Genet Resour Crop Evol (2010) 57:995–1005
DOI 10.1007/s10722-010-9539-7

RESEARCH ARTICLE

Investigation of recent population bottlenecks in Kenyan wild sorghum populations (*Sorghum bicolor* (L.) Moench ssp. *verticilliflorum* (Steud.) De Wet) based on microsatellite diversity and genetic disequilibria

M. M. Muraya · F. Sagnard · H. K. Parzies

Received: 17 October 2009 / Accepted: 25 January 2010 / Published online: 18 February 2010

Abstract Identifying populations that have recently suffered a severe reduction in size is particularly important for their conservation as they are likely to suffer an increased risk of genetic erosion. We investigated the presence of recent bottlenecks in two wild sorghum populations from different eco-geographical conditions in Kenya employing 18 microsatellite markers. Microsatellite analysis showed high allelic diversity in the two populations, with a mean of 4.11 and 6.94 alleles per locus in the North-West wild sorghum population (NWWSP) and the South-East wild sorghum population (SEWSP), respectively. The mean observed heterozygosity was 0.34 and 0.56 in NWWSP and SEWSP, respectively. A large long-term effective populations size for both populations was observed assuming either an infinite allele model or a stepwise mutation model. There was no apparent loss of genetic variability for either of the populations. Test of heterozygosity excess indicated that a recent bottleneck in the two populations is highly unlikely. Furthermore, analysis of the allele frequency distribution revealed an L-shaped distribution which would not have been observed in case a recent bottleneck had reduced genetic variability in the two populations. The fact that most loci displayed a significant heterozygosity deficiency could be explained by population subdivision and the mixed mating system exhibited by wild sorghum populations. Furthermore, the possibility of a historical expansion of wild sorghum populations and presence of null alleles could not be ruled out.

Keywords Bottleneck · Genetic diversity · Linkage disequilibrium · Microsatellite markers · Null allele · *Sorghum bicolor* ssp. *verticilliflorum* · Wild sorghum population

M. M. Muraya · H. K. Parzies
Institute of Plant Breeding, Seed Science and Population Genetics, University of Hohenheim, Fruwirthstrasse 21, 70599 Stuttgart, Germany

M. M. Muraya (✉)
Leibniz Institute of Plant Genetics and Crop Plant Research, Corrensstrasse 3, 06466 Gatersleben, Germany
e-mail: mahugu2002@yahoo.com

F. Sagnard
CIRAD-UMR Développement et Amélioration des Plantes, c/o ILRI, PO Box 30709, Nairobi, Kenya

F. Sagnard
International Crops Research Institute for the Semi-Arid Tropics (ICRISAT-Nairobi), P.O. Box 39063-00623, Nairobi, Kenya

Introduction

Population bottlenecks can cause increased rates of inbreeding, loss of genetic variation and fixation of mildly deleterious alleles, and thereby reduce the adaptive potential and increase the risk of extinction due to lack of adaptive flexibility (Mills and Smouse 1994; Frankham 1998; Keller and Waller 2002).

Furthermore, identification and understanding population bottlenecks is important in evolutionary biology. A founder effect can occur if small populations become reproductively isolated from the main population, leading to speciation.

Due to the ever-increasing anthropogenic changes to natural ecosystems, it is becoming increasingly important to monitor unintended effects of such changes on natural populations. Many national and international organisations have established principles and strategies for monitoring biodiversity (UNEPCD 2003). However, little use has been made of the benefits afforded by molecular genetic markers, which are more sensitive and reliable than traditional methods such as phenotypic traits (Schwartz et al. 2007). New laboratory techniques and statistical methods are now available to enable the use of molecular markers for genetic monitoring of wild populations (Cornuet and Luikart 1997; Luikart et al. 2003; Beaumont and Rannala 2004). Historical populations' sizes are rarely known, therefore, methods for detecting bottlenecks in the absence of ancient historical data are useful. Cornuet and Luikart (1997) described and evaluated methods for detecting recent historical bottlenecks using allele frequency data. These methods utilize the level of polymorphism detectable by molecular markers (e.g., microsatellites) and the theory of the effects of bottlenecks on the loss of alleles and heterozygosity at selectively neutral loci (Nei et al. 1975; Maruyama and Fuerst 1985).

Neutral genetic markers have proven very useful when describing genetic diversity both within and among populations, and infer their demographic history (Saillant et al. 2004; Goossens et al. 2006). The amount of neutral genetic variation in species is due to two primary factors: (1) drift, which decreases variability and (2) mutation, which increase it (Kimura 1983). A population that experiences a dramatic reduction in size will lose genetic variation as a function of its population size, growth rate and the duration of the population contraction (Nei et al. 1975). However, the separations of ancient and recent demographic events is crucial for an efficient management of endangered populations or species (Chikhi and Bruford 2005). For instance, a low genetic diversity could be the result of small long-term effective population size.

The loss of genetic variability during a recent bottleneck event can be identified by examining the current population for evidence of heterozygosity excess at neutral loci (Luikart and Cornuet 1998; Luikart et al. 1998a; Spencer et al. 2000). When a bottleneck occurs in a population, allelic diversity is reduced faster than is heterozygosity (Nei et al. 1975). Consequently, a transient excess of heterozygosity occurs (Maruyama and Fuerst 1985). This condition occurs due to loss of rare alleles from the population since rare alleles that were lost contributed little to the overall heterozygosity. Populations that have recently lost genetic diversity may be identified by testing for a heterozygosity excess irrespective of severity of the bottleneck or pre-bottleneck levels of genetic variability (Cornuet and Luikart 1997).

