

ORIGINAL ARTICLE

Past climate changes explain the phylogeography of *Vitellaria paradoxa* over AfricaF Allal¹, H Sanou², L Millet¹, A Vaillant¹, L Camus-Kulandaivelu¹, ZA Logossa^{1,3}, F Lefèvre⁴ and J-M Bouvet¹¹Cirad-Bios Department, Research Unit 'Genetic Diversity and Breeding of Forest Tree Species', International Campus of Baillarguet, Montpellier, France; ²Institut d'Economie Rurale, Forest Resource Programme, Regional Agronomical Research Centre of Sotuba, Bamako, Mali; ³University of Lomé, Lomé, Togo and ⁴INRA, URFM, Ecologie des Forêts Méditerranéennes, Avignon, France

The evolution of the savanna biome has been deeply marked by repeated contraction/expansion phases due to climate perturbations during the Quaternary period. In this study, we investigated the impact of the last glacial maximum (LGM) on the present genetic pattern of *Vitellaria paradoxa* (shea tree), a major African savanna tree. A range-wide sampling of the species enabled us to sample 374 individuals from 71 populations distributed throughout sub-Saharan Africa. Trees were genotyped using 3 chloroplasts and 12 nuclear microsatellites, and were sequenced for 2 polymorphic chloroplast intergenic spacers. Analyses of genetic diversity and structure were based on frequency-based and Bayesian methods. Potential distributions of *V. paradoxa* at present, during the LGM and the last interglacial period, were examined using DIVA-GIS ecological niche modelling (ENM). Haplotypic and allelic richness varied significantly across the range

according to chloroplast and nuclear microsatellites, which pointed to higher diversity in West Africa. A high but contrasted level of differentiation was revealed among populations with a clear phylogeographic signal, with both nuclear ($F_{ST} = 0.21$; $R_{ST} = 0.28$; $R_{ST} > R_{ST}$ (permuted)) and chloroplast simple sequence repeats (SSRs) ($G_{ST} = 0.81$; $N_{ST} = 0.90$; $N_{ST} > N_{ST}$ (permuted)). We identified a strong geographically related structure separating western and eastern populations, and a substructure in the eastern part of the area consistent with subspecies distinction. Using ENM, we deduced that perturbations during the LGM fragmented the potential eastern distribution of shea tree, but not its distribution in West Africa. Our main results suggest that climate variations are the major factor explaining the genetic pattern of *V. paradoxa*. *Heredity* (2011) **107**, 174–186; doi:10.1038/hdy.2011.5; published online 16 March 2011

Keywords: intergenic spacers; microsatellite markers; last glacial maximum; phylogeography; sub-Saharan Africa; *Vitellaria paradoxa*

Introduction

Today the savanna biome occupies a fifth of the earth's land surface, and is characterized by a mixture of grasslands and woodlands (Sankaran *et al.*, 2005). It dates from the late Miocene, about eight million years ago (Cerling *et al.*, 1997; Beerling and Osborne, 2006). During the Quaternary, the Pleistocene epoch (from 2.5 million to 11 000 years ago) underwent successive cooling and warming periods. During the last glacial maximum (LGM), around 20 000 years ago (Clark *et al.*, 2009), 30% of the Earth's surface was covered by ice, the African continent was drier and the Sahara desert was bigger than today (Lioubimtseva *et al.*, 1998). Although many phylogeographic studies have reconstructed the history of holarctic and rain forest species distributions, identifying refugia and routes of postglacial colonization (Bekessy *et al.*, 2002; Petit *et al.*, 2003, 2008; Jaramillo-Correa *et al.*, 2008; Dick, 2010), little attention has been

paid to the ecological consequences of these perturbations for savanna ecosystems. Up to now, most studies on this biome have concentrated on animal and plant species in the South American Cerrado, and did not identify a general trend in the impacts of perturbations that occurred during the Quaternary. Confirming paleopalynological studies (Behling, 2002), some savanna species showed evidence of range expansions during the LGM (Bonaccorso *et al.*, 2006), while others showed evidence of fragmentation (Bonaccorso *et al.*, 2006; Ramos *et al.*, 2007, 2009). Cumulating evidence indicates that plant and animal populations in the African rain forest underwent historical genetic contractions and bottlenecks, surviving in putative refugia during the LGM (Dauby *et al.*, 2010; Lowe *et al.*, 2010), and expanded during postglacial periods to colonize wider areas (Maley, 1989, 1996). In contrast, on the basis of data concerning the Pliocene glacial event 2.5 million years ago (Poumot, 1989; Morley and Richards, 1993), Maley (1996) proposes that savanna species underwent similar expansions during the cold and arid periods of the Quaternary (Flenley, 1998). To test these assumptions and to help understand the impact of the LGM on African savanna tree species, we analyzed the range-wide phylogeography of *Vitellaria paradoxa* C.F. Gaertn (commonly known as the shea tree), an IUCN Red List

Correspondence: Dr F Allal, Cirad-Bios Department, Research Unit 'Genetic Diversity and Breeding of Forest Tree Species', International Campus of Baillarguet TA-A 39/C, BP 5035, 34398 Montpellier Cedex 5, France.
E-mail: francois.allal@cirad.fr; francois.allal@gmail.com
Received 14 July 2010; revised 2 December 2010; accepted 13 January 2011; published online 16 March 2011

vulnerable but economically important endemic African savanna tree species exploited in agroforestry systems. Mostly reproducing sexually and mainly pollinated by insects, shea tree fruits are mainly disseminated barochorously and secondarily zoochorously (by birds, monkeys, rodents and humans). Its natural range is a 6000-km-wide belt through sub-Saharan Africa covering a contrasted ecological environment. *V. paradoxa* is currently subdivided into two non-overlapping subspecies, *paradoxa* (from Senegal to Central African Republic) and *nilotica* (in Sudan and Uganda). A gap of 175 km between the two subspecies is reported because of the different watersheds in the area: the Lake Chad basin is home to *paradoxa* and the Nile drainage basins are home to *nilotica* (Hall *et al.*, 1996). However, this distinction is based on slight variations in a few morphological traits (the length and width of leaves and stamina) compiled from different studies (Hall *et al.*, 1996), and no data are available concerning their natural or controlled hybridization. Therefore, *nilotica* and *paradoxa* are referred as 'putative' subspecies throughout this article. The 'butter', extracted from its kernels, is the main source of edible vegetable fat in the region. This oil, exported throughout Africa and used in the food, confectionary, cosmetic and pharmaceutical industries (Lovett and Haq, 2000), displays a clear trend, becoming increasingly harder from east to west due to variable ratios of saturated to unsaturated fatty acids (Davrieux *et al.*, 2010).

An initial evaluation of the genetic diversity of *V. paradoxa* (Fontaine *et al.*, 2004) suggested a genetic differentiation between eastern and western Africa in shaping the genetic structure of the species by emphasizing the role of the Dahomey Gap, a north to south savanna corridor that appeared in the late Holocene (~3000 years ago) in Togo and Benin, disrupting the West African rain forest cover (Maley, 1996). However, this study was performed on a very small sample using RAPD markers, and the results thus need to be considered with caution (Bagley *et al.*, 2001).

To be able to draw general conclusions on the phylogeography of this emblematic African tree, we based our study on a representative sample of the species range using a combination of molecular markers to infer different evolutionary processes. To this end, we used sequence polymorphism of intergenic chloroplast fragments from 58 individuals from 40 populations; and length variation of 3 chloroplast and 12 nuclear microsatellites genotyped on 364 trees from 71 populations. We used climatic models to assess the distribution of the species in three key periods in the Quaternary era: the modern period, the LGM, and the last interglacial (LIG). Using this experimental design, we tried to answer the following questions: (i) what are the impacts of past perturbations on the genetic structure of the species? (ii) Can we propose a reliable theory concerning the origin of current populations, with special focus on the differentiation process between the subspecies *paradoxa* and *nilotica*?

Materials and methods

Plant material

In this study, a total of 364 trees were sampled at 71 sites in 14 countries, covering the whole distribution range of

the two subspecies: Senegal, Mali, Ivory Coast, Ghana, Burkina-Faso, Togo, Benin, Niger, Nigeria, Cameroon, Chad and Central African Republic for *paradoxa*, and Sudan and Uganda for *nilotica*. A subset of samples including 58 individuals from 40 populations was used for the sequence-based chloroplast experiments. Spatial coordinates of the populations are given in Table 1. Total genomic DNA was extracted from dried leaves (50 mg) following the Dolezel method (Ky *et al.*, 2000).

Molecular markers

Genotyping of chloroplast sequences: We successfully amplified and sequenced 11 chloroplast loci on *V. paradoxa*, including exclusively non-coding intronic or intergenic regions. Universal primer pairs and annealing temperatures are summarized in Supplementary Appendix S1 (Supplementary information). The polymorphism of these loci was assessed in a subset of 10 individuals belonging to 10 different populations, including *nilotica* and *paradoxa* subspecies (Table 1). Among these 11 chloroplast loci, which correspond to 7 kb of chloroplast genome, only 2 loci presented nucleotide polymorphism: *trnQ-trnS* (Dumolin-Lapegue *et al.*, 1997) and *matK-trnK* spacers (Johnson and Soltis, 1994). These polymorphic loci were sequenced using the above-mentioned subsample (Table 1). Polymerase chain reaction (PCR) was performed in a total of 20 μ l, consisting in 20 ng of total DNA, 10 μ l of standard QIAGEN PCR buffer, 100 μ M of dNTP, 0.2 μ M of each primer and 1 U of Taq DNA polymerase (QIAGEN, Courtaboeuf, France). Amplifications were performed in a MasterCycler Eppendorf thermocycler (Hamburg, Germany) with an initial 4-min denaturation step at 94 °C, followed by 30 cycles (30 s of denaturation at 94 °C, 1 min at 61.7 °C for *matK-trnK* and at 68 °C for *trnQ-trnS*, 1 min of extension at 72 °C), then a final extension step at 72 °C for 5 min for *matK-trnK* or for 10 min for *trnQ-trnS*. PCR products were purified and sequenced by GATC Biotech (Konstanz, Germany). Using a similar protocol, we amplified and sequenced *trnQ-trnS* and *matK-trnK* from five individuals of *Baillonella toxisperma* Pierre, a relative of the Sapotaceae family, used as outgroup species.

