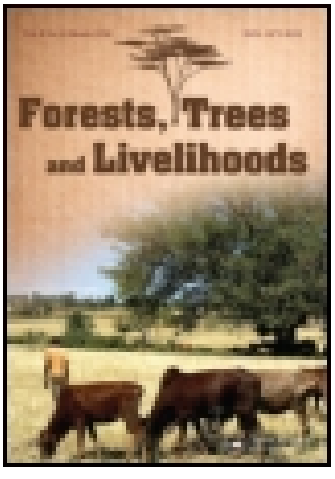

Forests, Trees and Livelihoods
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tftl20

Genetic diversity in shea tree (Vitellaria paradoxa subspecies nilotica) ethno-varieties in Uganda assessed with microsatellite markers

Samson Gwali[ab], Alexandre Vaillant[c], Grace Nakabonge[a], John Bosco Lamoris Okullo[a], Gerald Eilu[a], Alice Muchugi[d] & Jean-Marc Bouvet[c]

[a] School of Forestry, Environment and Geographical Sciences, College of Agricultural Sciences, Makerere University, P. O. Box 7062 Kampala, Uganda

[b] Tree Improvement and Germplasm Research Programme, National Forestry Resources Research Institute (NaFORRI), P. O. Box 1752 Kampala, Uganda

[c] UMR AGAP, Équipe GFP, Campus Lavalette - TA A-108/1-34398 Montpellier Cedex 5, France

[d] Genetic Resources Unit, World Agroforestry Centre, P. O. Box 30677-00100 Nairobi, Kenya

Published online: 18 Sep 2014.

To cite this article: Samson Gwali, Alexandre Vaillant, Grace Nakabonge, John Bosco Lamoris Okullo, Gerald Eilu, Alice Muchugi & Jean-Marc Bouvet (2014): Genetic diversity in shea tree (Vitellaria paradoxa subspecies nilotica) ethno-varieties in Uganda assessed with microsatellite markers, Forests, Trees and Livelihoods, DOI: 10.1080/14728028.2014.956808

To link to this article: http://dx.doi.org/10.1080/14728028.2014.956808

Forests, Trees and Livelihoods, 2014
http://dx.doi.org/10.1080/14728028.2014.956808

Genetic diversity in shea tree (*Vitellaria paradoxa* subspecies *nilotica*) ethno-varieties in Uganda assessed with microsatellite markers

Samson Gwali[a,b]*, Alexandre Vaillant[c], Grace Nakabonge[a], John Bosco Lamoris Okullo[a], Gerald Eilu[a], Alice Muchugi[d] and Jean-Marc Bouvet[c]

[a]*School of Forestry, Environment and Geographical Sciences, College of Agricultural Sciences, Makerere University, P. O. Box 7062 Kampala, Uganda;* [b]*Tree Improvement and Germplasm Research Programme, National Forestry Resources Research Institute (NaFORRI), P. O. Box 1752 Kampala, Uganda;* [c]*UMR AGAP, Équipe GFP, Campus Lavalette - TA A-108/1-34398 Montpellier Cedex 5, France;* [d]*Genetic Resources Unit, World Agroforestry Centre, P. O. Box 30677-00100 Nairobi, Kenya*

Shea trees (*Vitellaria paradoxa* C. F. Gaertn.) are classified locally into several folk or ethno-varieties by farmers in Uganda. It is, however, not clear whether this folk classification is supported by genetic differentiation between ethno-varieties. Genetically linked traits from folk classification are useful in conservation and breeding programmes. A total of 118 individual shea trees constituting 28 ethno-varieties sampled from three farming systems of Uganda were analysed using microsatellite markers. The number of alleles amplified per microsatellite locus ranged from 6 to 13 with an average of 9.5, with a total of 106 alleles identified. Observed (H_o) and expected heterozygosity (H_e) per locus ranged from 0.366 to 0.934 and 0.580 to 0.840, respectively. Mean H_o and H_e values for all loci across all ethno-varieties were 0.633 and 0.727, respectively. Analysis of molecular variance indicated that most of the variation (86.28%) occurred within individual trees; 11.25% was found among individual trees within ethno-varieties while 2.47% was found among ethno-varieties. The in-breeding index ($f = 0.130$), fixation index ($\theta_P = 0.025$), gene flow value ($N_m = 6.56$) and cluster analysis show that all shea tree ethno-varieties were a single out-crossing population with very low genetic differentiation and high gene flow. The low differentiation in shea tree ethno-varieties was most likely due to the utilization of non-genetic traits in folk classification. However, while ethno-variety genetic structure was very weak, overall spatial population structure indicated the presence of three populations (West Nile, Northern and Teso). The West Nile population was more distantly related to the other two most likely due to isolation barriers such as the Rift Valley, Lake Albert and River Nile.

Keywords: farming system; genetic variation; gene flow; heterozygosity; folk classification; population structure; SSR fingerprinting

Introduction

Vitellaria paradoxa C. F. Gaertn. ('shea butter tree' or 'shea tree' in English and 'karite' in French) is an indigenous African tree species of the Sudano-Sahelian belt (Hall et al. 1996). Its distribution spans 20 African countries, from Senegal in West Africa to Uganda in eastern Africa (Hall et al. 1996). The species is taxonomically divided into two subspecies, *paradoxa* (found in West and central Africa) and *nilotica* (found in east Africa). Shea trees bear fruits whose nuts are cherished locally and internationally for their high oil content. It is the most important source of oil for the people in the shea belt (Gwali, Okullo, et al. 2012). The shea fruit pulp is eaten as a food supplement, especially during

*Corresponding author. Email: gwalis@yahoo.co.uk

the dry season when the fruits mature and ripen. In some African countries such as Burkina Faso, shea fruit pulp is locally processed into fruit jam (Sanou & Lamien 2011). Shea trees are also utilized for medicinal and cultural purposes, such as in the treatment of wounds, making of funeral beds, pregnancies, births and weddings (Hall et al. 1996; Gwali, Okullo, et al. 2012). There are reports of the use of shea oil in war rituals by some traditional communities in Uganda (Sturges 2008). Shea trees are therefore economically important to the local economies and lifestyles of many African traditional communities.

