Article (refereed)

Omondi, Stephen F.; Kireger, Eliud; Dangasuk, Otto G.; Chikamai, Ben; **Odee, David W.**; **Cavers, Stephen**; Khasa, Damase P.. 2010. Genetic Diversity and Population Structure of Acacia senegal (L) Willd. in Kenya. *Tropical Plant Biology*, 3 (1). 59-70. 10.1007/s12042-009-9037-2

Running head: Diversity and structure of *A. senegal*

Article type: Research Paper

Title: Genetic diversity and population structure of *Acacia senegal* (L) Willd. in Kenya.

Authors: STEPHEN F. OMONDI[1,3], ELIUD KIREGER[1], OTTO G. DANGASUK[2], BEN CHIKAMAI[3], DAVID W. ODEE[3, 4] STEPHEN CAVERS[4] AND DAMASE P. KHASA[5]

[1]Department of Forestry and Wood Science Technology, Moi University P.O Box 1125 Eldoret Kenya

[2]Department of Biological sciences, Moi University, P.O Box 1125 Eldoret

[3]Kenya Forestry Research Institute, P.O Box 20412-00200 Nairobi

[4]NERC Centre for Ecology and Hydrology, Bush Estate, Penicuik, Midlothian, EH26 0QB, UK

[5]Centre for Forest Research (CEF), Canada Research Chair in Forest and Environmental Genomics, Faculty of Forestry and Geomatics, Pavillon Eugene Marchand, Université Laval, Sainte-Foy, Québec, Canada G1K 7P4

Correspondence: Stephen Cavers, Centre for Ecology and Hydrology, Bush Estate, Penicuik, Midlothian, EH26 0QB, UK. Tel 00 44 (0)131 445 4343, Fax: 00 44 (0)131 445 3943, email: scav@ceh.ac.uk

Abstract (250 words)

The level of genetic diversity and population structure of *Acacia senegal* variety *kerensis* in Kenya was examined using seven polymorphic nuclear microsatellite loci and two chloroplast microsatellite loci. In both chloroplast and nuclear datasets, high levels of genetic diversity were found within all populations and genetic differentiation among populations was low, indicating extensive gene flow. Analysis of population structure provided support for the presence of 2 groups of populations, although all individuals had mixed ancestry. Groups reflected the influence of geography on gene flow, with one representing Rift Valley populations whilst the other represented populations from Eastern Kenya. The similarities between estimates derived from nuclear and chloroplast data suggest highly effective gene dispersal by both pollen and seed in this species, although population structure appears to have been influenced by distributional changes in the past. The few contrasts between the spatial patterns for nuclear and chloroplast data provided additional support for the idea that, having fragmented in the past, groups are now thoroughly mixed as a result of extensive gene flow. For the purposes of conservation and *in situ* management of genetic resources, sampling could target a few, large populations ideally distributed among the spatial groups identified. This should ensure the majority of extant variation is preserved, and facilitate the investigation of variation in important phenotypic traits and development of breeding populations.

Keywords: *Acacia senegal*, microsatellite, chloroplast, population structure, gene flow

Abbreviations:

ASAL -arid and semi-arid lands

RAPD - randomly amplified polymorphic DNA

ISSR - Inter-simple sequence repeats

HWE - Hardy-Weinburg Equilibrium

PCR - Polymerase chain reaction

DNA - Deoxyribonucleic acid

IAM - Infinite alleles model

SMM - Stepwise mutation model

TPM - Two-phase mutation model

Introduction

The arid and semi-arid lands (ASALs) of Africa are threatened by both land degradation (due to unsustainable agriculture, deforestation, overgrazing) and climate change. Across this vast area (approximately 55% of Africa's surface, Wickens *et al.* 1995) low and erratic rainfall (annually, <100-600 mm), high temperatures and poor soil water and nutrient availability limit agricultural productivity and place a high value on perennial, multipurpose tree species that provide a means to maximise agricultural potential and stabilise yields in stressful, unpredictable growing conditions (Fagg & Allison 2004). In the ASALs, the currently under-utilised tree species *Acacia senegal* (L.) Willd. (Fabaceae, Mimosoideae, *Aculeiferum*) has been recognised as having great potential to increase and diversify agricultural production as well as stabilise and restore degraded, vulnerable agroecosystems. In Kenya, where a new market is developing for an exudate from the species, commonly known as gum arabic, there is currently keen interest in assessing, protecting and improving local *A. senegal* resources, for the benefit of the large silvopastoral dryland community.

Acacia senegal is a small, deciduous tree, most highly valued for the production of gum arabic, a commodity of international trade since ancient times. Gum arabic is used in food, pharmaceuticals, cosmetic products and lithographic ink, because of its unique emulsification, film forming and encapsulation properties, making it a very important economic resource in the ASALs (Goodrum *et al*. 2000; Motlagh *et al*. 2006; Al-Assaf *et al*. 2007). In addition, the tree improves soil fertility through nitrogen fixation and provides shade, fodder, traditional medicine, fuel wood and substrate stabilisation (Arce

& Blanks 2001; Raddad *et al*. 2005). It is distributed in arid and semi-arid zones, mostly in tropical and sub-tropical regions, across the whole of sub-Saharan Africa and as far as India and Pakistan (White 1983; FAO 1985; ICRAF 1992). Four varieties are recognised within the species: *senegal*, *kerensis*, *rostrata* and *leiorhachis* (Brenan 1983; Fagg & Allison 2004), of which three (*senegal*, *leiorhachis* and *kerensis*) are found in Kenya. The latter, variety *kerensis*, is most widely distributed in Kenya and is very highly valued for the quality of gum it produces, forming the basis of an active international market (Booth & Wickens 1988; Fagg & Allison 2004; Chretin et al. 2008).

As part of an effort to promote sustainable management and conservation of forests for the production of wood and non-wood products, and increased tree-planting in ASALs, the Kenyan government has called for research on priority species (Sessional Paper no 9, 2005, Forest Policy), of which *Acacia senegal* is one. A science-based strategy for management and exploitation of genetic resources would help to identify the factors underlying gum quality variation and improve sustainability of production for the benefit of dryland inhabitants (Chikamai & Banks 1993; Chikamai & Odera 2002; Motlagh *et al*. 2006). An important focus for research and development, one which limits adoption and domestication of *A. senegal*, is variation in chemical composition of the gum. This determines gum quality and ultimately, its end use (Chikamai & Odera 2002; Motlagh *et al*. 2006). It is known that the chemical and molecular structure of gum is highly variable depending on source and this has significant effects on trade (Chikamai & Banks 1993; Jurasek *et al*. 1994; Islam *et al*. 1997; Chikamai & Odera 2002; Motlagh *et al*. 2006). If this geographic variation is a genetically controlled trait then there is the potential for

productivity improvement by sourcing and distribution of superior germplasm. However, in this case it is important, for maintenance of extant genetic resources and avoidance of genetic pollution, that the distribution and sources of genetic variation are well-characterised and the physical dimensions of gene flow understood.

Despite its biological, social and economic importance, little information exists on the genetic resources of *A. senegal* variety *kerensis* in Kenya (Chikamai & Banks 1993; Chikamai & Odera 2002). Indeed, only limited information is available on the extent, distribution and nature of its genetic variability across its entire distribution range (Fagg & Allison 2004). Genetic diversity and population structure of variety *senegal* was evaluated using isoenzymes in Senegal and low diversity was reported (H=0.175, Bergonzini & Joly 1992; Chevallier *et al*. 1994). Another study of population differentiation in variety *senegal* with samples from Kenya, Sudan, Pakistan, Mali and Mauritania found a clear separation between West and East African populations (Chevallier & Borgel 1998), a distinction most probably due only to distance. The only study on Kenyan populations of *A. senegal* was done using random amplified polymorphic DNA (RAPD) and inter-specific simple sequence repeat (ISSR), which reported moderate diversity (H = 0.283, Chiveu *et al*. 2008). However this study made no distinction between varieties; as commercial gum production in Kenya is entirely from variety *kerensis*, this is a division that must be recognised. Knowledge of genetic diversity and population structure of variety *kerensis* in its natural range in Kenya is crucial for formulation of appropriate management strategies directed towards conservation and maximum utilization of the genetic resource (Hueneke 1991; Malligan

et al. 1994). Furthermore, information on population differentiation will help to develop efficient sampling strategies for the variety (Bonnin *et al*. 1996) and enable conservation and improvement initiatives.

