilst stabilising and/or augmenting the incomes of the rural poor (McNeely and Scherr 2001; Leakey et al. 2005).

Case studies of other non-timber forest products (NTFPs) being brought into cultivation indicate, however, that careful management of the domestication process is required if it is to bring environmental benefits. Cultivation of NTFPs will not necessarily take pressure off natural populations and may indeed have unintended detrimental consequences for associated flora and fauna, as well as for different human communities (Vantomme 2004; Marshall et al. 2006). The example of oil palm—the cultivation of which for edible fat and biofuel has led to significant deforestation—is salutatory (Donald 2004). These issues are a concern for the World Conservation Union (IUCN), who form part of a novel private–public partnership that also involves Unilever, ICRAF, national agricultural and forestry research institutions, farming communities, market traders and others who are working on the *Allanblackia* agri-business as a case study for sustainable development (Attipoe et al. 2006).

Key knowledge gaps that need to be addressed for the proper use of *Allanblackia* include a lack of information on genetic structure within the genus and the absence of data on gene flow within and between populations. In the context of domestication, information on genetic structure is necessary for formulating proper germplasm sampling strategies, whilst knowledge on gene flow is required for designing appropriate farm management approaches, especially (as species are dioecious) in determining the right sex ratio and appropriate female–male configurations during tree planting. Here, we assess the genetic composition of populations of five *Allanblackia* species considered to be of commercial importance. Our objective was to employ amplified fragment length polymorphism (AFLP) markers (widely used in tropical trees, e.g. Adin et al. 2004) to describe genetic variation within and amongst species, in order to enhance the current very limited understanding of species relationships and to begin to define genetic management strategies. In addition, through comparison with AFLP data, we wished to develop and test the value of simple sequence repeat (SSR) markers (widely used in tropical trees, e.g. Hollingsworth et al. 2005) for detecting genetic variation in the genus, with a view to subsequently undertaking gene flow studies. Our rationale here was to develop *de novo* SSRs for one species and see if these could be applied effectively not only to this species but to other important taxa in the genus. Previous experience has demonstrated that SSRs show some level of cross-species transferability in plants, but this level varies with the family and genus in question (Woodhead et al. 2005), and ascertainment biases are possible (Almeida and Penha-Goncalves 2004). If amplifications across *Allanblackia* species are successful, this will allow considerable savings in resources, as the cost of designing a specific suite of SSR

primers for each separate species will be saved. Furthermore, use of a common set of primers would allow direct comparison of genetic structure across species.

Materials and methods

The genus *Allanblackia*

The limited information currently available suggests that *Allanblackia* consists of nine species, although the boundaries between taxa are sometimes indeterminate and multiple synonyms are apparent (primary mapped distribution in parentheses): *Allanblackia floribunda* (Nigeria to Democratic Republic of Congo [DRC]), *Allanblackia gabonensis* (Cameroon and Gabon), *Allanblackia kimbilensis* (Kivu region of DRC, Uganda), *Allanblackia kisonghi* (DRC), *Allanblackia marienii* (DRC), *Allanblackia parviflora* (Sierra Leone to Ghana), *Allanblackia stanerana* (Angola, Cameroon, DRC), *Allanblackia stuhlmannii* (Eastern Arc Mountains of Tanzania) and *Allanblackia ulugurensis* (Eastern Arc Mountains of Tanzania) (van Rompaey 2003). Two members of the genus thus appear to be endemic to DRC and two to Tanzania. In some cases, the distributions of species appear, at least in part, to overlap on a local scale, e.g. *A. floribunda* and *A. stanerana* appear to be sometimes sympatric, as apparently are *A. floribunda* and *A. gabonensis*, in Cameroon. In addition, the Tanzanian species *A. stuhlmannii* and *A. ulugurensis* occasionally co-occur. According to IUCN criteria, *A. gabonensis*, *A. stuhlmannii* and *A. ulugurensis* are considered to be vulnerable from a conservation perspective due to habitat loss and degradation and/or small initial population areas (www.iucnredlist.org/). The impacts of harvesting and domestication on populations of these species are therefore a particular concern. Trees are single stemmed, grow to 40 m tall, have whorled branches and are long living and long fruiting (over several decades). Natural regeneration appears to be episodic and favoured by partially shaded conditions. *Allanblackia* fruit are large (*A. stuhlmannii* has one of the biggest fruits of all plants in African rainforests, weighing 6 kg or more) and contain between 30 and 70 seed. Small mammals such as giant rats appear to be important seed dispersers, whilst flowers are pollinated by bees, other flying insects, nectar-feeding birds and, possibly, bats (Cordeiro et al. in preparation; Mpanda et al. in preparation).

Sampling of *Allanblackia* and DNA extraction

Our concern in the present study was to sample *Allanblackia* populations of five species (*A. floribunda, A. parviflora, A. stanerana*, *A. gabonensis* and *A. stuhlmannii*) from Cameroon, Ghana and Tanzania. These countries were chosen because seed collection volumes in nascent market value chains are currently highest in Tanzania (*A. stuhlmannii*), followed by Ghana (*A. parviflora*), with impact assessments, ecological studies and domestication also currently underway in both countries. In addition, significant domestication research is underway in Cameroon, primarily on *A. floribunda* (also important for domestication in Nigeria), but with some work also on *A. stanerana* and *A. gabonensis* (Atangana et al. 2006). Leaf samples were collected from mature trees in natural stands or as farm remnants (retained during forest clearance). Within stands, sampled trees were normally a minimum of 100 m apart. Self-indicating silica gel in snap-top plastic bags was used to dry and preserve material. Samples were subsequently stored at −20°C until DNA extraction. Leaf samples were assessed from a total of 11 populations (Table 1). In two cases in Cameroon, sampling was undertaken on different species coincident at the same site: at Sangmelima (*A. floribunda* and *A. gabonensis*) and at Yalpenda (*A. floribunda* and *A. stanerana*). Total genomic DNA was extracted based on minor modifications to the protocol of Doyle and Doyle (1987) and purified using a QIAGEN DNeasy Kit. Briefly, dried leaves were ground in liquid nitrogen, 3% CTAB buffer added, and samples incubated at 65°C for 30 min. Following chloroform–isoamyl alcohol purification, supernatant was transferred to a QIAshredder mini-column and manufacturer's instructions followed in subsequent steps of extraction.