Genotyping of chloroplast SSRs: Among the 10 microsatellite primer pairs tested, 9 showed PCR amplification on our samples: 8 universal microsatellites, Ccmp1, Ccmp2, Ccmp3, Ccmp4, Ccmp5, Ccmp6, Ccmp7 and Ccmp9 (Weising and Gardner, 1999), and 1 tobacco microsatellite, Ntcp9 (Bryan *et al.*, 1999). For determination of polymorphism, these nine microsatellite markers were tested on 10 individuals covering the whole distribution range (Table 1). Three chloroplast microsatellites (cpSSR) were polymorphic (Ccmp3, Ccmp5 and Ntcp9) and were chosen to lead the global analysis of the 364 individuals from 71 populations. The protocol used for amplification and genotyping is described in Müller *et al.* (2009).

Genotyping of nuclear SSRs: Among the 16 polymorphic nuclear SSR markers (nucSSRs), 4 were discarded as many genotypes were missing. Finally, 12 polymorphic nuclear SSR markers (Allal *et al.*, 2008) were used (mCIRVp08, mCIRVp14, mCIRVp28, mCIRVp70, mCIRVp113, mCIRVp134, mCIRVp159, mCIRVp167, mCIRVp168, mCIRVp175, mCIRVp181 and mCIRVp191). PCR

Table 1 Characterization of sampled populations

Region	Country	Site	ID	Longitude	Latitude	Subsp. species	Ncp sequences	NcpSSRs	NnucSSRs	
West Africa	Senegal	Bandafassi	1	W 12.32	N 12.53	ssp. <i>paradoxa</i>	2	5	5	
		Samecouata	2	W 12.13	N 12.60	ssp. <i>paradoxa</i>	1	4	4	
	Mali	Mahina	3	W 10.85	N 13.78	ssp. <i>paradoxa</i>	1	5	5	
		Torodo	4	W 8.84	N 14.50	ssp. <i>paradoxa</i>		4	4	
		Badougou	5	W 8.32	N 11.26	ssp. <i>paradoxa</i>	1	4	4	
		Kodougouni	6	W 6.66	N 12.97	ssp. <i>paradoxa</i>	1	4	4	
		Mperesso ^a	7	W 5.46	N 12.31	ssp. <i>paradoxa</i>	1	6	6	
		Songo	8	W 3.70	N 14.40	ssp. <i>paradoxa</i>	1	2	2	
		Tenndeli	9	W 3.45	N 14.09	ssp. <i>paradoxa</i>	2	4	4	
	Ivory Coast	Napieoledougou	10	W 5.57	N 9.28	ssp. <i>paradoxa</i>	2	4	4	
		Korhogo	11	W 4.39	N 9.62	ssp. <i>paradoxa</i>	1	2	2	
	Burkina-Faso	Peni	12	W 4.45	N 10.97	ssp. <i>paradoxa</i>		6	6	
		Karo	13	W 3.30	N 12.36	ssp. <i>paradoxa</i>	1	6	6	
		Bonogo ^a	14	W 1.60	N 12.06	ssp. <i>paradoxa</i>	1	6	6	
		Pô	15	W 1.15	N 11.18	ssp. <i>paradoxa</i>	1	6	6	
		Konia	16	W 0.99	N 13.10	ssp. <i>paradoxa</i>	1	6	6	
		Fada N'Gourma	17	E 0.41	N 12.06	ssp. <i>paradoxa</i>	1	6	6	
	Ghana	Larabanga	18	W 1.85	N 9.19	ssp. <i>paradoxa</i>	2	6	6	
		Kawampe ^a	19	W 1.56	N 8.43	ssp. <i>paradoxa</i>	1	6	6	
		Kadia	20	W 0.86	N 9.90	ssp. <i>paradoxa</i>	2	6	6	
	Togo	Yendi	21	W 0.04	N 9.45	ssp. <i>paradoxa</i>	2	6	6	
		Tabango	22	E 0.20	N 10.73	ssp. <i>paradoxa</i>		4	4	
		Payô	23	E 0.69	N 10.23	ssp. <i>paradoxa</i>	1	4	4	
		Koudassi	24	E 0.86	N 6.64	ssp. <i>paradoxa</i>		6	6	
		Kante ^a	25	E 1.03	N 9.98	ssp. <i>paradoxa</i>	1	6	6	
	Benin-Niger	Djangbassou	Kolina	26	E 1.14	N 9.10	ssp. <i>paradoxa</i>	1	6	6
			Djangbassou	27	E 1.19	N 7.81	ssp. <i>paradoxa</i>		6	6
		Cocota ^a	Cocota ^a	28	E 1.36	N 10.41	ssp. <i>paradoxa</i>	1	6	6
			Kpale	29	E 1.73	N 9.65	ssp. <i>paradoxa</i>	1	4	4
		Pira	Pira	30	E 1.78	N 8.48	ssp. <i>paradoxa</i>	1	6	6
			Kandi	31	E 2.95	N 11.16	ssp. <i>paradoxa</i>	1	6	6
			Kalale	32	E 3.36	N 10.27	ssp. <i>paradoxa</i>	2	6	6
		Nigeria	Boumba	33	E 2.86	N 12.41	ssp. <i>paradoxa</i>		2	2
			Wawa	34	E 4.42	N 9.90	ssp. <i>paradoxa</i>		10	10
			Rijau	35	E 5.25	N 11.10	ssp. <i>paradoxa</i>		10	10
	Paiko		36	E 6.60	N 9.43	ssp. <i>paradoxa</i>		11	11	
	Abuja		37	E 7.42	N 9.08	ssp. <i>paradoxa</i>		10	10	
	Kuru		38	E 8.81	N 9.68	ssp. <i>paradoxa</i>		6	6	
	Shendam		39	E 9.53	N 8.88	ssp. <i>paradoxa</i>		6	6	
	Kaltungo		40	E 11.32	N 9.82	ssp. <i>paradoxa</i>		6	6	
	Central Africa	Cameroon	Mubi	41	E 13.27	N 10.27	ssp. <i>paradoxa</i>		6	6
Bangangte			42	E 10.52	N 5.15	ssp. <i>paradoxa</i>	2	6	6	
Foumban			43	E 10.90	N 4.73	ssp. <i>paradoxa</i>	2	6	6	
Ngong			44	E 13.50	N 9.02	ssp. <i>paradoxa</i>		5	5	
Ngaoundere			45	E 13.58	N 7.32	ssp. <i>paradoxa</i>		4	4	
Mafa-Kilda ^a			46	E 13.75	N 9.03	ssp. <i>paradoxa</i>	2	5	5	
Figuil			47	E 13.95	N 9.75	ssp. <i>paradoxa</i>		5	5	
Chad		Mendeo	48	E 14.43	N 10.37	ssp. <i>paradoxa</i>		4	4	
		Kelo	49	E 15.78	N 9.20	ssp. <i>paradoxa</i>	1	5	5	
		Mondou	50	E 16.02	N 8.62	ssp. <i>paradoxa</i>	1	5	5	
		Sarh ^a	51	E 18.37	N 9.17	ssp. <i>paradoxa</i>	1	4	4	
Central African Republic		North	52	E 20.45	N 8.58	ssp. <i>paradoxa</i>	2	3	3	
		Central ^a	53	E 20.59	N 8.22	ssp. <i>paradoxa</i>	2	6	6	
East Africa		Sudan	South	54	E 20.85	N 7.88	ssp. <i>paradoxa</i>	3	6	6
			Kuajena	55	E 28.25	N 7.52	ssp. <i>nilotica</i>		5	5
	Cuebet		56	E 29.21	N 7.03	ssp. <i>nilotica</i>		5	5	
	Wulu		57	E 29.63	N 6.49	ssp. <i>nilotica</i>		3	3	
	Kuel Kuac		58	E 29.64	N 6.58	ssp. <i>nilotica</i>		2	2	
	Matengai		59	E 29.64	N 6.83	ssp. <i>nilotica</i>		5	5	
	Mbara		60	E 30.17	N 5.45	ssp. <i>nilotica</i>		2	2	
	Uganda	Mundri	61	E 30.32	N 5.34	ssp. <i>nilotica</i>		4	4	
		Arua	62	E 30.90	N 3.03	ssp. <i>nilotica</i>		4	4	
		Eria	63	E 31.64	N 3.62	ssp. <i>nilotica</i>		4	4	
		Adumara	64	E 33.05	N 2.53	ssp. <i>nilotica</i>	2	4	4	
		Otongere ^a	65	E 33.15	N 2.72	ssp. <i>nilotica</i>	2	4	4	
		Okwongo	66	E 33.25	N 2.50	ssp. <i>nilotica</i>	2	5	5	
		Lacankweri	67	E 33.31	N 2.80	ssp. <i>nilotica</i>		4	4	
Omwoko	Omwoko	68	E 33.50	N 2.43	ssp. <i>nilotica</i>		5	5		
	Otuke ss ^a	69	E 33.70	N 2.51	ssp. <i>nilotica</i>	2	4	4		
	Abarilela	70	E 33.81	N 1.98	ssp. <i>nilotica</i>		5	5		
	Akoboi	71	E 33.99	N 1.82	ssp. <i>nilotica</i>		4	4		

Abbreviations: ID, population code; Ncp sequences, number of individuals used for chloroplast sequence analyses; NcpSSRs, number of individuals used for chloroplast SSR analyses; NnucSSRs, number of individuals used for nuclear SSR analyses.

^aPopulations used for chloroplast polymorphism prescreening.

conditions and genotyping methods were identical to those described by Allal *et al.* (2008).

Statistical analyses

Chloroplast sequences: The chloroplast genome is assumed to be non-recombinant, and was thus treated as a single haploid locus. Haplotypes were defined as distinct combinations of alleles found at the two loci, *matK-trnK* and *trnQ-trnS*. Sequences were concatenated for each individual, and global alignment was performed using the MUSCLE algorithm implemented in CODONCODE ALIGNER 3.5.7 (Codon Code Corporation, Dedham, MA, USA). Phylogenetic relationships among haplotypes were assessed by reconstructing the haplotype network with NETWORK 4.516 software (<http://www.fluxus-engineering.com/>) using the median-joining network algorithm (Bandelt *et al.*, 1999).

