Genetic Structure of Prunus africana Rosaceae (Hook.f.) Kalkm. in East Africa, as Inferred from Nuclear and Chloroplast DNA

Maria Mbatudde1* • Sylvester Nyakaana1 • Stefanie Plöß2 • Helmut Dalitz3

1 College of Science, School of Biological Sciences, Makerere University, Uganda
2 Institute of Zoology 220 and 3 Institute of Botany 210, University of Hohenheim, D-70593 Stuttgart, Germany

Corresponding author: * mmbatudde@yahoo.com

INTRODUCTION

Genetic diversity represents the heritable variation within and between populations of organisms, and in the context of this paper, among populations of Prunus africana species. This pool of genetic variation within an inter-mating population is the basis for selection as well for plant improvement. Thus, conservation of this plant genetic diversity is essential for present and future human well-being. Diversity patterns inferred by such studies can have important implications for forest conservation and management (Newton et al. 1999).

Prunus africana (Hook.f.) Kalkml., the African cherry, is an evergreen climax vegetation tree species typically reaching 25–30 m in height and occurs primarily in montane and submontane forests. In this study, the genetic structure of P. africana was analyzed using a coding chloroplast DNA region; Megakaryocyte-Associated Tyrosine Kinase Gene (MATK) and nuclear ribosomal internal transcribed spacer (nrITS) with 63 individuals in East Africa. This study detected low levels of genetic diversity as revealed by nucleotide diversity in nrDNA (π = 0.00529) and cpDNA (0.00448). As revealed by the results of AMOVA analysis, genetic differentiation for cpDNA (FST = 0.0275) was obviously lower than for nrDNA data (FST = 0.237) in P. africana. Gene flow among populations based on nrDNA data (Nm = 2.641) was significantly higher than that based on cpDNA (Nm = 0.82). Mantel test revealed a significant correlation between genetic and geographic distances for cpDNA (r2 = 3.0 × 10–5) and nrDNA (r2 = 7.0 × 10–5). Demographic history analyses based on pair-wise nucleotide sequence mismatch distributions revealed that only the Kakamega population was in mutation-drift disequilibrium. Tajima’s D neutrality test, however, revealed significant signatures of recent population expansion in only the Kakamega population (D = 1.85646; P < 0.05). This study therefore proposes immediate ex situ and in situ conservation of P. africana populations in Mabira and Elgon forests coupled with in situ conservation of Budongo, Kakamega and Monduli forests.

Keywords: African cherry, demographic history, haplotype, population diversity

Abbreviations: AMOVA, analyses of molecular variance; cpDNA, chloroplast DNA; FST, genetic differentiation among populations; h, haplotype diversity; IT, total genetic diversity; ITS, Internal transcribed spacer; MATK, megakaryocyte-associated tyrosine kinase; Nm, gene flow among populations; nrDNA, nuclear DNA; PCA, Principal Components Analysis; π, nucleotide diversity

AUTHOR’S ADDRESS

Maria Mbatudde
College of Science, School of Biological Sciences
Makerere University
P.O. Box 7072
Kampala
Uganda

Sylvester Nyakaana
College of Science, School of Biological Sciences
Makerere University
P.O. Box 7072
Kampala
Uganda

Stefanie Plöß
Institute of Zoology 220
University of Hohenheim
D-70593 Stuttgart
Germany

Helmut Dalitz
Institute of Botany 210
University of Hohenheim
D-70593 Stuttgart
Germany

Received: 26 June, 2012. Accepted: 24 July, 2012.