AFLP analysis

AFLP analysis was undertaken based on modifications to the protocol of Vos et al. (1995). Approximately 100 ng of DNA was digested using the restriction enzymes *Eco*RI and *Mse*I, and adapters (*Eco*RI, 5′ CTC GAT GAC TGC GTA CC, 5′ AAT TGG TAC GCA GTC; *Mse*I, 5′ GAC GAT GAG TCC TGA G, 5′ TAC TCA GGA CTC AT) were ligated to the resulting DNA fragments. Pre-amplification was performed in 20-μL reactions on diluted restriction/ligation mixture with the primers E00 (5′ GAC TGC GTA CCA ATT C) and M00 (5′ GAT GAG TCC TGA GTA A), using a PE Applied Biosystems 9700 thermocycler. Four selective primer pair combinations shown to be useful for other plant species in the laboratory were used to generate fingerprints (E00-CCG, M00-ATT; E00-CCT, M00-ATT; E00-CCG, M00-AGA; E00-CCT, M00-AGA). In each case, the *Eco*RI-selective primer was labelled with the fluorescent dye FAM (Operon Biotechnologies GmbH). Selective polymerase chain reaction (PCR) products were separated using an ABI Prism 3730 automated sequencer (Applied Biosystems) with a GeneScan Rox 500 internal size standard. The resultant profiles were described and screened for quality using GeneMapper 3.7 software

Table 1 *Allanblackia* sampled from three countries in Africa for an assessment of genetic variation, using 411 AFLPs and four SSR loci

		AFLPs		SSRs		
Tested material	Species	*N*	*PL*	*N*	*A*	A_5
By site						
Tanzania						
Amani	*A. stuhlmannii*	4	110	12.50	10	9.82
Manyangu	*A. stuhlmannii*	7	165	9.00	12	10.56
Mazumbai	*A. stuhlmannii*	7	204	13.50	10	8.51
Mufindi	*A. stuhlmannii*	8	147	10.25	13	11.12
Uluguru	*A. stuhlmannii*	4	121	8.75	11	9.49
Ghana						
Composite[a]	*A. parviflora*	11	203	18.25	6	4.58
Cameroon						
Bangangte	*A. gabonensis*	9	246	1.75	7	NA
Sangmelima	*A. gabonensis*	2	76	–	–	–
Sangmelima	*A. floribunda*	15 (13)	252	6.75	10	9.05
Yalpenda	*A. floribunda*	12	203	11.75	11	8.27
Yalpenda	*A. stanerana*	9 (7)	170	8.25	9	7.98
Totals		88 (84)	172[b]	100.75	25	8.82[b]
By country						
Tanzania	*A. stuhlmannii*	30	285	54.00	19	9.90[b]
Ghana	*A. parviflora*	11	203	18.25	6	4.58[b]
Cameroon	*A. gabonensis*, *A. floribunda*, *A. stanerana*	47 (43)	357	28.50	16	8.43[b]
By country and species						
Tanzania	*A. stuhlmannii*	30	285	54.00	19	9.90[b]
Ghana	*A. parviflora*	11	203	18.25	6	4.58[b]
Cameroon	*A. gabonensis*	11	259	1.75	7	NA
Cameroon	*A. floribunda*	27 (25)	279	18.50	14	8.66[b]
Cameroon	*A. stanerana*	9 (7)	170	8.25	9	7.98[b]

In total, five species and 11 populations (some stands sympatric) were collected. *N* denotes the number of individuals tested for AFLPs (values in parentheses are the number of individuals used for AMOVA and to determine *PL* values, if different; see "Results" for explanation) and the mean number of individuals scored across loci for SSRs, *PL* the number of polymorphic AFLPs, *A* the total number of alleles revealed at SSR loci, A_5 the total number of SSR alleles corrected by rarefaction to account for varying stand sizes (see Leberg 2002; values corrected to the lowest number of individuals typed from a stand, *N*=5 [i.e. ten haploid samples], NA=not available, excluded from analysis because of limited data)

[a] Collected across an area of~150 km (N/S) by 75 km (E/W) in southwestern Ghana

[b] The arithmetic mean of individual stands

(Applied Biosystems), and unambiguous AFLPs that are expected to be reproducible were scored as present (1) and absent (0) character states. Any marker that occurred in only a single individual was excluded from subsequent analysis, in order to eliminate possible artefacts of PCR amplification.

SSR analysis

SSR analysis involved *de novo* development of a suite of primers for *A. stuhlmannii* according to the method of Brennan et al. (2002). Briefly, 50 µg of DNA from dried leaves of *A. stuhlmannii* (sampled from Amani in Tanzania) was digested overnight at 65°C with Tsp509 (New England Biolabs, Inc.). Tsp509 site-specific adaptors were ligated onto size-fractionated digested DNA and PCR undertaken with specific primers. Product was hybridised to Hybond N+ membranes (Amersham Pharmacia) carrying the following oligonucleotides: $[CA]_{15}$, $[GA]_{15}$, $[AAG]_8$ and $[ATG]_8$. Enriched DNA was ligated into the pGEM-T Easy vector (Promega) and used to transform DH10B cells (Life Technologies). Inserts were sequenced on an ABI Prism 3730 using the M13 reverse primer and Big Dye Terminator 2.0 chemistry (Applied Biosystems). Sequences were quality-scored using the Phred package (www.phrap.com/) and the Blast algorithm (www.ncbi.nlm.nih.gov/blast/Blast.cgi) was used to search non-redundant nucleotide data-