To identify levels and structure of genetic variation and initiate development of long term strategies for sustainable use of resources, a widespread analysis of variation in the species in Kenya was undertaken. To achieve this, a combination of nuclear and chloroplast microsatellites was used. This approach takes advantage of data from different parts of the genome to provide insight into different aspects of the species' evolution, at different timescales and reflecting different modes of gene dispersal (nuclear DNA is dispersed via pollen and seed, whilst organelle DNA is dispersed by seed only). The aim was to characterize and understand the genetic diversity and structure of natural populations of *A. senegal* variety *kerensis* in Kenya, detect population structure and patterns of gene flow and identify factors influencing diversity across the range and in individual populations.

Results

Genetic diversity

A total of 48 alleles were observed across the 7 nuclear microsatellite loci used in this study with a total number of alleles detected per locus varying from 4 (*Ame02* and *Ab06*) to 13 (*Ab26*) and an average number of alleles of 6.9 per locus (Table 1). Gene diversity (H_E) values per locus ranged from 0.530 (*Ame05*) to 0.868 (*Ab26*) with a mean value over all loci at 0.697, while observed heterozygosity (H_O) values ranged from 0.516 (*Ame05*) to 0.800 (*Ame03*) with a mean value of 0.709 (Table 1). At individual loci, four departures from HWE were noted: loci *Ame07* and *Ab26* showed significant ($P < 0.05$) heterozygote excess, whilst *Ame02* and *Ab18* showed significant ($P < 0.001$) heterozygote deficits. No linkage disequilibrium was detected between different genotypes with the Fisher exact test among the different microsatellite loci ($P > 0.05$) indicating that all seven loci segregate independently of each other.

Among the 11 populations surveyed, the mean multilocus observed heterozygosity (H_O) based on nuclear data was highest at Ngarendare (0.824, Table 2) and lowest at Daaba (0.613), whilst allelic richness (A_R) per population ranged from 5.05 at Serolipi to 4.27 at Daaba based on a minimum sample size of 20 individuals per population. Private alleles were identified in three of the 11 populations. One private allele per population was found in Ngarendare, Ngurunit and Serolipi.

A total of 8 alleles were found across the two chloroplast loci, characterising 9 haplotypes. Mean within population diversity (H_S) was 0.679, whilst overall diversity (H_T) was 0.807.

At a population level, significant deviations from HWE were observed at three sites (Table 2): heterozygote deficiency - Daaba (F_{IS} = 0.039, $P < 0.01$), heterozygote excess - Ngarendare (-0.207, $P < 0.001$), Merille (-0.136, $P < 0.05$). However, across all loci and all populations there was no significant deviation from HWE.

Population differentiation and structure

Population differentiation and structure was evaluated using both F_{ST} and R_{ST}. No significant differentiation was detected using F_{ST} across all populations and loci (0.045, $P > 0.05$), whilst R_{ST} was significant 0.125 ($P < 0.05$). If the stepwise mutation model is assumed, as seems reasonable for microsatellite evolution, there is a low but significant differentiation among populations, although in general high levels of gene flow are prevalent. Theory suggests that population differentiation is more accurately estimated by SMM, because this best accounts for the high mutation rate of microsatellite markers. In contrast, IAM often underestimates population differentiation at microsatellite loci (Hedrick, 1995).

Based on variation in chloroplast haplotype, population differentiation (G_{ST}) across all populations was 0.159. Comparing R_{ST} and G_{ST} (Pons & Petit, 1996) for the chloroplast data, no evidence for phylogeographic structure was found, reflecting the fact that

populations are largely mixed for chloroplast haplotypes. In keeping with these results, analysis of molecular variance (AMOVA) found that most of the genetic variation was partitioned within (91 %) rather than among (9 %) the populations (Table 3), as is characteristic of an outcrossing tree species (Hamrick et al, 1992).

The genetic distance between populations based on Cavalli-Sforza and Edward's chord distance (Dc) ranged from 0.152 (Archers' Post and Merille) to 0.347 (Ngarendare and Lokichar) with an average distance of 0.264 (Table 4). Mantel tests showed a positive but non-significant correlation of both F_{ST} (r = 0.099, $P > 0.05$) and R_{ST} (r = 0.2314, $P > 0.05$) with geographic distance for the 55 pairwise comparisons among the 11 populations under the isolation-by-distance model.

Using the delta K criterion (Evanno et al 2005), Bayesian clustering (Pritchard et al 2000; Falush et al 2003, 2007) suggested the presence of two groups as most probable (Figure 2), although all individuals showed mixed ancestry. This separation placed populations Kakuma, Lokichar, Lokitaung and Marigat within one group (defining group membership as most individuals within a population having >60% ancestry of that group), whilst Archer's Post, Kargi, Merille and Ngarendare fell within a second group. Populations Daaba, Ngurunit and Serolipi were highly mixed and not clearly within one group or another. Spatial analysis of chloroplast variance also suggested most likely presence of two groups, although not completely coincident with the groups identified in the nuclear microsatellite data. Groups were: 1. Kakuma, Lokichar, Lokitaung and Ngurunit and 2. Marigat, Merille, Ngarendare and Serolipi (Figure 1).

Tests for bottlenecks

Using the Wilcoxon's signed rank test, all 11 populations deviated significantly from mutation-drift equilibrium ($P < 0.05$) under the assumption of the infinite allelic mutation model (IAM), whilst only two of the populations (Kargi, Ngurunit) showed significant deviations under the TPM and SMM (Table 2, results shown for TPM only). In two populations (Kargi, Kakuma) evidence of a mode-shift was detected.

Discussion

Genetic diversity

From nuclear microsatellite data, all populations showed high levels of genetic diversity. The levels of heterozygosity (mean H_E = 0.667) detected were comparable to estimates obtained for other tropical tree species such as *Ceiba pentandra* (H_E = 0.85, Brondani *et al*., 2003), *Carapa guianensis* (H_E = 0.61, Dayanandan *et al*., 1999), *Swietenia humilis* (H_E = 0.53, White *et al*., 1999) and *Swietenia macrophylla* (H_E = 0.66, Novick *et al*., 2003) and are in keeping with expectations for widespread, long lived and outcrossing tree species (Hamrick & Godt 1990; Hamrick *et al*. 1992; Figueira *et al*. 2006). The estimates were similar to those recorded for the species from which the microsatellites were transferred, *A. brevispica* (H_E = 0.72, Otero-Arnaiz *et al*., 2005) and *A. mellifera* (H_E = 0.67, Ruiz-Guajardo *et al*., 2007). Diversity estimates from the chloroplast data were also high (H_T = 0.807), although similar to those found in other species (e.g. *Vitellaria paradoxa,* $H_e$cp = 0.71, Fontaine *et al*. 2004; *Adansonia digitata*, G_{ST} = 0.97, Tsy *et al*, 2009).

Such findings confirm earlier studies of the species by Chevallier *et al*. (1994) using isoenzymes, where significant levels of heterozygosity and partitioning of the majority of variation within populations was attributed to its outcrossing mating system. Doligez & Joly (1997), reviewing the outcrossing rates reported for 28 species of tropical forest trees in natural populations, found high genetic diversity to be common. Such levels of diversity are maintained by high levels of gene flow facilitated by efficient pollen

movement by pollinators and the widespread occurrence of efficient self-incompatibility mechanisms (Dick et al, 2008; White *et al*., 2002).