Original Research Paper


© 2013 Global Science Books

1545-7703 (Print) 1545-7711 (Online)
MATERIALS AND METHODS

Population sampling

From July to September 2009, wild populations of *P. africana* were sampled within five forests in East Africa; Mabira, Budongo, Elgon, Kakamega and Moduli forests (Fig. 1), using random sampling. Random sampling is a sampling technique where a group of subjects (a sample) is selected for study from a larger group (a population). Each individual is chosen entirely by chance and each member of the population has a known, but possibly non-equal, chance of being included in the sample (Easton and McColl 1997). In each of the five populations; 16 individuals were sampled from the five forests (Fig. 1), using random sampling. Random sampling is a sampling technique where a group of subjects (a sample) is selected for study from a larger group (a population). Each individual is chosen entirely by chance and each member of the population has a known, but possibly non-equal, chance of being included in the sample (Easton and McColl 1997). In each of the five populations; 16 individuals were sampled from Mabira, 5 individuals from Budongo, 7 individuals from Elgon, 26 individuals from Kakamega and 8 individuals from Monduli forests. Leaves were collected from individual trees at least 15-20 m apart to increase the possibility of detecting the potential variations within each population. Fresh leaves were collected, pressed and dried between blotters. A 1 × 1 cm² leaf area was crushed.

Genomic DNA was extracted using InnuPREP Plant DNA kit (Analytik Jena AG, Jena, Germany). DNA quality and quantity were assessed on a spectrophotometer by absorbance (BioPhotometer, Eppendorf, Hamburg, Germany). Working stocks of DNA were then prepared based on estimates, and stored in 0.1 × TE buffer. Total genomic DNA was extracted using InnuPREP Plant DNA kit (Analytik Jena AG, Jena, Germany). DNA concentration was assessed using a spectrophotometer by absorbance (BioPhotometer, Eppendorf, Hamburg, Germany). Working stocks of DNA were then prepared based on estimates, and stored in 0.1 × TE buffer. After preliminary screening, polymorphism was observed in the coding region and the internal transcribed spacer (ITS). These were amplified via polymerase chain reactions (PCRs) using primers; MatK-Kim-f reverse (5’-cgtacagactttggttagta-3’) and ITS-4R reverse (5’-CACCAACCCACCCACRC-3’).

DNA amplifications were carried out in reaction volumes of 25.0 μl containing 2.0 μl DNA, 0.4 μM of each 2 mmol/L dNTP (BIO-39028, Bioline Germany), 1 μL 5 × K buffer (BIO-39028, Bioline Germany), 0.5 μL Taq Polymerase (5 U/μL, MangoTag™), and 0.8 μg/ml, water (ddH₂O) up to 50 μl. PCR was carried out in an Eppendorf Mastercycler gradient that was programmed for one cycle of 1 min at 94°C followed by 35 cycles of 30 sec at 94°C, 20 sec at 56°C and 1 min at 72°C, 4 min at 95°C followed by 35 cycles of 40 sec at 95°C, 40 sec at 56°C and 1 min at 72°C and with a cycle of 1 min at 72°C. PCR products were purified using the Geniluista kit (GE Healthcare, life technologies, USA).

Sequencing was done using the same primers as used for PCR amplification by a commercial sequencing provider (GATC Biotech, Germany). The DNA sequences were edited manually based on the chromatograms and aligned by Clustal X (Thompson et al. 1997).

Data analysis

1. Chloroplast DNA (cpDNA)

Haplotypes of nuclear gene were inferred via “haplotype subtraction” (Clark 1990; Zhou et al. 2007). Haplotype diversity (h) and nucleotide diversity (n) were calculated for each population (hS, nS) and at the species level (hT, nT) using DNASP (version 4.0; Rozas et al. 2003). Gene flow (Nm) (Hudson et al. 1992) among populations was also calculated by DNASP (Librado and Rozas 2009). Analyses of molecular variance (AMOVAs) were used to calculate genetic variance components and their significance levels among populations and within populations by ARLEQUIN (version 3.1; Excoffier and Lischer 2010). To see if the obtained cpDNA and nrDNA sequences satisfied the assumption of neutrality, we calculated Tajima’s D (Tajima 1989) and Fu and Li’s D* (Fu and Li 1993) for the entire species and groups of populations, using DNASP ver.4.90.1 (Librado and Rozas 2009). Statistical significance of D and D* was estimated with coalescent simulations as implemented in this program. In general, significant negative departures of these statistics from zero indicate deviation from neutrality, but might also be taken as evidence of recent demographic expansions or population bottlenecks when markers are otherwise assumed to be independent of selection (Tajima 1989; Fu 1997). To further infer demographic processes, we explicitly tested the null hypotheses of a spatial expansion and of a pure demographic expansion in DNASP by comparing observed and expected distributions of pairwise sequence differences (mismatch distributions). For both cpDNA and nrDNA data, Mantel tests were carried out to investigate the correlation between genetic and geographic distances. A Principal Component Analysis (PCA) was carried out, using GenAlEx 6.4 (Peakall and Smouse 2006).