Chloroplast SSRs: Haplotypes (chlorotypes) were defined as distinct combinations of alleles at the three cpSSRs (see Supplementary Appendix S2 in supporting information). For each population, we computed haplotypic diversity (H_{cp}) (Pons and Petit, 1995), the number of chlorotypes (na_{cp}) and rarefied haplotypic richness (H_R) using the CONTRIB 1.02 program (Petit *et al.*, 1998). H_R is the expected number of different chlorotypes found in each population using a standardized sample size (n) fixed as the smallest available number of genotyped individuals, following the formula given by Comps *et al.* (2001). In our case, the smallest population contained only two individuals. Although this statistic is not biased by a small sample ($n=2$), we eliminated seven populations containing fewer than four individuals to better account for rare alleles and to improve precision. The variation trend of haplotypic richness was then examined at different levels of population stratification (between subspecies, between regions and within regions) using Kruskal–Wallis’s test implemented in XLSTAT 2009 (Addinsoft, Paris, France). This nonparametric ‘one-way analysis of variance’ enables comparison of independent samples that do not follow a normal distribution. Following Kruskal–Wallis’s test, pairwise comparisons were performed using Dunn’s multiple comparison test with Bonferroni correction, XLSTAT 2009 (Addinsoft).

To assess differentiation among subspecies and populations within subspecies, and differentiation among regions and populations within regions, we used hierarchical analysis of molecular variance implemented in ARLEQUIN version 3.5 (Excoffier *et al.*, 2005) set up for haploid data.

Two measurements of genetic differentiation described by Pons and Petit (1996) were used for the chlorotypes: G_{ST} differentiation accounting for frequencies and N_{ST} differentiation accounting for genetic distances between chlorotypes (Pons and Petit, 1996). If the contribution of stepwise mutation to genetic differentiation is negligible, the expected G_{ST} and N_{ST} values are equal. If the differentiation follows a phylogeographic pattern, stepwise mutation would result in $N_{ST} > G_{ST}$. Indeed, under a phylogeographic pattern, phylogenetically related chlorotypes are found more often within than among populations (Pons and Petit, 1996). To compare G_{ST} with N_{ST} , we used a simple test procedure implemented in the SPAGEDI program (Hardy and Vekemans, 2002). The test consists in comparing the observed N_{ST} with the

distribution of N_{ST} obtained with 1000 estimates (pN_{ST}) under random permutation of the chlorotypes in the distance matrix between chlorotypes. In addition to the global test of N_{ST} vs pN_{ST} , we applied the test in pairwise comparison at different levels of population stratification (between regions and between populations within regions) in order to identify at which scale phylogeographic differentiation occurs.

A matrix of pairwise genetic distances was generated from population chlorotype frequencies using Cavalli–Sforza’s chord measure (Cavalli–Sforza, 1967) implemented in the GENDIST program (PHYLIP, version 3.6; Felsenstein, 1993). From this matrix, a dendrogram was constructed using the NJ method of Saitou and Nei (1987) with the NEIGHBOR program of PHYLIP. The significance of each node was evaluated by bootstrapping frequencies over chlorotypes with 1000 replications using the SEQBOOT program of PHYLIP. The consensus tree was obtained with the CONSENSE program of PHYLIP with a cutoff set at 50%, and connections with geographic locations were displayed using GENGIS (Parks *et al.*, 2009).

In the same way as for chloroplast sequences, a network was reconstructed using the median-joining network algorithm (Bandelt *et al.*, 1999) implemented in the NETWORK 4.516 program (<http://www.fluxus-engineering.com/>).

Nuclear SSRs: To evaluate polymorphism using nuclear microsatellite markers (nucSSRs), we recorded allele frequencies, number of alleles per locus (na_{nuc}), observed heterozygosity (Ho_{nuc}) and expected heterozygosity (He_{nuc}) using GENETIX 4.05 (Belkhir *et al.*, 2004). Allelic richness (A_R) is defined as the expected number of alleles in each population for a sample size set at the smallest number of individuals genotyped. A_R was calculated with an adaptation of the rarefaction index using FSTAT version 2.9.3.2 (Goudet, 1995). In the same way as for assessment of haplotypic richness, we computed A_R considering only populations with at least four individuals. Like for haplotypic richness, variation trends in allelic richness were examined at different levels of population stratification (between subspecies, between regions and within regions) using Kruskal–Wallis’s test followed by Dunn’s multiple comparison test with Bonferroni correction, implemented in XLSTAT 2009 (Addinsoft).

Like for cpSSRs, nuclear-based genetic structure was assessed using ARLEQUIN version 3.5 (Excoffier *et al.*, 2005) and analysis of molecular variance. We tested for a phylogeographic signal using the SPAGEDI program (Hardy and Vekemans, 2002), evaluating the contribution of the stepwise mutation in the differentiation pattern by comparing observed R_{ST} with R_{ST} obtained after 1000 allele size permutations (pR_{ST}). If stepwise mutations do not contribute to differentiation, F_{ST} and R_{ST} values are equal, and R_{ST} is expected to be significantly higher than mean permuted R_{ST} under a phylogeographic pattern. In the same way as for cpSSRs, this permutation test was applied to global R_{ST} vs pR_{ST} and in pairwise R_{ST} vs pR_{ST} comparisons at different levels of population stratification (between regions and between populations within regions).

To assess the overall structure of nucSRR diversity, we used the STRUCTURE program version 2.3.3 (Pritchard *et al.*, 2000). We used an admixture model with correlated

allele frequencies, performing 10 independent repetitions for each K (ranging from 1 to 20), with 100 000 Markov chain Monte Carlo repetitions and an 80 000 burn-in period. To identify the optimal K value, we examined absolute values of the second-order rate of change of the likelihood distribution divided by the standard deviation of likelihoods (K), as described by Evanno *et al.* (2005). To explore divergence between inferred clusters, we examined the genetic differentiation (F_{ST}) between these clusters and their ancestral population, predicted by STRUCTURE (Pritchard *et al.*, 2000). Phylogenetic relationships among populations were assessed using population allele frequencies with the PHYLIP package (Felsenstein, 1993) using a similar protocol as for cpSSRs.

Bioclimatic data and ecological niche modelling (ENM): The ENM implemented in DIVA-GIS software version 7.2 (Hijmans *et al.*, 2001) enables prediction of the spatial distributions of a species on a bioclimatic basis using the BIOCLIM algorithm (Busby, 1991). From the climate characteristics of given species occurrences, this algorithm gives a percentile distribution (between 0 and 100 percentile) for each climatic variable. On the basis of homocline matching (Booth *et al.*, 1987), the BIOCLIM algorithm then compares the climatic characteristics of each grid cell in a geographic layer with the percentile distributions of variables inferred from species occurrences. The program, run with the BIOCLIM 'classic' option (http://www.diva-gis.org/docs/DIVA-GIS5_manual.pdf), determined six types of area according to their suitability: 'not suitable', that is, one or more climate variables are outside the 0–100 percentile envelope; 'excellent', that is, within the 20–80 percentile envelope for each variable; 'very high', that is, within the 10–90 percentile envelope; 'high', that is, within the 5–95 percentile envelope; 'medium', that is, within the 2.5–97.5 percentile envelope, and marginal areas of 'low' suitability, for which all variables are at the boundaries of the 0–100 envelope (that is, between 0–2.5% and 97.5–100%). Considering a 5% threshold, we did not include marginal areas of 'low' suitability. We used the occurrence data of *V. paradoxa* from 126 sample sites (see Supplementary Appendix S3 in Supporting information) and considered seven climatic variables at spatial resolutions of 30 arc-seconds (about 1×1 km): (1) annual mean temperature; (2) mean diurnal temperature range; (3) maximum temperature of the coldest month; (4) minimum temperature of the coldest month; (5) annual precipitation; (6) precipitation seasonality; and (7) precipitation of the wettest quarter. Percentile distributions of these variables were created from current climate conditions and projected onto the layers of current, LGM (~ 21 000 years BP), and the LIG period (~ 120 000–140 000 years BP). Using current climate conditions (1950–2000) derived from the WorldClim global climate database (Hijmans *et al.*, 2005), ENM enabled us to check that our 126 sample sites were suitable to recover the known actual distribution of *V. paradoxa* (Hall *et al.* 1996) and, thus, to describe its climatic preferences. Using data on the LGM obtained from the Paleoclimate Modelling Intercomparison Project Phase II (PMIP2) database and achieved under the community climate system model (CCSM3) (Collins *et al.*, 2006), we illustrated the distribution shift of *V. paradoxa* during this period of major climate

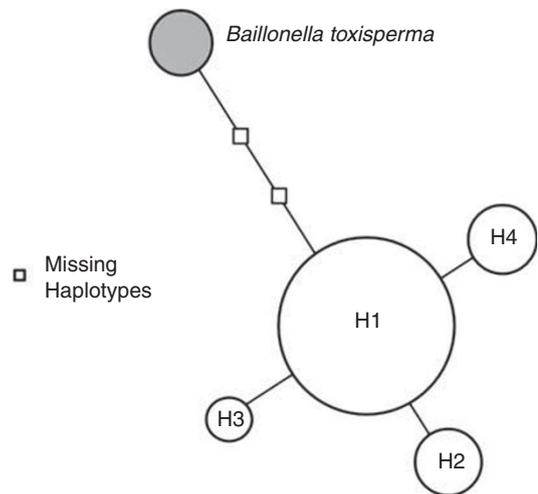


Figure 1 Phylogenetic relationships among haplotypes (H1, H2, H3 and H4; *B. toxisperma* used as outgroup species) using the median-joining haplotype network (NETWORK 4.516 program).

perturbations (Maley, 1996). Finally, to describe the potential distribution of the species before LGM perturbations, we predicted the suitability of *V. paradoxa* using data on the LIG period generated by Otto-Bliesner *et al.* (2006) and provided by the WorldClim global climate database. The performance of the model under current conditions was tested by measuring the area under a relative operating characteristic curve (AUC) (Pearce and Ferrier, 2000) using DIVA-GIS software (Hijmans *et al.*, 2001). AUC is commonly used to evaluate ENM models by measuring their ability to distinguish locations where the species is present from those where it is absent (Hanley and Mcneil, 1982). ENM with $AUC > 0.7$ is considered to perform well (Fielding and Bell, 1997).