bases. SSRs were identified using the Sputnik programme (espressosoftware.com/pages/sputnik.jsp) and oligonucleotide primers were designed using Primer 3 software (www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi). Primer pairs were tested for their ability to amplify microsatellite loci. For each primer pair, one oligonucleotide was fluorescently labelled with TAMRA, FAN or JOE, and PCR was performed on a PE Applied Biosystems 9700 thermocycler using 10-ng DNA, 1 μM of both forward and reverse primers, 200-μM dNTPs, 1-μL 10×Taq buffer and 0.5 U Taq polymerase (Roche), in a reaction volume of 10 μL. A 'touchdown' programme was used: 94°C 5 min; 94°C 30 s, 65°C decreasing to 58°C at 1°C per cycle 30 s, 72°C 30 s, eight cycles; 94°C 30 s, 58°C 30 s, 72°C 30 s, 25 cycles; 72°C 7 min. Products were diluted tenfold before sizing according to standard protocols on an ABI Prism 3730 sequencer using ABI GeneScan and Genotyper software. Initially, a test sample of 24 individuals, including representative individuals from species, country and site, was screened with 51 primer pairs. Four primer pairs (Table 2) that appeared to reveal clear single-locus amplification products across species were chosen for genotyping all individuals.

Data analysis

Our primary objective was to provide an overview of genetic variation detected by AFLP and SSR markers within and amongst species and populations of *Allanblackia*. To provide an initial visual representation of genetic variation, we undertook a principal coordinate analysis (PCoA) of AFLP and SSR phenotypes, based on the PAST 0.82 software package (Hammer et al. 2002) and Sørensen's similarity coefficient for AFLPs, and GENSTAT Release 10.2 (Rothamsted Experimental Station) and the simple matching coefficient for SSRs (see Legendre and Legendre 1998 for coefficients). To partition AFLP and SSR variation amongst individuals, populations, species and countries, we employed the ARLEQUIN 2.000 software package (Schneider et al. 2000) to undertake an analysis of molecular variance (AMOVA, Excoffier et al. 1992), using 5,000 permutations to assign significance values to estimates. To compare genetic diversity levels by population, country and species, for AFLPs we counted the total number of polymorphic loci (*PL*) revealed, and for SSRs the total number of alleles observed across loci (*A*). In addition, FSTAT 2.9.3.2 (Goudet 2002) was employed to generate SSR allelic richness values for populations based on a rarefaction procedure to account for varying sample sizes (estimates from highly variable loci such as SSRs are very sensitive to differences in sample sizes, and correction is therefore essential; Leberg 2002).

Results

The four primer combinations employed in AFLP analysis revealed 411 clear polymorphisms across 88 individuals, with a mean product presence frequency of 0.256. Initial screening of 51 putative SSR primer pairs on a test sample showed uneven amplification across species. In total, nine primer pairs successfully amplified *A. stuhlmannii* (the species to which primers were designed); of these nine, four pairs gave good amplification across species. Genotyping all individuals with these four pairs (mean of 101 individuals typed across loci) revealed between two (*ABCK43*) and 12 (*ABNC04*) alleles at individual loci, with a total of 25 alleles (Table 2; see "Appendix" for raw data).

Principle coordinate analysis

AFLPs The first two axes of a PCoA of AFLP phenotypes accounted for 27% of variation and illustrated three distinct groups of individuals (Fig. 1a). One group consisted of *A. stuhlmannii* from Tanzania, another of *A. floribunda* and *A. stanerana* from Cameroon (*A. stanerana* only sampled from one location, Yalpenda, where it was coincident with *A. floribunda*), and a third of *A. gabonensis* and *A.*

Table 2 *De novo* primers developed and employed for the amplification of SSR markers in *Allanblackia* sampled from three countries in Africa

Locus	Primer sequences (from 5′)	Repeat motif	No. of alleles	Range (bp)
ABCK09	F: GGC GTT ACA TCA TTG TCG AA R: CCA AAA GCA AAA CGG TCA CT	$(AGT)_6$	7	202–228
ABCK39	F: AAC ACT ATG GCC AGG GTC AC R: GGA TTA TTT CCC GCC AAA AT	$(TC)_{10}$	4	180–192
ABCK43	F: TTG TTT CCC TTT CGT CTG CT R: CAG AGA CAG GGC AGG AAG TC	$(TTTCC)_2$	2	177–179
ABNC04	F: CTG CAT TTC AGG CCA AAG TT R: GAC AGC CGT TGT TTG ATG TG	$(TC)_9$	12	206–228

The repeat motif identified in the original clone, the total number of alleles revealed in typed individuals (mean $N=101$) and the size range of amplified products are indicated. The same annealing temperature (58°C) was used for all primer pairs