2. Nuclear DNA (nrDNA)

Percent identity and G: C content was calculated with Bioedit. Sequence diversity within a subpopulation was calculated as π, which is the average pairwise difference between sequences from within a taxon (Tajima 1983). The standard deviation of π over both the stochastic and sampling process is given in Nei (1975). We included both indels and ambiguous sites (except ‘N’s) in the calculation of π and its standard deviation.

RESULTS

cpDNA diversity and population structure

Chloroplast DNA (cpDNA) regions sequenced in *P. africana* (63 individuals, 5 populations) showed a consensus sequence length of 578 bp after alignment. The mean haplotype diversity and nucleotide diversity was  h = 0.24153 and  π = 4.48 × 10⁻⁴, respectively. Gene flow among populations based on combined cpDNA data was (Nm) = 0.82. Nucleotide diversity (π) among 5 populations ranged from 0 to 9.42×10⁻⁴ and haplotype diversity (h) varied between 0.06 and 0.163. Highest nucleotide and haplotype diversity was in Kakamega, followed by Mabira while there was no diversity in Budongo, Elgon and Monduli (Table 1).

Based on these polymorphisms, five cpDNA haplotypes (A-E) were identified across the material surveyed (Table 1). Neighbor joining tree of haplotypes revealed that out of the five haplotypes, haplotype D was distantly related to the other four haplotypes. The DNA sequences were edited manually based on the chromatograms and aligned by Clustal X (Thompson et al. 1997).

Fig. 1 Locations of *Prunus africana* populations sampled for this study.

Table 1 Overview of population diversity for *Prunus africana* (P. africanum), with cpDNA diversity and population structure.

©2013 Global Science Books
The haplotype diversity and nucleotide diversity were 0.8106, \( \pi T = 5.29 \times 10^{-3} \). Nucleotide diversity (\( \pi \)) among 5 populations ranged from 3.38 \( \times 10^{-3} \) to 10.27 \( \times 10^{-3} \) and haplotype diversity (\( h \)) varied between 0.643 and 0.857 (Table 1). The sequences of \( P. \) africana samples have been deposited in GenBank database under accession numbers (JN412001-JN412062).

Gene flow among populations based on nrDNA (\( Nm = 2.641 \)) was obviously higher than that based on cpDNA. Maximum Parsimony tree of haplotypes revealed that there was no distinct or outstanding group; there was intermingling of haplotypes from populations. Branches corresponding to partitions reproduced in less than 50% trees are collapsed.

AMOVA revealed that the greatest percentage of variation 97.25% was distributed within populations as compared to 23.7% among populations (Table 2). Significant Mantel test though negative was detected for cpDNA at the species-wide range (\( r = 0.30 \times 10^{-3} \)). Tajima’s \( D \) and Fu and Li’s \( D^* \) statistics for deviation from neutrality were not significant for each geographic population and the whole species (all \( P > 0.10 \)). Furthermore, the observed mismatch distributions of haplotypes from the whole species was not different significantly from mismatches expected under models of both spatial and sudden demographic expansion (\( r = 0.3742 \)). Chloroplast DNA also revealed a high level of intermingling with two individuals far away. The first three components of PCA explained 86.93, 9.19 and 1.6% of the total variation.

### Nuclear (nrDNA) diversity and population structure

The aligned sequences of the ITS were 456 bp in length. The haplotype diversity and nucleotide diversity were \( hT = 0.8106, \pi T = 5.29 \times 10^{-3} \). Nucleotide diversity (\( \pi \)) among 5 populations ranged from 3.38 \( \times 10^{-3} \) to 10.27 \( \times 10^{-3} \) and haplotype diversity (\( h \)) varied between 0.643 and 0.857 (Table 1). The sequences of \( P. \) africana samples have been deposited in GenBank database under accession numbers (JN412001-JN412062).