Due to the high degree of socio-economic and cultural significance of shea trees, a system of folk classification below the subspecies level has evolved among local communities in Uganda. While folk classifications are crucial in guiding selection and breeding studies, they are usually influenced by various socio-economic factors. A recent survey of folk classification of subspecies *nilotica* in Uganda by Gwali et al. (2011) indicated that local people classify shea trees into 'ethno-varieties' based on criteria such as fruit pilosity; fruit/nut size and shape; nut colour; pulp colour, quantity, taste and eubstance (hardness or softness) as well as habitat preferences. Shea trees that produce fruits with sweet pulp are, for example, classified as a distinct variety from trees that produce fruits with astringent pulp. Similarly, trees that bear large fruits are classified as different from trees that bear small fruits. Fruit traits, such as size and shape, are, however, very plastic due to influence from environmental factors (De Smedt et al. 2011). Since these phenotypic traits are usually products of long-term adaptation, it is important to understand how genetically fixed they are.

Knowledge about genetic diversity within and relationships among shea tree ethno-varieties would be useful for designing effective conservation and breeding programmes. Genetic markers have shown important and critical application in revealing the genetic structure of plant varieties. Microsatellites are excellent molecular markers in characterizing genetic diversity, population history and signature of selection in a range of plant species. Microsatellites are highly polymorphic, co-dominant, reproducible and easy to interpret. Unfortunately, shea tree genetics, especially of farmers' varieties, is still poorly understood. Most studies that have characterized the genetic variation in shea trees have not focussed on unravelling the genetic basis of farmers' (folk) classifications. So far, several authors have studied the existing diversity and pattern of genetic variation within shea trees at different geographic levels using isozymes (Lovett & Haq 2000), neutral DNA markers (Bouvet et al. 2004; Kelly et al. 2004; Allal et al. 2008) or a few genes involved in oil production pathways (Allal et al. 2011).

Since farmers' criteria and perceptions in varietal selection are pivotal to crop improvement programmes, it is necessary to identify and understand the genetic basis of shea tree ethno-variety classifications. This work represents the first molecular characterization of shea tree ethno-varieties. The main objective of this study was to characterize shea tree ethno-varieties in Uganda using microsatellite markers. The specific objectives were (1) to describe the distribution of genetic variation within and between shea tree ethno-varieties and (2) investigate whether the pattern of genetic variation is concordant with shea tree folk classification in Uganda. Our hypothesis was that there is no genetic basis for shea tree folk classification and as such, these ethno-varieties are subgroups of a single large randomly mating population.

Materials and methods

Study area

Uganda is located astride the equator, stretching from latitudes 4° 12′ N to 1° 29′ S and longitudes 29° 34′ W to 35° 0′ E with an area of 241,551 km^2. The country can be divided into

seven broad farming systems on the basis of cropping patterns, climate, livelihoods, soil and natural resources (NEMA 1996). Within each of these farming systems, the ecological conditions (soil types, topography and rainfall) and agricultural practices are fairly homogeneous. This study was conducted in three farming systems that form the extent of shea tree distribution in Uganda, i.e. Teso, Northern and West Nile (Figure 1). Shea trees form a significant tree component of the vegetation in the three farming systems, hence their classification as *Combetrum/Butyrospermum* and grass savannas of Uganda (Langdale-Brown et al. 1964). According to NEMA (1996), the Teso farming system is characterized by bi-modal rainfall (mean annual rainfall of 1300 mm) on sandy-loam soils of medium to low fertility. The main staple crops are cassava (*Manihot esculenta* Crantz), finger millet (*Eleusine coracana* (L.) Gaertn.) and maize (*Zea mays* L.). There is short grassland ideal for grazing. The Northern farming system receives mono-modal annual rainfall ranging from 800 to 2000 mm. There is a long dry season and drought tolerant annuals such as finger millet, sesame (*Sesanum indicum* L.), cassava and sorghum (*Sorghum bicolor* (L.) Moench) are cultivated. Tobacco (*Nicotiana tabacum* L.) and cotton (*Gossypium hirsutum* L.) are major cash crops. Communal grazing abounds in the extensive grasslands. In the West Nile system, rainfall patterns are similar to the Northern system with greater rain at higher elevation. Tobacco is the main cash crop, although intercropping is common with a wide variety of crops such as sesame and sunflower (*Helianthus annuus* L.). The system is in the sub-humid zone and livestock activities are limited by the presence of tsetse fly.

Plant materials

Individual trees of putative shea tree ethno-varieties (Gwali et al. 2011) were identified by key informants during field visits. Shea trees identified under a single name within a

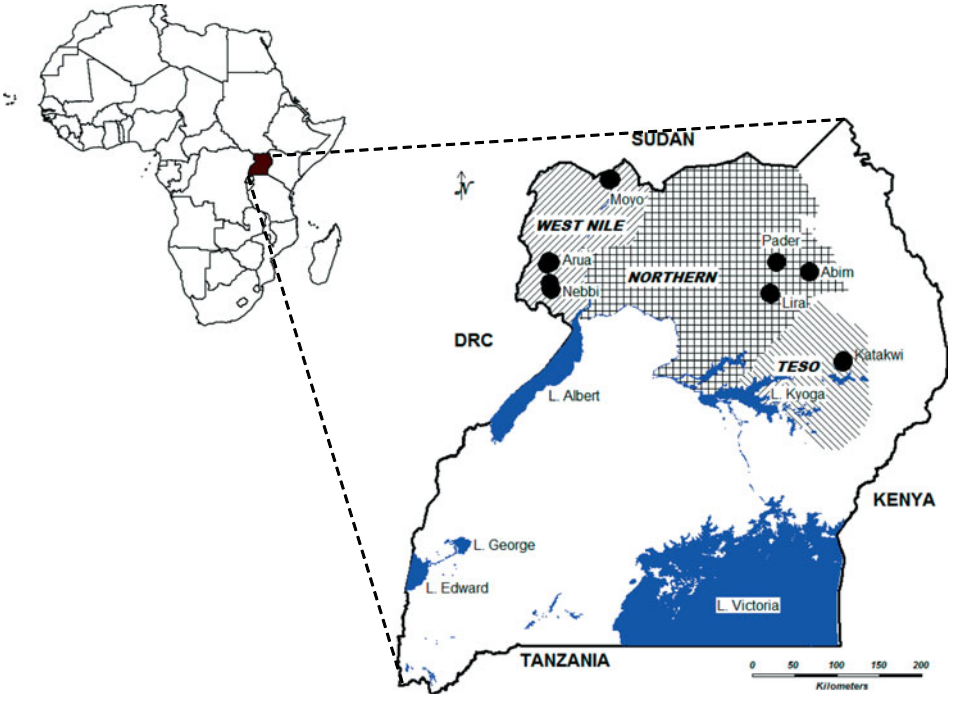

Figure 1. Study area indicating location of study sites in three farming systems.

particular ethnic group were considered as belonging to the same ethno-variety. In this study, an ethno-variety was therefore understood to refer to infra-specific diversity as understood and managed by farmers. A total of 118 trees representing 28 ethno-varieties and separated by a minimum distance of 100 m were randomly sampled for analysis. Sample sizes were uneven for the different ethno-varieties because of the difficulty in identifying an equal number of trees of each ethno-variety. Only ethno-varieties with at least three individuals were considered for inclusion in this analysis. Two or three young leaves of healthy appearance were collected from each tree, cleaned with ethanol and placed in zip-lock bags containing silica gel in order to desiccate them. The specimen samples were stored at ambient temperature until further use.