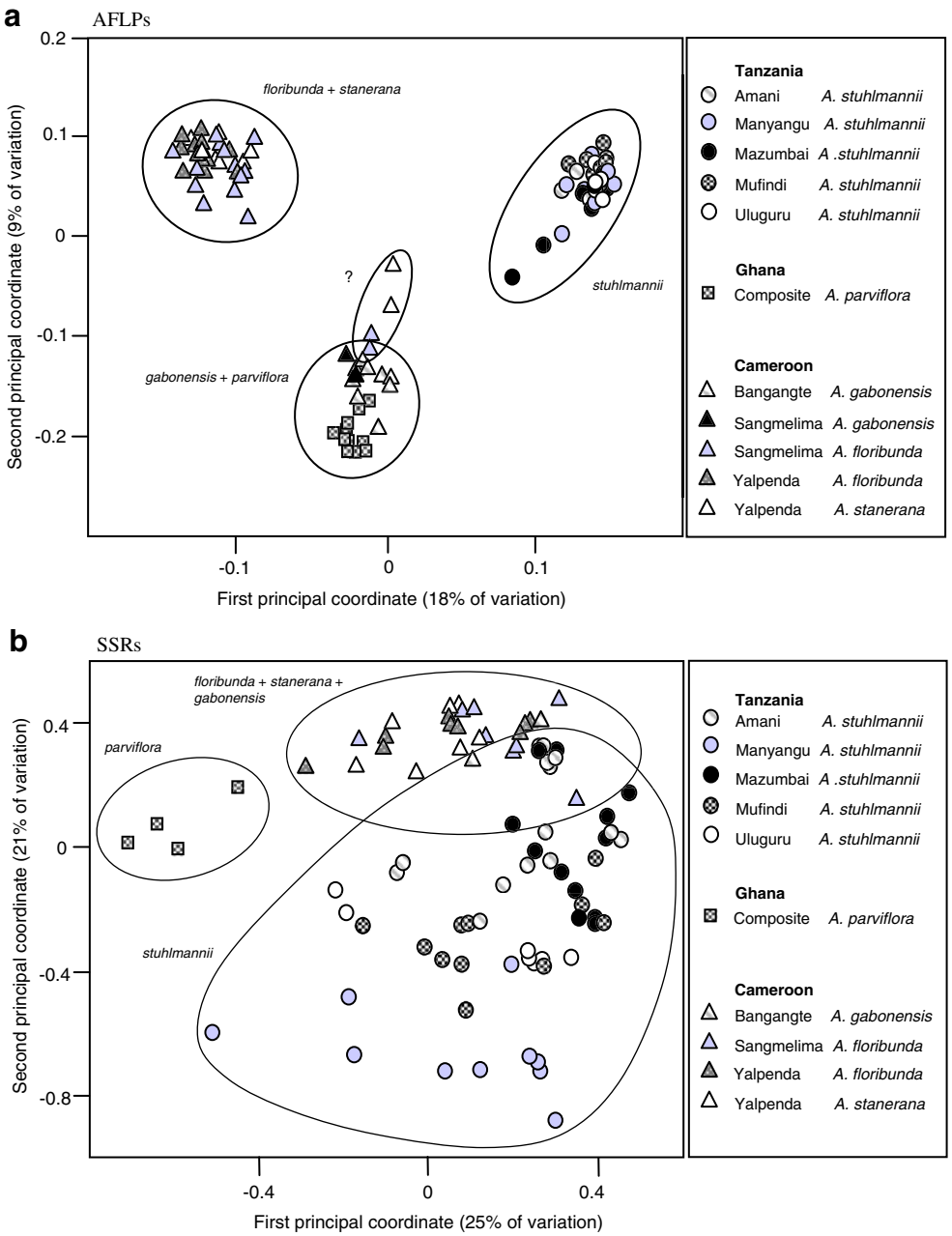

Fig. 1 Principle coordinate analysis of **a** AFLP (N=88) and **b** SSR (N=112) phenotypes of *Allanblackia* sampled from three countries in Africa, using 411 AFLPs and four SSR loci. Groups of individuals are circumscribed by species designation. In **b**, many individuals of *A. parviflora* (Ghana) are coincident in analysis. Four individuals that placed unusually during AFLP ordination are indicated by "?"

parviflora from Cameroon and Ghana, respectively. Separation of *A. stuhlmannii* from other species corresponded with the significant geographical distance (>2,500 km) between the Eastern Arc Mountains of Tanzania and the locations in central and west Africa from which other taxa were sampled. The grouping of *A. parviflora* with *A. gabonensis* was unexpected because known distribution ranges of the two species do not overlap or border geographically (*A. gabonensis*=Cameroon and Gabon only, *A. parviflora*=Sierra Leone to Ghana). If anything, *A. parviflora* was expected to show a closer relationship with *A. floribunda*, as the latter species has a distribution that appears to extend west from Cameroon (into Nigeria), and overall distributions are therefore more proximate. Individuals identified during sampling as *A. floribunda* and *A. gabonensis* and that were collected from the same location (Sangmelima in Cameroon) generally appeared distinct. However, two individuals identified during collection at Sangmelima as *A. floribunda* grouped with *A. gabonensis*, and it appears likely that these individuals were misidentified during leaf sampling as the alternate taxon (although *Allanblackia* species can normally be distinguished by fruit shape, identification is more difficult when trees are not fruiting or are immature and are out of season or male). In addition, two individuals identified during collection as *A. stanerana* (from Yalpenda) occupied an intermediate posi-

tion in PCoA. It is possible that these individuals were also misidentified during leaf sampling and may represent another species or, potentially, inter-specific hybrids. Alternatively, though less likely, these individuals may represent 'true' *A. stanerana*, in which case the majority of apparent *A. stanerana* individuals sampled at Yalpenda may have been misidentified and could in fact represent the coincident *A. floribunda*. Due to the indeterminate status of these four individuals—two '*A. floribunda*' and two '*A. stanerana*'—they were excluded from subsequent analyses of AFLP data. Regardless of these individuals, it is evident that significant between species variation exists in Cameroon, including in situations where distributions are sympatric.

SSRs Interpretation of SSR ordination should be cautious because it is based on four loci only. The first two axes of a PCoA of SSR phenotypes accounted for 46% of variation. Groups of individuals defined by species were apparent, although these groups were not as distinct as those revealed by AFLP markers, with some overlap in ordination (Fig. 1b). This difference was not unexpected because of the small number of SSR loci used and the high allelic variation at individual loci (the presence of many alleles at a locus means that differences are unlikely to be 'absolute'). One group consisted of *A. stuhlmannii* from Tanzania, another of *A. floribunda*, *A. gabonensis* and *A. stanerana* from Cameroon, and a third of *A. parviflora* from Ghana. Apart from the generally looser associations of individuals within groups, the main differences from AFLP analysis were: (a) the different position of *A. gabonensis* (now grouped 'by country' with other Cameroonian species; NB, only two individuals of *A. gabonensis* included in SSR analysis); and (b) the particularly high spread in coordinates for *A. stuhlmannii* individuals compared to other species (especially compared to *A. parviflora*, for which many individuals were coincident in SSR analysis). In the case of *A. stuhlmannii*, some differentiation by population was evident (e.g. Manyangu vs. Mazumbai). Since *A. floribunda*, *A. gabonensis* and *A. stanerana* grouped in SSR ordination, the issue of possible misidentification explored above with AFLPs could not be addressed here, and no individuals were excluded from subsequent analyses of SSR data.