Wild sorghum populations, the progenitors of cultivated sorghum, are found in Africa. They are classified together with cultivated sorghum into the species *Sorghum bicolor* (L.) Moench (de Wet 1978). The species is further subdivided into *S. bicolor* ssp. *bicolor* (cultivated sorghum) and *S. bicolor* (L.) Moench ssp. *verticilliflorum* (Steud.) De Wet (wild sorghum). Sorghum domestication began $\sim$5,000 years ago at the Northern-east quadrant of Africa (Mann et al. 1983). Since the transition from wild species to domesticates, cultivated sorghum has continued to evolve due to forces of evolution, i.e., selection, migration, mutation and genetic drift. As a result of both sorghum domestication and evolution, today's cultivated sorghum are differentiated from their wild progenitors by an assortment of morphological characteristics, called the domestication syndrome.

Wild sorghum forms an important genetic resource, but due to changing agricultural and natural ecosystems it may be at risk of genetic erosion. Wild sorghum populations are rapidly being depleted and are increasingly under threat. The rapid increase of human and livestock populations have resulted in the loss of natural habitats of wild sorghum populations as they are successively being converted to farmland. Unlike traditional agriculture, modern agricultural practices have made it difficult to maintain large fallow land where wild sorghum populations can freely grow and replenish. These factors dictate that the population sizes will never grow very large. Thus, it is expected that small wild sorghum populations will result. Consequently, resulting in inbreeding, loss of genetic variation, fragmentation, reduction in genetic adaptation and eventually extinction, which are the critical issues in conservation and management of endangered populations.

Hence, wild sorghum populations may require human intervention to save them from extinction or at least to optimize their management. Intervention may take the form of habitat protection, habitat restoration and ex-situ conservation. For effective conservation of wild sorghum populations, frequent monitoring of its genetic changes is needed. It is useful to predict the patterns of allelic loss and thus genetic erosion. To identify conservation objectives properly, it is important to identify populations that have lost genetic variability recently, as they may be more susceptible to demographic stochasticity (Mills and Smouse 1994). The objective of this study was to investigate if recent population bottlenecks occurred in wild sorghum populations in Kenya. We used microsatellite markers because they are useful for estimating long-term effective population size (N_e) and detecting populations that have undergone a recent bottleneck event (Cornuet and Luikart 1997; Spencer et al. 2000). Estimates of N_e in natural populations are important in order to evaluate whether threatened populations are able to maintain sufficient genetic variation to adapt to future environmental changes (Franklin 1980).

Materials and methods

Two wild sorghum populations representing different geographical and ecological ranges of wild sorghum in Kenya were used in this study. One population, designated North-West wild sorghum population (NWWSP), was collected at the North-West part of Kenya which is within the region of the centre of sorghum origin and diversity (Fig. 1). The second population, designated South-East wild sorghum population (SEWSP), was collected at the South-East part of Kenya which forms a region of wild sorghum diffusion from its centre of origin. For each population 24 individuals were sampled.

DNA was extracted from lyophilized leaf tissues using a modified CTAB protocol (Mace et al. 2003). The concentration and quality of the DNA was assessed using electrophoresis of 1 µl of DNA on 0.7% agarose gel followed by normalization of the concentration at 50 ng/µl.

Eighteen unlinked microsatellite loci were employed to estimate allele frequencies and the level of heterozygosity in wild sorghum populations (Table 1). Forward primers were labelled with FAM, HEX, or TET,

Fig. 1 Sampling sites (encircled) of the two wild sorghum populations: North-West wild sorghum population (NWWSP) and South-East wild sorghum population (SEWSP)

allowing PCR multiplexing of the 18 primer products into six groups of three differently fluorescent labelled primer products. PCR reactions were performed in a volume of 20 µl each containing 1 × PCR buffer (20 mM Tris–HCl (pH 8.4), 50 mM KCl), 1.5 mM $MgCl_2$, 0.25 µM of each primer, 0.2 mM dNTPs, 0.5 unit per reaction of Taq polymerase and 2 µl template DNA.

The amplification reaction consisted of a denaturing step of 3 min at 94°C, followed by 40 cycles beginning with 94°C for 1 min, annealing reaction of 1 min at 55 or 62°C (depending on microsatellite primers; Table 1), extension at 72°C for 1 min followed by one terminal step at 72°C for 10 min and consecutive storage at 4°C. All PCR reactions were performed on a MJResearch icl PTC-100 thermocycler. Fragment analysis was carried out on an automated laser fluorescence sequencer (MegaBACE).

Statistical analysis

Polymorphic information content (PIC) and unbiased heterozygosity (Nei 1987) were estimated for each locus. Fisher's exact test, as implemented in GENEPOP 4.0 (Raymond and Rousset 1995), was used to determine if populations were in Hardy–Weinberg equilibrium. In addition, exact tests of heterozygote

