npg

Genetic diversity of the shea tree (*Vitellaria paradoxa* C.F. Gaertn), detected by RAPD and chloroplast microsatellite markers

C Fontaine[1], PN Lovett[2], H Sanou[3], J Maley[4] and J-M Bouvet[1]

[1]Cirad-Forêt, Campus international de Baillarguet TA10/C, BP 5035, 34398 Montpellier cedex, France; [2]Technoserve-Ghana, PO Box 135, Accra, Ghana; [3]Institut d'Economie Rurale, Programme Ressources Forestières; Centre Régional de la Recherche Agronomique de Sotuba, BP 258, Bamako, Mali; [4]Institut des Sciences de l'Evolution Université de Montpellier II, cc 065, Université Montpellier 2, Place Eugène Bataillon, 34095 Montpellier cedex 05, France

RAPDs and chloroplast microsatellites were used to quantify the genetic variation of *Vitellaria paradoxa* (an economically important tree species in sub-Saharan Africa, north of the equator) and to analyse the geographic distribution of diversity in relation to the refuge theory. A total of 13 locations were sampled in eight countries, covering most of the natural range from Senegal to Uganda. A total of 67 polymorphic and 15 monomorphic RAPD loci were detected in 179 individuals. No relationship was identified between diversity and longitude or latitude. An unrooted neighbour-joining tree suggested a western group and an eastern group, representing 7% ($P = 0.000$) of the total variation. A Mantel test suggested that genetic distances between populations were correlated to geographic distances ($R = 0.88$, $P = 0.001$). The three-chloroplast microsatellite primers, assayed in 116 individuals, revealed 10 different alleles and seven chlorotypes. Most of the populations comprised a single haplotype. It is proposed from these results that the difference between western and eastern populations results from putative refugia separated by the current 'Dahomey Gap' (a semiarid zone that meets the coast around the Ghana–Togo–Benin–Nigeria borders), an area that may have been exceptionally dry during glacial periods. In addition, it is suggested that the haplotype distribution and frequency in the western populations could be due to the more recent impact of humans, particularly shea tree selection and dispersal during traditional agroforestry.

Introduction

Present patterns of genetic variation of tree species have been determined by numerous factors such as the effects of humans and climate change. Although the theory has been debated, one of the most consistent factors, suggested to have had a strong influence on the partition of the genetic diversity of tree species, is the last glaciation and the subsequent migration from various refugia (Willis and Whittaker, 2000). In Europe (Vendramin *et al*, 1998) or in South America (Bekessy *et al*, 2002), the genetic variability of tree species has been analysed in connection with the last glaciation, 15–20 000 years ago. In the soudano-sahelian region of Africa, this period has also been suggested as having a strong impact on vegetation. Although the debate continues as to the exact timing, nature and causes of vegetation changes, conservative predictions from various studies all suggest large-range changes – in the range of hundreds of kilometres, that is, northern boundaries of wet forest are thought to have moved at least up to 400 km south following drying after the Holocene climatic optimum between ca. 4000–7000 bp most notably around the Dahomey gap (Maley, 1991; Lézine and Vergnaud-Grazzini, 1994; Salzmann and Waller, 1998). To allow for the survival of plant and animal species during the particularly cold and arid periods, Maley (1996) has proposed putative refugia along the West African coast from Sierra Leone to Ivory Coast, in southwest Cameroon to western Gabon, in the eastern part of the Democratic Republic of Congo (former Zaïre) and along the Zaire River. Although few studies are concerned with tree species of the Central Africa (Lowe *et al*, 2000; Muloko-Ntoutoume *et al*, 2000), no studies, to our knowledge, have attempted to establish the relationship between the molecular genetic variation of a tree species and the possible evolution of vegetation, after the last glaciation, in the soudano-sahelian region of Africa.

Different methodologies using molecular markers are widely used to analyse the pattern of variation within and among natural populations of tree species. Among the various maker systems, RAPDs are one of the most popular DNA-based approaches (Martin and Hernandez Bermejo, 2000; Bekessy *et al*, 2002). They are the least technically demanding and offer a fast method for providing information from a large number of loci, particularly in species where no study has previously been undertaken. Moreover, the diversity assessed with

RAPDs is comparable with that obtained with allozymes or RFLP (Wu *et al*, 1999). Some limitations exist, however, owing to their lack of reproducibility and the identifical patterns produced by homozygous and heterozygous individuals. Chloroplast DNA markers are often used to study genetic diversity and structure, especially in combination with nuclear markers (Viard *et al*, 2001; Newton *et al*, 2002), and have provided useful information on plant colonisation and dispersal. Chloroplasts are maternally inherited in most angiosperms and paternally inherited in gymnosperms, so the level of differentiation is greater than at loci with biparental inheritance (Ennos, 1994). Chloroplast DNA polymorphism can be specifically employed to infer seed flow among populations (Muloko-Ntoutoume *et al*, 2000). Most studies over the last decade have used cleaved amplified polymorphisms (CAPS) generated by a PCR–RFLP procedure (Mohanty *et al*, 2000; Raspé *et al*, 2000), although the polymorphism detected by this marker is sometimes very low. Chloroplast microsatellite markers (cpSSRs) have higher polymorphism and are now frequently used in phylogeographic analyses of forest tree (Palmé and Verdramin, 2002; Collevatti *et al*, 2003; Grivet and Petit, 2003) and other species (Grivet and Petit, 2002; Mengoni *et al*, 2001, 2003).

Few studies have been undertaken on trees in the tropics, particularly on African trees. In this study, we used both RAPDs and chloroplast microsatellite markers to investigate the pattern of variation at the natural range scale of the shea tree.

The shea tree, *Vitellaria paradoxa* C.F. Gaertn (family Sapotaceae, syn. *Butyrospermum paradoxum* Hepper), is one of the major components of the agroforestry parklands in the soudano-sahelian region (Lovett and Haq, 2000a) and is the main oil-producing plant of this region. The oil, extracted from dried kernels, is utilised for local domestic uses such as cooking and is exported for western food, confectionary, cosmetic and pharmaceutical industries. Compared to other forest tree species in the natural environment, the shea tree has been strongly influenced by man. For centuries, during fallow cycles, healthy trees have been maintained and it has been proposed that unconscious selection has resulted in semi-domestication of this species (Lovett and Haq, 2000a).