Analysis of molecular variance

Analysis of AFLPs and SSRs indicated significant variation structured by population, species and country (Table 3).

Table 3 Analysis of molecular variance (AMOVA) for 411 AFLPs and four SSR loci amongst *Allanblackia* individuals (N=84 and 112, respectively) sampled from three countries in Africa

	AFLPs					SSRs				
Source of variation	*df*	MSD	Variance component	Percentage of total	*P* value	*df*	MSD	Variance component	Percentage of total	*P* value
All stands nested by species										
Amongst species	4	423.2	22.45	35.5	<0.001	4	16.51	0.310	29.9	<0.001
Amongst stands within species	6	63.5	4.08	6.4	<0.001	5	6.51	0.257	24.7	<0.001
Amongst individuals within stands	73	36.8	36.79	58.1	<0.001	214	0.47	0.471	45.4	0.035
All stands nested by country										
Amongst countries	2	634.5	20.56	30.9	<0.001	2	32.17	0.387	36.0	<0.001
Amongst stands within countries	8	100.6	9.09	13.7	<0.001	7	4.89	0.219	20.3	<0.001
Amongst individuals within stands	73	36.8	36.79	55.4	<0.001	214	0.47	0.471	43.7	0.001
Tanzanian stands only										
Amongst stands	4	69.9	6.07	15.1	<0.001	4	7.88	0.296	33.2	<0.001
Amongst individuals within stands	25	34.2	34.20	84.9		119	0.60	0.596	66.8	
Cameroonian stands only										
Amongst stands	4	131.30	11.27	22.2	<0.001	3	0.90	0.029	5.6	0.123
Amongst individuals within stands	38	39.5	39.46	77.8		58	0.48	0.484	94.4	

Nested analysis was undertaken by species and country. Separate analysis was also undertaken for stands from Tanzania and Cameroon only. Degrees of freedom (*df*), mean squared deviations (MSD) and the significance (*P*) of the variance components are shown. Significance values were based on the random permutation (5,000 times) of individuals assuming no genetic structure

Consistent with differences revealed during ordination, AFLPs revealed more variation amongst Cameroonian stands (22%, three species: *A. floribunda*, *A. gabonensis* and *A. stanerana*) than amongst Tanzanian stands (15%, single species only: *A. stuhlmannii*). Consistent with the difference observed between AFLPs and SSRs in ordination (less distinct species groupings for the latter), an analysis of stands nested by species revealed relatively more variation amongst stands within species with SSRs (AFLPs=6%, SSRs=25%). Unlike AMOVA of AFLPs (see above), but consistent with ordination (particularly high spread in coordinates for *A. stuhlmannii* individuals in Tanzania), SSRs revealed much less variation amongst Cameroonian stands (6%) than amongst Tanzanian stands (33%). SSRs in Cameroon represented the only AMOVA comparison where the 'amongst stands' component was not considered significant (P=0.123).

Diversity estimates

The number of polymorphic AFLPs (*PL*) indicated that material sampled from Cameroon contained high diversity (*PL*=357) compared to Tanzania (*PL*=285; Table 1). This may partly be due to sample size differences (N=43 and 30, respectively). However, consistent with ordination and AMOVA, it also appears to be attributable to the differentiation between collected species in Cameroon (*A. floribunda* and *A. stanerana* vs. *A. gabonensis*). Furthermore, it appears to relate to the high diversity of certain populations in Cameroon (*A. gabonensis* from Bangangte, *PL*=246; *A. floribunda* from Sangmelima, *PL*=252), although the estimates for single stands should be treated with particular caution because of varying sample sizes. Material sampled from Ghana did not have high variation (*PL*=203) but nor was diversity low if considering the relatively small sample size tested (N=11). Consistent with ordination (particularly high spread in coordinates for *A. stuhlmannii* individuals in Tanzania and coincidence of *A. parviflora* individuals in Ghana), SSR allelic richness values corrected by rarefaction, to a sample size equivalent to five individuals per population (A_5), indicated that Tanzanian stands were on average most diverse, followed by Cameroonian stands and then (much less diverse) Ghanian material (mean A_5 of 9.9, 8.4 and 4.6, respectively). The very low SSR diversity of Ghanian material was despite the fact that it is a 'composite' sample collected from a relatively wide geographic area (Table 1). Considering the total number of SSR alleles (A) revealed by country, Tanzania was more diverse (A=19) than Cameroon (A=16), even though the estimate for the former nation is based on only one species (*A. stuhlmannii*) and the latter nation on three (*A. floribunda*, *A. gabonensis* and *A. stanerana*). Again, Ghana (*A. parviflora* only) had a much lower value (A=6).