Table 1 Microsatellite markers used in the genotyping work

Microsatellite name	Motif	Forward primer (5′–3′)	Reverse primer (5′–3′)	T_a[6] (°C)	Size range (bp)
Xtxp057[1]	(GT)21	GGAACTTTTGACGGGTAGTGC	CGATCGTGATGTCCCAATC	55	199–223
Xtxp136 (kaf3)[1]	(GCA)5	GCGAATAGCATCTTACAACA	ACTGATCATTGGCAGGAC	55	240–243
Xtxp145[1]	(AG)22	GTTCCTCCTGCCATTACT	CTTCCGCACATCCAC	55	208–244
Xtxp273[1]	(TTG)20	GTACCCATTTAAATTGTTTGCAGTAG	CAGAGGAGGAGGAAGAGAAGG	55	169–199
Xcup14[2]	(AG)10	TACATCACAGCAGGGACAGG	CTGGAAAGCCGAGCAGTATG	62	211–225
Xtxp015[1]	(TC)16	CACAAACACTAGTGCCTTATC	CATAGACACCTAGGCCATC	55	199–223
mSbCIR262[3]	(CATG)3.25	GCACCAAAATCAGCGTCT	CCATTTACCCGTGGATTAGT	62	208–220
sb6-84 = Xgap94[4]	(AG)14	CGCTCTCGGGATGAATGA	TAACGGACCACTAACAAATGATT	55	183–217
Xcup61[2]	(CAG)7	TTAGCATGTCCACCACAACC	AAAGCAACTCGTCTGATCCC	55	198–201
Xcup02[2]	(GCA)6	GACGCAGCTTTGCTCCTATC	GTCCAACCAACCCACGTATC	55	192–204
Xtxp012[1]	(CT)22	AGATCTGGCGGCAACG	AGTCACCCATCGATCATC	55	161–205
Xcup53[2]	(TTTA)5	GCAGGAGTATAGGCAGAGGC	CGACATGACAAGCTCAAACG	62	186–198
Xcup62[2]	(GAA)6	CGAGAAGATCGAGAGAACCC	TGAAGACGACGACGACAGAC	55	190–193
Xtxp021[1]	(AG)18	GAGCTGCCATAGATTTGGTCG	ACCTCGTCCCACCTTTGTTG	55	169–199
Xtxp141[1]	(GA)23	TGTATGGCCTAGCTTATCT	CAACAAGCCAACCTAAA	55	135–167
SbAGB02[5]	(AG)35	CTCTGATATGTCGTTGTGCT	ATAGAGAGGATAGCTTATAGCTCA	55	96–154
Xcup63[2]	(GGATGC)4	GTAAAGGGCAAGGCAACAAG	GCCCTACAAAATCTGCAAGC	55	133–145
xtxp040[1]	(GGA)7	CAGCAACTTGCACTTGTC	GGGAGCAATTTGGCACTAG	55	129–141

[1] Kong et al. (2000); [2] Schloss et al. (2002); [3] Unpublished, Agropolis-CIRAD-Genoplante; [4] Brown et al. (1996); [5] Taramino et al. (1997); [6] T_a = annealing temperature

deficiency and excess were conducted (Raymond and Rousset 1995). The program LINKDIS (Garnier-Gere and Dillmann 1992), as implemented in GENEPOP 4.0, was used to examine gametic phase disequilibrium among the microsatellite loci. GENEPOP 4.0 was also used to estimate the frequency of null alleles. Genepop uses an iterative EM algorithm to find the maximum likelihood estimate of null allele frequencies (Dempster et al. 1977). Long-term effective population size was calculated from estimates of unbiased expected heterozygosity under both the infinite alleles (IAM) and the stepwise mutation (SMM) model (Nei 1987; Lehmann et al. 1998) as shown below:

$$\text{IAM} : N_e = \frac{H_e}{4\mu(1 - H_e)} \quad (1)$$

$$\text{SMM} : N_e = \frac{[1/(1 - H_e)]^2}{8\mu} \quad (2)$$

where N_e is the effective population size, H_e is the expected heterozygosity and μ is the mutation rate. A mutation rate of 2.34×10^{-4} was anticipated in our study, based on the mutation rate of microsatellites reported for a related *Poaceae* species *Triticum turgidum* (L.) ssp. *durum* (Desf.) Husn. (Thuillet et al. 2002).

The program BOTTLENECK 1.2.02 was used to investigate the presence of recent bottlenecks in wild sorghum populations according to Cornuet and Luikart (1997). The program tests for departure from mutation-drift equilibrium based on heterozygosity excess or deficiency under both the IAM and SMM models. For each population sample and for each locus the distribution of heterozygosity expected from the observed number of alleles (k), given a sample size (n) under the assumption of mutation-drift equilibrium was computed. This distribution was obtained through simulating the coalescent process of n genes under both the IAM and SMM model. This enabled computation of the unbiased heterozygosity to establish weather there is heterozygosity excess or deficiency at each locus. The average and standard deviations of these distributions were estimated, as well as the

proportion of heterozygosity values that were larger than the average. The distribution obtained through simulation further enables the computation of a probability value for the observed heterozygosity. After all loci in the population sample were processed, significance was determined by Sign and Wilcoxon statistical tests as explained in Cornuet and Luikart (1997) and Luikart et al. (1998b). Moreover, using Bottleneck 1.2.02, the allele frequency distribution of the microsatellite loci was examined for a model shift (Luikart et al. 1998b), which indicates whether a recent genetic bottleneck has occurred or not.

Results

A total of 74 alleles with a mean of 4.11 alleles per locus were observed in NWWSP, while a total of 125 alleles with a mean of 6.94 alleles per locus were observed in SEWSP (Table 2). Overall, the observed heterozygosity ranged from 0.04 to 0.67 with a mean of 0.34 in NWWSP and 0.08 to 0.93 with a mean of 0.56 in SEWSP. The NWWSP displayed relatively low heterozygosity compared to the SEWSP. All microsatellites scored were polymorphic, displaying PIC values ranging from 0.04 to 0.62 with a mean of 0.31 in NWWSP and 0.08 to 0.91 with a mean of 0.51 in SEWSP. PIC values were slightly lower in comparison to observed heterozygosity (Table 2). The frequency of null alleles was generally high in both populations (Table 2). It ranged from 0.00 to 0.98 with a mean of 0.39 in NWWSP and ranging from 0.06 to 0.98 with a mean of 0.36 in SEWSP. The number of loci showing evidence of heterozygote deficiency, Hardy–Weinberg equilibrium and heterozygote excess in NWWSP were 12, 4, and 2, respectively. In SEWSP, 17 and 1 loci showed evidence of heterozygote deficiency and Hardy–Weinberg equilibrium, respectively (Table 2). None of the loci in SEWSP displayed heterozygote excess. Gametic phase disequilibrium was significant in 77 and 124 in NWWSP and SEWSP pairwise comparisons of a total of 152, respectively (Table 3).