In general, most populations showed no significant deviation from HWE, although slight heterozygote excess was observed in some. Significant deviations from HWE were observed in three populations: Ngarendare and Merille showed heterozygote excess, whilst Daaba showed significant heterozygote deficiency. *Acacia senegal* var. *kerensis* is found distributed across the whole of the northern part of Kenya and is known to be long-lived (Fagg and Allison, 2004). The high levels of allelic diversity and heterozygosity coupled with low and negative values of F_{IS} found in this study support botanical and experimental data suggesting that *A. senegal* is a predominantly outcrossing species. In pollination trials, Obunga (1995) and Tandon *et al.* (2001) found *A. senegal* to be almost exclusively outcrossed and self incompatible. The numerous flowers produce abundant pollen, grouped into polyads of 16 pollen grains (Guinet, 1969) and the cup-shaped non-papillate stigma generally only accommodates one polyad. While self-pollen can be deposited on the stigma, self-incompatibility appears to operate inside the embryo sac and may function either pre- or post-fertilization (Obunga, 1995). The significant deviations noted at Ngarendare, Merille and Daaba therefore probably reflect local site history, although, due to a lack of detailed historical information, explanations can only be speculative. There are various potential causes of heterozygote excess: e.g. heterozygote advantage, negative assortative mating, clonal growth, reduced presence of selfed progeny due to small population size, population mixing (Stoeckel *et al*, 2006). In the absence of good supporting data to allow rejection of the first three and as there is no

evidence for clonal reproduction in this species, the most parsimonious explanation seems to be demographic. Certainly, the sample taken from the site at Ngarendare and Merille were necessarily (due to low density) drawn from across a much greater area than at other sites and so may represent mixing of discrete demographic units. In contrast, heterozygote deficiency is most commonly the result of inbreeding, due either to selfing or mating between close relatives (biparental inbreeding). The population at Daaba was notably even-sized (80 % of trees of the same height and DBH) suggesting they may be a single or few cohorts, possibly establishing together as a result of some disturbance event. The area is known to be prone to flooding which would provide a mechanism for significant, biased seed dispersal. If the population is more closely related than it would be if seed dispersal was random, a slight reduction in heterozygosity might be expected.

Population differentiation and structure

At nuclear markers, the degree of differentiation among populations was low (F_{ST} = 0.0447), even taking the evolution of microsatellites into account (R_{ST} = 0.12483) although the latter estimate should most accurately reflect the true situation (Hedrick 1995). Similarly, for chloroplast data, differentiation was low (reflecting the effectiveness of gene flow via seed dispersal), though slightly higher than at nuclear markers (G_{ST} = 0.159). In contrast to patterns observed for many tree species, the levels of differentiation at nuclear and chloroplast markers were of a similar order of magnitude: it is expected that differentiation at chloroplast markers will be higher, due to the more limited dispersal capability of seeds versus pollen, the slower evolution of the chloroplast molecule and lower effective population size. However, in *A. senegal*, it appears that

dispersal via pollen and seed may be equally effective and highly efficient. The species is both wind (Tybirk, 1997) and insect pollinated (Fagg & Allison, 2004) and, as the landscapes in which it commonly occurs tend to be structurally open, the potential for long distance pollen dispersal is high. Open, low density forest structures facilitate longer pollen dispersal distances, not just because pollinators must travel further to find flowering conspecifics but also because visibility of flowering trees is heightened, drawing pollinators across longer distances. Indeed, it has been observed that fragmentation of forested populations often drives increases in pollen dispersal distances (Dick *et al* 2008; White *et al* 1999, 2002). If wind dispersal also plays a role in pollination, dispersal distances should also be increased. The seeds of *A. senegal* are relatively large and attractive as a food source to animals in arid environments (Fagg & Allison, 2004). Dispersal following ingestion has a high potential to be long distance (Hamrick *et al*., 1991), particularly as the animals in question may be domestic herds belonging to pastoralist communities in transit between sparsely distributed pastures or water sources and potentially between regions during seasonal movements of livestock.

Generally speaking, however, these data suggest the existence of a single, weakly structured population, with few physical barriers to gene flow and extensive outcrossing (9 % of variation among populations - AMOVA and no significant isolation by distance). Very low levels of differentiation were observed in the earlier isoenzyme and RAPD/ISSR studies (Chevallier *et al*. 1994; Chiveu *et al*. 2008), with over 80 % of variation partitioned within population. This is in keeping with general observations for other tropical tree species with similarly high outcrossing rates and wide distributions,

(Muona 1990; Hamrick & Godt 1990; Hamrick *et al*. 1991, Dayanandan, *et al*. 1999; White *et al*. 1999; Collevatti *et al*. 2001; Lemes *et al*. 2002) and suggests species ecology, such as pollen and seed dispersal mechanisms and demographic history, rather than biogeography, is the major driver of population structure in Kenyan *A. senegal* (Alvarez-Buylla *et al*. 1996).

In studies of other *Acacia* species, geographic patterns have been similar. Joly *et al*. (1992), in a study of the population genetics of *Faidherbia albida* (Delile) A. Chev.- a species with similar seed and pollen dispersal characteristics to *A. senegal* - found strong differentiation among geographic zones, although the comparison was between West African and Zimbabwean populations. Within West Africa, where population differentiation might be expected to more directly reflect seed and pollen dispersal characteristics, genetic differentiation among populations bore no relation to geography, and the influence of grazing animals was proposed. In a survey of allozyme variation in *Acacia melanoxylon* two genetically distinct regions were observed (Playford *et al*. 1993), and similar differentiation was found by Brain (1986) in a study of the influence of environment on the geographic distribution of *Acacia karroo*. These studies indicated the role of geographic distance in the partitioning of genetic variation.

Nevertheless, some structuring of the Kenyan populations was noted, with two groupings detected in both nuclear and chloroplast data. Although groups in the two datasets do not precisely coincide, there appears to have been division in the past between populations to the northwest of Lake Turkana and those in Eastern Kenya. Effective gene flow

following this division appears to have more or less eliminated differentiation between these groups but traces still remain. Clustering of populations using nuclear data suggests the Eastern Rift Valley formation may have played a role in shaping gene flow with Marigat, Lokichar, Kakuma and Lokitaung - all located within or to the west of the Eastern Rift - showing predominant common ancestry, whilst to the east populations were either predominantly of single ancestry or highly mixed; the chloroplast data was largely in agreement. The intermediate case, Marigat, which grouped with eastern populations for chloroplast data but with Rift Valley populations for nuclear data, reflects the effective gene dispersal capability of the species and indicates how effectively populations are now mixed. As to the potential driving forces for creating the two groupings observed, a more rigorous analysis of the habitat requirements of *A. senegal* var. *kerensis* might shed some light. Certainly the zone between northwest Turkana and Eastern Kenya represents something of a corridor between the extensive drylands of Sudan and those of the Horn of Africa, and is interrupted not just by the Lake itself, but also by numerous upland areas such as the Samburu, Lenkiyo, Ndoto and Nyiru hills, all of which rise above common altitudinal limits for *A. senegal* var. *kerensis.* In this region, following an increase in aridity which began >2Mya (Bonnefille, 1976) the environment has remained relatively stable despite global climatic shifts (Gibernau & Montuire 1996). It is possible that more extensive forest vegetation and changes in the Lake levels could have been sufficient to restrict *A. senegal* to disjunct eastern and western populations, that have subsequently expanded, but this will require more detailed analysis, in particular bringing in other varieties of the species which co-occur with var. *kerensis* in Kenya.

Population bottlenecks

Bottleneck effects were detected in all the populations of *Acacia senegal* under the infinite alleles model (IAM) but in only two populations under the stepwise mutation (SMM) and the two-phase mutation (TMM) models . Under the IAM, microsatellite loci may exhibit heterozygosity excess even in stable populations (Cornuet & Luikart, 1996; Luikart & Cornuet 1998; Maruyama & Fuerst, 1985), so it seems conservative to assume that the finding of bottlenecks in all populations under this model is excessive and that the results of assuming a TPM or SMM (which were broadly in agreement) are closer to reality. In the latter cases, two of the populations show significant effects deviations from equilibrium heterozygosity suggesting recent bottlenecks; this is not to say that other populations have not undergone bottlenecks, but indicates that there are two cases in which the effect is detectable. If gene flow to bottlenecked populations is extensive, as seems likely from other results, any signature of recent bottlenecking will be obscured.

In any case, regardless of the mutation model assumed, two populations show significant signatures of recent bottlenecks. The possible reasons for this are many. In the dry, northern part of Kenya, the destruction of forests through clearance for agriculture, charcoal burning and settlement has produced landscapes of patchily distributed forests. Such landscape level change may have severe effects on diversity including changes in competitive regimes of plant species, dispersal patterns of pollen and seeds and effective population size (Nason *et al*. 1997). Furthermore, the arid and semi-arid lands are susceptible to regular natural disturbances such as fire and therefore individual populations may well have experienced substantial recent size changes. Based on the

observed efficiency of gene flow evident in other results, it seems likely that the two populations showing evidence of bottlenecking have experienced the effect recently and represent post-disturbance generations. Given the clear evidence that bottlenecking is occurring in some of these populations, a systematic study of multiple generations within single populations is merited to assess the impacts of recent human degradation of these forests.