Discussion

Distribution of genetic variation in *Allanblackia* and implications for management

Genetic diversity data for indigenous African fruit trees subject to active domestication are relatively scarce (e.g. Lowe et al. 2000; Kadu et al. 2006; Jamnadass et al. 2009) and only limited comparison with other studies is therefore possible. In an assessment of genetic variation in bush mango (*Irvingia gabonensis* and *Irvingia wombolu*) in the humid forests of central Africa, Lowe et al. (2000) found significant variation amongst species that are sometimes sympatric and occasional field misidentification of samples. Our study on *Allanblackia* indicated similar results, with AFLPs showing significant differentiation between some species and a similar level of apparent taxonomic misidentification during field collection. The level of field misidentification observed is not unusual for related taxa that co-occur and, indeed, appears to be relatively low, indicating that field classification is generally accurate. However, our results are cautionary if a domestication strategy is to be based on a single taxon in areas where species distributions overlap, with care being required in germplasm sampling. This is especially the case if collecting material vegetatively rather than as seed (as species-distinguishing characteristics are more evident during fruit production and seed collection) in Cameroon and, possibly, in Tanzania (where *A. stuhlmannii* and *A. ulugurensis* are occasionally sympatric; *A. ulugurensis* not tested here but should be the subject of future study).

Whereas AFLPs indicated clear distinction between some *Allanblackia* species, other taxa overlapped in ordination. Sometimes, geographic proximity and genetic relatedness corresponded, e.g. sympatric *A. floribunda* and *A. stanerana* were genetically similar; well-separated *A. stuhlmannii* was clearly genetically different from other tested species. On other occasions, proximity and relatedness did not clearly correspond, e.g. sympatric *A. floribunda* and *A. gabonensis* were clearly genetically distinct; well-separated *A. gabonensis* and *A. parviflora* were genetically similar. AFLP data indicate, therefore, that the relationships between *Allanblackia* taxa are not simple and likely reflect complex evolutionary processes related to migration and dispersal. Although overland migration (related to past climate change) and long-distance seed dispersal have not been widely studied using genetic techniques in African trees (although see e.g. Muchugi et al. 2006 and 2008 for the medicinal trees *Prunus africana* and *Warburgia*, respectively), such research would be very relevant for deriving present-day management strategies. In any case, it appears evident that a simple 'sampling-by-distance' model for assessing variation in *Allanblackia* will

not be adequate, and other points such as ecological boundaries and divisions between phytochoria should be taken into consideration (White 1993).

Within the framework of a national management programme, Cameroon clearly presents particular opportunities and challenges for domestication and conservation, with high overall levels of genetic variation that may be exploited in the development of *Allanblackia* cultivars, sympatric distributions and possible misattributions during classification. It is not known, however, whether the differences revealed by AFLPs correspond to differences in important commercial characteristics of *Allanblackia* such as oil composition (e.g. are the fat profiles of *A. floribunda* and *A. gabonensis* very different?). These aspects should be explored through genomic approaches, in order to assess possibilities for the selection of superior types.

Developing SSR markers and implications for gene flow studies

Due to the generally high allelic diversity revealed, SSRs are the method of choice for detailed population genetic research on gene flow and parentage (Woodhead et al. 2005). Such studies are crucial for *Allanblackia* because the degree of connectivity between extant forest and neighbouring cultivated populations will determine the way in which planted stands are established. In particular, it will help determine the number of male trees that need to be established in the farm landscape and the locations in which they should be placed relative to female fruit-producing trees.

AFLPs provide a baseline to assess the cross-taxa performance of SSR markers developed from *A. stuhlmannii*. Initial screening of genomic library-derived SSRs indicated that only around half of the primer pairs (four of nine) that successfully amplified *A. stuhlmannii* also gave good amplification of all four other tested *Allanblackia* species. Furthermore, whilst these four SSR primer pairs revealed common features with AFLP analysis within and amongst taxa, differences were also evident, e.g. in the level of discrimination between species and the relative genetic diversity of stands. Most importantly, the 'source' species for SSR primers appeared relatively more diverse for SSRs than for AFLPs, when compared to other taxa and countries. Comparisons based on only four SSR loci should be treated with caution; however, these results, in combination with the relatively low overall success rate for cross-species amplification, indicate limitations in primer transferability.

Possibly, low diversity values for SSRs compared to AFLPs outside Tanzania indicate the presence of null (non-amplifying) SSR alleles in species other than *A. stuhlmannii*; such alleles are a relatively common occurrence at SSR regions (Woodhead et al. 2005). Null alleles result in biassed estimates of allelic and genotypic frequencies and underestimation of diversity. Another explanation for relatively low SSR diversity outside Tanzania may be an ascertainment bias in the initial detection of repeats in *A. stuhlmannii*, and similar effects have been observed in other species (Almeida and Penha-Goncalves 2004). Overall, our observations suggest that, whilst the SSR primers typed here revealed quite high allelic variation in *A. stuhlmannii* (total $A=19$, or an average of 4.75 alleles per locus) and should, therefore, be suitable for detailed population genetic studies of this species, they are unlikely to be very useful when extended to the analysis of other commercially important taxa in the genus, for which detailed genetic research is also crucial. Our small population sizes preclude any more formal analysis for the presence of null alleles, although larger samples could be tested with software packages that check for homozygous excesses, such as MICRO-CHECKER (van Oosterhout et al. 2004).

Final considerations

AFLP data presented here provide an initial overview of genetic variation in an important indigenous African fruit tree newly subject to domestication, and we discuss the structuring of this variation in the context of management. More detailed sampling of species and populations is required, and research should consider the consequences of migration and dispersal in determining genetic structure, since these processes may have important implications for domestication and conservation of specific taxa. SSR markers derived from a genomic library to one species do not appear to be applicable for the analysis of other taxa, suggesting that an alternative way to develop 'generic' SSR primers needs to be found. This may be possible through the expressed sequence tag method based on cDNA library production, since such primer sequences appear to be more conserved across taxa (Woodhead et al. 2005). Although this method for developing primers is more expensive than one based on genomic libraries, it appears worth testing in an important genus such as *Allanblackia* that has significant future commercial potential. Otherwise, SSR primer suites may need to be designed specifically for each species.