Microsatellite marker analysis

Approximately 50 mg of silica dried leaves was placed in a 20-ml tube and flash frozen in liquid nitrogen. The super-frozen leaf samples were ground to fine powder using an automatic grinder (Retsch® MM301, Haan, Germany) at a frequency of 30 Hz for 1–2 minutes. Total genomic DNA was extracted using the Doležel–Matab method (Ky et al. 2000). Up to 1.8 ml of Doležel buffer (15 mM Tris-HCl pH = 9, 2 mM Na_2EDTA, 80 mM KCl, 20 mM NaCl, 0.2% β-mercaptoethanol and 0.5% × 100-Triton™) was autoclaved and then added to the powder samples before centrifugation for 20 minutes at 4000*g*. Following centrifugation, the supernatant was discarded and the pellets re-suspended in 1.3 ml of lysis buffer (0.1 M Tris-HCl pH = 8, 0.02 M Na_2EDTA pH = 8, 1.25 M NaCl, 4% Matab and 0.2% β-mercaptoethanol). DNA quantitation was performed on a Shimdazu BioSpec-Mini spectrophotometer (Shimadzu Corporation, Tokyo, Japan) at a wavelength of 260 and 280 nm. All DNA samples were diluted with de-ionized water to a working concentration of 50 ng/μl. The samples were stored at − 20°C until further use.

Ten SSR markers (*mCIRVp167*, *mCIRVp168*, *mCIRVp134*, *mCIRVp113*, *mCIRVp175*, *mCIRVp191*, *mCIRVp14*, *mCIRVp28*, *mCIRVp159*, *mCIRVp181*) (Allal et al. 2008) were used for this study. Polymerase chain reaction (PCR) amplification reactions were carried out in a total volume of 10 μl containing: 3.64 μl de-ionized H_2O, 5-μl 1 × Qiagen Kit Multiplex buffer, 0.08 μM forward and 0.1 μM reverse primers, 0.1 μM Oligo M13 (BIOTROP) dye and 0.5 ng/μl of DNA. PCR amplifications were performed using an Eppendorf® Mastercycler under the following reaction conditions: step 1 (denaturation at 95°C for 15 minutes, hybridization at 67°C for 90 seconds and elongation at 94°C for 90 seconds); step 2 (denaturation at 94°C for 30 seconds, hybridization at 65°C for 90 seconds, and elongation at 72° for 60 seconds, running 8 cycles with a drop of 2 degrees per cycle); and step 3 (denaturation at 94°C for 30 seconds, hybridization at 51°C for 90 seconds and elongation at 72°C for 60 seconds, running 24 cycles). An additional elongation step was performed at 60°C for 30 minutes.

Electrophoresis of amplification products

After amplification, a 2 μl aliquot of the amplified SSR samples was combined with 6 μl of loading mix (denaturing bromophenol blue solution + desmiling mix containing markers of the following sizes: 363 – 305 – 247 – 198 – 174 – 140 – 111 – 71 bp). The SSR sample mixes were then denatured on an Eppendorf® thermocycler for three minutes at 74°C before loading on a 6.5% polyacrylamide gel in a Li-Cor® 4300 sequencer. Electrophoresis was run at 1500 V in 1 × TBE buffer. Genotyping was performed using $SAGA^{GT}$ version 3.1 software (LI-COR®).

Data Analysis

The mean number of alleles (A), mean number of alleles per polymorphic loci (A_p), expected heterozygosity (H_e), observed heterozygosity (H_o) and fixation indices were computed using the Genetic Data Analysis (GDA) software (Lewis & Zaykin 2001). In GDA, Wright's conventional F-statistics, i.e. F_{is} (the inbreeding coefficient of an individual relative to the sub-population), F_{it} (the inbreeding coefficient of an individual relative to the total population) and F_{st} (the inbreeding among sub-populations relative to the total population) are computed as f, F and θ_P, respectively. The GDA notation is henceforth used to describe these statistics. Polymorphic information content (PIC), which indicates the percentage contribution of each locus to the polymorphism in the analysis was calculated using the Microsatellite toolkit (Botsten et al. 1980) in Excel 2007 (Microsoft® Corporation). Deviations from Hardy–Weinberg equilibrium were tested using the Markov chain method in Genepop version 4.1.0 (Rousset 2008). Markov chain parameters were set at 10,000 dememorization steps for 20 batches at 5000 iterations per batch. Gene flow was estimated using the private alleles method of Barton and Slatkin (1986):

$$N_m = \frac{1 - F_{st}}{4F_{st}}$$

Relationships between individual trees were assessed based on a dissimilarity matrix constructed using the neighbour – joining method as implemented in DARwin 5.0 (Perrier & Jacquemoud-Collet 2006). Missing data were handled by deletion of all units with missing data. Dissimilarity was calculated based on simple matching coefficients. Cluster analysis was performed using the un-weighted neighbour-joining method (Nei 1978) in DARwin 5.0.158. To test the representativeness of the resulting dendrogram with the genetic structure in the data matrix, an analysis of fit using the fit criterion option in DARwin 1.0.158 was performed. The resulting dendrogram was then visualized in TreeView version 1.6.6 (Page 1996).

To test for genetic variability, an Analysis of molecular variance (AMOVA) was performed using Arlequin version 3.5.1.3 (Excoffier & Lischer 2010). Population genetic structure was assessed using a Bayesian Monte Carlo Markov Chains (MCMC) method implemented in the Geneland package version 6.0 under the R software as described by Guillot et al. (2005). The Geneland software utilizes geographic data in the analysis of genetic structuring in order to optimize the delineation of sub-populations. In our Geneland analysis, 10 independent MCMC runs were initially performed with the following settings: 100,000 iterations with 100 thinning intervals and a burn-in of 250, using the correlated allele frequencies and null allele model options. During the initial analysis, the number of populations (K) was allowed to vary between 1 and 28 (the number of ethno-varieties identified by farmers). A map of posterior probabilities was obtained by PostProcessChain function. A final run was then performed based on the highest average posterior probability and the topography of population membership was computed for each pixel of the spatial domain (250 × 250 pixels). Since it is known that *V. paradoxa* is an out-crossing tree species with high gene flow (Lovett & Haq 2000), the spatial coordinates utilized on the Geneland analysis were treated as uncertain so as to allow the programme to assign samples with the same coordinates to different populations (Guillot et al. 2005). In effect, this reduces the over-arching influence of geographical coordinates on the resulting population structure.