Conclusions

This study has detected high genetic diversity, low inter-population genetic differentiation and occurrence of gene flow at long distances in *Acacia senegal* var. *kerensis* in Kenya, valuable information for conservation and improvement purposes. Understanding how genetic variation is partitioned within and among populations is a prerequisite for decision-making in the management of natural populations of valuable tree species (William & Hamrick 1996), particularly where development of the species as a commercial crop may initiate movements of germplasm. For the establishment of breeding programs or seed orchards, sampling can focus on a few populations of *A. senegal* variety *kerensis* as most variation should be accessible this way (Newton *et al*. 2003). However, it will also be important to assess the variation in quantitative traits across the range as the current data can only advise as to patterns of gene flow and likely distributions of gene diversity. Selectively influenced traits may differ substantially in their distribution. The current study indicates that, historically, reproductive isolation among populations has been low, although the extent to which more recent population fragmentation has increased isolation is not clear; this will require further investigation of

seed and seedling generations. Although variation in neutral markers does not necessarily reflect the adaptive potential of different populations, the allelic richness and the variance in allelic frequencies are useful for estimating the contribution of each population to the total genetic diversity of a species and to orientate conservation priorities (Petit *et al*. 1998). Future work will consider the sources of variation observed in the current data in more detail, assess patterns of quantitative variation and, in particular, examine relationships between genetic variation and gum arabic biochemical composition to complement the efforts towards improving gum arabic quality production from Kenya.

Methods

Population sampling

Three hundred individual trees of *Acacia senegal* variety *kerensis* were sampled from eleven natural populations in Kenya representing gum arabic production potential areas (Table 2, Figure 1). Leaf tissue was collected from 20-30 adult trees per population at a distance of between 150-600 metres apart depending on the size of the population and distribution of trees within the population. Sample sizes also varied among populations due to ease of accessibility. The leaves were dried on silica gel and stored at –20 °C until DNA extraction.

DNA extraction and PCR

Total genomic DNA was isolated from the leaves following a modified cetyltrimethyl ammonium bromide (CTAB) procedure (Fernandez *et al*. 2000) and DNA quantification performed through both Bio-photometer readings and comparison with low DNA mass ladder (Invitrogen) in ethidium bromide-stained 2 % agarose gels. Polymerase chain reaction (PCR) amplifications of seven polymorphic microsatellite loci (Otero-Arnaiz *et al*. 2005; Ruiz-Guajardo *et al*. 2007) were carried out in reaction volumes of 10 µl containing approximately 20 ng of genomic DNA, 1x PCR buffer (10 mM Tris-HCL pH 8.3, 50 mM KCl,), 1.5 mM $MgCl_2$, 200 µM of each dNTP, 0.05 µM of each primer, and 1 unit of *Taq* DNA polymerase (Invitrogen). Amplification profile used a touchdown program, with annealing temperatures ranging from 60-50, 58-48 to 55-45 (Table1), using a Peltier Thermal Cycler PTC-225 (MJ research). The program was 95 °C for 3 min, 20 cycles at 95 °C for 30 s (denaturation), either 60-50, 58-48 or 55-45 °C

(decreased by 0.5 °C, annealing) for 30 s, 72 °C for 30 s (extension), 10 cycles of 95 °C for 30 s, 50, 48 or 45 °C for 30 s, 72 °C for 30 s, followed by 72 °C for 10 min. The PCR products were then separated on 0.25 mm thick polyacrylamide gels containing 10 % ammonium persulfate (APS), tetramethylethylenediamine (TEMED)-Omnipur and 8 % acrylamide (Gene-PAGE plus)-Amresco, electrophoresed in 1x TBE buffer on a Li-Cor IR^2 4200 DNA sequencer. Band sizes and genotypes were determined by referencing the standard IRD 800 (50-350 bp and 50-700 bp) molecular marker from Li-Cor (Lincoln) using $SAGA^{GT}$ version 2.1 software (Li-Cor) and double checked by eye (Khasa *et al*. 2005).

For 8 of the 11 populations for which microsatellite genotyping was undertaken (Kakuma, Lokichar, Lakitaung, Marigat, Merille, Ngarendare, Ngurunit, Serolipi), two universal chloroplast microsatellite loci ccmp5 and 10 (Weising & Gardner 1999) were amplified (selected following screening of all ccmp loci for variation; data prepared by D. Odee as part of a rangewide study). PCR amplification followed Weising & Gardner (1999) and electrophoresis was carried out on 8% non-denaturing polyacrylamide gels in a Hoefer SE600 electrophoresis unit (300V) using Tris borate EDTA buffer (1X). All individuals were characterised for cpDNA haplotype at both loci, scoring band sizes against a 1 kb standard (Microzone).

Data analysis

Genetic diversity at nuclear microsatellite loci was estimated per population and overall. For each locus, the total number of alleles per locus and per population was determined and the distribution of allele frequencies calculated. The average number of alleles per

locus (A) for each population over all loci and unbiased value of expected (H_E) and observed (H_O) frequency of heterozygote were determined using F_{STAT} software version 2.9.3 (Goudet 2002). For each population, deviation from Hardy-Weinberg Equilibrium (HWE) and linkage disequilibrium tests were performed using Fisher exact tests employing GenePop 4.0 (Raymond & Rousset 1998), in which significant levels were determined using the Markov chain method (Guo & Thompson 1992). The heterozygote deficiency for each locus and population was tested using GenePop 4.0 software.

The extent and significance of genetic differentiation among populations was quantified using unbiased estimates of Wright's F_{ST} (Weir & Cockerham 1984) and its analogue R_{ST} (Slatkin 1995) using two models for microsatellite evolution (Infinite allele model (Kimura & Crow 1964) and stepwise mutation model (Ohta & Kimura 1973; Kimura & Ohta 1978)). A randomization test of pairwise F_{ST} differentiation with 1000 interactions (Goudet 2002) was conducted through F_{STAT}. Unbiased estimates of R_{ST} and their significance were determined after 1000 bootstraps with 95 % nominal confidence interval and permutation tests (Lynch & Crease 1990), using R_{ST}CALC version 2.2 (Goodman 1997). Fixation indices (Wright 1965); degree of inbreeding within population (F_{IS}) and overall inbreeding coefficient (F_{IT}) were also determined. The statistical significance of F_{IS}, F_{IT} and F_{ST} were tested by bootstrapping over loci with 95 % confidence interval and significance tests of multilocus pairwise differentiation were done using F_{STAT} software version 2.9.1 (Goudet 2002) with Bonferroni corrections.

A pairwise geographic distance matrix was calculated based on the latitude and longitude of each population, using the Geographic Distance Matrix Generator version 1.2.2 (Ersts, Internet). The hypothesis, that populations are differentiated because of isolation by distance, was tested by correlating pairwise F_{ST} and R_{ST} matrices against the geographical distance matrix. Spearman's rank correlation coefficient was calculated and significance determined with 10000 permutations using a Mantel procedure (Mantel 1967) available in F_{STAT} software version 2.9.1 (Goudet 2002). The co-existence of two mutational processes operating on different loci or even different alleles within a microsatellite locus raises the question of the appropriate measure of genetic distance between populations (Estoup *et al*. 2002). It has been suggested that Cavalli-Sforza and Edwards'chord distance (*Dc*) (Cavalli-Sforza & Edwards 1967; Goldstein & Pollock 1997) should be more appropriate, because it relies on the geometric disposition of populations in a multidimensional sphere delimited by allele frequencies, rather than on a given mutational model. This distance has also been shown to be the most suitable for analysis of microsatellite data, even in the presence of strong bottlenecks (Takezaki & Nei 1996). Genetic distances between populations were calculated using Cavalli-Sforza and Edwards' chord distance (D_c) (Cavalli-Sforza & Edwards 1967; Goldstein & Pollock 1997) using the GENDIST program (PHYLIP, version 3.6, Felsentein 1993). To analyze intra- and inter-population genetic variation, analysis of molecular variance (AMOVA) was carried out using GenAlex version 6 software (Peakall & Smouse 2006). Population structuring was analysed using Bayesian inference (implemented in STRUCTURE v2.2, Pritchard et al 2000; Falush et al 2003, 2007). After preliminary testing, the optimal number of groups (K) was determined using a 50,000 cycle burn-in period and 500,000

Monte-Carlo Markov Chains, using the admixture model (which assumes individuals may have mixed ancestry), assuming correlated allele frequencies among subpopulations and without any prior information on clustering of samples. Simulations for each value of K were repeated 20 times to provide stable probability estimates. The optimal number of groups was determined using the second order rate of change approach of Evanno *et al* (2005).