Gene flow among populations based on nrDNA (\( Nm = 2.641 \)) was obviously higher than that based on cpDNA. Maximum Parsimony tree of haplotypes revealed that there was no distinct or outstanding group; there was intermingling of haplotypes from populations. Branches corresponding to partitions reproduced in less than 50% trees are collapsed.

AMOVA revealed that the greatest percentage of variation 76.3% was distributed within populations as compared to 23.7% among populations (Table 2). Significant
Mantel test though negative was detected for cpDNA at the species-wide range ($r^2 = 7.0 \times 10^{-6}$). Tajima’s $D$ and Fu and Li’s $D^*$ statistics for deviation from neutrality were not significant for each geographic population and the whole species; $D = -2.068248$ (all $P > 0.10$). Furthermore, the observed mismatch distributions of haplotypes from the whole species was different significantly from mismatches expected under models of both spatial and sudden demographic expansion ($r = 0.0204$).

The first three components of PCA explained 82.3% of the variation; 62.98, 10.02 and 9.30% ($r^2$). PCA revealed that there was intermingling of individuals from different populations.

**Demographic history of the populations**

The mismatch frequency spectra for the five populations are shown in Fig. 4A-E. The observed unimodal mismatch frequency distributions of most of the populations (Mabira, Budongo, Elgon and Monduli) sampled showed excesses of mutations that appeared in only a few individuals and a deficiency of mutations shared by many individuals in the population. This is a scenario in conformity with the recent expansion model. On the other hand, the mismatch distribution of the Kakamega population (Fig. 4D) was bimodal and fitted poorly with the corresponding distribution expected under the recent expansion model.

Tajima’s $D$ neutrality test, revealed significant signatures of recent population expansion in only the Kakamega population; ($D = -1.85646; P < 0.05$). However, Fu’s $F$ test and SSD statistics were in contrast not significant (Table 3). There was no population with mismatch frequency spectra significantly deviated from what would be expected under the sudden expansion model (Table 3).

**DISCUSSION**

**Genetic diversity**

The natural populations of *P. africana* have decreased tremendously as a result of the overexploitation of the species for its medicinal extract, yet little is known about its genetic diversity. The results of using both cpDNA and nrDNA indicate that there are low levels of genetic diversity in all 5 investigated natural populations of *P. africana* as revealed by nucleotide diversity within nrDNA and cpDNA. Total levels of cpDNA haplotype diversity in *P. africana* ($h_T = 0.24153$) were lower than other seed plants for maternally inherited markers (Huang et al. 2002; Gao et al. 2007). Still total cpDNA diversity was lower than for *P. africana* in Africa ($h = 0.886$; Kadu et al. 2010), and comparable to *Fragaria* species all in family Rosaceae; range of (0.28 to 0.78; Wambui 2010). However, this value is comparable to the mean population diversity ($h_s = 0.234$; Kadu et al. 2010).

For nrDNA, species-wide levels of haplotype diversity in *P. africana* ($h_T = 0.81055$) were significantly high by comparison with other woody plant species with similar life history traits and geographical range reviewed by Hamrick et al. (1992) (mean $h_T = 0.211$, for out crossing, animal-pollinated species). However, haplotype diversity was lower than for *Ficus carica* L. $h_T = 0.983$ (Baraket et al. 2009). Generally, geographic distribution, breeding system and size of population are all associated with genetic diversity in plant species (Hamrick and Godt 1989). Species with out-crossing and mixed breeding have significantly higher diversity than species with self-crossing, geographically widespread species usually have higher genetic diversity than limited distributions, large size of population have higher than small diversity (Hamrick and Godt 1989; Ham-