Results

Microsatellite diversity

The mean percentage of polymorphic loci was 100% (0.99 criterions). All loci exhibited high PIC values (range, 0.510–0.817) (Table 1). A total of 106 alleles were detected over the 10 SSR loci. The number of alleles (A) per locus ranged from 6 (*mCIRVp28*) to 13 (*mCIRVp14*) with an average of 9.5 alleles. Observed (H_o) and expected heterozygosity (H_e) were consistently high (≥ 0.400) in all loci, except *mCIRVp134* ($H_o = 0.366$). H_o values ranged from 0.366 at locus *mCIRVp134* to 0.934 at locus *mCIRVp168* while H_e values ranged from 0.580 at locus *mCIRVp134* to 0.840 at locus *mCIRVp181* (Table 1). The mean H_o and H_e values for all loci were 0.633 and 0.727, respectively. Six loci departed significantly ($p < 0.05$) from Hardy–Weinberg equilibrium when tested for heterozygous deficiency: *mCIRVp113*, *mCIRVp134*, *mCIRVp167*, *mCIRVp175*, *mCIRVp191* and *mCIRVp14*. Heterozygote excess was not observed at any locus.

The values of f, F and θ_P ranged from -0.203 (*mCIRVp168*) to 0.383 (*mCIRVp191*), -0.189 (*mCIRVp168*) to 0.387 (*mCIRVp191*) and 0.001 (*mCIRVp113*) to 0.074 (*mCIRVp167*), respectively (Table 1). Heterozygote deficiency was reflected in the mean inbreeding coefficient ($f = 0.130$). Global f, F and θ_P values were 0.130, 0.137 and 0.025, respectively. Bootstrap analysis at 95% confidence interval with 20,000 replicates showed that F and θ_P values were significantly ($p < 0.05$) greater than zero while f was not, suggesting a panmictic mating system among the sampled trees.

Genetic differentiation

A cluster analysis to assess the relationships among individual trees generated a phenogram (Figure 2) with an edge length sum of 17.07, mean error of 0.0008, mean absolute error of 0.0586, maximum absolute error of 0.26, mean square error of 0.0055 and co-phenetic r of 0.8496. A co-phenetic value greater than 0.8 indicates the phenogram truly represents the genetic structure in the data matrix. The cluster analysis did not reveal any clearly separated clusters implying that the ethno-varieties as perceived by farmers is not congruent with the genetic diversity of shea trees. Although the phenogram (Figure 2)

Table 1. Genetic variability of 10 SSR loci assessed in this study.

Locus	n	A	H_e	H_o	f	F	θ_P	PIC
mCIRVp167	111	8	0.676	0.460	0.321	0.339	0.074	0.628
mCIRVp168	106	10	0.777	0.934	−0.203	−0.189	0.033	0.742
mCIRVp134	93	7	0.580	0.366	0.371	0.373	0.010	0.510
mCIRVp113	105	8	0.720	0.619	0.141	0.141	0.001	0.670
mCIRVp175	98	10	0.656	0.561	0.145	0.152	0.022	0.597
mCIRVp191	96	12	0.826	0.510	0.383	0.387	0.016	0.798
mCIRVp14	113	13	0.832	0.779	0.064	0.066	0.007	0.808
mCIRVp28	113	6	0.700	0.611	0.128	0.131	0.010	0.646
mCIRVp159	112	9	0.664	0.670	−0.009	−0.004	0.013	0.629
mCIRVp181	113	12	0.840	0.823	0.020	0.039	0.056	0.817
All	106	9.5	0.727	0.633	0.130	0.137	0.025	

Note: $n =$ sample size per locus; $A =$ number of alleles per locus; $H_e =$ Nei's (1978) un-biased estimate of the expected proportion of heterozygous loci per individual; $H_o =$ observed heterozygosity; $f =$ inbreeding coefficient of an individual relative to the sub-population; $F =$ inbreeding coefficient of an individual relative to the total population; $\theta_P =$ inbreeding among sub-populations relative to the total population.