Table 2 Genetic diversity of wild sorghum populations

Locus	NWWSP					SEWSP				
	T_A[1]	E_t[2]	N_A[3]	H_O[4]	PIC[5]	T_A[1]	E_t[2]	N_A[3]	H_O[4]	PIC[5]
mSbCIR262	2	D*	0.20	0.08	0.08	8	D*	0.34	0.85	0.81
sb6-84	5	D*	0.38	0.64	0.57	9	D*	0.29	0.44	0.42
SbAGB02	10	D*	0.20	0.60	0.56	19	D*	0.28	0.93	0.91
Xcup02	2	E*	0.98	0.04	0.04	2	D*	0.20	0.08	0.08
Xtxp012	4	HWE*	0.00	0.12	0.12	4	D*	0.32	0.61	0.53
Xcup14	3	D*	0.55	0.58	0.50	9	D*	0.26	0.79	0.75
Xcup53	3	HWE*	0.00	0.23	0.21	4	D*	0.35	0.57	0.46
Xcup61	3	D*	0.25	0.20	0.18	14	D*	0.20	0.91	0.87
Xcup62	4	D*	0.27	0.30	0.27	5	D*	0.22	0.40	0.37
Xcup63	3	D*	0.14	0.08	0.08	2	D*	0.98	0.08	0.08
Xtxp015	3	D*	0.26	0.50	0.42	6	D*	0.33	0.72	0.65
Xtxp021	7	E*	0.47	0.64	0.56	9	D*	0.28	0.75	0.70
xtxp040	5	D*	0.39	0.62	0.53	5	D*	0.27	0.3	0.29
Xtxp057	3	HWE*	0.94	0.12	0.12	6	D*	0.34	0.57	0.50
Xtxp136	4	D*	0.31	0.38	0.34	4	HWE*	0.06	0.30	0.27
Xtxp141	5	D*	0.22	0.24	0.22	6	D*	0.76	0.62	0.57
Xtxp145	2	HWE*	0.96	0.12	0.11	6	D*	0.67	0.70	0.64
Xtxp273	6	D*	0.43	0.67	0.62	7	D*	0.26	0.37	0.36
Mean	4.11		0.39	0.34	0.31	6.94		0.36	0.56	0.51

[1] T_A = total number of alleles; [2] E_t = Exact test (*D* heterozygosity deficiency; *HWE* Hardy–Weinberg equilibrium; *E* heterozygosity excess); [3] N_A = frequency of null alleles; [4] H_O = observed heterozygosity; [5] PIC = Polymorphic information content

* $P < 0.05$

Table 3 Correlation coefficients resulting from test of gametic phase disequilibrium: NWWSP in upper diagonal and SEWSP in lower diagonal

Loci[1]	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1		0.15*	0.07	0.05	0.05	0.14	0.1	0.06	0.07	0.05	0.13	0.16*	0.11	0.06	0.08	0.05	0.06	0.17**
2	0.24**		0.27**	0.15	0.15*	0.23**	0.12	0.12	0.14*	0.15	0.21**	0.60**	0.16**	0.15*	0.17**	0.12	0.18*	0.18
3	0.20**	0.21**		0.86**	0.53**	0.15**	0.22**	0.52**	0.1	0.08	0.21**	0.37**	0.14**	0.15**	0.44**	0.37**	0.44**	0.12
4	0.13**	0.05	0.10**		0.94**	0.14	0.44**	0.27*	0.07	0.05	0.13	0.74*	0.13	0.06	0.08	0.86**	0.93**	0.09
5	0.23**	0.38**	0.25**	0.26**		0.27**	0.29**	0.13	0.08	0.05	0.12	0.35**	0.09	0.60**	0.13	0.38*	0.67**	0.38**
6	0.21**	0.31**	0.19**	0.18**	0.22**		0.16	0.19	0.24**	0.25**	0.17	0.19*	0.19**	0.19*	0.18*	0.19**	0.12	0.26**
7	0.19**	0.56**	0.22**	0.15	0.24**	0.19**		0.12	0.14	0.32*	0.15	0.2	0.15	0.44**	0.14	0.18	0.37*	0.19*
8	0.17**	0.22**	0.30**	0.19**	0.27**	0.20**	0.27**		0.09	0.06	0.1	0.12	0.1	0.09	0.60**	0.1	0.23**	0.13*
9	0.18**	0.36**	0.19**	0.08	0.14*	0.31**	0.14	0.22**		0.94**	0.19**	0.09	0.47**	0.09	0.17**	0.15**	0.09	0.26**
10	0.1	0.05	0.39**	0.05	0.11	0.1	0.15	0.14**	0.08		0.13	0.12	0.11	0.06	0.08	0.05	0.06	0.23**
11	0.20**	0.21**	0.18**	0.36**	0.30**	0.18**	0.22**	0.22**	0.18**	0.13		0.25**	0.19**	0.12	0.19**	0.38**	0.09	0.16*
12	0.19**	0.48**	0.17**	0.63**	0.26**	0.27**	0.45**	0.16**	0.14**	0.09	0.21**		0.14*	0.13	0.18**	0.28*	0.34**	0.16**
13	0.18**	0.47**	0.18**	0.06	0.17**	0.31**	0.44**	0.24**	0.58**	0.06	0.22**	0.15**		0.13	0.46**	0.36**	0.13	0.20**
14	0.20**	0.43**	0.19**	0.08	0.32**	0.22**	0.60**	0.30**	0.14*	0.08	0.24**	0.41**	0.38**		0.12	0.08	0.09	0.45**
15	0.18**	0.11	0.21**	0.08	0.17	0.24**	0.19	0.14*	0.41**	0.08	0.13	0.18*	0.43**	0.17		0.40**	0.08	0.14**
16	0.19**	0.28**	0.25**	0.09	0.25**	0.24**	0.42**	0.29**	0.15**	0.79**	0.23**	0.22**	0.38**	0.50**	0.15		0.39**	0.11
17	0.20**	0.36**	0.21**	0.1	0.29**	0.32**	0.38**	0.19**	0.15**	0.22**	0.20**	0.47**	0.16**	0.37**	0.23**	0.31**		0.09
18	0.16**	0.52**	0.24**	0.06	0.17**	0.36**	0.52**	0.25**	0.60**	0.06	0.19**	0.40**	0.60**	0.41**	0.37**	0.31**	0.40**	