Its natural range spans from Senegal to Uganda, a semiarid zone that forms a band of ca. 6000 km long and averaging ca. 500 km wide. Two subspecies have been defined, characterised by different morphological criteria (leaves and flowers) and a rather ill-defined geographic segregation. *V. paradoxa* subsp *paradoxa* (Kotschy) is present in the western part of the range, from Senegal to Central Africa, whereas *V. paradoxa* subsp *nilotica* (Kotschy) is found in the eastern part, including Zaire, Uganda, Sudan and Ethiopia. Leaves and flowers of subsp *nilotica* are bigger but the distinction between the two subspecies is vague and needs clarification (Hall *et al*, 1996).

Despite its social and economic importance, little research has been undertaken that provides information for sustainable management. Limited studies have permitted some improvements to the definitions of ecological and silvicultural patterns (Lovett and Haq, 2000b) and little is known about genetic variation across the species range. The aims of this study are then (i) to quantify the genetic variation within and between populations using these two molecular markers and (ii) to analyse the geographic distribution of diversity in relation to locations proposed as part of the 'refuge theory'.

Material and methods

Plant material and DNA extraction

A total of 13 locations in eight countries have been identified, that covering much of the natural range from Senegal to Uganda (Table 1), provide a sample of the two subspecies. Within each location, between 15 and 30 adult trees, with a diameter greater than 20 cm, were chosen randomly in agroforestry parklands (defined as 'land-use systems in which woody perennials are deliberately preserved in association with crops and/or animals in a spatially dispersed arrangement and where there is both ecological and economic interaction

Table 1 Characteristics of the *V. paradoxa* population sampled in the natural range

Country	Population	Nb ind	Latitude	Longitude	Rainfall (mm)	Temperature[a] (°C)	Dry season
Benin	Dikouhan	14	10°32 N	1°29 E	1300[b]	26	November–February
Burkina	FasoSapone	13	12°34 N	1°34 W	500–1400[c]	29	October–April
	Peni	7	10°58 N	4°27 W	500–1400	29	October–March
	Toukoum	7	13°11 N	0°56 W	500–1400	29	October–March
Ivory Coast	Dolekaha	21	9°50 N	6°10 W	2000	26	November–April
Mali	Massala	9	13°16 N	6°30 W	600–1300	27	November–April
	Badougou	5	11°15 N	8°18 W	600–1300	27	November–April
	Mperesso	21	12°17 N	5°21 W	600–1300	27	November–April
Senegal	Samecouta	16	12°38 N	12°30 W	700–1700	28	November–May
Central Africa	Ndélé	13	8°00 N	20°00 E	1500	26	November–February
Cameroon	Mafakilda	14	9°20 N	13°50 E	1500	23	November–March
Uganda	Amoya	18	2°32 N	33°03 E	1000	22	November–February
	Angetta	21	2°26 N	33°30 E	1000	22	November–February
Total		179					

[a]Mean average temperature.
[b]Estimation of the mean annual rainfall.
[c]Range of variation of the annual rainfall observed in a short period.

between trees and other components of the system'). A minimum distance of 50 m was respected to avoid selection of closely related individuals. Although sampling was sparse, current analyses (microsatellite markers not published) have shown that within each country populations are weakly differentiated. One sampled population per country is expected to give an acceptable estimate of the allele frequency in the country. Five leaves were collected from each tree and dried rapidly in the field using silica gel.

DNA was extracted from dried leaves. A 100 mg protein of leaves was ground to a fine powder with a mortar and pestle in a 1.5 ml Eppendorf tube under liquid nitrogen. A 5 ml volume of DNA extraction buffer was added (100 mM Tris-HCl (pH 8.0), 20 mM EDTA, 1.4 M NaCl, 1% PEG 6000, 2% MATAB, 0.5% sodium sulphite). The tube was then incubated at 74°C for 20 min. Samples were washed with wet chloroform (CIAA, 24:1) to remove cellular debris and protein. After 15 min of centrifugation at 5000 g, the liquid phase was transferred to 15 ml tubes. A 5 ml portion of isopropanol was added and mixed gently to precipitate the DNA. The resulting DNA pellets were resuspended in 400 μl of sterile water overnight at 37°C and stored at −20°C until required.

RAPD method

PCR amplification was carried out in a 20 μl reaction mixture with 5 μl of DNA (3 ng/μl), 2 × buffer, 0.2 μM of each primer (OPB7, OPB11, OPN15, OPR15, OPW9, OPW12, OPW13, OPX3, OPX6, OPX11, OPY6, OPY13, OPY20, OPW5, OPW19), 5 U/μl of polymerase DNA Taq and completed with sterile water. The reaction mixture was overlaid with 40 μl of sterile mineral oil to prevent fluid evaporation. All reactions were performed in Techne Cyclogene. Optimal amplification conditions for RAPDs were one cycle of 3 min at 94°C (initial denaturation), followed by 45 cycles of 4 min at 94°C (denaturation), 1 min at 36°C (annealing) and 2 min at 72°C (extension). A final step of 10 min at 72°C ensured full extension of all amplified products. RAPD bands were separated in 1.5% agarose gel, stained in ethidium bromide and visualised by UV transillumination.

To reduce errors when comparing RAPD profiles between different PCR runs, the same 10 individuals were included in all PCR runs. Only RAPD bands that could be unequivocally scored were used in the analysis. In addition, low (<280 bp) and high MW (>1700 bp) weight bands have been eliminated because they show a low repeatability. This reduced the potential number of bands that could have been scored but avoided misscoring.

Chloroplast microsatellite method

A total of 10 universal microsatellite primers (Ccmp1 to Ccmp10) described by Weising and Gardner (1999), six rice microsatellite primers (Rtc3, Rtc4, Rtc5, Rtc6, Rtc9, Rtc10) described by Ishii and McCouch (2000) and five tobacco microsatellites (Ntcp7, Ntcp9, Ntcp19, Ntcp23, Ntcp28) described by Bryan et al (1999) were tested over a subset of the total population. Among the 21 primer pairs tested on a sample of eight individuals, 10 amplified and only four (Ccmp3, Ccmp5, Ntcp8, Ntcp9) were polymorphic. For the primers Ccmp3 and 5, PCR

amplifications were conducted in an 8 μl reaction mixed with 2 μl of DNA, 2 × buffer, 10 μM of each primer (R and F), 5 U/μl of polymerase DNA Taq and completed with sterile water. All reactions were performed in 'Stratagene' Thermocyclor. Optimal amplification conditions were one cycle of 4 min at 94°C (initial denaturation), followed by 30 cycles of 30 s at 94°C (denaturation), 1 min at 56°C (annealing) and 1 min at 72°C (extension). A final step of 5 min at 72°C ensured full extension of all amplified products.