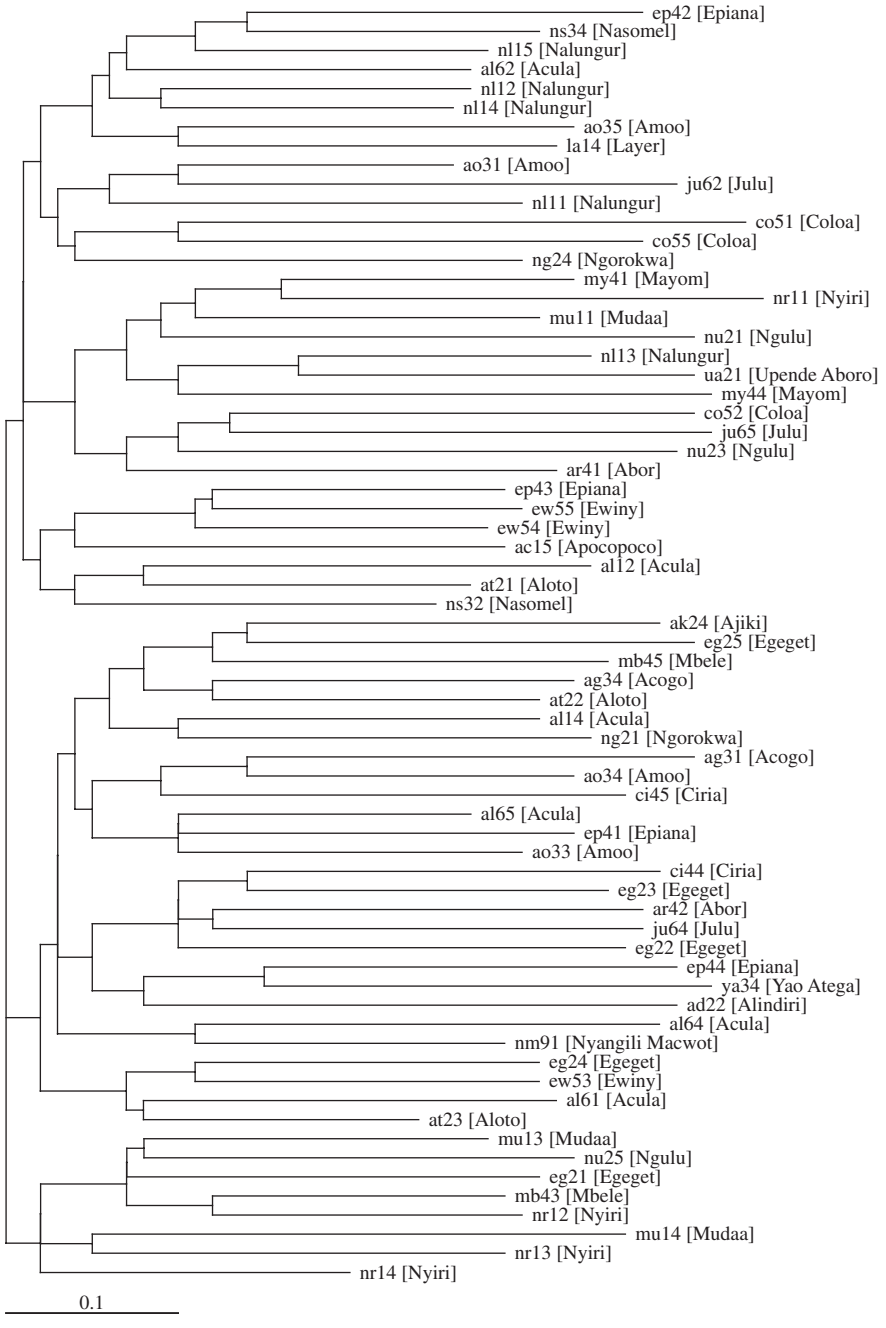

Figure 2. Neighbour-joining phenogram (based on a dissimilarity matrix and simple matching coefficients) showing relationships of 67 shea tree genotypes classified by local communities as distinct ethno-varieties. Names in parenthesis refer to local ethno-variety names (Table 1) while the codes before the ethno-variety names refer to individual tree identification codes.

shows that the tree genotypes were separable into three broad clusters, there was mixed ethno-variety membership in all the clusters indicating little or no genetic differentiation between trees.

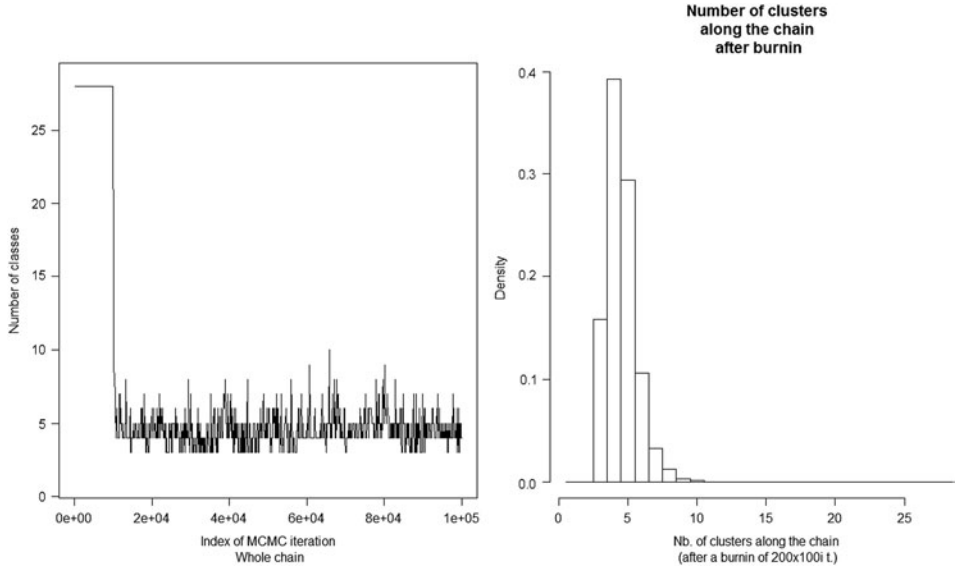

Figure 3. Number of populations (clusters) represented by the highest average density of samples as determined by the initial 10 runs in Geneland.

The initial run of the Bayesian clustering algorithm of Geneland inferred four populations (Figure 3). However, one of the inferred populations (Figure 4) had no samples assigned to it, representing a case of a 'ghost' population, which has been described as an artefact of rarely visited states in the Markov chain, and hence should be ignored (Guillot et al. 2005). Three sub-populations ($K = 3$) were therefore retained. The final run (based on $K = 3$) assigned the samples to each sub-population that were labelled as 'Teso' (cluster 1 in Figure 4), 'West Nile' (cluster 3 in Figure 4) and 'Northern' (cluster 4 in Figure 4).

The proportion of genetic variation attributed to differences in the populations ($\theta_P = 0.025$, Table 1) indicates that more than 97% of the genetic variation occurs within rather than between the populations. Results from AMOVA analysis (Table 2) showed that within the aforementioned (97%) variation, 11.2% is found among individual trees within the populations while 86.3% is found within individual trees. In addition, evidence for low differentiation is further indicated by the high level of gene flow ($N_m = 6.56$). Computation of Nei's (1978) genetic distances showed that the West Nile population was separated from the other two populations (Northern and West Nile).

Discussion

Microsatellite markers have demonstrated their superiority in screening for genetic differentiation and diversity in several tree species (Fares et al. 2009). The present investigation focused on the utilization of microsatellite markers on *V. paradoxa* ethno-varieties and shows that all microsatellite loci were highly polymorphic (51–82%) and within the range (37–85%) found by Allal et al. (2008), indicating a fairly wide and diverse genetic base. The levels of genetic diversity revealed by the present study are concordant to those found for shea tree populations in Mali (Kelly et al. 2004) and Ghana (Lovett & Haq 2000). Such high levels of genetic diversity are common in out-crossing