**Table 3** Demographic history summary statistics of the five *Prunus africana* populations using nrDNA

<table>
<thead>
<tr>
<th>Population</th>
<th>SSD</th>
<th>$D$ (P-value)</th>
<th>$F_r$ (P-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mabira</td>
<td>0.941</td>
<td>-0.62592 (&gt; 0.10)</td>
<td>-1.493 (0.123)</td>
</tr>
<tr>
<td>Budongo</td>
<td>0.800</td>
<td>-0.97256 (&gt; 0.10)</td>
<td>-0.829 (0.244)</td>
</tr>
<tr>
<td>Elgon</td>
<td>0.857</td>
<td>-1.35841 (&gt; 0.10)</td>
<td>-0.237 (0.299)</td>
</tr>
<tr>
<td>Kakamega</td>
<td>1.160</td>
<td>-1.85646 (&lt; 0.05)*</td>
<td>-1.462 (0.115)</td>
</tr>
<tr>
<td>Monduli</td>
<td>0.250</td>
<td>-1.05482 (&gt; 0.10)</td>
<td>-0.182 (0.354)</td>
</tr>
<tr>
<td>Total sample</td>
<td>0.766</td>
<td>-2.78761 (&lt; 0.05)</td>
<td>-5.095 (&lt; 0.005)</td>
</tr>
</tbody>
</table>

*P < 0.05; SSD = sum of squared differences, $D = $ Tajima’s statistic, $F_r = $ Fu’s statistic.

**Fig. 4** Observed and expected pair-wise mismatch distributions in the five *Prunus africana* populations under the sudden population expansion model. The number of pair-wise nucleotide differences between haplotypes is represented on the abscissa whereas their frequencies are represented on the ordinate axis.

---

rick et al. 1992; Hamrick and Godt 1996; Nybom 2004). As a widespread species in East Africa, *P. africana* is an out crossing insect pollinated species with small population sizes. Although, their population size decrease dramatically, the detected responses in genetic diversity may be due to overharvesting of the species from the wild.

Widespread tree species generally have high levels of genetic diversity overall (Hamrick and Godt 1989); like *Eucalyptus* (Horvitz et al. 1987; Medlock 1992). As with most tree species the majority of genetic diversity occurs within populations (Table 2). Among the populations investigated for cpDNA, Kakamega (*h* = 0.163, *π* = 9.42 × 10⁻³) had the highest haplotype diversity and nucleotide diversity; for nrDNA, Elgon (*h* = 0.85714), had the highest haplotype diversity while Budongo (*π* = 10.27 × 10⁻³) had the highest nucleotide diversity, Mabira has the least nucleotide diversity (*π* = 3.38 × 10⁻³) yet it is under high pressure from the surrounding human populations.

Overall, *P. africana* contains low levels of variation within populations.

### Genetic structure

nrDNA data demonstrated significantly population differentiation within *P. africana* (*FST* = 0.23698), this was due to the fixation of particular haplotypes in single populations. In contrast to the significant population genetic structure obtained with cpDNA, there was a moderate level of genetic differentiation for cpDNA haplotype (*FST* = 0.275). A high pollen flow among populations was considered to be the explanation for the high population differentiation in *P. africana*. NtDNA differentiation was lower compared to *Angelica palustris* (Apiaceae) *ΦST* = 0.34 (Ditthrenner et al. 2005), *Gentiana germanica* *ΦST* = 0.42 (Fisher and Matthens 1998), *Dictamnus albus* *ΦST* = 0.62 (Hensen and Oberprieler 2005) and *Pulsatilla vulgaris* *ΦST* = 1.22 (Hensen et al. 2005).

Tajima’s D statistic together with the multimodal mismatch distribution also suggested that population expansion did not occur but instead, population reduction revealed by the negative values. Maximum parsimony tree of nrDNA haplotypes revealed that there was no distinctive group; haplotypes occurred in more than one collection site, with haplotype 19 forming the ancestral stock for nrDNA. Neighbor-Joining of cpDNA revealed that also there is no distinct group, with Haplotype D forming the ancestral stock for cpDNA. Although, the pollen flow was very strong compared to seed flow, the gene flow might play a main role in nearby populations.