For the primers Ntcp8 and Ntcp9, PCR amplifications were carried out in a 20 μl reaction mix with 5 μl of DNA, 10 × Buffer, 2 μM of each primer (R and F), 5 mM of dNTPs, 50 mM of MgCl₂, 99% of glycerol, 5 U/μl polymerase DNA Taq and completed with sterile water. Amplification conditions were similar, except that the annealing temperature was 55°C. Bands were separated and visualised in polyacrylamide gel.

Due to difficulties in reading the differences between alleles on the gel, Ntcp9 was excluded from the analyses.

Data analysis

Initially the number of individuals sampled within each of the populations was about 30. But the number of individuals that gave a good DNA decreased strongly during laboratory activities especially for populations from Burkina and Mali. For those populations, the leaves appear to have been too old.

In the case of RAPD data, amplified DNA marker bands were scored in a binary manner as either present (1) or absent (0) and entered into a binary data matrix. Each PCR product was assumed to represent a single locus because homology is generally high at the intraspecific level (Païvi, 2000). The frequency of each band and the percentage of polymorphic loci (%P) were calculated in each population.

Some specific genetic analyses were conducted with the POGGENE software (Yeh and Boyle, 1997). To assess molecular variation, Shannon's diversity index was used. This parameter, defined as $I_{RAPD} = -\sum_{i=1}^{i=2} p_i \log_2 p_i$, where p_i is the frequency of the RAPD phenotype (presence (1) or absence (0) of the band), is frequently used without the need to make assumptions regarding Weinberg equilibrum (Aide and Rivera, 1998; Martin and Hernandez Bermejo, 2000). It was calculated for each locus averaged over loci to quantify the degree of variation within each population, $I_{RAPDpop}$. Shannon's index was also estimated for the whole sample considered as a single population, $I_{RAPDtot}$. The expected genetic heterozygosity Her was estimated by assuming Hardy–Weinberg equilibrium (index of fixation $F = 0$). This assumption relies on estimates obtained in other studies with nuclear microsatellite markers (Kelly et al, 2004).

For cpSSR markers, because of the non-recombining nature of the chloroplast genome, cpDNA haplotypes were treated as alleles at a single locus. Chloroplast haplotype variation within populations was calculated by estimating the effective number of haplotypes $ne = 1/\sum_{i=1}^{n} p_i^2$ (where p_i is the frequency of the chlorotype i in a population) and the haplotype diversity, by using the formula $He_{cp} = 1 - \sum_{i=1}^{n} p_i^2$, with the SPAGEDI software (Hardy and Vekemans, 2002).

To analyse genetic structure, for RAPD and chloroplast microsatellite markers, a matrix of pairwise distance measures was analysed using the analysis of molecular variance, (AMOVA) with the software ARLEQUIN version 2000 (Schneider *et al*, 2000). The percentages of variance within and among populations were estimated to partition the variation. A matrix of genetic distances was calculated using the AMOVA-derived F_{st}. The F_{st} distance matrix was used for a cluster analysis using the neighbour-joining method from the software package DARwin 3.6 (Perrier *et al*, 2003). The degree of genetic differentiation among populations was also estimated using the parameter G_{st} ($G_{st_{RAPD}}$ for RAPDs and $G_{st_{cp}}$ for chloroplast microsatellites). It was computed by the ratio $G_{st} = (H_t - H_s)/H_t$, where H_t is the total gene diversity and H_s is the mean of the within-population gene diversities. The percentage of variation among populations was then calculated with the Shannon's diversity index by $I_{st} = (I_{tot} - I_{pop})/I_{tot}$.

The levels of differentiation among populations estimated from AMOVA for nuclear markers ($F_{st_{RAPD}}$) and chloroplast markers ($F_{st_{cp}}$) could be used to derive the pollen to seed migration ratio, using the following formula (Ennos, 1994):

$$r = m_p/m_s$$
$$= [((1/F_{st_{RAPD}}) - 1) - 2((1/F_{st_{cp}}) - 1)]/[(1/F_{st_{cp}}) - 1)]$$

where m_p and m_s are pollen and seed migration rates, respectively. In the case of RAPDs, this value is biased, and probably overestimated, due to the amplification of the cytoplasmic genome. Association between geographic and genetic distance was estimated as a Spearman's rank correlation coefficient. The association was tested with the Mantel Test using Genepop software (Raymond and Rousset, 1995).

Minimum spanning networks (each network embedding all minimum spanning trees for a given distance matrix) were computed with MINSPNET (Excoffier and Smouse, 1994), provided with the software ARLEQUIN version 2000 (Schneider *et al*, 2000). The distance matrix between haplotypes was calculated according to two methods: distance matrix calculated by the number of different alleles between haplotypes or by the square in the difference of microsatellite size with the formula

$$D_{ij} = \sum_{l=1}^{L} (a_{il} - a_{jl})^2$$

where a_{ij} and a_{jl} give the allele size in base pairs at the *l*th locus of individuals *i* and *j*, respectively.

Results

Genetic diversity and structure with RAPD markers

The 16 random primers generated a total of 67 RAPD polymorphic and 15 monomorphic loci ranging in size from 1670 to 280 bp. This set of loci is expected to give a good sampling of the total genome and a good assessment of the genetic diversity. The number of bands per primer varied from 1 to 6. Shannon's diversity parameter for the total population was equal to 0.40 (standard deviation (SD) = 0.24) and varied from 0.24 (SD = 0.27) for the population of Badougou in Mali to 0.35 (SD = 0.36) for the population of Dolekaha in Ivory Coast. The diversity parameter He_{RAPD} followed the variation of Shannon's diversity parameter, with a value for the total population equal to 0.26 (SD = 0.17). It varied among the populations from 0.16 (SD = 0.19) in Badougou to 0.23 (SD = 0.18) in Dolekaha. The percentage of polymorphic loci varied from 50% in Peni up to 71% in Dolekaha (Table 2) and the total population value was 81%.

No pattern of variation for I_{RAPD}, He_{RAPD} and %P was observed among the set of populations. For example, the relationship between diversity parameters and latitude was in low correlation between latitude and He_{RAPD} ($R = -0.28$, associated *P*-value = 0.21); a similar result was obtained with longitude ($R = -0.18$, $P = 0.29$). In addition, the differences between populations were smaller than the standard deviation (Table 2).