Acknowledgements This paper benefited considerably from discussions with colleagues at ICRAF, particularly Alain Atangana and Alice Muchugi. Research on *Allanblackia* domestication at ICRAF is supported by a number of partners: funding in this instance was provided from the World Bank through the CGIAR System-wide Genetic Resources Programme. CACK was supported through the Rothamsted International African Fellows Programme to undertake laboratory research at SCRI. This programme supports agriculture in sub-Saharan Africa by helping to develop strategic alliances that develop local scientific capacity for sustainable agricultural production. NJC was supported through the Boyd Postdoctoral Fellowship at The Field Museum.

Appendix

Table 4 Frequencies of alleles at four SSR loci in *Allanblackia* stands sampled from three countries in Africa

Country	Tanzania	Tanzania	Tanzania	Tanzania	Tanzania	Ghana	Cameroon	Cameroon	Cameroon	Cameroon	
Site	Amani	Manyangu	Mazumbai	Mufindi	Uluguru	Composite	Bangangte	Sangmelima	Yalpenda	Yalpenda	
Species	*A. stuhlmannii*	*A. stuhlmannii*	*A. stuhlmannii*	*A. stuhlmannii*	*A. stuhlmannii*	*A. parviflora*	*A. gabonensis*	*A. floribunda*	*A. floribunda*	*A. stanerana*	Total
ABCK09	*N*=6	*N*=9	*N*=10	*N*=10	*N*=8	*N*=19	*N*=1	*N*=8	*N*=11	*N*=9	*N*=91
Allele202	0.417	0.389	1.000	0.550	0.500	0.000	0.000	0.063	0.000	0.000	0.286
Allele204	0.000	0.111	0.000	0.200	0.000	0.000	0.000	0.000	0.000	0.000	0.033
Allele205	0.417	0.500	0.000	0.250	0.500	0.974	0.500	0.188	0.682	0.667	0.522
Allele207	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.188	0.000	0.000	0.016
Allele208	0.167	0.000	0.000	0.000	0.000	0.026	0.500	0.375	0.273	0.278	0.115
Allele210	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.125	0.000	0.000	0.011
Allele211	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.063	0.045	0.056	0.016
ABCK39	*N*=15	*N*=6	*N*=15	*N*=10	*N*=9	*N*=19	*N*=2	*N*=6	*N*=12	*N*=8	*N*=102
Allele180	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.005
Allele188	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.186
Allele190	0.733	0.333	0.767	0.100	0.111	0.000	1.000	1.000	1.000	1.000	0.534
Allele192	0.267	0.667	0.200	0.900	0.889	0.000	0.000	0.000	0.000	0.000	0.275
ABCK43	*N*=15	*N*=12	*N*=15	*N*=11	*N*=9	*N*=19	*N*=2	*N*=8	*N*=12	*N*=9	*N*=112
Allele178	0.000	0.958	0.000	0.045	0.111	0.000	0.000	0.000	0.000	0.000	0.116
Allele179	1.000	0.042	1.000	0.955	0.889	1.000	1.000	1.000	1.000	1.000	0.884
ABNC04	*N*=14	*N*=9	*N*=14	*N*=10	*N*=9	*N*=16	*N*=2	*N*=5	*N*=12	*N*=7	*N*=98
Allele206	0.000	0.000	0.000	0.000	0.000	0.969	0.000	0.000	0.083	0.000	0.168
Allele207	0.000	0.000	0.000	0.000	0.167	0.000	0.000	0.700	0.625	0.429	0.158
Allele209	0.000	0.000	0.071	0.000	0.056	0.000	0.000	0.000	0.167	0.071	0.041
Allele210	0.000	0.000	0.000	0.000	0.000	0.000	0.250	0.000	0.000	0.000	0.005
Allele211	0.250	0.111	0.393	0.200	0.056	0.000	0.250	0.300	0.042	0.429	0.184
Allele213	0.286	0.222	0.214	0.100	0.611	0.000	0.500	0.000	0.042	0.000	0.173
Allele215	0.179	0.389	0.214	0.000	0.000	0.031	0.000	0.000	0.042	0.000	0.102
Allele217	0.000	0.000	0.000	0.000	0.111	0.000	0.000	0.000	0.000	0.071	0.015
Allele219	0.000	0.056	0.000	0.100	0.000	0.000	0.000	0.000	0.000	0.000	0.015
Allele221	0.000	0.000	0.000	0.250	0.000	0.000	0.000	0.000	0.000	0.000	0.026
Allele226	0.286	0.222	0.107	0.300	0.000	0.000	0.000	0.000	0.000	0.000	0.107
Allele228	0.000	0.000	0.000	0.050	0.000	0.000	0.000	0.000	0.000	0.000	0.005

N denotes the number of individuals typed for a locus

References

Adin A, Weber JC, Sotelo Montes C, Vidaurre H, Vosman B, Smulders MJM (2004) Genetic differentiation and trade among populations of peach palm (*Bactris gasipaes* Kunth) in the Peruvian Amazon—implications for genetic resource management. Theor Appl Genet 108:1564–1573

Almeida P, Penha-Goncalves C (2004) Long perfect dinucleotide repeats are typical of vertebrates, show motif preferences and size convergence. Mol Biol Evol 21:1226–1233

Amanor K, Ghansah W, Hawthorne WD, Smith G (2003) Sustainable wild harvesting: best practices document for wild harvesting of *Allanblackia* seeds from forest and farmlands with some additional notes for sustainable establishment and management of smallholder plantation and agroforestry systems that incorporate a significant *Allanblackia* component. Ghana report. World Conservation Union (IUCN), Gland and Unilever, Rotterdam

Atangana AR, Tchoundjeu Z, Asaah EK, Simons AJ, Khasa DP (2006) Domestication of *Allanblackia floribunda*: amenability to vegetative propagation. For Ecol Manag 237:246–251

Attipoe L, van Andel A, Nyame SK (2006) The Novella project: developing a sustainable supply chain for *Allanblackia* oil. In: Ruben R, Slingerland M, Nijhoff H (eds) Agro-food chains and networks for development. Springer, Amsterdam, pp 179–189