In general, *A. senegal* populations have suffered significant degradation over approximately the past 50 years, due to human impacts. As aging *A. senegal* trees is difficult, and precise population histories were not available, it was not clear whether or not the trees sampled in this study represented pre- or post-distrubance generations. In addition, natural disturbance effects, such as fire, are a frequent occurrence in arid and semi-arid zones and may have played a role at any of sites studied. Therefore, the populations were assessed for possible bottlenecks, following the procedure of Cornuet and Luikart (1996), implemented in the program BOTTLENECK v1.2 (Piry *et al*. 1999). The infinite alleles model (IAM), stepwise mutation model (SMM) and two-phase mutation model (TPM, using 95% SMM) were applied. This approach compares observed and expected gene diversities based on the observed number of alleles under mutation-drift equilibrium. Tests for mode-shift (change in allele frequency distribution) were also applied. If a recent bottleneck has occurred, allelic diversity is likely to have been lost faster than heterozygosity and may be manifested as a shift in the allele frequency distribution away from rare / low-frequency alleles.

For chloroplast data, within-population (H_S) and total (H_T) diversity and the level of population subdivision (G_{ST}) were estimated using Permut v2.0 (Pons & Petit, 1996). A test for phylogeographic structure was carried out by comparing the estimate of population differentiation using ordered alleles, assuming stepwise evolution, (R_{ST}) with that for unordered alleles (G_{ST}). If phylogeographic structure exists, population differentiation based on ordered alleles (i.e. reflecting phylogenetic relationships) should exceed that based on unordered alleles. Spatial structuring of variation at chloroplast loci was examined using a simulated annealing procedure - spatial analysis of molecular variance (SAMOVA, Dupanloup et al, 2002). For user-defined numbers of groups (K), the SAMOVA algorithm identifies geographical groups of populations such that the component of variation partitioned among groups (F_{CT}) is maximised. Values of K from 2-8 were tested, with 10,000 iterations, starting from 100 initial conditions.

Acknowledgements

This work was funded by the Department of Foreign Affairs and International Trade Canada (DFAIT) through Canadian Bureau for International Education (CBIE) in the framework of a Graduate Student Exchange Program (GSEP) for SFO, a Marie Curie Fellowship (ASGEN, No. 39216) at CEH for DWO and the ACACIAGUM project (EC FP6 contract 032233, http://inco-acaciagum.cirad.fr/). We thank Mr John Gicheru of KEFRI for helping in sample collection and Andre Gagne of Forestry research centre, Université Laval for coordinating the acquisition of research materials and all the colleges at the informatiqué office, Université Laval for their advice and support.

References

Al-Assaf S, Phillips GO, Aoki H, and Sasaki Y (2007) Characterization and properties of *Acacia senegal* (L.) Willd. var. *senegal* with enhanced properties (Acacia (sen) SUPER GUMTM): Part 1. Controlled maturation of *Acacia senegal* var. senegal to increase viscoelasticity, produce a hydrogel form and convert a poor into a good emulsifier. Food Hydrocolloids, 21:319-328

Aldrich PR, Hamrick JL, Chavarriaga P, Kochert G (1998) Microsatellite analysis of demographic genetic structure in fragmented populations of the tropical tree, *Symphonia globulifera*. Molecular Ecology, 7: 933–944.

Alvarez-Buylla ER, Garcia-Barrios R, Lara-Moreno C, Martinez-Ramos M (1996) Demographic and genetic models in conservation biology: Applications and perspectives for tropical rainforest tree species. Annual Review of Ecology and Systematics, 27: 387–421.

Arce LR, Banks H (2001) A preliminary survey of pollen and other morphological characters in neotropical *Acacia* subgenus *Aculeiferum* (Leguminosae: Mimosoideae). Botanical Journal of the Linnean Society, 135: 263-270.

Bergonzini JC, Joly HI (1992) Raspport de mission au Senegal et au Burkina Faso au 11-20 Mai 1992. CIRAD/CTFT 1-17.

Bonnefille R (1976) Implications of pollen assemblage from the Koobi Fora formation, East Rudolf, Kenya. Nature. 264: 403-407.

Bonnin I, Huguet T, Gherardi M, Prosperi JM, Olivieri I (1996) High level of polymorphism and spatial structure in a selfing plant species, *Medicago*

trancatula (Leguminosae), shown using RAPD markers. American Journal of Botany, 83: 843-855.

Booth FEM, Wickens GE (1988) Non-timber use of selected and zone trees and shrubs in Africa. FAO Conservation Guide 19. Rome, FAO.

Brenan JPM (1983) Manual on Taxonomy of, *Acacia* Species. Rome, FAO Forestry Division.

Brain P (1986) Leaf peroxidase types in *Acacia karroo*. Geographical distribution and influence of the environment. South African Journal of Botany, 26: 365-379.

Brondani RPV, Brondani C, Tarchini R, Grattapaglia D (1998) Development, characterization and mapping of microsatellite markers in *Eucalypts grandis* and *E. urophylla*. Theory and Applied Genetics, 97: 816–827.

Brondani RPV, Gaiotto FA, Missiaggia AA, Kirst M, Gribel R, Grattapaglia D (2003) Microsatellite markers for *Ceiba pentandra* (Bombacaceae), an endangered tree species of the Amazon forest. Molecular Ecology Notes, 3:177-179.

Cavalli-Sforza LL, Edwards AWF (1967) Phylogenetic analysis models and estimation procedures. American Journal of Human Genetics, 19, 233–257.

Chase M, Kesseli R, Bawa K (1996) Microsatellite markers for population and conservation genetics of tropical trees. American Journal of Botany, 83, 51–57.

Chevallier MH, Borgel A (1998) Genetic diversity of acacias. In: Lacacia au Senegal (eds, Campa, C., Grignon, C., Guey, M. and Hamon, S.) L' Orstom collection colloques et seminaries, Paris, France 287-308.

Chevallier MH, Brizard JP, Diallo I, Leblanc JM (1994) Genetic diversity in the *Acacia senegal* complex. Bois et Forets des Tropiques, No. 240 5-12.

Chretin M, Chikamai B, Loktari PE, Ngichili J, Loupa N, Odee D, Lesueur D (2008) The current situation and prospects for gum arabic in Kenya: a promising sector for pastoralists living in arid lands. International Forestry Review, 10: 14-22.

Chikamai BN, Banks WB (1993) Gum arabic from *Acacia senegal* (L)Willd. in Kenya. Food Hydrocolloids, 7 (6): 521-534.

Chikamai BN, Odera JA (2002) Commercial plant gums and resins in Kenya. Nairobi, Kenya.

Chiveu CJ, Dangasuk OG, Omunyin ME, Wachira FN (2008) Genetic diversity in Kenyan populations of *Acacia senegal* (L) Willd. revealed by combined RAPD and ISSR markers. African Journal of Biotechnology, 7 (14): 2333-2340.

Collevatti RG, Brondani RV, Grattapaglia D (1999) Development and Characterization of microsatellite markers for genetic analysis of a Brazilian endangered tree species *Caryocar brasiliense*. Heredity, 83: 748-756.

Collevatti RG, Grattapaglia D, Hay JD (2001) Population genetic structure of the endangered tropical tree species *Caryocar brasiliense*, based on variability at microsatellite loci. Molecular Ecology, 10: 349–356.

Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics, 144:2001-2014.

Dayanandan S, Dole J, Bawa K, Kesseli R (1999) Population structure delineated with microsatellite markers in fragmented populations of a tropical tree, *Carapa guianensis* (Meliaceae). Molecular Ecology, 8: 1585–1592.