The differentiation assessed among populations was not marked ($G_{st_{RAPD}} = 0.23$) showing that 77% of the variation was present within populations. The unrooted neighbour-joining tree obtained with RAPD markers (Figure 1) exhibited two clusters formed by the

Table 2 Population size (N), Shannon's index (I_{RAPD}), per cent of polymophic RAPD loci (%P) and RAPD diversity (He_{RAPD})

Country	Population	N	ne	I_{RAPD}	He_{RAPD}	%P
Benin	Dikouhan	14	1.31 (0.34)	0.30 (0.25)	0.20 (0.18)	68
Burkina	Sapone	13	1.35 (0.36)	0.32 (0.27)	0.21 (0.19)	63
	Peni	7	1.31 (0.37)	0.26 (0.28)	0.18 (0.20)	50
	Toukoum	7	1.29 (0.35)	0.27 (0.28)	0.18 (0.20)	52
Ivory Coast	Dolekaha	21	1.38 (0.35)	0.35 (0.26)	0.23 (0.18)	71
Mali	Massala	9	1.35 (0.37)	0.31 (0.28)	0.21 (0.17)	61
	Badougou	5	1.27 (0.36)	0.24 (0.27)	0.16 (0.19)	61
	Mperesso	21	1.33 (0.33)	0.32 (0.26)	0.21 (0.18)	70
Senegal	Samecouta	16	1.29 (0.35)	0.27 (0.26)	0.18 (0.19)	61
Central Africa	Ndele	13	1.36 (0.38)	0.31 (0.28)	0.21 (0.20)	63
Cameroon	Mafakilda	14	1.39 (0.37)	0.34 (0.27)	0.23 (0.20)	67
Uganda	Amoya	18	1.36 (0.38)	0.32 (0.27)	0.21 (0.20)	66
	Angetta	21	1.30 (0.31)	0.30 (0.25)	0.19 (0.18)	70
Total population		179	1.42 (0.33)	0.40 (0.24)	0.26 (0.17)	81

Standard deviations are in parentheses.

populations of Burkina Faso, Ivory Coast, Mali and Senegal, and an eastern group comprised of populations from Cameroon, Central Africa and Uganda. Analyses of
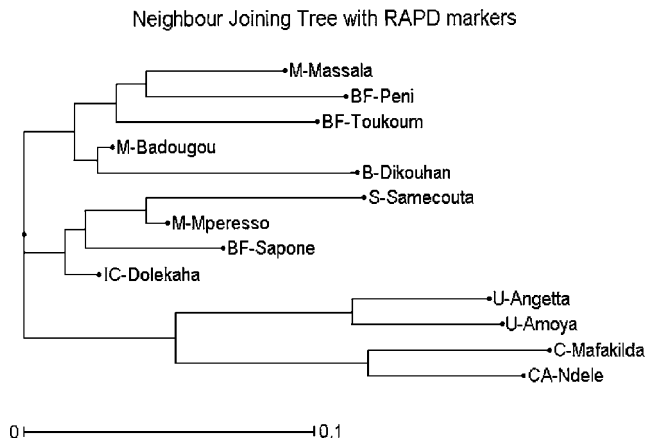

Figure 1 Unrooted neighbour-joining tree based on RAPD markers illustrating the phylogenetic relationship between the populations of Burkina Faso (BF), Ivory Coast (IC), Mali (M) and Senegal (S) and the east one with the populations of Benin (B), Cameroon (C), Central Africa (CA) and Uganda (U).
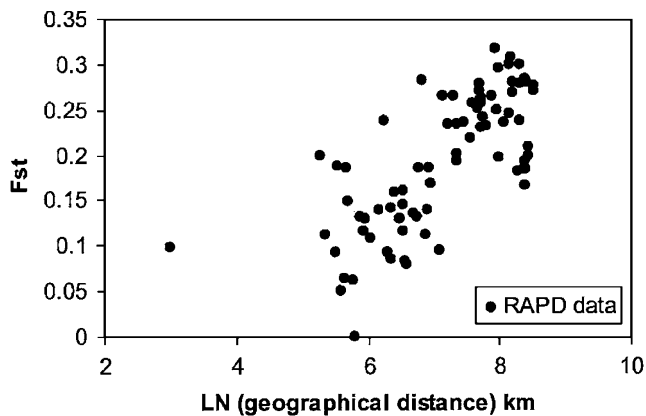

Figure 2 Relation between genetic and Neperian logarithm of geographical distances for the RAPD markers. Matrix of genetic distances was calculated using the AMOVA-derived F_{st}.

molecular variances indicated that variance among the groups represented 7% ($P = 0.000$) of the total variation. The among-populations within-groups variance represented 15% ($P = 0.000$) and the among-individuals within-populations variance represented 78%. A Mantel test suggested that genetic distances between populations were correlated to geographic distances ($R = 0.88$, $P = 0.001$), and this relationship is illustrated in Figure 2.

Genetic diversity and structure with chloroplast microsatellite markers

The three chloroplast microsatellite primers assayed on the 116 individuals gave 10 different alleles: Ccmp3, three alleles; Ccmp5, three alleles; and Ntcp8, four alleles. The combination of the three loci and the 10 alleles gave seven chlorotypes (Table 3). Most of the populations showed a single haplotype. In the total population, the haplotypes A and B exhibited high frequencies, 34 and 38% respectively while others were smaller than 15%.

Haplotype diversity for the total population was equal to $He_{cp} = 0.71$ and ranged from 0.00 to 0.49 (Samecouta in Senegal) among the populations (Table 4). The effective number of haplotypes (ne) followed the same pattern (Table 4). The haplotypic diversity was not related to the size of the population.

The unrooted neighbour-joining tree for cpSSR exhibited a slightly different pattern from RAPD data (Figure 3). One cluster concerned the group of populations from Mali, Burkina Faso and Ivory Coast, the second concerned the population from Benin and Uganda, the third the population from Central Africa and Senegal and the fourth the population from Cameroon. This strong differentiation among populations was confirmed by the high value of structure parameter ($G_{st_{cp}} = 0.84$) and the analysis of molecular variance. The variation among groups represented 91% ($P = 0.0000$) of total variation, the variation among populations within group 3% ($P = 0.000$) and the variation between individuals within population 6% ($P = 0.000$).
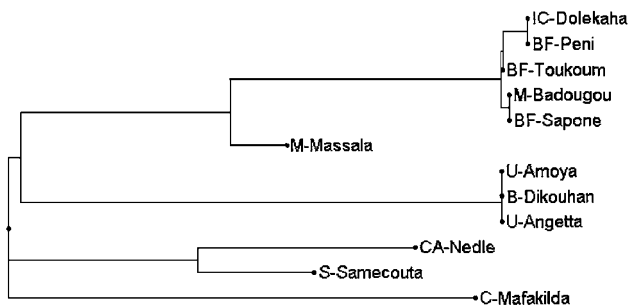
Table 3 Allelic characteristics and frequencies of the chlorotypes present in the 12 populations and in the total population of V. paradoxa