Bergl RA, Oates JF, Fotso R (2007) Distribution and protected area coverage of endemic taxa in West Africa's Biafran forests and highlands. Biol Conserv 134:195–208

Brennan R, Jorgensen L, Woodhead M, Russell J (2002) Development and characterisation of SSR markers in *Ribes* species. Mol Ecol Notes 2:327–330

Burgess ND, Butynski TM, Cordeiro NJ, Doggart NH, Fjeldså J, Howell KM, Kilahama FB, Loader SP, Lovett JC, Mbilinyi B, Menegon M, Moyer DC, Nashanda E, Perkin A, Rovero F, Stanley WT, Stuart SN (2007) The biological importance of the Eastern Arc Mountains of Tanzania and Kenya. Biol Conserv 134:209–231

Donald PF (2004) Biodiversity impacts of some agricultural commodity production systems. Conserv Biol 18:17–37

Doyle JJ, Doyle JL (1987) CTAB DNA extraction in plants. Phytochem Bull 19:11–15

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479–491

Goudet J (2002) FSTAT, a program to estimate and test gene diversities and fixation indices, version 2.9.3.2. www2.unil.ch/izea/softwares/fstat.html

Hammer O, Harper DAT, Ryan PD (2002) PAST—palaeontological statistics, version 0.82. Palaeontological Museum, University of Oslo, Oslo

Hollingsworth PM, Dawson IK, Goodall-Copestake WP, Richardson JE, Weber JC, Sotelo Montes C, Pennington RT (2005) Do farmers reduce genetic diversity when they domesticate tropical trees? A case study from Amazonia. Mol Ecol 14:497–501

Jamnadass R, Lowe A, Dawson IK (2009) Molecular markers and the management of tropical trees: the case of indigenous fruits. Trop Plant Biol 2:1–12

Kadu CAC, Imbuga M, Jamnadass R, Dawson IK (2006) Genetic management of indigenous fruit trees in southern Africa: a case study of *Sclerocarya birrea* based on nuclear and chloroplast variation. S Afr J Bot 72:421–427

Kremen C (2005) Managing ecosystem services: what do we need to know about their ecology? Ecol Lett 8:468–479

Leakey RRB, Tchoundjeu Z, Schreckenberg K, Shackleton SC, Shackleton CM (2005) Agroforestry tree products (AFTPs): targeting poverty reduction and enhanced livelihoods. Int J Agr Sustain 3:1–23

Leberg PL (2002) Estimating allelic richness: effects of sample size and bottlenecks. Mol Ecol 11:2445–2449

Legendre P, Legendre L (1998) Numerical ecology, 2 English editionth edn. Elsevier Science, Amsterdam

Lowe AJ, Gillies ACM, Wilson J, Dawson IK (2000) Conservation genetics of bush mango from central/west Africa: implications from random amplified polymorphic DNA analysis. Mol Ecol 9:831–841

Marshall D, Schreckenberg K, Newton AC (eds) (2006) Commercialisation of non-timber forest products: factors influencing success. Lessons learned from Mexico and Bolivia and policy implications for decision-makers. United Nations Environment Programme-World Conservation Monitoring Centre, Cambridge

McNeely JA, Scherr SJ (2001) Common ground, common future: how ecoagriculture can help feed the world and save wild biodiversity. The World Conservation Union (IUCN), Gland

Mittermeier RA, Robles-Gil P, Hoffmann M, Pilgrim JD, Brooks TM, Mittermeier CG, Lamoreux JL, da Fonseca GAB (eds) (2004) Hotspots revisited: earth's biologically richest and most endangered eco-regions, 2nd edn. CEMEX, Mexico City

Muchugi A, Lengkeek AG, Kadu CAC, Muluvi GM, Njagi ENM, Dawson IK (2006) Genetic variation in the threatened medicinal tree *Prunus africana* in Cameroon and Kenya: implications for current management and evolutionary history. S Afr J Bot 72:498–506

Muchugi A, Muluvi GM, Kindt R, Kadu CAC, Simons AJ, Jamnadass RH (2008) Genetic structuring of important medicinal species of genus *Warburgia* as revealed by AFLP analysis. Tree Genet Genomes 4:787–795

Schneider S, Roessli D, Excoffier L (2000) Arlequin, version 2.000: a software for population genetic data analysis. Genetics and Biometry Laboratory, University of Geneva, Geneva

van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Mol Ecol Notes 4:535–538

van Rompaey R (2003) Distribution and ecology of *Allanblackia* spp. (Clusiaceae) in African rain forests, with special attention to the development of a wild picking system of the fruits. ECOSYN Consulting, Wageningen and Unilever Research Laboratories, Vlaardingen

Vantomme P (2004) Extraction of nonwood forest products. In: Vinceti B, Amaral W, Meilleur B (eds) Challenges in managing forest genetic resource for livelihoods: examples from Argentina, Brazil. Bioversity International (formally the International Plant Genetic Resources Institute), Rome, pp 51–69

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–4414

White F (1993) Refuge theory, ice-age aridity and the history of tropical biota: an essay in plant geography. Fragm Florist Geobot suppl. 2:385–409

Wily LA (2003) Participatory forest management in Africa: an overview of progress and issues. In: Second international workshop on participatory forestry in Africa (ed) Defining the way forward: sustainable livelihoods and sustainable forest management through participatory forestry. Arusha, United Republic of Tanzania, 18–22 February 2002. Food and Agriculture Organisation, Rome, pp 31–58

Woodhead M, Russell J, Squirrell J, Hollingsworth PM, Mackenzie K, Gibby M, Powell W (2005) Comparative analysis of population genetic structure in *Athyrium distentifolium* (Pteridophyta) using AFLPs and SSRs from anonymous and transcribed gene regions. Mol Ecol 14:1681–1695