Results

Chloroplast sequences

From alignments of the two polymorphic chloroplast sequence intergenic spacers *trnQ-trnS* (601 bp) and *matK-trnK* (452 bp), we identified only 3 single-nucleotide polymorphisms among the 40 populations (58 individuals) of *V. paradoxa*. The combination of these mutations (one located in *trnQ-trnS* and two in *matK-trnK*) resulted in a total of four haplotypes: H1, H2, H3 and H4. H1 was the predominant haplotype with a global frequency of 74.1%, and was shared by West African and Ugandan populations. H2, with a global frequency of 10.3%, was found in Cameroon. H3, found in 5.2% of the individuals, was limited to populations in Chad. Finally, H4, with a global frequency of 10.3%, was found in most individuals in Central African Republic (one tree in southern Central African Republic displayed the H1 haplotype). Therefore, most of the populations were fixed for a single haplotype. We identified a single haplotype among the five individuals of our outgroup taxon *B. toxisperma*.

The haplotype network (Figure 1) indicates that haplotypes of *V. paradoxa* are poorly divergent, and suggests that H2, H3 and H4 are derived from the H1 haplotype through a single mutation. The haplotype of

Table 2 Diversity parameters calculated per population and averaged per country

Region	Country	Ns	N	Chloroplast microsatellites			Nuclear microsatellites			Subsp. species
				na _{cp}	H _{cp} (s.d.)	Ne _{cp} (s.d.)	na _{nuc}	Ho _{nuc} (s.d.)	He _{nuc} (s.d.)	
West Africa	Senegal	2	9	1 L	0.00 (0.00)	1.00 (0.00)	3.42	0.53 (0.26)	0.55 (0.20)	ssp. <i>paradoxa</i>
	Mali	7	29	2 CL	0.00 (0.00)	1.00 (0.00)	2.68	0.47 (0.33)	0.43 (0.23)	ssp. <i>paradoxa</i>
	Ivory Coast	2	6	1 C	0.00 (0.00)	1.00 (0.00)	3.00	0.67 (0.26)	0.54 (0.18)	ssp. <i>paradoxa</i>
	Burkina-Faso	6	36	3 BCF	0.24 (0.21)	1.11 (0.27)	3.49	0.59 (0.30)	0.55 (0.23)	ssp. <i>paradoxa</i>
	Ghana	4	24	3 BCN	0.31 (0.22)	1.42 (0.50)	4.23	0.60 (0.26)	0.59 (0.21)	ssp. <i>paradoxa</i>
	Togo	6	32	4 CFGH	0.15 (0.24)	1.11 (0.27)	3.28	0.54 (0.33)	0.48 (0.23)	ssp. <i>paradoxa</i>
	Benin-Niger	6	30	4 CFGH	0.28 (0.32)	1.11 (0.27)	3.49	0.55 (0.36)	0.51 (0.25)	ssp. <i>paradoxa</i>
	Nigeria	8	65	4 GMNO	0.27 (0.09)	1.71 (0.12)	4.57	0.56 (0.27)	0.57 (0.22)	ssp. <i>paradoxa</i>
Central Africa	Cameroon	7	35	2 AM	0.06 (0.15)	1.14 (0.38)	2.74	0.53 (0.32)	0.48 (0.20)	ssp. <i>paradoxa</i>
	Chad	3	14	1 A	0.00 (0.00)	1.00 (0.00)	2.28	0.55 (0.37)	0.42 (0.22)	ssp. <i>paradoxa</i>
	Central African Republic	3	15	3 DIJ	0.22 (0.19)	1.33 (0.58)	3.86	0.59 (0.25)	0.62 (0.14)	ssp. <i>paradoxa</i>
East Africa	Sudan	7	26	4 DEJK	0.24 (0.31)	1.24 (0.42)	2.49	0.52 (0.30)	0.46 (0.18)	ssp. <i>nilotica</i>
	Uganda	10	43	1 E	0.00 (0.00)	1.00 (0.00)	2.60	0.54 (0.31)	0.47 (0.20)	ssp. <i>nilotica</i>

Abbreviations: Ns, number of populations; N, number of individuals; na_{cp}, number of haplotypes per country; H_{cp}, haplotypic diversity; Ne_{cp}, number of effective haplotypes per population; na_{nuc}, mean number of alleles per population and per locus; Ho_{nuc}, observed heterozygosity; He_{nuc}, expected heterozygosity on Hardy–Weinberg equilibrium.

the outgroup species was phylogenetically close to *V. paradoxa* haplotypes, and H1 was closer to the outgroup haplotype than H2, H3 and H4. ‘H1’ populations were located in both West Africa and Uganda, suggesting either a West African or East African origin for existing populations of *V. paradoxa*.

Chloroplast SSR markers

Diversity patterns: We identified three polymorphic loci, ccmp3, ccmp5 and ntcp9, including, respectively, three, two and seven variant alleles. These variants were combined into 15 chlorotypes, resulting in high total haplotypic diversity with H_{cp} = 0.9. However, Table 2 shows very poor haplotypic diversity at the population level, with an average of H_{cp} = 0.17, because of the small sample size. Figure 2a shows variation in haplotypic richness (H_R) on a geographic basis. We observed major variations in haplotypic richness H_R among populations, from 0 to 2. No significant differences were detected between subspecies (P = 0.079), or between regions (P = 0.054), in H_R values using the Kruskal–Wallis test. However, in West Africa, we observed a central zone (around Togo) exhibiting high haplotypic richness values and bounded by two zones with lower H_R (Figure 2a). With the Kruskal–Wallis test, we detected significant differences in haplotypic richness between these areas (P < 0.001), and Dunn’s pairwise comparisons with Bonferroni correction showed that the central zone around Togo had significantly higher haplotypic richness (H_R = 1.12) than the eastern (H_R = 0.44) and western (H_R = 0.09) flanking zones (Bonferroni-corrected P = 0.017) (Figure 2a).

Differentiation: We identified 15 chlorotypes in the natural area (Figure 3a), with frequencies ranging from 1.1 (D) to 19.8% (C). Eleven chlorotypes were found in the subspecies *paradoxa*, including eight haplotypes (B–C–F–G–H–L–N–O) limited to West African populations, and three (A–I–M) restricted to Central Africa. Two private haplotypes (E–K) were limited to *nilotica* subspecies. Two haplotypes (D–J) were shared between the subspecies in Central African Republic (*paradoxa*) and Sudan (*nilotica*).

Evaluation of differentiation using hierarchical analysis of molecular variance revealed populations to be significantly differentiated at subspecies level, with 37% of the variation detected among subspecies (F_{CT} = 0.37, P = 0.00), 57% among populations within subspecies (F_{SC} = 0.91, P = 0.00) and only 6% of variation within populations. Differentiation at regional level (West Africa, Central Africa and East Africa) was even stronger, with 58% of genetic variation detected among geographical groups (F_{CT} = 0.58, P = 0.00), 37% among populations within groups (F_{SC} = 0.87, P = 0.00) and only 5% within populations.

Phylogeographic patterns: Overall differentiation among populations was very high and significant with G_{ST} = 0.81 (P = 0.00) based on unordered alleles, and N_{ST} = 0.90 (P = 0.00) based on ordered alleles. The global test of phylogeographical structure was significant (N_{ST} > pN_{ST}, P < 0.001) with pN_{ST} = 0.80, revealing a clear phylogeographical pattern in chloroplast variation. However, this phylogeographical signature was no longer detectable within regions (West: P = 0.18; Central: P = 0.60; and East: P = 0.49), indicating that mutation did not contribute to chloroplast differentiation within geographical zones. Pairwise comparisons of N_{ST} vs pN_{ST} between West and East Africa revealed a highly significant phylogeographical signal (N_{ST} = 0.69 and pN_{ST} = 0.35 with P = 0.00), but not between West and Central Africa (N_{ST} = 0.36 and pN_{ST} = 0.23 with P = 0.13), or between Central and East Africa (N_{ST} = 0.38 and pN_{ST} = 0.4 with P = 0.47).

Phylogenetic relationships between populations were supported by strong bootstrap values (Figure 3a). The frequency-based NJ tree displayed four clades (I–IV), which perfectly match the geography. Clade I contains marginal western populations (populations 1–4 and 6), suggesting their isolation from other West African populations (populations 5 and 7–42). Clade II contains other West African populations. Clade III includes most Central African sites, apart from Central African Republic. Clade IV contains eastern populations, combining the subspecies *paradoxa* (Central African Republic) and *nilotica* (Sudan and Uganda).

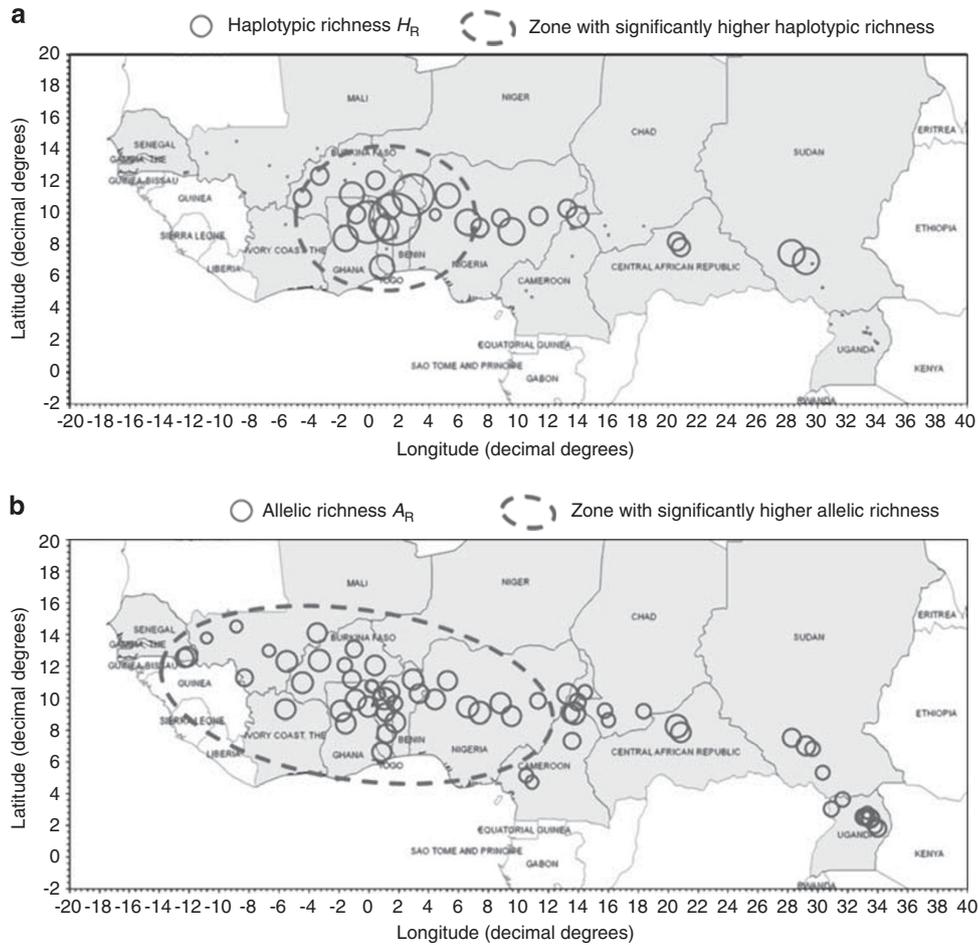


Figure 2 Trends in haplotypic (cpSSRs) and allelic (nucSSRs) richness. Diameters of circles vary proportionally with richness values. Areas bounded by a red dotted line had significantly higher values than others with Dunn's comparison test with Bonferroni correction. (a) Trends in haplotypic richness H_R . (b) Trends in allelic richness A_R . A full color version of this figure is available at the *Heredity* journal online.