[1] Microsatellite markers: 1 = mSbCIR262, 2 = sb6-84, 3 = SbAGB02, 4 = Xcup02, 5 = Xtxp012, 6 = Xcup14, 7 = Xcup53, 8 = Xcup61, 9 = Xcup62, 10 = Xcup63, 11 = Xtxp015, 12 = Xtxp021, 13 = xtxp040, 14 = Xtxp057, 15 = Xtxp136, 16 = Xtxp141, 17 = Xtxp145 and 18 = Xtxp273

**, * Correlation coefficients significant at $P = 0.01$ and $P = 0.05$, respectively

Average long-term effective population size (N_e) in NWWSP was 898 and 2,176 for IAM and SMM, respectively, whereas corresponding values in SEWSP were 1,709 and 5,674 (Table 4). The estimates of long-term N_e were larger for SMM relative to the IAM in both population. Paired-samples Wilcoxon signed ranked test (SEWSP (N_e)–NWWSP (N_e) for each locus) were highly significant for N_e derived by either IAM or SMM ($P = 0.001$ and $P = 0.0004$, respectively).

On a locus specific basis, most of the loci showed evidence of heterozygosity deficiency (Table 5). The sign and Wilcoxon tests on the difference (observed—expected heterozygosity) of heterozygosities across all loci in the populations sample are presented in Table 4. Under both mutation models, the results indicated that the deviation from mutation-drift equilibrium was not upheld using either the sign test or the Wilcoxon test (Table 5) in both populations. Furthermore, considering both mutation models, results indicated that there was a heterozygosity deficiency resulting from too many alleles, using either the sign test or the Wilcoxon test (Table 5) for both populations. In both populations, the expected number of loci displaying heterozygosity excess was substantially greater than the observed number of loci displaying a heterozygosity excess. On the other hand, they were substantially less than the observed number of loci displaying a heterozygosity deficiency (Table 5). Although the results of the test for standardized difference under both models are in agreement with the sign and Wilcoxon tests they are not reported here because at least 20 polymorphic loci are required for the standardized differences test (Cornuet and Luikart 1997). An analysis of allele frequency distribution (Figs. 2, 3) revealed an L-shaped distribution.

Table 4 Long-term effective population size estimates for wild sorghum populations

Locus	NWWSP[1]				SEWSP[2]			
	IAM[3]		SMM[4]		IAM[3]		SMM[4]	
	H_e[5]	N_e[6]	H_e[5]	N_e[6]	H_e[5]	N_e[6]	H_e[5]	N_e[6]
mSbCIR262	0.22	294	0.27	455	0.73	2918	0.82	15771
sb6-84	0.57	1440	0.69	5061	0.77	3497	0.85	21990
SbAGB02	0.79	4068	0.86	28347	0.92	12629	0.94	158261
Xcup02	0.23	310	0.28	485	0.21	286	0.26	441
Xtxp012	0.49	1035	0.61	2960	0.48	990	0.61	2978
Xcup14	0.37	619	0.49	1520	0.77	3557	0.84	21416
Xcup53	0.37	622	0.49	1504	0.48	1002	0.61	3014
Xcup61	0.37	622	0.48	1472	0.88	7618	0.91	70042
Xcup62	0.49	1039	0.61	3014	0.57	1422	0.69	5061
Xcup63	0.38	660	0.48	1464	0.22	298	0.26	431
Xtxp015	0.37	633	0.49	1480	0.65	1958	0.75	8151
Xtxp021	0.69	2378	0.79	11240	0.76	3346	0.84	20864
xtxp040	0.58	1451	0.69	5061	0.57	1428	0.69	4953
Xtxp057	0.38	652	0.48	1426	0.64	1899	0.75	7746
Xtxp136	0.49	1026	0.61	2996	0.48	971	0.62	3070
Xtxp141	0.57	1434	0.69	5134	0.63	1835	0.75	7945
Xtxp145	0.23	319	0.26	428	0.63	1827	0.75	7878
Xtxp273	0.63	1843	0.75	7746	0.69	2378	0.79	11240
Mean	0.46	898	0.56	2176	0.62	1709	0.71	5674
SE	0.04	218	0.04	1551	0.05	702	0.04	8925