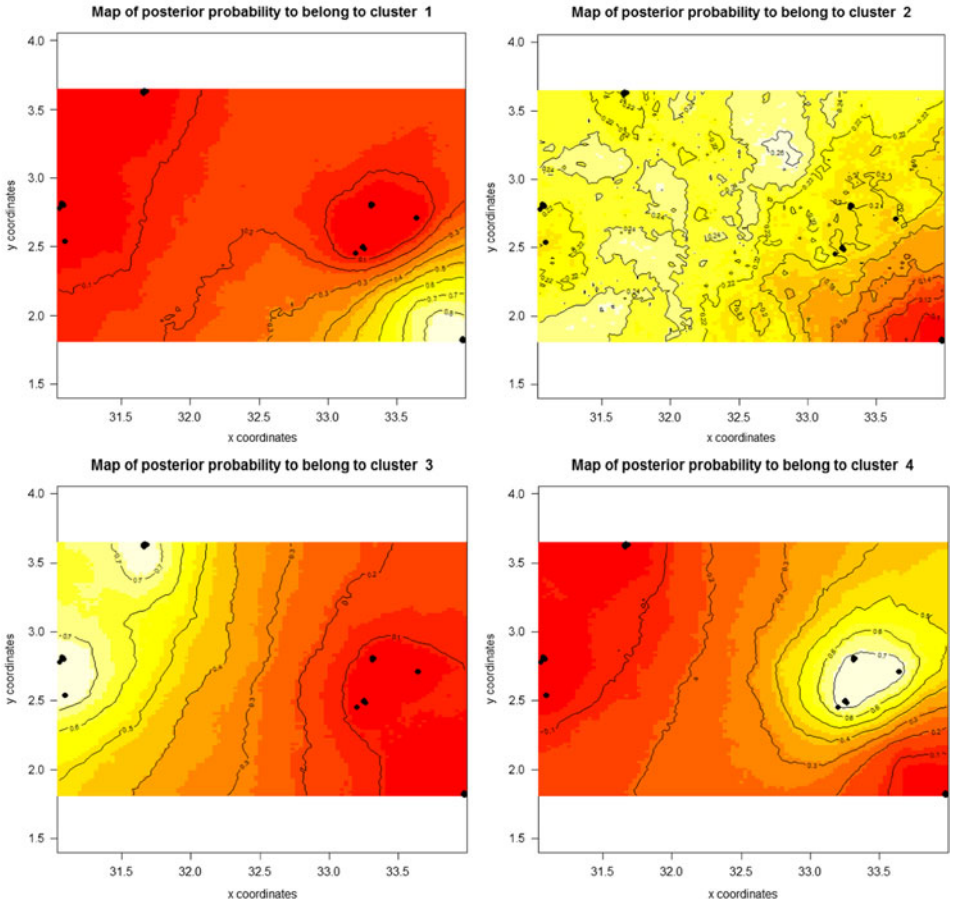

Figure 4. Maps showing posterior probabilities of cluster membership and location of genetic discontinuities for each of the inferred clusters (cluster 1 = Teso; cluster 3 = West Nile; cluster 4 = Northern) as determined by Geneland. Cluster 2 does not show genetic discontinuities among the samples and has been ignored from further analysis (see text for further explanation). Lighter shading indicates higher probability of cluster membership.

species. High genetic diversity has, therefore, been documented in economically important, out-crossing fruit tree varieties, such as *Olea europaea* L. (Rekik et al. 2008), *Pistacia vera* L. (Fares et al. 2009) and *Mangifera indica* L. (Kashkush et al. 2001).

The occurrence of individual trees of the same ethno-variety in different clusters (Figure 2) shows the existence of large gene pool. Generally, the grouping pattern clearly

Table 2. AMOVA for 118 individuals from 28 *V. paradoxa* subspecies *nilotica* ethno-varieties based on 10 SSR markers.

Source of variation	Sum of squares	Variance components	Percentage variation
Among populations	20.667	0.09060	2.46921
Among individuals within populations	409.586	0.41268	11.24719
Within individuals	339.000	3.16587	86.28359
Total	769.253	3.66914	

indicates that irrespective of differences in fruit colour, shape and sizes that exist among the ethno-varieties used, and which form the basis of folk classification, they were still interspersed with each other in different clusters (Figure 2). This may be attributed to the use of a single rather than a combination of locally observed traits to describe an ethno-variety (Gwali et al. 2011), which creates a folk classification system that is not congruent with the genetic composition of the individual shea trees. For example, the trees with a sweet fruit pulp are described as an ethno-variety (e.g. *Ewiny* and *Limi*), yet this does not exclude the possibility that the same trees may also produce fruits which are oval or round in shape, which are also classified ethno-varieties (e.g. *Acula* or *Nalungur*). In addition, traits used in folk classification may be influenced by environmental factors thereby contributing to differences in local perceptions of variation across different localities. A few private alleles were found to have some consensus for some ethno-varieties. It is worth looking further into these private alleles using a larger sample size to validate any genetic differentiation among the ethno-varieties. In addition, it will be important to establish these private alleles which may be linked to important functions that can be targeted in the tree improvement.

The lack of clear grouping in cluster analysis indicates that shea tree ethno-varieties as perceived by farmers are basically similar genotypes described with different names in different locations, since farmers name their varieties according to their own experiences. This implies that the numerous ethno-variety names used in shea tree folk classification are only phenotype expressions, corroborating recent findings from fatty acid analysis (Gwali, Nakabonge, et al. 2012), and therefore do not reflect significant genetic distinctiveness. Lack of concordance between folk classification and genetic variation has been observed in *Adansonia digitata* L. in West Africa (Assogbadjo et al. 2009). In addition, several studies have shown that phenotypic traits may not necessarily be correlated to genetic variation (Perrie & Brownsey 2005). Maranz and Wiesman (2003) reported a relationship between environmental variables (rainfall and temperature) and variation in shea tree phenotypic traits, such as fruit size and shape, pulp sweetness and oil content.

Groups of individuals in a geographical area may be linked by gene flow and therefore genetic markers are very useful in analysing their population structure (Coulon et al. 2008). The present study revealed that the population structure of *V. paradoxa* is not strongly differentiated throughout its range in Uganda. The AMOVA results in this study (Table 2) show that there are low levels of differentiation among populations ($\theta_P = 0.025$). In addition, the estimated gene flow ($N_m = 6.56$) was high reflecting a substantial amount of gene flow among the ethno-varieties. The low inbreeding coefficient ($f = 0.088$) which was not significantly different from zero therefore suggests that this is a largely panmictic (randomly mating) species, i.e. selfing is either very weak or absent. Since *V. paradoxa* can be described as a semi-domesticated tree species that is continually managed by humans (Sanou et al. 2006), this implies that such human management may be partly responsible for the high gene flow and the lack of significant genetic differences in phenotypes (Houehanou et al. 2013). Similar levels of genetic variation (0.026) have been observed in *V. paradoxa* in West Africa (Kelly et al. 2004). Low levels of population differentiation have been reported for other savanna parkland tree species such as *A. digitata* (Assogbadjo et al. 2006, 2009), *Milicia excelsa* (Ouinsavi et al. 2009), *Tamarindus indica* (Nyadoi 2005). Assogbadjo et al. (2006) found that some genotypes of *A. digitata* L. populations growing in West Africa belonged to more than one gene pool, suggesting some form of human introductions to the areas involved.