1 Dick CW, Hardy OJ, Jones FA, Petit RJ (2008) Spatial scales of pollen and seed-
2 mediated gene flow in lowland tropical rain forest trees. Tropical Plant
3 Biology. 1: 20-33.
4 Doligez A, Joly HI (1997) Genetic diversity and spatial structure within a natural stand of
5 a tropical forest tree species, *Carapa procera* (Meliaceae), in French Guiana.
6 Heredity, 79: 72–82;
7 Dupanloup I, Schneider S, Excoffier L (2002) A simulated annealing approach to define
8 the genetic structure of Populations. Molecular Ecology, 11: 2571-2581.
9 http://web.unife.it/progetti/genetica/Isabelle/samova.html
10 Ersts PJ [Internet] Geographic Distance Matrix Generator (version 1.2.3). American
11 Museum of Natural History, Center for Biodiversity and Conservation.
12 http://biodiversityinformatics.amnh.org/open_source/gdmg.
13 Estoup A, Jarne P, Cornuet JM (2002) Homoplasy and mutation model at microsatellite
14 loci and their consequences for population genetics analysis. Molecular
15 Ecology, 11**:** 1591–1604.
16 Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals
17 using the software STRUCTURE: a simulation study. Molecular Ecology,
18 14: 2611-2620
19 Fagg CW, Allison GE (2004) *Acacia senegal* and Gum Arabic Trade. Oxford Forestry
20 Institute, Tropical forestry papers, No. 42.
21 Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using
22 multilocus genotype data: Linked loci and correlated allele frequencies.
23 Genetics, 164:1567–1587.

Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles Molecular Ecology Notes, 7:574-578

FAO (1985) Sand dune stabilization, shelterbelts and Afforestation in dry zones. FAO Rome.

Figueira A, Rafael MA, Alexandre MS, Angela SA (2006) Microsatellite loci transferability from *Theobroma cacao* to *Theobroma grandiflorum* Molecular Ecology Notes, 6 (4): 1219 - 1221

Fontaine C, Lovett PN, Sanou H, Maley J, Bouvet JM (2004) Genetic diversity of the shea tree (Vitellaria paradoxa C.F. Gaertn), detected by RAPD and chloroplast microsatellite markers. Heredity, 93: 639–648.

Gaiotto FA, Brondani RPV, Grattapaglia D (2001) Microsatellite markers for Heart of Palm — *Euterpe edulis* and *E. oleracea* Mart. (*Arecaceae*). Molecular Ecology Notes, 1(2): 86–88.

Gibernau M, Montuire S (1996) Mammal diversity and environment evolution during the Plio-Pleistocene in East Africa. Human Evolution. 11: 193-204.

Goldstein DB, Pollock DD (1997) Launching microsatellites: a review of mutation processes and methods of phylogenetic inference. Heredity, 88: 335–342.

Goodman SJ (1997) *R*ST calc: A collection of computer programs for calculating estimates of genetic differentiation from microsatellite data and determining their significance. Molecular Ecology, 6: 881–885.

Goodrum LJ, Patel A, Leykam JF, Kieliszewski JM (2000) Gum arabic glycoprotein contains glycomodules of both extension and arabinogalactan-glycoproteins. .Phytochemistry, 54: 99-106.

Goudet J (2002) FSTAT, Version 2.9.3.2: a Computer Package for Pcs Which Estimates and Tests Gene Diversities and Differentiation Statistics from Codominant Genetic Markers. http://www.unil.ch/izea/software/fstat.html.

Guinet PH (1969) Les Mimosacées. Etude de palynologie fondamental, corrélations, évolutions. *In:* Travaux de la Section Scientifique et Techniques, 9. Pondichéry. 293 p.

Guo SW, Thompson EA (1992) Performing the exact test for Hardy–Weinberg proportions for multiple alleles. Biometrics, 48: 2868–2872.

Hamrick JL, Godt MJW, Sherman-Broyles SL (1992) Factors influencing genetic diversity in woody plant species. New forests, 6: 95-124.

Hamrick JL, Godt MJW, Murawski DA, Loveless MD (1991) Correlation between species traits and allozyme diversity: Implication for conservation biology. In Falk DA, Holsinger KE (Eds), Genetics and conservation of rare plants. Oxford University press, New York, pp 75-86.

Hamrick JL, Godt MJW (1990) Allozyme diversity in plant species. In Brown AHD, Clegg MT, Kahler AL, Weir BS (Eds), Plant population genetics, breeding and genetic resources, 43-63. Sinauer, Sunderland, MA.

Hueneke FL (1991) Ecological implication of genetic variation in plant populations. In: Falk, D..A. and Holsinger, K..E. (eds) Genetics and Conservation of Rare Pants. Oxford University Press, New York, pp 31-44.

International Centre for Research in Agroforestry (ICRAF) (1992) A selection of useful trees and shrubs for Kenya. ICRAF, Nairobi, Kenya, 43p.

Islam AM, Phillips GO, Sljivo A, Snowden MJ, Williams PA (1997) A review of recent developments on the regulatory, structural and functional aspects of gum arabic. Food Hydrocolloids, 11 (4): 493–505.

Joly HI, Zeh-Nlo M, Danthu P, Aygalent C (1992) Population genetics of an African acacia, *Acacia albida* I. Genetic diversity of populations from West Africa. Australian Journal of Botany, 40: 59–73.

Jurasek P, Philips GO, Varga S, Chikamai BN, Banks WB (1994) The classification of natural gums: VI. Gum arabic derived from *Acacia senegal* var. *kerensis* from Kenya. Food hydrocolloids, 8 (6): 567-588.

Khasa PD, Pollefeys P, Navarro-Quezada A, Perinet P, Bousquet J (2005) Species-specific microsatellite markers to monitor gene flow between exotic poplars and their natural relatives in eastern North America. Molecular Ecology Notes, 5: 920-923.

Kimura M, Crow JF (1964) The number of alleles that can be maintained in a finite population. Genetics 4 725-738.

Kimura, M, Ohta T (1978) Stepwise mutational model and distribution of allelic frequencies in a finite population, Proc Natl Acad Sci USA 75 2868–2872.

Lemes MR, Brondani RPV, Grattapaglia D (2002) Multiplexed systems of microsatellite markers for genetic analysis in mahogany, *Swietenia macrophylla* King (Meliaceae), a threatened Netropical timber species. Heredity, 93: 287–290.

Luikart G, Cornuet JM (1998) Empirical evaluation of a test for identifying recently bottlenecked populations from allele frequency data. Conservation Biology, 12: 228–237.

Lynch M, Crease TJ (1990) The analysis of population survey data on DNA sequence variation. Molecular Biology and Evolution, 7: 377–394.

Mantel N (1967) The detection of disease clustering and a generalized regression approach. Cancer Research, 27: 209–220.

Maruyama T, Fuerst PA (1985) Population bottlenecks and non equilibrium models in population genetics. II. Number of alleles in a small population that was formed by a recent bottleneck. Genetics 111: 675-689.

Milligan BG (1992) Plant DNA isolation. In Hoelzel, A.R. (eds) Molecular Genetic Analysis of Population. Oxford, UK IRL Press. pp. 59 – 88

Mouna O (1990) Population Genetics in Forest Tree Improvement. In Brown AHD, Clegg MT, Kahler AL, Weir BS (Eds), Plant population genetics, Breeding and Genetic, 282-298 Sinauer, Sunderland, MA.Motlagh S, Ravines P, Karamallah KA, Qifeng M (2006) The analysis of Acacia gums using electrophoresis. Food hydrocolloids, 20 (6): 848-854.

Nason JD, Aldrich PR, Hamrick JL (1997) Dispersal and the dynamics of genetic structure in fragmented tropical tree populations. In Laurence WF, Bierregard R (Eds), Tropical Forest Remnants; Ecology, Management, and Conservation of Fragmented Communities, pp. 304–320. The University of Chicago Press, Chicago.

Newton AC, Allnutt TR, Gillies ACM, Lowe AJ, Ennos RA (2003) Molecular phylogeography, intraspecific variation and the conservation of tree species. Trends in Ecology and Evolution, 14:140–145.