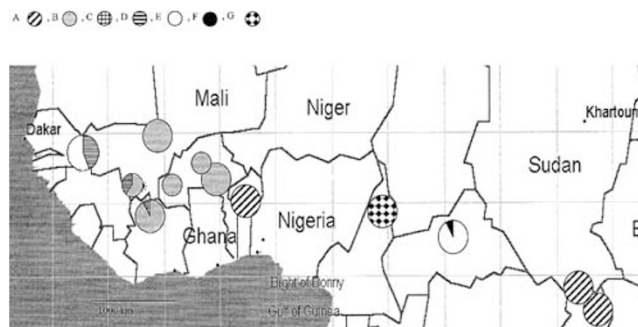|	Country	Population	Chlorotype	Ccmp3	Ccmp5	Ntcp8	Frequency
1	Benin	Dikouhan	A	114	112	244	1.00
2	Burkina	Sapone	B	113	112	243	1.00
		Peni	B	113	112	243	1.00
		Toukoum	B	113	112	243	1.00
3	Ivory Coast	Dolekaha	B	113	112	243	0.92
			C	113	112	242	0.08
4	Mali	Massala	B	113	112	243	1.00
		Badougou	B	113	112	243	0.60
			D	115	112	245	0.40
5	Senegal	Samecouta	D	115	112	245	0.43
			E	115	112	244	0.57
6	Central Africa	Ndele	E	115	112	244	0.92
			F	115	113	244	0.08
7	Cameroon	Mafakilda	G	115	106	242	1.00
8	Uganda	Amoya	A	114	112	244	1.00
		Angetta	A	114	112	244	1.00
Total population	A(0.34), B(0.38), C (0.01), D (0.04), E(0.14), F(0.01), G (0.08)						

Table 4 Population size (*N*), Shannon's index (I_{cp}), haplotypic diversity (He_{cp}) and effective number of haplotypes (ne)

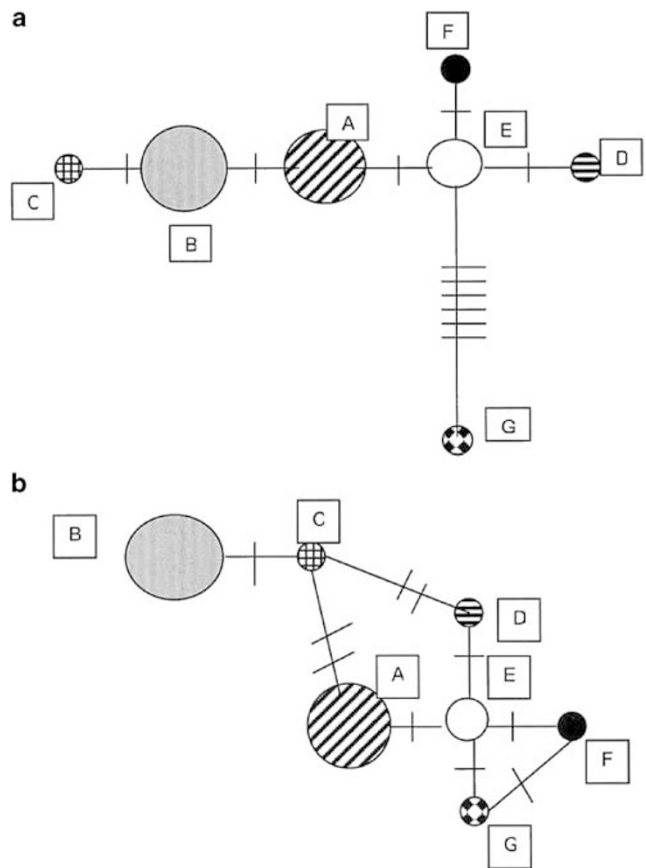| Country	Population	n	ne	He_{cp}	I_{cp}
Benin	Dikouhan	14	1.00	0.00	0.00
Burkina	Sapone	8	1.00	0.00	0.00
	Peni	5	1.00	0.00	0.00
	Toukoum	6	1.00	0.00	0.00
Ivory Coast	Dolekaha	13	1.17	0.14	0.27
Mali	Masala	10	1.00	0.00	0.00
	Badougou	5	1.92	0.48	0.67
Senegal	Samecouta	7	1.96	0.49	0.68
Central Africa	Ndele	13	1.17	0.14	0.27
Cameroon	Mafakilda	9	1.00	0.00	0.00
Uganda	Amoya	13	1.00	0.00	0.00
	Angetta	13	1.00	0.00	0.00
Total population		116	7.00	0.71	1.42

Neighbour-joining tree with microsatellite chloroplast markers

Figure 3 Unrooted neighbour-joining tree based on chloroplast microsatellite markers illustrating the phylogenetic relationships between the populations of Burkina Faso (BF), Ivory Coast (IC), Mali (M), Senegal (S), Benin (B), Cameroon (C), Central Africa (CA) and Uganda (U).

Figure 4 Geographic repartition of the seven chlorotypes in the natural range of *V. paradoxa*. The size of the circle is proportional to the size of the sample.

Figure 5 Minimum spanning network among the seven haplotypes found in the natural range of *V. paradoxa*. The size of the circle is proportional to the frequency of the haplotype in the total population. Connection lengths between haplotypes are represented by the number of marks. (**a**) Minimum spanning tree using a distance matrix based on the square difference of base pairs between haplotypes (connection length between E and G is 36). (**b**) Minimum spanning network using a distance matrix based on number of different alleles between haplotypes.

Phylogenetic relation and geographic distribution of the haplotypes

The distribution of the seven haplotypes across the natural range shows a complex pattern (Figure 4). Haplotypes A and E were scattered among distant populations (A was present in Uganda and Benin, E in Central Africa and Senegal), haplotypes B and D were present in various populations but in the same region (Burkina Faso, North Ivory Coast and Mali), and C, F, and G were restricted to a single population. Haplotype relatedness represented by MINSPNET programme (Figure 5a and b) does not identify clear groups. When using the square difference in size matrix (Figure 5a), the

minimum spanning tree shows that E occupied a central position, the other haplotypes except for G being connected by one mutational difference. G was connected to E by 6 bp. The minimum spanning network using a distance matrix based on number of different alleles between haplotypes (Figure 5b) presents a more complex network, exhibiting new connections between haplotypes such as between G and F, and C and D.