Phylogenetic relationships among chlorotypes illustrated in the median-joining network (Figure 3b) suggested global differentiation following a phylogeographic pattern, but with some inconsistent reticulations. Private West African chlorotypes were closely related, defining group 1 in the network. Within this group, each chlorotype was separated from 2 to 4 bp and others by only 1 bp, explaining the lack of a phylogeographic pattern in West Africa. Group 2 in the network includes the three East African *nilotica* chlorotypes. These East African chlorotypes were connected to West African chlorotypes: the network showed that L (found in Senegal) and K (found in Sudan) were separated by only 1 bp. Central African chlorotypes were split into two distant groups in the network. Haplotype M (group 3) was the only chlorotype from southern Cameroon related to 'N', a West African chlorotype found in Nigeria. However, the four missing chlorotypes between M and N suggest long-term isolation of populations located in southern Cameroon from West African populations. The other Central African chlorotypes (group 4) were found to be closely related to East African haplotypes, suggesting a common origin.

Nuclear SSR markers

Diversity patterns: A total of 173 alleles were detected at 12 nuclear SSR loci in our collection, with an average of

14.41 alleles per locus. Low variation was observed in mCIRVp08 and mCIRVp175 with five alleles, while mCIRVp167 was highly variable with 25 alleles. Expected and observed (in parenthesis) heterozygosity varied markedly from $H_{e_{nuc}} = 0.27$ ($H_{o_{nuc}} = 0.26$) in Mahina (Mali) to $H_{e_{nuc}} = 0.65$ ($H_{o_{nuc}} = 0.74$) in Kanté (Togo). Average values per country are listed in Table 2.

Allelic richness (A_R) varied from 1.8 to 3.4 among populations, but with no clear trend (Figure 2b). However, the Kruskal–Wallis test revealed significant differences between *paradoxa* ($A_R = 2.8$) and *nilotica* ($A_R = 2.4$) subspecies ($P = 0.016$), with significantly higher *paradoxa* A_R with the Dunn's pairwise comparison test with Bonferroni correction (Bonferroni-corrected $P = 0.05$). In addition, the Kruskal–Wallis test detected significant differences in allelic richness between geographical regions ($P = 0.006$), and Dunn's pairwise comparisons with Bonferroni correction highlighted significantly higher allelic richness in West Africa ($A_R = 2.9$) than in Central ($A_R = 2.5$) and Eastern ($A_R = 2.4$) Africa (Bonferroni-corrected $P = 0.017$) (Figure 2b).

Differentiation: In the same way as with chloroplast SSRs, with nuclear SSR data, analysis of molecular variance revealed that the subspecies *nilotica* and *paradoxa* were significantly differentiated: 22% of variation was detected between subspecies ($F_{CT} = 0.22$, $P = 0.00$), and

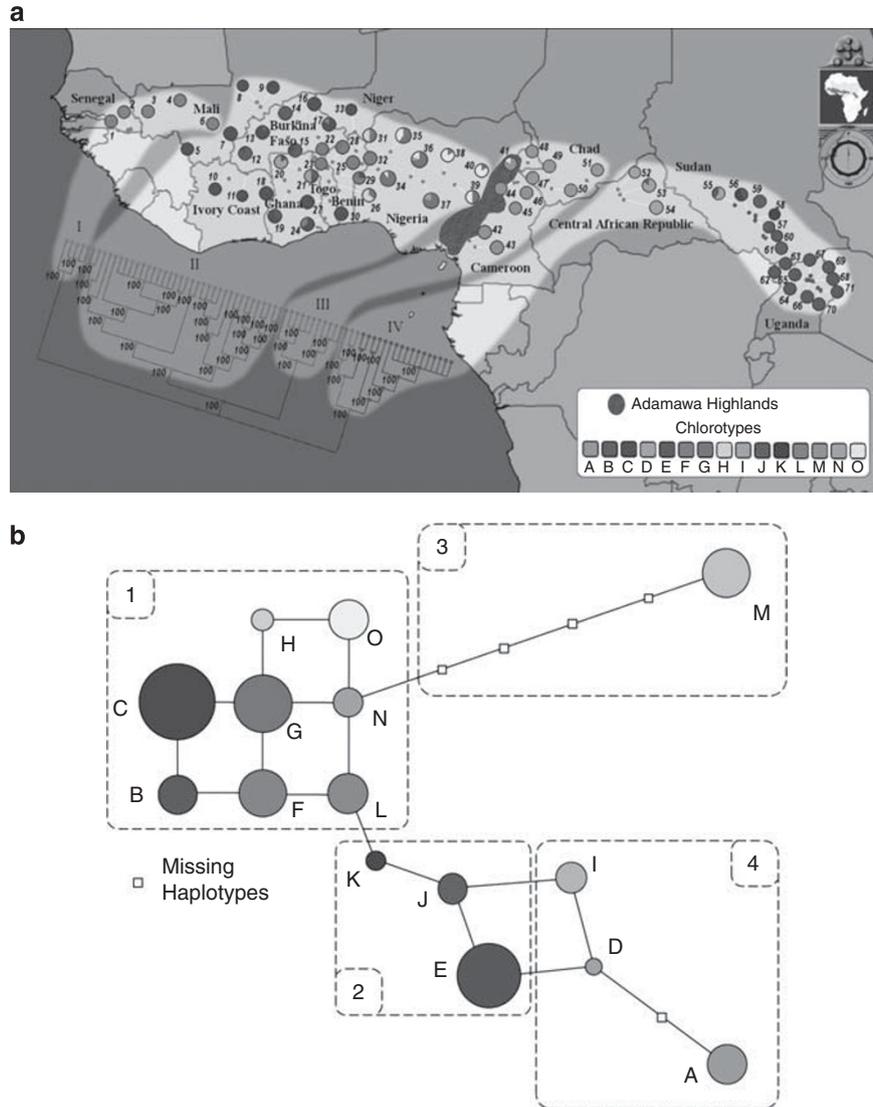


Figure 3 Genetic structure of chloroplast microsattellites across 374 individuals from 71 populations. (a) Geographical distribution of the 15 chlorotypes of *V. paradoxa*. Unrooted neighbor-joining tree of populations drawn with GENGIS software (Parks *et al.*, 2009) using Cavalli-Sforza genetic distances. Numbers at the base of the branches are percentages corresponding to the bootstrap values after 1000 replications. (b) Median-joining network of chlorotypes computed using NETWORK (<http://www.fluxus-engineering.com/>), with 1, West African chlorotypes; 2, East African chlorotypes; 3 and 4, Central African chlorotypes. A full color version of this figure is available at the *Heredity* journal online.

17% between populations within subspecies ($F_{SC} = 0.21$, $P = 0.00$). Likewise, differentiation was significant at the scale of geographic regions, with 19% of variance explained by the regions ($F_{CT} = 0.19$, $P = 0.00$) and 16% among populations within regions ($F_{SC} = 0.19$, $P = 0.00$). We can thus assume that geographic distribution is a strong determinant of the nuclear-based genetic structure of *V. paradoxa*. However, up to 65% of the variation was detected within populations, in contrast to 5% with chloroplasts.

Using the Bayesian-based approach of Pritchard (Pritchard *et al.*, 2000), two peaks of K for $K = 2$ and $K = 4$ were detected according to the method of Evanno *et al.* (2005). The clustering obtained for $K = 2$ (Figure 4a) separated populations into two geographically consistent groups with low levels of admixed individuals: the 'West' group included populations from West Africa and the 'East' group included the remaining populations. This clustering in two groups constituted the upper level

of the structure in *V. paradoxa*. Although $K = 2$ was identified as the best model, $K = 4$ (Figure 4b) obtained high probability and revealed very clear substructures in higher groups. Within the 'East' group, we detected a substructure that distinguished a 'Central' (C, *ssp. paradoxa*) from an 'Eastern' group (E, *ssp. nilotica*), consistent with the distribution of the two subspecies in this zone. In West Africa, populations were also split into two groups consistent with their geographical distribution, 'West 1' (W1) including most western populations and 'West 2' (W2) populations from Nigeria. However, W1 included a large number of admixed individuals. Bayesian analysis revealed migrants and admixed individuals between the 'Central' and 'Eastern' clusters of STRUCTURE, and thus between *nilotica* and *paradoxa* (Figures 4b and c). In comparison to the ancestral population predicted by STRUCTURE, clusters W1 ($F_{ST} = 0.06$) and W2 ($F_{ST} = 0.05$) displayed the least differentiation, whereas clusters C ($F_{ST} = 0.12$) and