[1] NWWSP = North-West wild sorghum population; [2] SEWSP = South-East wild sorghum population; [3] IAM = infinite alleles model; [4] SMM = stepwise mutation model; [5] H_e = expected heterozygosity; [6] N_e = long-term effective population size

Table 5 Test for population bottlenecks under the mutation-drift equilibrium hypothesis

Model		NWWSP	SEWSP
IAM	Expected loci with heterozygosity excess	9.7	10.42
	Observed loci with heterozygosity excess	5	8
	Observed loci with heterozygosity deficiency	13	10
	Sign test (Heterozygosity excess)[a]	$P = 0.97769$	$P = 0.82208$
	Sign test (Heterozygosity deficiency)[a]	$P = 0.02231$	$P = 0.17792$
SMM	Expected loci with heterozygosity excess	10.3	10.31
	Observed loci with heterozygosity excess	2	2
	Observed loci with heterozygosity deficiency	16	16
	Sign test (Heterozygosity excess)[a]	$P = 0.99994$	$P = 0.99995$
	Sign test (Heterozygosity deficiency)[a]	$P = 0.00006$	$P = 0.00005$
IAM	Wilcoxon test (two-tailed)[b]	$P = 0.01041$	$P = 0.26453$
	Wilcoxon test (one-tailed H excess)[b]	$P = 0.99552$	$P = 0.8769$
	Wilcoxon test (one-tailed H deficiency)[b]	$P = 0.0052$	$P = 0.13226$
SMM	Wilcoxon test (two-tailed)[b]	$P = 0.00004$	$P = 0.00005$
	Wilcoxon test (one-tailed H excess)[b]	$P = 0.99999$	$P = 0.99998$
	Wilcoxon test (one-tailed H deficiency)[b]	$P = 0.00002$	$P = 0.00003$

[a] Sign and [b] Wilcoxon tests: *P*-values given

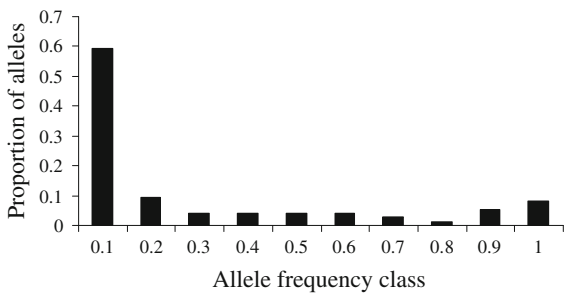

Fig. 2 Allele frequency distribution for all polymorphic microsatellite loci examined for the North-Western wild sorghum population (NWWSP). An L-shaped distribution is obtained. The values along the *x*-axis represent the maximum value for each respective allele frequency class (e.g., the first class represents alleles with frequencies between 0 and 0.1; the second, between 0.1 and 0.2; etc.)

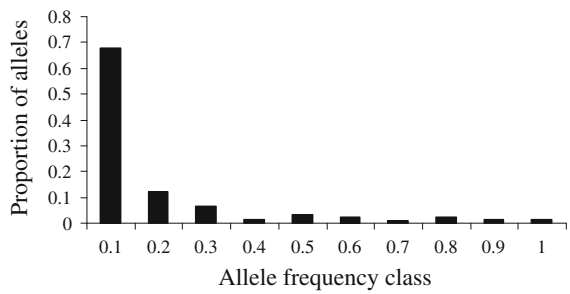

Fig. 3 Allele frequency distribution for all polymorphic microsatellite loci examined for the South-Eastern wild sorghum population (SEWSP). An L-shaped distribution is obtained. The values along the *x*-axis represent the maximum value for each respective allele frequency class (e.g., the first class represents alleles with frequencies between 0 and 0.1; the second, between 0.1 and 0.2; etc.)

Discussion

Lack of variability can be used as evidence for a bottleneck. However, in order to evaluate the lack of variation in reference to background heterozygosity, variation at the same loci in a population that did not experience a bottleneck needs to be compared to the presumed bottleneck population (Houlden et al. 1996). Unfortunately, this is not possible for wild sorghum populations because such reference populations do not exist.

Low levels of heterozygosity have been taken as evidence for bottlenecks in populations known to have undergone severe demographic declines (Houlden et al. 1996). Hence, the relatively high estimates of average heterozygosity observed for microsatellite loci (Table 2) suggest that the wild sorghum populations of our study may have maintained a large long-term effective population size and may not have

experienced a genetic bottleneck. However, comparisons of heterozygosity *per se* are not sufficient for studying bottlenecks, mainly because a high level of heterozygosity may be maintained following a bottleneck (Nei et al. 1975). The number of alleles at a give locus can be a better indicator of a population bottleneck because it is more sensitive to demographic fluctuations (Nei et al. 1975; Maruyama and Fuerst 1985). Furthermore, due to the limitation of heterozygosity, an analysis of allelic diversity in reference to heterozygosity excess is even more appropriate for testing if a recent bottleneck has occurred in a given population (Cornuet and Luikart 1997).

In our study most of the mocrosatellite loci showed evidence of heterozygote deficiency, while some of the loci showed evidence of heterozygosity excess and Hardy–Weinberg equilibrium. The heterozygosity deficiency may indicate a historical population expansion (Cornuet and Luikart 1997). The results of sign and Wilcoxon tests (Table 5) confirm that the populations are under expansion under both mutation models, except SEWSP under the IAM model. Too few alleles will result in an excess of heterozygosity, which is indicative of a bottleneck or founder event. In the case of wild sorghum populations, the results from the tests do not support the hypothesis of a recent bottleneck. Furthermore, analysis of the allele frequency distribution did not reveal evidence of a recent genetic bottleneck (Figs. 2, 3), as a model shift from an L-shaped distribution would have resulted in the event that a bottleneck reduced genetic variability in a given population (Luikart et al. 1998b).