Novick RR, Dick CW, Lemes MR, Navarro C, Caccone A, Bermingham E (2003) Genetic structure of Mesoamerican populations of big-leaf mahogany (*Swietenia macrophylla*) inferred from microsatellite analysis. Molecular Ecology, 12: 2885-2893.

Obunga EO (1995) A study of genetic systems of four African species of *Acacia*. DPhil. Thesis, School of Biological science, University of Sussex , pp1-142.

Ohta T, Kimura M (1973) A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. Genetics Research, 22: 201-204.

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes, 6: 288-295.

Petit RJ, El Mousadik A, Pons O (1998) Identifying populations for conservation on the basis of genetic markers. Conservation Biology, 12: 844–855.

Piry S, Luikart G, Cornuet JM (1999) Bottleneck: a computer program for detecting recent reductions in the effective population size using allele frequency data. Heredity, 90: 502–503.

Playford J, Bell JC, Moran GF (1993) A major disjunction in genetic diversity over the geographic range of *Acacia melanoxylon* R. Br. Australian Journal of Botany, 41: 355-368.

Pons O, Petit R (1995) Estimation, variance and optimal sampling of gene diversity.1. haploid locus. Theoretical and Applied Genetics, 90: 462-470.

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

Raddad EY, Salih AA, Elfadl MA, Kaarakka V, Luukkanen O (2005) Symbiotic nitrogen fixation in eight *Acacia senegal* provenances in dryland clays. Plant and Soil, 275: 261–26.

Raymond M, Rousset F (1998) Genepop (Version 3.1c) an updated, Version of Genepop V.1.2 (1995): population genetics software for exact tests and ecumenicism. Heredity, 86: 248–249.

Ruiz-Guajardo JC, Otero-Arnaiz A, Taylor T, Stone G, Glenn CT, Otero-Arnaiz NA, Miller JT, Preuss S, Schnabel A (2007) Isolation of polymorphic microsatellite markers in the sub-Saharan tree, *Acacia (senegalia) mellifera* (Fabaceae: Mimosoideae). Molecular Ecology Notes, 7: 1138-1140.

Stoeckel S, Grange J, Fernandez-Manjarres JF, Bilger I, Frascaria-Lacoste N, Mariette S (2006) Heterozygote excess in a self-incompatible and partially clonal forest tree species - *Prunus avium* L. Molecular Ecology, 15: 2109-2118.

Tsy J-MLP, Lumaret R, Mayne D, Ould A, Vall M, Abutaba YIM, Sagna M, Raoseta SOR, Danthu P (2009) Chloroplast DNA phylogeography suggests a West African origin for the baobab, *Adansonia digitata* L. (Bombacoideae, Malvaceae). Molecular Ecology, 18, 1707-1715.

Otero-Arnaiz A, Schnabel A, Glenn TC, Schnabel NA, Hagent C, Ndong L (2005) Isolation and characterization of microsatellite markers in the East African tree, *Acacia brevispica* (Fabaceae: Mimosoideae). Molecular Ecology Notes, 5: 366-368.

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. Genetics, 139: 457–462.

Tandon R, Shivanna KR, (2001) Pollination biology and breeding system of *Acacia senegal*. Botanical Journal of the Linnean Society, 135: 251-262.

Takezaki N, Nei N (1996) Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. Genetics, 144: 389–399.

Tybirk K. (1997) Reproductive biology and evolution of the genus *Acacia*. International Group for the Study of Mimosoideae. Bulletin No.20. p.45-53/

Weir BS, Cockerham CC (1984) Estimating P-statistics for the analysis of population structure. Evolution, 38: 1358-1370.

Weising K, Gardner RC (1999) A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms. Genome, 42: 9–19

Wickens GE, Seif El Din AG, Sita G, Nahal I (1995) Role of Acacia species in the ruraleconomy of dry Africa and the Near East. FAO Conservation Guide no. 27, Roma, Italy, 56p.

Williams CG, Hamrick JL (1996) Elite populations for conifer breeding and gene conservation. Canadian Journal of Forests research, 26:453-461.

White F (1983) The Vegetation of Africa. UNESCO. Paris. 356 p

White G, Powell W (1997) Isolation and characterization of microsatellite loci in *Swietenia humilis* (Meliaceae): an endangered tropical hardwood species. Molecular Ecology, 6: 851–860.

White GM, Boshier DH, Powell W (1999) Genetic variation within a fragmented population of *Swietenia humilis* Zucc. Molecular Ecology, 11: 1899–1910.

White GM, Boshier DH, Powell W (2002) Increased pollen flow counteracts fragmentation in a tropical dry forest: an example from *Swietenia humilis* Zuccarini. Proceedings of the National Academy of Sciences USA, 99: 2038–2042.

Wright S (1965) The interpretation of population structure by *F* statistics with special regard to systems of mating. Evolution, 19: 395–420.

Figure 1: Distribution of the 11 populations of *A. senegal* var. *kerensis* in Kenya. Large black or white circles show populations assigned to SAMOVA group (K=2) based on chloroplast haplotype variation. Populations represented by triangles were only analysed for nuclear SSR variation.

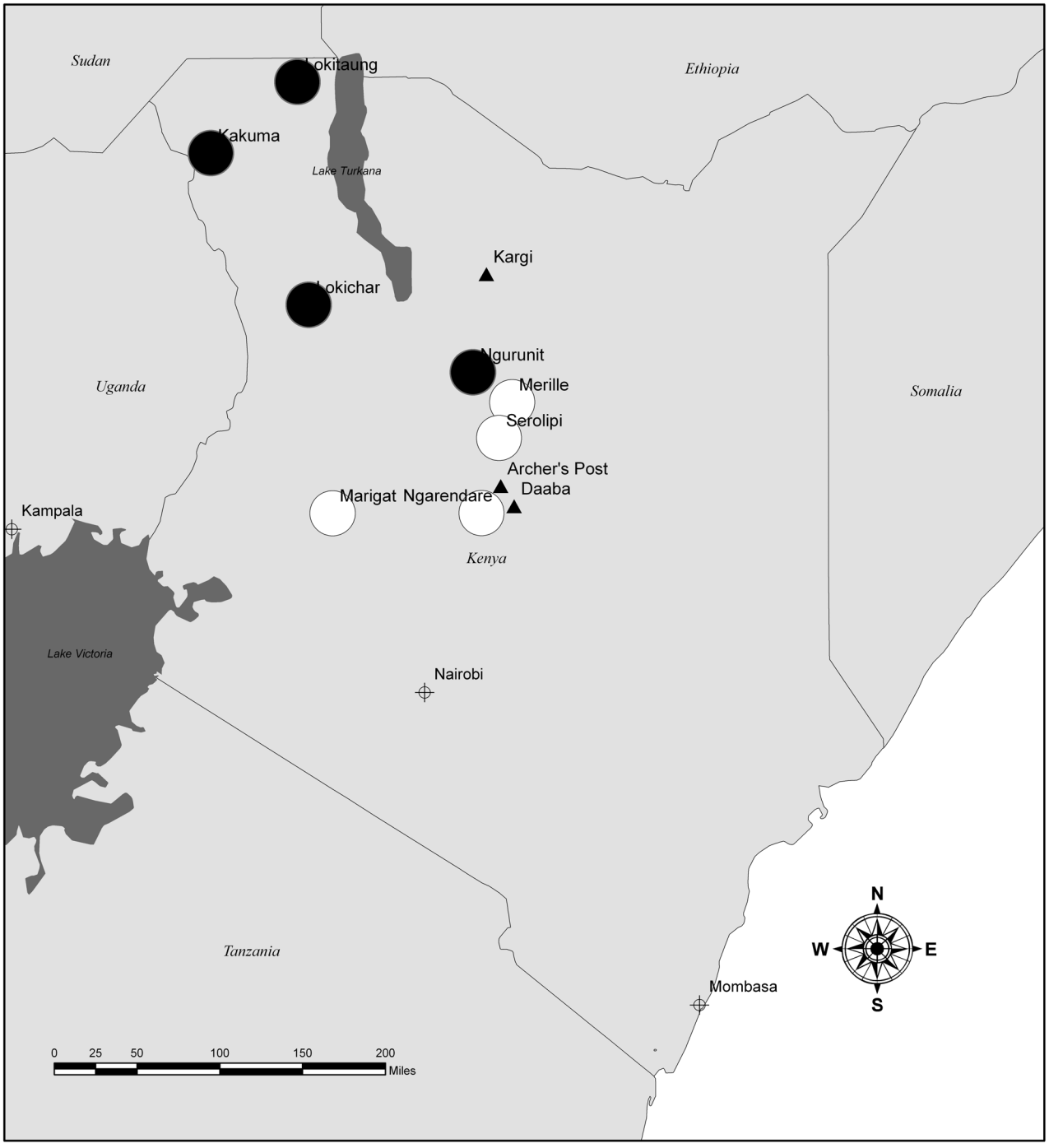

Figure 2: Plots of proportional group membership for the 300 trees genotyped at nuclear microsatellite loci, for K = 2. Each line represents a single tree, with colour representing proportion of ancestry derived from each group. Gray lines indicate the division between populations. Populations are 1 – Archers Post, 2 - Daaba, 3 - Kakuma, 4 - Kargi, 5 - Lokichar, 6 - Lakitaung, 7 - Marigat, 8 - Merille, 9 - Ngarendare, 10 - Ngurunit, 11 – Serolipi.