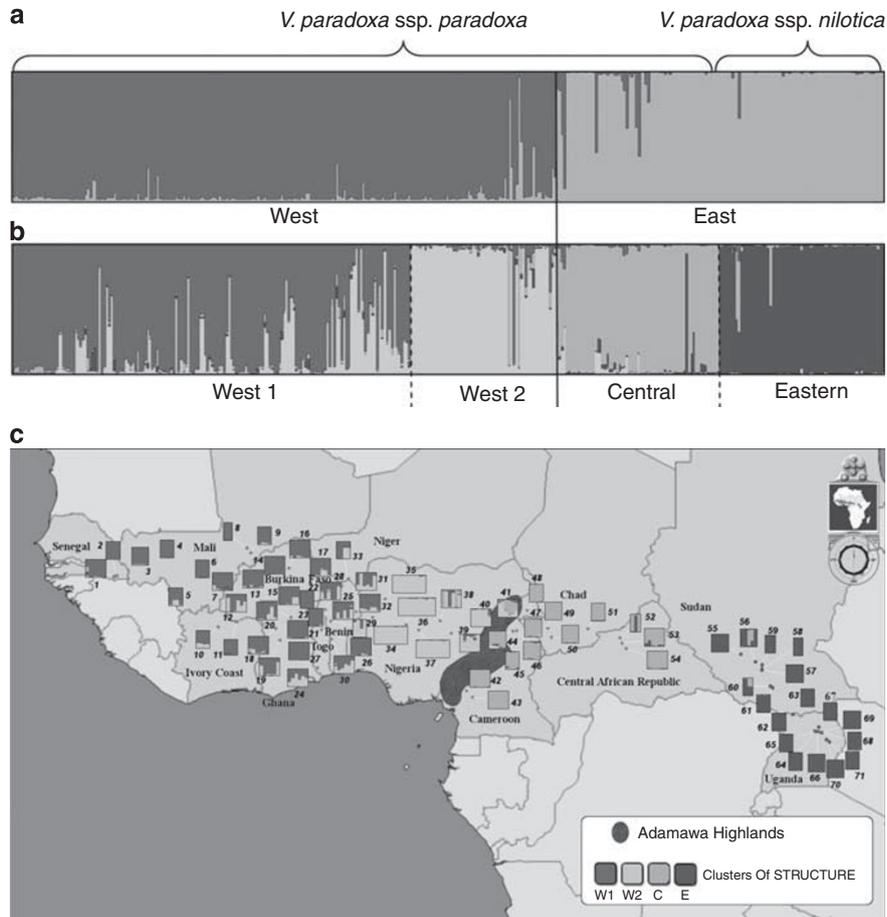


Figure 4 Genetic structure of nuclear microsatellites across 374 individuals from 71 populations. (a) Bar plot showing clustering of individuals by STRUCTURE with $K=2$ (Pritchard *et al.*, 2000). (b) Bar plot showing clustering of individuals by STRUCTURE WITH $K=4$. (c) Geographical plot of clustering of individuals by STRUCTURE WITH $K=4$: W1: 'West 1'; W2: 'West 2'; C: 'Central'; E: 'Eastern'. A full color version of this figure is available at the *Heredity* journal online.

E ($F_{ST}=0.23$) diverged further from the potential ancestral population. A PCA analysis on nuclear microsatellite allele frequencies was highly congruent with the STRUCTURE analysis (results not shown).

Phylogeographic patterns: Overall differentiation among populations assessed with nuclear SSRs was high and significant for both the index based on unordered alleles ($F_{ST}=0.21$, $P=0.00$) and that based on ordered alleles ($R_{ST}=0.28$, $P=0.00$). Comparisons of R_{ST} and pR_{ST} revealed that the nuclear microsatellite diversity pattern of *V. paradoxa* had a significant phylogeographic signature with $pR_{ST}=0.21$ and $R_{ST} > pR_{ST}$ ($P=0.004$). The phylogeographic signal was highly significant for each pairwise regional comparison: between West and Central Africa ($R_{ST}=0.16$ and $pR_{ST}=0.07$ with $P=0.005$), between West and East Africa ($R_{ST}=0.27$ and $pR_{ST}=0.07$ with $P=0.00$) and between Central and East Africa ($R_{ST}=0.25$ and $pR_{ST}=0.08$ with $P=0.00$). However, like for chloroplasts, variation trends within geographic regions did not reveal a phylogeographic pattern (West: $P=0.49$; Central: $P=0.25$; East: $P=0.34$), suggesting a major contribution of migration drift within geographical regions. Finally, we performed pairwise comparison between cluster groups 'W1' and 'W2' inferred by STRUCTURE

(Figure 4b), but detected no significant phylogeographic signal ($R_{ST}=0.06$ and $pR_{ST}=0.04$ with $P=0.11$).

The degree of admixture of each population linked to their geographical location confirmed that the clustering obtained with the Bayesian approach fitted the geographical distribution (Figure 4c). Moreover, higher levels of admixture were found at geographical junctions between clusters. Within the W1 group, admixed populations were most often found in Togo, Benin and Niger (populations 24, 25, 28, 29, 30, 31, 32 and 33). Populations 52, 53, 54, 56 and 60, at the boundary between *paradoxa* and *nilotica* zone, were also admixed (Figure 4c). Phylogenetic relationships at population level based on allele frequency were poorly informative, as nodes were not significant (data not shown).

Present and previous theoretical distribution of *V. paradoxa*

From our database containing 126 sample sites, we checked that the zone theoretically suitable for the species under current conditions (Figure 5a) was consistent with its known distribution (Hall *et al.*, 1996). A decrease in suitability was observed at the junctions between the geographic regions (West, Central and East), suggesting lower species density, and extending areas suitable for *V. paradoxa* to regions where the

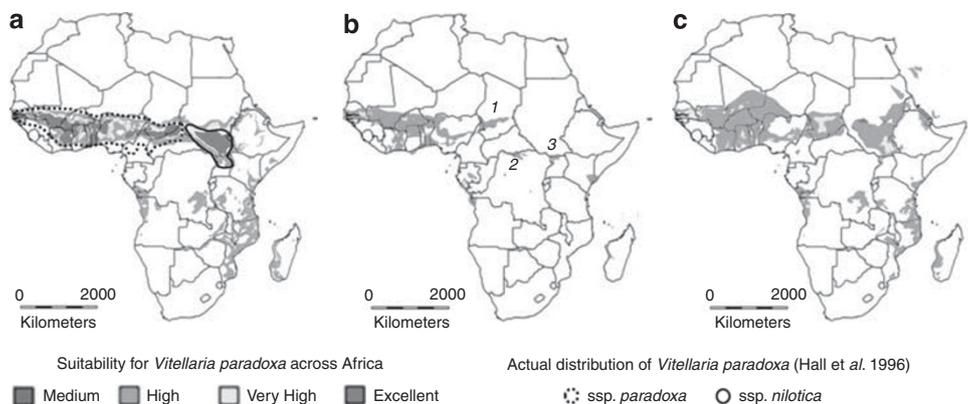


Figure 5 ENM of *Vitellaria paradoxa* at three different periods: (a) today's predicted and reported natural range; (b) potential distribution during the LGM (20 000 years BP) and refugia (1: around Lake Chad; 2: northern Congo; 3: northern Uganda); (c) potential natural range during the LIG period (~120 000–140 000 years BP). Colors represent the probability of suitable climatic conditions from red (excellent) to green (medium). Envelopes were produced using the Worldclim/PMIPII database and DIVA-GIS software version 7.2 (Hijmans *et al.*, 2001). A full color version of this figure is available at the *Heredity* journal online.

species is in fact absent (below the equator). The AUC, averaged over the seven climatic variables, was scored at 0.79, indicating that the accuracy of the model was reasonably good for ENM predictions. ENM of the species during the LGM (~21 000 years BP) showed a much smaller suitable zone (Figure 5b): while West Africa remained highly suitable, the putative distribution of *V. paradoxa* from Cameroon to Uganda was reduced and fragmented. In this part of Africa, three putative areas were detected where the species could have been limited: near Lake Chad, in southern Central African Republic, and in northern Uganda (see zones 1, 2 and 3 in Figure 5b). Finally, we predicted probable suitable areas for the shea tree during the LIG period around 120 000–140 000 BP (Figure 5c). Results showed a potential distribution pattern divided into two main zones. The first zone extended from present-day Senegal to the north of Central African Republic, but the zone was almost completely disrupted at the Nigerian border. The second zone area mainly comprised Sudan and northern Uganda. These zones were separated by a large gap (~200 km) at the border between Sudan and Central African Republic, consistent with the current geographical gap between the two subspecies *nilotica* and *paradoxa* (Hall *et al.*, 1996).

Discussion

Genetic diversity of *V. paradoxa*

In our study, nuclear and chloroplast SSRs genetic diversity was higher than that observed in previous studies on *V. paradoxa* (Fontaine *et al.*, 2004; Sanou *et al.*, 2005). This novel assessment based on a larger sample is consistent with diversity detected on angiosperms in other studies, (Müller *et al.*, 2009) and meets expectations concerning the diversity of widespread tree species with similar biological attributes (Hamrick *et al.*, 1993). However, we noted that the level of SSR diversity varied markedly among zones in our study (Figure 2). Possible explanations for higher levels of diversity are the presence of ancient glacial refugia or the confluence of migration routes (Petit *et al.*, 2003), and we propose some hypotheses below. In addition to SSR markers, polymorphism within chloroplast intergenic spacers revealed

very weak diversity with only four haplotypes. This poor variability was comparable to the diversity detected with chloroplasts in other widespread tree species, such as *Pinus pinea* L. (Vendramin *et al.*, 2008) and *Adansonia digitata* (baobab), another widespread savanna tree (Pock Tsy *et al.*, 2009). Therefore, in contrast with species harboring high levels of intergenic spacer polymorphisms (Dick and Heuertz, 2008), these markers are not an effective tool to study the phylogeography of *V. paradoxa*. However, the low polymorphism detected in these sequences was consistent with the low rate of nucleotide substitution in chloroplast genomes, around 10^{-9} (Wolfe *et al.*, 1987), compared with SSR mutation rates of from 10^{-4} to 10^{-2} (Vigouroux *et al.*, 2002) with nuclear SSRs and from 10^{-5} to 10^{-3} with chloroplast SSRs (Navascués and Emerson, 2005).