Discussion

Genetic diversity of *V. paradoxa*
Although comparisons are not always straightforward with RAPDs, the genetic diversity parameters exhibit values comparable to other species distributed across a wide continuous range (Bekessy *et al*, 2002; Newton *et al*, 2002).

There are few published studies concerning the diversity of chloroplast microsatellite markers in angiosperm tree species. These few studies show results that vary according to the species and the number of primers used (Table 5). With three primers, 116 individuals distributed in 12 populations, the seven haplotypes identified in *V. paradoxa* are consistent with the previous studies on angiosperms. The haplotypic heterozygosity $He_{cp} = 0.71$ is high compared to other tree species (Palmé and Verdramin, 2002). This suggests that *V. paradoxa* populations may not have experienced a recent bottleneck, like those proposed for species showing a low level of variation in expanding populations (Palmé and Verdramin, 2002; Grivet and Petit, 2003).

Genetic structure and gene flow
Hamrick *et al* (1991) suggest that species that are long-lived woody perennials, out-crossing insect pollinated and widespread are likely to have high within-population variability, all of which are characteristics typical of *V. paradoxa*. Its status of being a socially and economically important species may also explain the high RAPD variability within populations. Protection afforded to shea trees during traditional long-term management in parklands (from fire, competition, etc) would be expected to allow diversity to increase, particularly if coupled with amplified gene flow by transportation of fruits from village to village by humans or other animals such as elephants, birds, bats or primates (Lovett and Haq, 2000b; Maranz and Wiesman, 2003). Although cases of deliberate human planting are rare and seeds of the Shea tree are recalcitrant, as they do not maintain their viability during long transportation, models of population genetics have shown that a small number of seeds could be sufficient to reduce genetic differentiation.

The genetic distance between populations obtained with RAPDs (F_{st} estimated with AMOVA) is correlated to the geographic distance. Such correlations are often observed in tree populations over long distances (Bekessy *et al*, 2002), even though no significant correlation is detected over short distances (Schierenbeck *et al*, 1997).

The pattern of diversity is very different for chloroplast microsatellites. We note a strong differentiation between populations or between groups, which is a classical result for forest trees especially angiosperms (Raspé *et al*, 2000), because of the maternal inheritance of the chloroplast DNA and because seeds are dispersed over shorter distances than with pollen. In *V. paradoxa*, although humans or animals may disperse seeds over long distances, seed dispersal is mainly barochore. There is no relationship between genetic distance and geographic distance for cpSSR (Figure 6). The estimates of r indicate that gene flow by pollen is 47 times more important than by seeds. This value is smaller than estimates from temperate species such as *Fagus silvatica* (84), *Quercus robur* (286), *Quercus petraea* (500) although this is expected since these are wind pollinated (King and Ferris, 1998).

Marker distribution and historical factors
The use of the two markers allows for a parallel approach to interpret the phylogeography pattern of *V. paradoxa*. With RAPDs, we observe genetic isolation by distance. Although variation between groups was low (7%), neighbour-joining trees obtained with RAPDs differentiate the clusters of western and eastern populations (Figure 1). A phylogenetic tree based on haplotypes (Figure 5) and their distribution across the natural range (Figure 4) do not conform to a simple pattern of migration, and various scenarios can be proposed. Although haplotype E occupied a central position in the phylogenetic trees (Figure 5a and b) and is present in two locations of the natural range (Central Africa and Senegal), it is not likely to be the ancestor of most of the other chlorotypes. The populations where it has been observed are not the putative refugia of the last glaciation (18 000 BP) proposed by Maley (1996). Northern areas of Central Africa and Eastern Senegal are actually very dry regions today, and therefore are expected to have been more arid during the last glacial maximum, and unlikely to have corresponded to appropriate ecological conditions for *V. paradoxa*. Another scenario could be proposed based on a combinational RAPDs and the phylogenetic tree of haplotypes (Figure 5b). The phylogenetic lineage shown in Figure 5b and the close relationship of the populations

Table 5 Number of haplotypes detected according to the number of populations, number of individuals and number of primers tested in wood and nonwood angiosperm species

Species	Populations	Indiv.	Primers	Haplotypes	Reference
Carpinus betulus[a]	41	322	4 Ccmp	7	Grivet and Petit (2003)
Caryocar brasliiense[a]	10	160	10 Ccmp	21	Colevatti *et al* (2003)
Silene paradoxa[b]	8	96	5 Ccmp	27	Mengoni *et al* (2001)
Alyssum bertolonii[b]	9	90	10 Ccmp	17	Mengoni *et al* (2003)

[a]Forest tree species.
[b]Nonwood species.

based on RAPDs (Figure 1) suggest that western populations are closely related to each other and could result from a migration pattern different from that of the eastern populations. The presence of haplotype F in Senegal and Central Africa, which are about 3000 km apart, could then be explained by homoplasy in size, which is plausible in the case of highly variable stepwise mutating markers such as microsatellites (Estoup *et al*, 2002).

The separation between western and eastern populations could result from refugia separated by the current 'Dahomey Gap', an area previously identified as an important biogeographic barrier (Jeník, 1994). When the climate conditions become very dry during the last glacial maximum (Maley, 1996), the Dahomey Gap would have become an extremely arid area not compatible with shea tree ecology. This zone, which is 300 km wide today, is expected to have occupied a far larger area during drier phases and to have separated the forest refuges of southwest Ghana and west Cameroon by at least 1200 km (Figure 7). In far wetter periods, the shea ecological zone w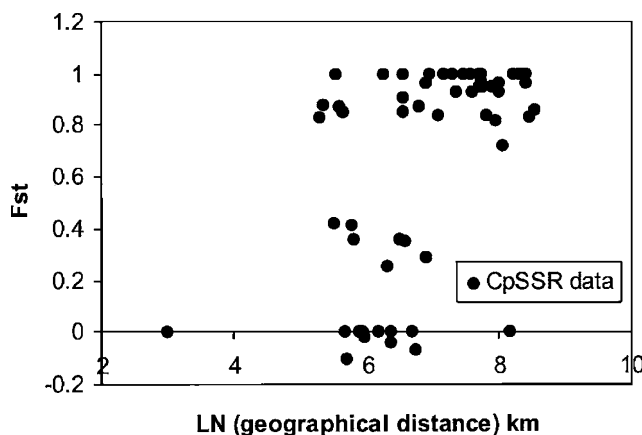ould have been shifted northwards into the Sahara and many areas would have lost shea in the south after it was outcompeted by wetter plant species. Towards the east, however, the shea ecological zone is currently narrower and the nature of the environment results in vast distances that the species would have to colonise in order to reach a potential refugia. It is therefore suggested that the species is more likely to have been eliminated completely from many of these areas, either by severe droughts or by wet periods that allowed rapid expansion of forest across the shea zone. These narrow parts of the shea zone, for example, between the Chad catchment area and the Nile valley, may have also caused population bottlenecks. The chlorotype G, which differs in size by 6 bp, could result from an isolated refugia in Cameroon.