Such a weak population differentiation could be explained by high gene flow and the dispersal mechanism of this species. Studies on the reproductive biology of *V. paradoxa*

indicate that the species is an out-crossing tree (Hall et al. 1996; Lovett and Haq 2000), pollinated mainly by insects such as bees and butterflies and dispersed barochorously and secondarily zoochorously by bats and monkeys, and also by birds (Hall et al. 1996). The fruits of *V. paradoxa* consist of a round or oval berry consisting of a sweet pulp, which highly enhances the barochorous seed dispersal mechanism. In addition to dispersal by animals, farmers (especially women and children) collect shea fruits for nut extraction and subsequent production of shea butter/oil. In many instances, the fruit pulp is eaten fresh (especially during famine periods) and when the nuts are discarded, this contributes to seed dispersal. These dispersal mechanisms greatly facilitate high gene flow resulting into high genetic diversity in the populations.

Dispersal probabilities which are utilized by spatial models such as those employed by Geneland software are very useful in complementing mainstream genetic analyses to reveal underlying population structure among individuals (Coulon et al. 2008). Although AMOVA results (Table 2) showed that 86% of genetic variation was found within individual trees, broad-scale patterns of genetic structure utilizing geographical locations as assessed in Geneland 6.0 showed three broad populations across the region (Figures 3 and 4). For the purposes of this study, these populations have been named according to the farming systems in which they occur. Based on Nei's (1978) genetic distances, the West Nile population is, however, more distantly placed compared to Teso and Northern populations. A probable hypothesis for this differentiation might be the physical isolation provided by the Rift Valley and the water bodies (Lake Albert and River Nile), which separate the West Nile population from the Northern and Teso populations. Previous studies on various plant and animal species have demonstrated the significance of the Rift Valley in determining population structure (e.g. Kebede et al. 2007). The Rift Valley and associated water bodies (Lake Albert and River Nile) are potential geographic barriers to the dispersal of *V. paradoxa* and may result in genetic differences between populations on the east and west sides of the valley. Such population differentiation due to physical isolation by water bodies in Uganda has been reported in other species. For example, Mulumba et al. (2012) reported that *Acacia senegal* L. populations in Uganda are genetically separated by Lake Kyoga, implying that such water bodies may provide some restrictions to pollen flow and seed dispersal and ultimately to species colonization. Nevertheless, the interpretation of the genetic differentiation found between the West Nile and Northern/Teso populations should take into account the limited number of loci examined, sample sizes and number of populations involved in this study. Much more sampling employing more loci are needed to establish a more reliable picture of the genetic relationships of these groups.

Conclusion

Shea trees in Uganda possess high levels of genetic diversity, which is useful in the long-term survival of the species. Most genetic variation was found within individual tree genotypes rather than between ethno-varieties. It was, therefore, not possible to distinguish the present shea tree ethno-varieties as perceived by farmers using SSR fingerprinting. This is most likely due to the use of single, non-genetic traits in ethno-variety classification. The classification of shea trees into ethno-varieties is therefore not congruent with the genetic variation present in shea trees in Uganda. We conclude that these ethno-varieties are subgroups of a single large randomly mating population. The observable variation is mainly phenotypic, continuous and plastic as a result of micro-environmental factors. In addition, spatial genetic structuring indicates the presence of three closely related populations, indicating some form of active gene exchange.

Acknowledgements

We are very grateful to the farmers in the shea parklands of Uganda for their support and cooperation during fieldwork. We are also greatly indebted to the field assistants (Richard Amuge, Henry Asindua, Joel Buyinza, Jasper Okello, Joseph Okello, Slim Abok Openythoo, Paul Okiror and Samuel Oyuko) for all their efforts during fieldwork and language translation. We are grateful for the technical assistance of the staff of the genetic improvement unit of CIRAD and the genotyping platform of the Montpellier Languedoc-Roussillon Genopole located in the UMR PIA of CIRAD.

Funding

Funding for this work was obtained from Makerere University School of Graduate Studies – Carnegie Corporation Competitive Research grant. Additional funding was provided by the European Union INCO – CT project no. 032037 (Innovative Tools and Techniques for Sustainable Use of the Shea Tree in Sudano – Sahelian zone – INNOVKAR) and the World Agroforestry Centre, Nairobi through the SII/World Agroforestry Project: 'Advancing Agroforestry Research and Development through Training and Education'.

References

Allal F, Sanou H, Millet L, Vaillant A, Camus-Kulandaivelu L, Logossa ZA, Lefevre F, Bouvet JM. 2011. Past climate changes explain the phylogeography of *Vitellaria paradoxa* over Africa. Heredity. 107(2):174–186.

Allal F, Vaillant A, Sanou H, Kelly B, Bouvet JM. 2008. Isolation and characterization of new microsatellite markers in shea tree (*Vitellaria paradoxa* C.F. Gaertn). Mol Ecol Resour. 8(4):822–824.

Assogbadjo AE, Kyndt T, Chadare FJ, Sinsin B, Gheysen G, Eyog-Matig O, Van Damme P. 2009. Genetic fingerprinting using AFLP cannot distinguish traditionally classified baobab morphotypes. Agroforest Syst. 75:157–165.

Assogbadjo AE, Kyndt T, Sinsin B, Gheysen G, van damme P. 2006. Patterns of genetic and morphometric diversity in Baobab (*Adansonia digitata*) populations across different climatic zones of Benin (West Africa). Ann Bot. 97:819–830.

Barton NH, Slatkin M. 1986. A Quasi-equilibrium theory of the distribution of rare alleles in a subdivided population. Heredity. 56:409–415.

Botsten D, White RL, Skolnick M, Davis RW. 1980. Construction of a genetic linkage map in man using Restriction Fragment Length Polymorphisms. Am J Hum Genet. 32(3):314–331.

Bouvet JM, Fontaine C, Sanou H, Cardi C. 2004. An analysis of the pattern of genetic variation in *Vitellaria paradoxa* using RAPD markers. Agroforest Syst. 60(1):61–69.