A phylogeographic pattern driven by climate changes

We showed that the genetic diversity of *V. paradoxa* has a clear east–west structure (Figures 3 and 4), and revealed a strong phylogeographic signature from chloroplast ($G_{ST}=0.81$; $N_{ST}=0.90$; $N_{ST}>pN_{ST}$) and nuclear SSRs ($F_{ST}=0.21$; $R_{ST}=0.28$; $R_{ST}>pR_{ST}$). We identified two major genetic groups geographically separated by the Adamawa Highlands: a 'West' group, which is relatively homogeneous over the 3000 km area comprising West Africa, and an 'East' group divided into two subgroups (Figure 3). We propose that the strong phylogeographic pattern of *V. paradoxa* is explained by past climate shifts and that LGM perturbation is a major factor in explaining this pattern. During the LGM, the climate in Africa was colder and drier; this shift was particularly marked in eastern sub-Saharan Africa and was characterized by small incursions of the Sahara desert into this region (Lioubimtseva *et al.*, 1998). Consistent with these climatic perturbations, ENM predictions indicated that the potential distribution of *V. paradoxa* in this zone was highly fragmented (Figure 5b). However, ENM results should be interpreted with caution because they depend on the reliability of the climatic models (current and ancient) resulting from mathematical interpolations and global smoothing of collected data (Hijmans *et al.*, 2005). For example, the poor quality of climatic documentation resulted in local over-estimation of the potential

distribution of a mountain bird species (*Cranioleuca henricae*) in Bolivia (Soria-Auza *et al.*, 2010). Consequently, we suggest using only ENM to describe general trends of species range in our study. In addition, ENM predictions assume a no evolution model, and thus do not consider the adaptation ability and the dispersal dynamics of species across ages. Having taken these limitations into account, using ENM on LGM data, we identified two reduced areas that could have formed glacial refugia for populations of *V. paradoxa* in this part of Africa: around Lake Chad and in northern Uganda, along with other marginal refugia (Figure 5b). We suggest that these refugia could be at the origin of the genetic divergence of *V. paradoxa* in eastern sub-Saharan Africa. In contrast to the eastern area, West Africa was less impacted by the climate change during the LGM, probably owing to the effect of the monsoon from the Gulf of Guinea (Braconnot *et al.*, 2000). As a result, although the eastern suitable area for *V. paradoxa* was probably deeply disturbed, ENM results suggest that distribution of the species was less impacted, and thus probably conserved in West Africa (Figures 5b and c). Although the limited LGM perturbation in West Africa explains the high level of diversity observed and the hotspot detected in its central area (around Togo) (Table 2 and Figure 2), the poor structure detected in this large region (Figure 4) could also result from the intense human activity in the area. Palynological studies reported human migrations and agropastoral activities involving shea trees 3000 years ago in Burkina (Neumann *et al.*, 1998), and the evidence of human selection of *V. paradoxa* has been found in this zone (Lovett and Haq, 2000; Maranz and Wiesman, 2003). In addition to past climate variations, we propose that the strong differentiation between West and Central African populations was probably preserved by topographic elevations. For instance, the Adamawa Highlands clearly represent a boundary between West African and other populations. Peaking at 1,300 m a.s.l. between Nigeria and Cameroon (Figure 3), this mountain range creates a colder climate resulting in a local barrier to *V. paradoxa* colonization, and limiting gene flow between West and Central Africa. Therefore, apart from the possible secondary role of Adamawa, we propose that the pancontinental phylogeographic pattern of *V. paradoxa* is fully consistent with perturbations during the LGM, and that this pattern was little affected by the dispersal dynamics of the species.

Differentiation between *paradoxa* and *nilotica*

Results in the literature argue for species differentiation in two non-overlapping subspecies based on slight variations in their leaf and stamina morphology (Hall *et al.*, 1996). We consequently expected to detect a previous genetic differentiation based on this distinction. However, several molecular results suggest that this taxonomic distinction should be reconsidered. The median-joining network obtained with chloroplast SSRs (Figure 3c) revealed that Central African chlorotypes (*paradoxa*) are closer to *nilotica* than to other *paradoxa* chlorotypes. In addition, the Bayesian approach applied for nuclear SSRs revealed a clear and strong differentiation of the species in two groups on the two sides of the Adamawa Highlands: a western group comprising

West African populations (ssp. *paradoxa*) and an eastern group including *nilotica* and *paradoxa* from Cameroon to Uganda (Figure 4a). The molecular distinction between subspecies *nilotica* and *paradoxa* appears only as a substructure of the eastern group for $K=4$ (Figure 4b). These results suggest that *paradoxa* populations in eastern sub-Saharan Africa (from Cameroon to Central African Republic) share a common evolutionary history with *nilotica* rather than with other *paradoxa* from West Africa, tightened by the inter-subspecies migrations and hybridizations revealed in our study (Figure 4). Moreover, migration events between subspecies were also supported by chloroplast microsatellites. Indeed, the *nilotica* migrant detected with nuclear SSRs in Central African Republic shared *nilotica* chlorotype ('J'), and two admixed individuals found in Sudan shared a *paradoxa* chlorotype ('D') (Figure 3). We cannot exclude that these shared chlorotypes are because of homoplasy or shared ancestral polymorphism rather than to migration, however, the chloroplast microsatellite mutation rate is likely to be low, reducing the effect of homoplasy (Navascués and Emerson, 2005). On the basis of this molecular differentiation, we suggest that the distinction between subspecies perhaps occurred spatially on the two sides of the Adamawa Highlands, which may be consistent with the trend observed in lipid composition, suggesting that shea butter in Cameroon is of the '*nilotica*' type (Maranz *et al.*, 2004).

Hypothesis on the origin of *V. paradoxa*

Although no evidence in the literature proposes an origin for *V. paradoxa*, combined results obtained with nuclear and chloroplast markers provided new insights into present populations. Using both chloroplast and nuclear SSRs, higher diversity was detected in West Africa, particularly around Togo (Figure 2, Table 2), suggesting ancient populations in this area (Petit *et al.*, 2003). This result has been confirmed by another study (Logossa, personal communication) on 674 individuals from 38 populations sampled throughout West Africa. In addition, Bayesian analysis of nuclear SSRs showed that West African groups displayed minimal divergence from their theoretical ancestor ($F_{ST}=0.05$), whereas other groups diverged more (Central: $F_{ST}=0.12$; East: $F_{ST}=0.23$). This result indicates a historically higher effective population size in West Africa, which could be compatible with the origin of *V. paradoxa* being in this zone. This hypothesis fits with the recently inferred West African origin of the baobab (Pock Tsy *et al.*, 2009), another African savanna tree whose distribution north of the equator is very similar to that of *V. paradoxa*. However, low F_{ST} relative to ancestral populations and higher diversity in West Africa does not necessarily result from a more ancient origin of the species compared with the other regions. To be able to conclude on the origin of *V. paradoxa*, additional DNA sequence data are required, as chloroplast sequence analysis indicates the presence of likely ancestral haplotype both in West Africa and in Uganda (Figure 1).

Outlook for conservation of *V. paradoxa* in the African savanna biome

The savanna is demographically, environmentally and economically a major African ecosystem that requires urgent management efforts (Sankaran *et al.*, 2005). Despite its importance, forecasts of anthropogenic climate

change on savanna biome are lacking. In our study, we identified West Africa as a diversity hotspot for *V. paradoxa*, possibly because of a weaker impact of climate shifts on the savanna tree in this zone. Given the greater past perturbations detected farther east, we suggest that current global warming could affect this part of the species area first, and by extension, the associated biotope. Consequently, major efforts should be made to protect the species in this zone. Owing to the possible buffering effect of the monsoon in Guinea, West African *V. paradoxa* populations should be less affected by the current climate shift, but we, nevertheless, suggest that efforts should be increased to maintain the diversity of this zone, including identifying and protecting the numerous *V. paradoxa* landraces in West Africa.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

This work is part of the PhD thesis of François Allal and forms one part of the tasks of the INCO project INNOVKAR funded by the European Union for 'Innovative Tools and Techniques for Sustainable Use of the Shea Tree in Sudano-Sahelian zone'. Experiments were conducted at CIRAD laboratory for genetic diversity and breeding of forest tree species in Montpellier, France. We thank our colleagues in Africa for their valuable help in sampling, without whom this work would not have been possible.

Author contributions: Pan-African Shea tree sampling was performed by HS, ZAL, AV, FA and other African colleagues; molecular data were collected by FA and LM, data were analyzed by FA, LC-K and J-MB, and FA led the writing. All authors contributed ideas, comments and revised manuscript versions.