On the western part of the range, more recent factors could in part explain the distribution of haplotype B and its high frequency. A number of studies have demonstrated the importance of traditional management systems for the abundance and diversity of *V. paradoxa* in the agroforestry parklands of West Africa (Lovett and Haq, 2000a) and archaeological evidence demonstrates that these systems for at least 1000 years old (Neumann *et al*, 1998). 'Unconscious selection' of superior landraces (high yields, early abscission, improved fruit palatability, etc) has been proposed during the cyclical farm-fallow system when unwanted individual shea trees are removed and intensive fruit harvesting occurs during annual crop cultivation periods (Lovett, 2000). According to Maranz and Wiesman (2003), other more active selection mechanisms are possible and they suggest that the prevalence of *V. paradoxa* north of the equator indicates pronounced and long-term human involvement in tree dispersal. They have compared the present-day distribution with historical range limits from 200-year-old records and demonstrate that range expansion by human migration has occurred. Data from their chemical analyses also indicate that the selection of desired fruit traits (kernel fat content, fat hardness and pulp sweetness) is occurring. Selection and migration could favour one specific haplotype if there is a linkage disequilibrium between favourable genes and the amplified chloroplast DNA region used for this study.
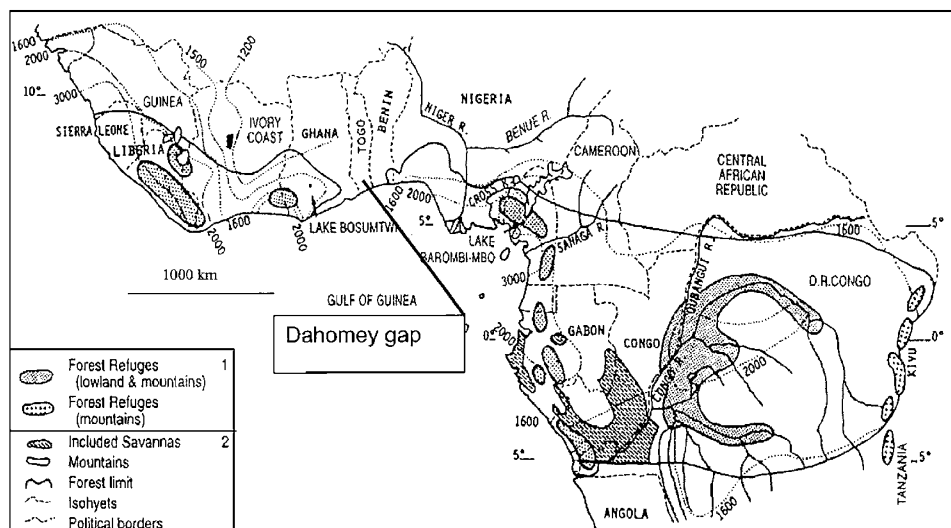

Figure 6 Relation between genetic distances and log geographical distances for the chloroplastic microsatellite markers. Matrix of genetic distances was calculated using the AMOVA-derived F_{st}.

Figure 7 Schematic map of the African forest refuges during the last glaciation maximum arid phase (around 18 000BP) (from Maley, 1996).

Burkina Faso and Mali, where haplotype B is very common, are among the regions where human selection and fruit transportation were and are still thought to be very pronounced.

In conclusion, the 'Dahomey Gap' may have been a factor leading to separation between western and eastern populations and human activity may have affected the structure in the western part of the range. Although further research is needed, these initial results would enable any programme for the conservation of genetic resources of this species.

Acknowledgements

References

Aide TM, Rivera E (1998). Geographic patterns of genetic diversity in *Poulsenia amata* (Moraceae): implications for the theory of Pleistocene refugia and the importance of riparian forest. *J Biogeogr* **25**: 695–705.

Bekessy SA, Allnutt TR, Premoli AC, Lara A, Ennos RA, Burgman MA *et al* (2002). Genetic variation in the vulnerable and endemic Monkey Puzzle tree, detected using RAPDs. *Heredity* **88**: 243–249.

Bryan GJ, McNicoll J, Ramsay G, Meyer RC, De Jong WS (1999). Polymorphic simple sequence repeat markers in chloroplast genomes of solanaceous plants. *Theor Appl Genet* **99**: 859–867.

Collevatti RG, Grattapaglia D, Hay JD (2003). Evidence for multiple lineages of Caryocar brasiliense populations in the Brazilian Cerrado based on the analysis of chloroplast DNA sequence and microsatellite haplotype variation. *Mol Ecol* **12**: 105–115.

Ennos RA (1994). Estimating the relative rates of pollen and seed migration among plant populations. *Heredity* **72**: 250–259.

Estoup A, Jarne P, Cornuet JM (2002). Homoplasy and mutation model at microstallite loci and their consequences for population genetics analysis. *Mol Ecol* **11**: 1591–1604.

Excoffier L, Smouse P (1994). Using allele frequencies and geographic subdivision to reconstruct gene trees within a species: molecular variance parsimony. *Genetics* **136**: 343–359.

Grivet D, Petit R (2002). Phylogeography of the common ivy (Hedera sp) in Europe: genetic differentiation through space and time. *Mol Ecol* **11**: 1351–1362.

Grivet D, Petit R (2003). Chloroplast DNA phylogeography of the hornbeam in Europe: evidence for the bottleneck at the outset of the postglacial colonisation. *Conserv Genet* **4**: 47–56.

Hall JB, Aebischer DP, Tomlinson HF, Osei-Amaning E, Hindle JR (1996). *Vitellaria paradoxa*. A Monograph School of Agricultural Sciences Publication Number 8. University of Wales: Bangor. 105pp.

Hamrick JL, Godt MJW, Murawski DA, Loveless MD (1991). Correlations between species traits and allozymes diversity: implications for conservation biology. In: Falk D, Holsinger K (eds) *Genetics and Conservation of Rare Plants*. Oxford University Press: New York, pp 75–86.