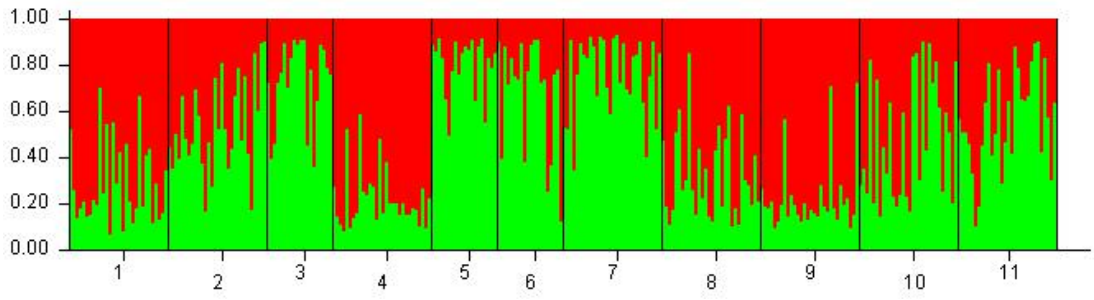

Table 1: Microsatellite loci in *A. senegal* var. *kerensis* obtained from cross amplification from *A. mellifera* and *A. brevispica* including primer sequence, GenBank accession number and annealing temperatures.

Locus accession no.	Primer sequence	Repeat motif	Ta (°C)	Size (bp)	A	H_O	H_E	HWE
Ame02	GAACCATCAGCGTAATAA	$(AC)_7(AG)_{11}$	55-45	117	4	0.670	0.543	***
DQ467674	GGTTTAGCAACATACTATCTC							
Ame03	GAACAATATCAGCAATCACT	$(AG)_9$	55-45	139	8	0.800	0.810	ns
DQ467673	CCTCATGCACACACAAGAT							
Ame05	CCCAACAAAGATCATCAT	$(ATC)_7$	58-48	203	5	0.516	0.530	ns
DQ467656	ATGGTTCAGTTTCTTTATTCT							
Ame07	ATAAAAACAAAAACCCAACTAAATG	$(GT)_{20}$	55-45	353	10	0.695	0.802	***
DQ467658	GTCCAAAACTCTTCAATGTCAA							
Ab06	CCTTCTTTGACGGTATTC	$(AC)_9(AG)_{10}$	58-48	147	4	0.785	0.745	ns
AY843537	TCATCTCTCTTCTCCATT							
Ab18	GAAGGGTCTGGCATTAC	$(AAG)_{15}$	60-50	212	4	0.717	0.579	***
AY843549	CGACGACGAAGATACT							
Ab26	ATATTCTGCTTTAGTCTA	$(AG)_8(AG)_9$	61-51	126	13	0.780	0.868	**
AY843557	GGGGCATAAATATGAG							

Ta, annealing temperature; bp, expected allelic size in base pairs; A, number of alleles observed per locus (in all cases, 30 individuals were analysed); H_O, observed heterozygosity; H_E, expected heterozygosity; HWE - departures from Hardy Weinburg Equilibrium (** P<0.01, ***P<0.001).

Table 2: Geographical locations and descriptive statistics over all loci for each population of *Acacia senegal* var. *kerensis* sampled in Kenya. Final column shows probabilities associated with Wilcoxon's sign rank test used to detect deviations from expected levels of gene diversity under the assumption of mutation-drift equilibrium using the two-phase mutation model (TPM).

Population	Latitude (°N)	Longitude/ (°E)	Altitude (m)	n	N_e	A_R	H_o	H_e	F_{IS}	TPM†
Ngarendare	0°33′39.9″	37°20′45.3″	972	30	5.30	5.00	0.824	0.685	-0.207***	0.2969
Daaba	0°32′00.2″	37°45′39.9″	941	30	4.40	4.27	0.613	0.637	0.039**	0.1094
Archer's Post	0°39′52.7″	37°38′47.0″	810	30	4.70	4.47	0.701	0.661	-0.061	0.2969
Serolipi	1°09′05.8″	37°35′51.9″	763	30	4.60	5.05	0.705	0.655	-0.063	0.1094
Kargi	2°38′35.8″	37°27′34.7″	454	30	4.60	4.45	0.724	0.639	-0.077	0.0078**
Merille	1°31′40.8″	37°45′23.7″	627	30	5.10	4.35	0.729	0.686	-0.136**	0.2969
Ngurunit	1°43′17.0″	37°17′24.3″	760	30	5.00	4.90	0.714	0.701	-0.019	0.0078**
Lokichar	2°21′57.2″	35°38′24.5″	786	20	4.40	4.43	0.693	0.677	-0.056	0.1094
Lokitaung	4°23′48.5″	35°31′40.5″	606	20	5.00	5.00	0.714	0.684	-0.045	0.5781
Kakuma	3°45′26.1″	34°39′59.5″	670	20	4.40	4.43	0.664	0.630	-0.025	0.0547
Marigat	0°28′20.4″	35°55′10.6″	1243	30	4.90	4.69	0.719	0.687	-0.047	0.0547
Average				**27.3**	**6.90**	**5.48**	**0.709**	**0.667**	**-0.0634**	

n -number of samples per population; N_e - mean number of alleles per locus per population; A_R - allelic richness; H_o - mean observed heterozygosity; H_e - mean expected heterozygosity; F_{IS} - fixation index; Significant *P*-values are followed by asterisks * - $P<0.05$; ** - $P<0.01$; † - TPM with 95% SMM.

Table 3: Analysis of molecular variance (AMOVA) for 11 populations of *Acacia senegal* var. *kerensis,* based on variation at nuclear microsatellite loci.

Source of variation	DF	SS	MS	Est. Var.	%Mol var.
Among Populations	10	167.103	16.710	0.453	9%
Within Populations	289	1268.467	4.389	4.389	91%
Total	299	1435.570	-	4.842	100%

DF, degrees of freedom; SS, sum of square; MS, mean square, Est.Var., estimated variance; % Mol. Var.; percentage molecular variance

Table 4: Matrix of pairwise genetic distances between populations of *Acacia senegal* var. *kerensis* in Kenya calculated using Cavalli-Sforza and Edwards chord distance, D_C (1967).

D	D	NN	AP	KI	ME	SE	Ng	KA	LOK	LO
NN	0.250									
AP	0.276	0.203								
KI	0.298	0.226	0.247							
ME	0.296	0.210	0.152	0.243						
SE	0.239	0.225	0.233	0.269	0.203					
Ng	0.278	0.264	0.242	0.288	0.207	0.264				
KA	0.236	0.283	0.294	0.304	0.264	0.202	0.229			
LOK	0.328	0.294	0.263	0.293	0.247	0.238	0.242	0.251		
LO	0.318	0.347	0.297	0.341	0.303	0.278	0.299	0.250	0.244	
MA	0.265	0.285	0.324	0.322	0.293	0.237	0.252	0.258	0.261	0.269

D, Daaba, NN, Ngarendare; AP, Archer's Post; KI, Kargi; ME, Merille; SE, Serolipi; Ng, Ngurunit; KA, Kakuma; LOK, Lokitaung; LO, Lokichar; MA, Marigat