Coulon A, Fitzpatrick JW, Bowman R, Stith BM, Makarewich CA, Stenzler LM, Lovette IJ. 2008. Congruent population structure inferred from dispersal behaviour and intensive genetic surveys of the threatened Florida scrub-jay (*Aphelocoma coerulescens*). Mol Ecol. 17:1685–1701.

De Smedt S, Alaerts K, Kouyaté A, Van Damme P, Potters G, Samson R. 2011. Phenotypic variation of baobab (*Adansonia digitata* L.) fruit traits in Mali. Agroforest Syst. 82(1):87–97.

Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour. 10(3):564–567.

Fares K, Guasmi F, Touil L, Triki T, Ferchichi A. 2009. Genetic diversity of Pistachio tree using Inter-Simple Sequence Repeat markers (ISSR) supported by morphological and chemical markers. Biotechnology. 8:24–34.

Guillot G, Estoup A, Mortier F, Cosson JF. 2005. A spatial statistical model for landscape genetics. Genetics. 170:1261–1280.

Gwali S, Okullo JBL, Eilu G, Nakabonge G, Nyeko P, Vuzi P. 2011. Folk classification of shea butter tree (*Vitellaria paradoxa* subsp. nilotica) ethno-varieties in Uganda. Ethnobot Res Appl. 9:243–256.

Gwali S, Okullo JBL, Eilu G, Nakabonge G, Nyeko P, Vuzi P. 2012. Traditional management and conservation of shea trees (*Vitellaria paradoxa* subspecies *nilotica*) in Uganda. Environ Dev Sustain. 14(3):347–363.

Gwali S, Nakabonge G, Okullo JBL, Eilu G, Forestier-Chiron N, Piombo G, Davrieux F. 2012. Fat content and fatty acid profiles of shea tree (*Vitellaria paradoxa* subspecies *nilotica*) ethno-varieties in Uganda. For Trees Live. 21(4):267–278.

Hall JB, Aebischer DP, Tomlison HF, Osei-Amaning E, Hindle JR. 1996. *Vitellaria paradoxa*: a monograph. Bangor: School of Agricultural and Forest Sciences, University of Wales.
Houehanou TD, Kindomihou V, Stevart T, Tente B, Houinato M, Sinsin B. 2013. Variation of Loranthaceae impact on *Vitellaria paradoxa* C.F. Gaertn. fruit yield in contrasting habitats and implications for its conservation. Fruits. 68(2):109–120.
Kashkush K, Jinggui F, Tomer E, Hillel J, Lavi U. 2001. Cultivar identification and genetic map of mango (*Mangifera indica*). Euphytica. 122(1):129–136.
Kebede M, Ehrich D, Taberlet P, Nemomissa S, Brochmann C. 2007. Phylogeography and conservation genetics of a giant lobelia (*Lobelia giberroa*) in Ethiopian and tropical East African mountains. Mol Ecol. 16:1233–1243.
Kelly BA, Hardy OJ, Bouvet JM. 2004. Temporal and spatial genetic structure in *Vitellaria paradoxa* (shea tree) in an agroforestry system in southern Mali. Mol Ecol. 13(5):1231–1240.
Ky CL, Barre P, Lorieux M, Trouslot P, Akaffou S, Louarn J, Charrier A, Hamon S, Noirot M. 2000. Interspecific genetic linkage map, segregation distortion and genetic conversion in coffee (*Coffea* sp.). Theor Appl Genet. 101:669–676.
Langdale-Brown I, Osmaston H, Wilson J. 1964. The vegetation of Uganda and its bearing on land use. Entebbe: Government Printer.
Lewis PO, Zaykin D. 2001. Genetic Data Analysis: Computer program for the analysis of allelic data. Version 1.0 (d16c). http://www.eeb.uconn.edu/people/plewis/software.php
Lovett PN, Haq N. 2000. Diversity of the Sheanut tree (*Vitellaria paradoxa* C.F. Gaertn.) in Ghana. Genet Resour Crop Ev. 47(3):293–304.
Maranz S, Wiesman Z. 2003. Evidence for indigenous selection and distribution of the shea tree, *Vitellaria paradoxa*, and its potential significance to prevailing parkland savanna tree patterns in sub-Saharan Africa north of the equator. J Biogeogr. 30(10):1505–1516.
Mulumba JW, Nyakana S, Jamnadass R. 2012. Genetic diversity and structure of *Acacia senegal* (L.) Willd. in Uganda. Int J Bot. 8(3):107–116.
Nei M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics. 89:583–590.
NEMA. 1996. State of the Environment Report for Uganda 1996. Kampala: National Environment Management Authority.
Nyadoi P. 2005. Population structure and socio-economic importance of *Tamarindus indica* in Tharaka district, eastern Kenya [MSc thesis]. Kampala, Uganda: Makerere University. 110 pp.
Ouinsavi C, Sokpon N, Khasa DP. 2009. Genetic diversity and population structure of a threatened African tree species, *Milicia excelsa*, using nuclear microsatellites DNA markers. Int J For Res. doi:10.1155/2009/210179.
Page RDM. 1996. TREEVIEW: An application to display phylogenetic trees on personal computers. Comput Appl Biosci. 12:357–358.
Perrie LR, Brownsey PJ. 2005. Genetic variation is not concordant with morphological variation in the fern *Asplenium hookerianum* sensu lato (Aspleniaceae). Am J Bot. 92(9):1559–1564.
Perrier X, Jacquemoud-Collet JP. 2006. DARwin software 5.0 edition. http://darwin.cirad.fr/darwin
Rekik I, Salimonti A, Kamoun NG, Muzzalupo I, Lepais O, Gerber S, Perri E, Rebai A. 2008. Characterization and identification of Tunisian olive tree varieties by microsatellite markers. HortScience. 43(5):1371–1376.
Rousset F. 2008. GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. Mol Ecol Resour. 8(1):103–106.
Sanou H, Lamien N. 2011. *Vitellaria paradoxa* (shea butter tree): conservation and sustainable use of genetic resources of priority food tree species in sub-Saharan Africa. Rome: Bioversity International.
Sanou H, Picard N, Lovett PN, Dembélé M, Korbo A, Diarisso D, Bouvet JM. 2006. Phenotypic variation of agromorphological traits of the shea tree, *Vitellaria paradoxa* C.F. Gaertn., in Mali. Genet Resour Crop Ev. 53(1):145–161.
Sturges P. 2008. Information and communication in bandit country: an exploratory study of civil conflict in northern Uganda (1986–2007). Info Dev. 24(3):204–212.