Hardy O, Vekemans X (2002). *Freeware program for spatial pattern analysis of genetic diversity*. Université Libre de Bruxelles.

Ishii T, McCouch SR (2000). Microsatellites and microsynteny in the chloroplast genomes of Oryza and eight other Graminae species. *Theor Appl Genet* **100**: 1257–1266.

Jenik J (1994). The Dahomey Gap, an important issue in West African phytogeography. *Mém Soc Biogéogr* **4**: 125–133.

Kelly BA, Hardy O, Bouvet J-M (2004). Temporal and spatial genetic structure of *Vitellaria paradoxa* (shea tree) in agroforestry system in southern Mali. *Mol Ecol* **13**: 1231–1240.

King RA, Ferris C (1998). Chloroplast DNA phylogeography of *Alnus glutinosa* (L) Gaertn. *Mol Ecol* **7**: 1151–1161.

Lézine A-M, Vergnaud-Grazzini G (1994). Evidence of forest extension in West Africa since 22,000 BP: a pollen record from the eastern tropical Atlantic. *Quaternary Sci Rev* **12**: 203–210.

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

Lovett PN (2000). The Genetic Diversity of the Sheanut tree (*Vitellaria paradoxa*) in the Farming Systems of Northern Ghana. PhD Thesis, University of Southampton. 187pp.

Lovett PN, Haq N (2000a). Evidence for anthropic selection of the sheanut tree (*Vitellaria paradoxa*). *Agroforestry Systems* **48**: 273–278.

Lovett PN, Haq N (2000b). Diversity of the sheanut tree (*Vitellaria paradoxa* CF Gaertn) in Ghana. *Genet Resour Crop Evol* **47**: 293–304.

Maley J (1991). The African Rain Forest vegetation and palaeoenvironments during late Quaternary. *Climatic Change* **19**: 79–98.

Maley J (1996). The African rain forest: main characteristics of changes in vegetation and climate from the upper Cretaceous to the Quaternary. *Proc R Soc Edinburgh* **104B**: 31–73.

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**: 1505–1516.

Martin JP, Hernandez Bermejo JE (2000). Genetic variation in the endemic and endangered *Rosmarinus tomentosus* Huber-Morath & Maire (Labiatae) using RAPD markers. *Heredity* **85**: 434–443.

Mengoni A, Barabesi C, Gonnelli C, Galardi F, Gabbrielli R, Bazziacalpo M (2001). Genetic diversity of heavy metal tolerant populations in *Silene paradoxa* L (Caryophyllaceae): chloroplast microsatellite analysis. *Mol Ecol* **10**: 1909–1916.

Mengoni A, Gonnelli C, Brocchini E, Galardi F, Pucci S, Gabbrielli R *et al* (2003). Chloroplast genetic diversity and biogeography in the serpentine endemic Ni-hyper-accumulator *Alyssum bertolonii*. *New Physiologist* **157**: 349–356.

Mohanty A, Martin JP, Aguinagalde I (2000). Chloroplast DNA diversity within and among populations of the allotetraploid *Prunus spinosa* L. *Theor Appl Genet* **100**: 1304–1310.

Muloko-Ntoutoume N, Petit RJ, White L, Abernethys K (2000). Chloroplast DNA variation in a rainforest tree (*Aucoumea klaineana*, Burseraceae) in Gabon. *Mol Ecol* **9**: 359–363.

Neumann K, Kahlheber S, Uebel D (1998). Remains of woody plants from Saouga, a medieval West African village. *Veg Hist Archaeobot* **7**: 57–77.

Newton AC, Allnutt TR, Dvorak WS, Del Castillo RF, Ennos RA (2002). Patterns of genetic variation in *Pinus chiapensis*, a threatened Mexican pine, detected by RAPD and mitochondrial DNA RFLP markers. *Heredity* **89**: 191–198.

648

Päivi H (2000). Genetic Basis of Adaptation: Bud Set Date and Frost Hardiness Variation in Scots Pine. Academic Dissertation, University of Oulu, Oulu, Finland Oulu University Library, 26pp.

Palme AE, Verdramin GG (2002). Chloroplast DNA variation, postglacial recolonization and hybridation in hazel, *Corylus avellana*. *Mol Ecol* **11**: 1769–1779.

Perrier X, Flori A, Bonnot F (2003). Data analysis methods. In: Hamon P, Seguin M, Perrier X, Glaszmann, JC (eds) *Genetic Diversity of Cultivated Tropical Plants*. Enfield Science Publishers, Montpellier, pp 43–76.

Raspé O, Saumitou-Laprade P, Cuguen J, Jacquemart A-L (2000). Chloroplast DNA haplotype variation and population differentiation in *Sorbus aucuparia* L (Rosaceae: Maloideae). *Mol Ecol* **9**: 1113–1122.

Raymond M, Rousset F (1995). *Genepop version 12*: population genetic software for exact test and eucumenism. *J Hered* **86**: 248–249.

Salzmann U, Waller M (1998). The Holocene vegetational history of Nigerian Sahel based on multiple pollen profiles. *Rev Paleobot Palynol* **100**: 39–72.

Schierenbeck KA, Skupski M, Lieberman D, Lieberman M (1997). Population structure and genetic diversity in four tropical tree species in Costa Rica. *Mol Ecol* **6**: 137–144.

Schneider S, Roessli D, Excoffier L (2000). *Arlequin version 2000: a software for population genetics data analysis*. Genetics and Biometrics Laboratory, Department of Anthropology, University of Geneva, Geneva, Switzerland.

Vendramin GG, Anzidei M, Madaghiele A, Sperisen C, Bucci G (1998). Distribution of genetic diversity in *Pinus pinaster* Ait as revealed by chloroplast microsatellites. *Theor Appl Genet* **97**: 456–463.

Viard F, El-Kassaby YA, Ritland K (2001). Diversity and genetic structure in populations of *Pseudotsuga menziesii* (Pinaceae) at chloroplast microsatellite loci. *Genome* **44**: 336–344.

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

Willis KJ, Whittaker RJ (2000). The refugial debate. *Science* **287**: 1406–1407.

Wu J, Krutovskii KV, Strauss SH (1999). Nuclear DNA diversity, population differenciation and phylogenetic relatioships in the California closed-cone pines based on RAPD and allozyme markers. *Genome* **42**: 893–908.

Yeh FC, Boyle TJB (1997). Population genetic analysis of co-dominant and dominant markers and quantitative traits. *Belg J Bot* **129**: 157.