Historical expansion might have occurred in wild sorghum, following the subdivision of a potentially large wild sorghum population, with domestication of sorghum and subsequent diffusion of wild sorghum from its centre of origin. Since these populations are connected through migration of farmers carrying with them their seeds, a higher effective population size within wild sorghum population may have resulted subsequent to their establishment in new environments. Consequently, growth of the effective population size could have been initiated, leaving a signature of population expansion. On the other hand, wild sorghum may have undergone periods of rapid changes caused by natural selection, in which most individuals suffered genetic load. Survivors of these evolutionary bottlenecks probably reproduced successfully resulting in large populations in successive generations. The effect of these bottlenecks may be the decline in genetic diversity of the populations, given that most inconsistency is lost at the time of occurrence of the bottlenecks.

The interpretation of the results from long-term N_e using microsatellite data from the wild sorghum populations are dependent on the mutation-drift equilibrium (Waples 1991), how well the assumed mutation models fit evolutionary of the microsatellites under study, and the accuracy of assumed mutation rates at these microsatellite loci (Lehmann et al. 1998; Thuillet et al. 2005). However, most difficulties in the interpretation of absolute long-term N_e values, such as incorrect mutation rates, are far less important in comparisons between populations using the same microsatellite loci. The average long-term N_e in SEWSP was 1.9 and 2.6 times higher than those of NWWSP for the IAM and SMM, respectively. The difference between populations was significant regardless of whether long-term N_e was calculated based on the SMM or the IAM.

The results of this study showed that most of the microsatellite loci were in gametic phase disequilibrium (Table 3). The probability of observing this pattern is greater than by chance alone ($P < 0.05$), and it is unlikely that size homoplasy (Estoup et al. 1995) could account for multiple loci. Several factors could be potentially responsible for observed gametic phase disequilibrium in pairwise comparison. Firstly, the wild sorghum populations may exhibit population subdivision. This was supported by the observed significant probability for overall test of gametic phase disequilibrium (Table 3; $P = 0.05$). Furthermore, family sub-structuring could cause population subdivision. Although, the two populations were sampled from two different ecosystems, they remained in contact through seed mediated gene flow (Muraya et al. 2009; unpublished data). Furthermore, wild sorghum exhibits a mixed mating system with high outcrossing (Muraya et al. 2009; unpublished data). Secondly, strong drift effects could be active as a consequence of a bottleneck and thereby produce genetic disequilibria. However, there is no evidence for a bottleneck from the microsatellite data. Third, selection may act to produce the gametic phase disequilibrium pattern observed if the loci in question are hitchhiking together with a gene or functional region under selection (Slatkin 1995). More evidence is needed to confirm this. Fourth, the microsatellite loci

in disequilibrium may be linked or could be a syntenic group. However, mapping studies have been carried out on all the microsatellite loci used, which showed that they were unlinked. Therefore, linkage is not the true cause of gametic phase disequilibrium in the affected microsatellite loci. Finally, the microsatellite loci in disequilibrium may contain null alleles. The frequency of null alleles in our study was generally high. However, this does not necessarily prove that null alleles are the true cause of heterozygosity deficiency because even some of the loci displaying heterozygosity excess and Hardy–Weinberg equilibrium had a high frequency of null alleles. The high frequency of null alleles found in this study could be explained by the fact that the microsatellites used were designed for cultivated sorghum. One potential cause of microsatellite null alleles is poor primer annealing due to nucleotide sequence divergence in one or both flanking primers (Kwok et al. 1990).

Conclusion

Assuming that SMM and IAM generally apply to the loci used in this study, a historical expansion of wild sorghum populations can be inferred, perhaps having taken place some time subsequent to the domestication of sorghum, ~5,000 years ago. We suspect that population subdivision, wild sorghum mating system and null alleles may be the cause of heterozygosity deficiency. The results of this study indicate that domestication and diffusion of wild sorghum may not have severely affected genetic variability of wild sorghum. Therefore, collection and ex-situ conservation of wild sorghum populations seems unnecessary provided the current situations remains constant. However, the ever-increasing anthropogenic changes to wild sorghum natural ecosystems need to be constantly monitored.

Acknowledgment This study was funded by the United States Agency for International Development (USAID) Biotechnology and Biodiversity Interface Program (BBI), the Institute of Plant Breeding and Population Genetics at the University of Hohenheim, Germany, and Germany Academic Exchange Service (DAAD: A0523923). We are grateful to Kenya Agricultural Research Institute and Ben Kanyenji who supervised the collection of genetic materials in full compliance with regulations according to the Convention on Biological Diversity (CBD).