The PCA resolved the five populations into mainly three differentiated groups which were not correlated with geography of the sampling sites. This was in support of the Maximum Parsimony analysis which did not reveal a distinct group. Lack of evidence for correlation between genetic and geographic distances was further supported by a Mantel test. The results indicate that distance is not responsible for the genetic structure of the species. The breeding system (1) maintains the genetic diversity of individuals and populations, and (2) promotes gene flow. The Nm value (*Nm* = 2.641) actually indicates a high rate of gene flow among populations. Apportion of genetic diversity of plant species usually reflects their breeding system (Hamrick and Godt 1996).

### Population demography

Tajima’s and Fu’s tests for neutrality of mutations used in this study have been reported to be unable to disentangle the effects of true selective departures from neutrality (such as those caused by purifying selection and genetic hitchhiking) from those due to demographic processes such as population expansion on mismatch distribution profiles. For example, both statistics can be significantly negative due to purifying selection, genetic hitchhiking or population expansion while both of them can be significantly positive under scenarios of population bottlenecks and balancing selection at linked loci (Tajima 1989a, 1999b). However, because the data used in this study are derived from the selectively neutral non-coding region of the nuclear genome (control region), it is highly unlikely that the observed significant deviations from neutrality are due to selection. The high haplotype diversity, low nucleotide diversity, unimodal mismatch distributions, and significantly negative values of *Tajima’s D* statistic (see Table 1, Fig. 4, Table 3) all support a recent population expansion from a smaller founder population as the most plausible explanation for the observed significant deviations from neutrality in the Kakamega populations. The observation that population expansion in the Kakamega population was detected by Tajima’s D statistic but not Fu’s Fs attests to the fact that Tajima’s D statistic is a more powerful tool for detecting deviations from neutrality when testing for population expansion and genetic hitchhiking in populations.

### CONCLUSION

This study has established that there is low gene diversity in both cpDNA (*π* = 0.0045) and nrDNA (*π* = 0.0053) of *P. africana* within East Africa. It also points out the uniqueness of the species within the Budongo population as exemplified within haplotype D within cpDNA.

### GUIDELINES FOR CONSERVATION

The estimate of genetic diversity and population genetic structure provide a basis for conservation and utilization of *P. africana*. It will help us in determining what to conserve and how to conserve this species. In this study, the results show that there are low levels of gene diversity in *P. africana*. The low levels of gene diversity make *P. africana* prone to genetic erosion. This implies that soon, populations of *P. africana* will not be viable, leading to extinction of the species if the present populations are not conserved. Due to this fact there is an urgent need to conserve the current populations of *P. africana* both in situ and ex situ. *In situ* conservation is first recommended to protect its original habitat from further destruction. Out of the five populations, Mabira is critically endangered because of habitats loss. Therefore, it should be protected immediately. For *ex situ* conservation, we suggest that the population which has high gene diversity should be selected as the conservation area. Elgon and Kakamega populations with high gene diversity are the best candidates for *ex situ* conservation. Also the Budongo population should be conserved since it is different in terms of cpDNA.

### ACKNOWLEDGEMENTS

Financial support was by the Federation of the Germany Government (BIOTA East Africa E15) under grant no. FKZ_01LC0625F1 in collaboration with the Makerere University Herbarium, Botany Department and the University of Hohenheim, Botany Department, Stuttgart, Germany. We are thankful to The Institute of Botany and Zoology, University of Hohenheim Germany for the labwork. We thank the Makerere University Herbarium (MHU), Uganda National Forestry Authority (NFA), Uganda Wildlife Authority (UWA), Kenya Wildlife Services (KWS) and the National Herbarium of Tanzania in Arusha for permission to collect and identify samples. The authors thank Dr. Paul Sseggawa (Mak), Dr. Gerald Eilu (Mak) for project planning and implementation of BIOTA E15and special thanks to Kiyemba Jalia for her help during the field trips for collecting *P. africana* samples.

### REFERENCES


Barakat G, Olfa S, Khaled C, Messaoud M, Mohamed M, Mohktar T, Amel...


Clark AG (1990) Inference of haplotypes from PCR-amplified samples of diploid populations. *Molecular Biology and Evolution* 7, 111-122