References

- Allal F, Vaillant A, Sanou H, Kelly B, Bouvet J-M (2008). Isolation and characterization of new microsatellite markers in shea tree (*Vitellaria paradoxa* c. F. Gaertn.). *Mol Ecol Resour* 8: 822–824.
- Bagley MJ, Anderson SL, May B (2001). Choice of methodology for assessing genetic impacts of environmental stressors: polymorphism and reproducibility of RAPD and AFLP fingerprints. *Ecotoxicology* 10: 239–244.
- Bandelt HJ, Forster P, Rohlf A (1999). Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16: 37–48.
- Beerling DJ, Osborne CP (2006). The origin of the savanna biome. *Glob Change Biol* 12: 2023–2031.
- Behling H (2002). South and Southeast Brazilian grasslands during Late Quaternary times: a synthesis. *Palaeogeogr Palaeoclimatol Palaeoecol* 177: 19–27.
- 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.
- Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2004). Genetix 4.05, logiciel sous windows tm pour la génétique des populations. Laboratoire génome, populations, interactions, CNRS UMR 5171. Université de Montpellier II, Montpellier (France).
- Bonaccorso E, Koch I, Peterson AT (2006). Pleistocene fragmentation of amazon species' ranges. *Divers Distributions* 12: 157–164.
- Booth TH, Nix HA, Hutchinson MF, Busby JR (1987). Grid matching: a new method for homoclimate analysis. *Agric Forest Meteorol* 39: 241–255.
- Braconnot P, Joussaume S, De Noblet N, Ramstein G (2000). Mid-Holocene and last glacial maximum African monsoon changes as simulated within the paleoclimate modelling intercomparison project. *Glob Planet Change* 26: 51–66.
- Bryan GJ, Mcnicoll J, Ramsay G, Meyer RC, De Jong WS (1999). Polymorphic simple sequence repeat markers in chloroplast genomes of solanaceous plants. *TAG Theor Appl Genet* 99: 859–867.
- Busby JR (1991). Bioclim—a bioclimate analysis and prediction system. In: CRaA Margules MP (ed) *Nature Conservation: Cost Effective Biological Surveys and Data Analysis*. CSIRO: Melbourne.
- Cavalli-Sforza L (1967). Phylogenetic analysis: models and estimation procedures. *Am J Hum Genet* 19: 233–257.
- Cerling TE, Harris JM, Macfadden BJ, Leakey MG, Quade J, Eisenmann V *et al.* (1997). Global vegetation change through the Miocene/Pliocene boundary. *Nature* 389: 153–158.
- Clark PU, Dyke AS, Shakun JD, Carlson AE, Clark J, Wohlfarth B *et al.* (2009). The last glacial maximum. *Science* 325: 710–714.
- Collins WD, Blackmon ML, Bonan GB, Hack JJ, Henderson TB, Kiehl JT *et al.* (2006). The community climate system model version 3 (ccsm3). *J Climate* 19: 2122–2143.
- Comps B, Gomory D, Letouzey J, Thiebaut B, Petit RJ (2001). Diverging trends between heterozygosity and allelic richness during postglacial colonization in the european beech. *Genetics* 157: 389–397.
- Dauby G, Duminil J, Heuertz M, Hardy O (2010). Chloroplast DNA polymorphism and phylogeography of a central African tree species widespread in mature rainforests: greenwayodendron suaveolens (Annonaceae). *Trop Plant Biol* 3: 4–13.
- Davrieux F, Allal F, Piombo G, Kelly B, Okulo JB, Thiam M *et al.* (2010). Near infrared spectroscopy for high-throughput characterization of shea tree (*Vitellaria paradoxa*) nut fat profiles. *J Agric Food Chem* 58: 7811–7819.
- Dick C (2010). Phylogeography and population structure of tropical trees. *Trop Plant Biol* 3: 1–3.
- Dick CW, Heuertz M (2008). The complex biogeographic history of a widespread tropical tree species. *Evolution* 62: 2760–2774.
- Dumolin-Lapegue S, Pemonge M-H, Petit RJ (1997). An enlarged set of consensus primers for the study of organelle DNA in plants. *Mol Ecol* 6: 393–397.
- Evanno G, Regnaut S, Goudet J (2005). Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol Ecol* 14: 2611–2620.
- Excoffier L, Laval G, Schneider S (2005). Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinform* 1: 47–50.
- Felsenstein J (1993). PHYLIP (*Phylogenetic inference package*). Department of Genetics, University of Washington, Seattle.
- Fielding AH, Bell JF (1997). *A Review of Methods for the Assessment of Prediction Errors in Conservation Presence/Absence Models*. Cambridge University Press: Cambridge, Royaume-Uni.
- Flenley JR (1998). Tropical forests under the climates of the last 30,000 years. *Climatic Change* 39: 177–197.
- 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.
- Goudet J (1995). Fstat (version 1.2): a computer program to calculate f-statistics. *J Hered* 86: 485.
- Hall JB, Aebischer DP, Tomlinson HF, Osei-Amaning E, Hindle JR (1996). *Vitellaria Paradoxa: A Monograph*. School of Agricultural and Forest Sciences, University of Wales Bangor: UK.
- Hamrick JL, Murawski DA, Nason JD (1993). The influence of seed dispersal mechanisms on the genetic structure of tropical tree populations. *Plant Ecol* 107–108: 281–297.
- Hanley J, Mcneil B (1982). The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology* 143: 29.

- Hardy OJ, Vekemans X (2002). Spagedi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol Ecol Notes* 2: 618–620.
- Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005). Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol* 25: 1965–1978.
- Hijmans RJ, Guarino L, Cruz M, Rojas E (2001). Computer tools for spatial analysis of plant genetic resources data: 1. Divagis. *Plant Genet Resour Newsl* 127: 15–19.
- Jaramillo-Correa JP, Aguirre-Planter E, Khasa DP, Eguiarte LE, Pinero D, Furnier GR et al. (2008). Ancestry and divergence of subtropical montane forest isolates: molecular biogeography of the genus *Abies* (Pinaceae) in southern Mexico and Guatemala. *Mol Ecol* 17: 2476–2490.
- Johnson LA, Soltis DE (1994). *matK* DNA sequences and phylogenetic reconstruction in Saxifragaceae s. str. *Syst Bot* 19: 143–156.
- Ky CL, Barre P, Lorieux M, Trouslot P, Akaffou S, Louarn J et al. (2000). Interspecific genetic linkage map, segregation distortion and genetic conversion in coffee (*Coffea* sp.). *Theor Appl Genet* 101: 669–676.
- Lioubimtseva E, Simon B, Faure H, Faure-Denard L, Adams JM (1998). Impacts of climatic change on carbon storage in the Sahara-Gobi desert belt since the Last Glacial Maximum. *Glob Planet Change* 16–17: 95–105.
- Lovett P, Haq N (2000). Evidence for anthropic selection of the sheanut tree (*Vitellaria paradoxa*). *Agroforestry Syst* 48: 273–288.
- Lowe A, Harris D, Dormontt E, Dawson I (2010). Testing putative African tropical forest refugia using chloroplast and nuclear DNA phylogeography. *Trop Plant Biol* 3: 50–58.
- Maley J (1989). Late quaternary climatic changes in the African rain forest: forest refugia and the major role of sea surface temperature variations. In: SME Leinen M (ed). *Paleoclimatology and Paleometeorology: Modern and Past Patterns of Global Atmospheric Transport*. NATO Advanced Sc. Inst. Series, C, Math. & Phys. Sc., 282. Kluwer: Dordrecht. pp 585–616.
- 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 Edinb B* 104: 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.
- Maranz S, Wiesman Z, Bisgaard J, Bianchi G (2004). Germplasm resources of *Vitellaria paradoxa* based on variations in fat composition across the species distribution range. *Agroforest Syst* 60: 71–76.
- Morley R, Richards K (1993). Gramineae cuticle: a key indicator of Late Cenozoic climatic change in the Niger delta. *Rev Palaeobot Palynol* 77: 119–127.
- Müller F, Voccia M, B A, Bouvet J (2009). Genetic diversity and gene flow in a Caribbean tree *Pterocarpus officinalis* Jacq.: a study based on chloroplast and nuclear microsatellites. *Genetica* 135: 185–198.
- Navascués M, Emerson BC (2005). Chloroplast microsatellites: measures of genetic diversity and the effect of homoplasy. *Mol Ecol* 14: 1333–1341.
- Neumann K, Kahlheber S, Uebel D (1998). Remains of woody plants from Saouga, a medieval West African village. *Veg Hist Archaeobot* 7: 57–77.
- Otto-Bliesner BL, Marshall SJ, Overpeck JT, Miller GH, Hu A, Members CLIP (2006). Simulating arctic climate warmth and icefield retreat in the last interglaciation. *Science* 311: 1751–1753.
- Parks DH, Porter M, Churcher S, Wang S, Blouin C, Whalley J et al. (2009). Gengis: a geospatial information system for genomic data. *Genome Res* 19: 1896–1904.
- Pearce J, Ferrier S (2000). Evaluating the predictive performance of habitat models developed using logistic regression. *Ecol Modell* 133: 225–245.
- Petit RJ, Aguinalde I, De Beaulieu J-L, Bittkau C, Brewer S, Cheddadi R et al. (2003). Glacial refugia: hotspots but not melting pots of genetic diversity. *Science* 300: 1563–1565.
- Petit RJ, Hu FS, Dick CW (2008). Forests of the past: a window to future changes. *Science* 320: 1450–1452.
- Petit RJ, Mousadik AE, Pons O (1998). Identifying populations for conservation on the basis of genetic markers. *Conserv Biol* 12: 844–855.
- Pock Tsy J-ML, Lumaret R, Mayne D, Vall AOM, Abutaba YIM, Sagna M et al. (2009). Chloroplast DNA phylogeography suggests a West African centre of origin for the baobab, *Adansonia digitata* L. (Bombacoideae, Malvaceae). *Mol Ecol* 18: 1707–1715.
- Pons O, Petit RJ (1995). Estimation, variance and optimal sampling of gene diversity. *TAG Theor Appl Genet* 90: 462–470.
- Pons O, Petit RJ (1996). Measuring and testing genetic differentiation with ordered versus unordered alleles. *Genetics* 144: 1237–1245.
- Poumot C (1989). Palynological evidence for eustatic events in the tropical Neogene. *Bulletin des Centres de Recherches Exploration-Production Elf Aquitaine, Boussems* 13: 437–453.
- Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.
- Ramos ACS, De Lemos-Filho JP, Lovato MB (2009). Phylogeographical structure of the neotropical forest tree *Hymenaea courbaril* (Leguminosae: Caesalpinioideae) and its relationship with the vicariant *Hymenaea stigonocarpa* from Cerrado. *J Hered* 100: 206–216.
- Ramos ACS, Lemos-Filho JP, Ribeiro RA, Santos FR, Lovato MB (2007). Phylogeography of the tree *Hymenaea stigonocarpa* (Fabaceae: Caesalpinioideae) and the influence of Quaternary climate changes in the Brazilian Cerrado. *Ann Bot* 100: 1219–1228.
- Saitou N, Nei M (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4: 406.
- Sankaran M, Hanan NP, Scholes RJ, Ratnam J, Augustine DJ, Cade BS et al. (2005). Determinants of woody cover in African savannas. *Nature* 438: 846–849.
- Sanou H, Lovett PN, Bouvet JM (2005). Comparison of quantitative and molecular variation in agroforestry populations of the shea tree (*Vitellaria paradoxa* c.f. Gaertn) in Mali. *Mol Ecol* 14: 2601–2610.
- Soria-Auza RW, Kessler M, Bach K, Barajas-Barbosa PM, Lehnert M, Herzog SK et al. (2010). Impact of the quality of climate models for modelling species occurrences in countries with poor climatic documentation: a case study from Bolivia. *Ecol Modell* 221: 1221–1229.
- Vendramin GG, Fady B, Gonzalez-Martinez SC, Hu FS, Scotti I, Sebastiani F et al. (2008). Genetically depauperate but widespread: the case of an emblematic Mediterranean pine. *Evolution* 62: 680–688.
- Vigouroux Y, Jaqueth JS, Matsuoka Y, Smith OS, Beavis WD, Smith JSC et al. (2002). Rate and pattern of mutation at microsatellite loci in maize. *Mol Biol Evol* 19: 1251–1260.
- 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.
- Wolfe KH, Li WH, Sharp PM (1987). Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc Natl Acad Sci USA* 84: 9054–9058.

Supplementary Information accompanies the paper on Heredity website (<http://www.nature.com/hdy>)