References

Beaumont MA, Rannala B (2004) The Bayesian revolution in genetics. Nat Rev Genet 5:251–261

Brown SM, Hopkins MS, Mitchell SE, Senior ML, Wang TY, Duncan RR et al (1996) Multiple methods for identification of polymorphic simple sequence repeats (SSRs) in sorghum (*Sorghum bicolor* (L.) Moench). Theor Appl Genet 93:190–198

Chikhi L, Bruford MW (2005) Mammalian population genetics and genomics. In: Ruvinsky A, Marshall-Graves J (eds) Mammalian genomics. CABI Publishing, Oxford

Cornuet JM, Luikart G (1997) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics 144:2001–2014

de Wet JMJ (1978) Systematics and evolution of *Sorghum* sect. Sorghum (Gramineae). Amer J Bot 65:477–484

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc B 39:1–38

Estoup A, Tailliez C, Cornuet JM, Solignac M (1995) Size homoplasy and mutational processes of interrupted microsatellites in two bee species, *Apis mellifera* and *Bombus terrestris* (Apidae). Mol Biol Evol 12:1074–1084

Frankham R (1998) Inbreeding and extinction: island populations. Conserv Biol 12:665–675

Franklin IR (1980) Evolutionary changes in small populations. In: Soulé ME, Wilcox BA (eds) Conservation biology: an evolutionary-ecological perspective. Sinauer Associates, Sunderland, Massachusetts, pp 135–150

Garnier-Gere P, Dillmann C (1992) A computer program for testing pairwise linkage disequilibria in subdivided populations. J Hered 83:239

Goossens B, Chikhi L, Ancrenaz M, Lackman-Ancrenaz I, Andau P, Bruford MW (2006) Genetic signature of anthropogenic population collapse in orang-utans. PLoS Biol 4:285–291

Houlden BA, England PR, Taylor AC, Greville WD, Sherwin WB (1996) Low genetic variability of the koala *Phascolarctos cinereus* in south-eastern Australia following a severe population bottleneck. Mol Ecol 5:269–281

Keller LF, Waller DM (2002) Inbreeding effects in wild populations. Trends Ecol Evol 17:230–241

Kimura M (1983) The neutral theory of molecular evolution. United Kingdom, Cambridge University Press

Kong L, Dong J, Hart GE (2000) Characteristics, linkage-map positions, and allelic differentiation of *Sorghum bicolor* (L.) Moench DNA simple-sequence repeats (SSRs). Theor Appl Genet 101:438–448

Kwok S, Kellog DE, McKinney N, Spasic D, Goda L, Levenson C et al (1990) Effects of primer-template mismatches on the polymerase chain reaction: human immunodeficiency virus 1 model studies. Nucleic Acids Res 18:999–1005

Lehmann T, Hawley WA, Grebert H, Collins FH (1998) The effective population size of *Anopheles gambiae* in Kenya: implications for population structure. Mol Biol Evol 15:264–276

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

Luikart G, Allendorf FW, Cornuet JM, Sherwin WB (1998a) Distortion of allele frequency distributions provides a test for recent population bottlenecks. J Hered 89:238–247

Luikart G, Sherwin WB, Steele BM, Allendorf FW (1998b) Usefulness of molecular markers for detecting population bottlenecks via monitoring genetic change. Mol Ecol 7:963–974

Luikart G, England PR, Tallomon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. Nat Rev Genet 4:981–994

Mace ES, Buhariwalla HK, Crouch JH (2003) A high throughput DNA extraction protocol for molecular breeding programs. Plant Mol Biol Rep 21:459a–459h

Mann JA, Kimber CT, Miller FR (1983) The origin and early cultivation of sorghums in Africa. Texas Agricultural Experiment Station, Bulletin 1454

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

Mills LS, Smouse PE (1994) Demographic consequences of inbreeding in remnant populations. Am Nat 144:412–431

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Nei M, Maruyama T, Chakraborty R (1975) The bottleneck effect and genetic variability in populations. Evolution 29:1–10

Raymond M, Rousset F (1995) An exact test for population differentiation. Evolution 49:1280–1283

Saillant E, Patton JC, Ross KE, Gold JR (2004) Conservation genetics and demographic history of the endangered Cape Fear shiner (*Notropis mekistocholas*). Mol Ecol 13: 2947–2958

Schloss SJ, Mitchell SE, White GM, Kukatla R, Bowers JE, Paterson AH et al (2002) Characterization of RFLP probe sequences for gene discovery and SSR development in *Sorghum bicolor* (L.) Moench. Theor Appl Genet 105:912–920

Schwartz MK, Luikart G, Waples RS (2007) Genetic monitoring as a promising tool for conservation and management. Trends Ecol Evol 22:25–33

Slatkin M (1995) Hitchhiking and associative overdominance at a microsatellite locus. Mol Biol Evol 12:473–480

Spencer CC, Neigel JE, Leberg PL (2000) Experimental evaluation of the usefulness of microsatellite DNA for detecting bottlenecks. Mol Ecol 9:1517–1528

Taramino G, Tarchini R, Ferrario S, Lee M, Pe' ME (1997) Characterisation and mapping of simple sequence repeats (SSR) in *Sorghum bicolor*. Theor Appl Genet 95:66–72

Thuillet A-C, Bru D, David J, Roumet P, Santoni S, Sourdille P et al (2002) Direct estimation of mutation rate for 10 microsatellite loci in durum wheat, *Triticum turgidum* (L.) Thell. ssp. *durum* Desf. Mol Biol Evol 19:122–125

Thuillet A-C, Bataillon T, Poirier S, Santoni S, David JL (2005) Estimation of long-term effective population sizes through the history of durum wheat using microsatellite data. Genetics 169:1589–1599

United Nations Environment Programme Convention on Biological Diversity (UNEPCD) SBSTTA (2003) Monitoring and indicators: designing national-level monitoring programmes and indicators, United Nations

Waples RS (1991) Genetic methods for estimating the effective population size of cetacean populations. In: Hoelzel AR (ed) Genetic ecology of whales and dolphins. International Whaling Commission, Cambridge, pp 279–300 (Spec Issue No 13)