Theor Appl Genet (2003) 106:1316–1325
DOI 10.1007/s00122-003-1202-7

R. Uptmoor · W. Wenzel · W. Friedt · G. Donaldson · K. Ayisi · F. Ordon

Comparative analysis on the genetic relatedness of *Sorghum bicolor* accessions from Southern Africa by RAPDs, AFLPs and SSRs

Received: 29 May 2002 / Accepted: 24 September 2002 / Published online: 6 March 2003

Abstract In order to get an overview on the genetic relatedness of sorghum (*Sorghum bicolor*) landraces and cultivars grown in low-input conditions of small-scale farming systems, 46 sorghum accessions derived from Southern Africa were evaluated on the basis of amplified fragment length polymorphism (AFLPs), random amplified polymorphic DNAs (RAPDs) and simple sequence repeats (SSRs). By this approach all sorghum accessions were uniquely fingerprinted by all marker systems. Mean genetic similarity was estimated at 0.88 based on RAPDs, 0.85 using AFLPs and 0.31 based on SSRs. In addition to this, genetic distance based on SSR data was estimated at 57 according to a stepwise mutation model ($\Delta\mu$-SSR). All UPGMA-clusters showed a good fit to the similarity estimates (AFLPs: $r = 0.92$; RAPDs: $r = 0.88$; SSRs: $r = 0.87$; $\Delta\mu$-SSRs: $r = 0.85$). By UPGMA-clustering two main clusters were built on all marker systems comprising landraces on the one hand and newly developed varieties on the other hand. Further sub-groupings were not unequivocal. Genetic diversity (H, DI) was estimated on a similar level within landraces and breeding varieties. Comparing the three approaches to each other, RAPD and AFLP similarity indices were highly correlated ($r = 0.81$), while the Spearman's rank correlation coefficient between SSRs and AFLPs was $r = 0.57$ and $r = 0.51$ between RAPDs and SSRs. Applying a stepwise mutation model on the SSR data resulted in an intermediate correlation coefficient between $\Delta\mu$-SSRs and AFLPs ($r = 0.66$) and RAPDs ($r = 0.67$), respectively, while SSRs and $\Delta\mu$-SSRs showed a lower correlation coefficient ($r = 0.52$). The highest bootstrap probabilities were found using AFLPs (56% on average) while SSR, $\Delta\mu$-SSR and RAPD-based similarity estimates had low mean bootstrap probabilities (24%, 27%, 30%, respectively). The coefficient of variation (CV) of the estimated genetic similarity decreased with an increasing number of bands and was lowest using AFLPs.

Keywords *Sorghum bicolor* · Genetic relatedness · Genetic diversity · Cluster analysis · Coefficient of variation

Communicated by H.C. Becker

R. Uptmoor · W. Wenzel · W. Friedt · F. Ordon (✉)
Institute for Crop Science and Plant Breeding I,
Justus-Liebig-University, Heinrich-Buff-Ring 26-32,
35392 Giessen, Germany,
e-mail: Frank.Ordon@agrar.uni-giessen.de

W. Wenzel
Agricultural Research Council (ARC),
Grain Crops Institute (GCI),
Private Bag X1251 Potchefstroom 2520, South Africa

G. Donaldson
Northern Province Department of Agriculture,
Land and Environment, Potgietersrus, South Africa,

K. Ayisi
Department of Plant Production, University of the North,
Private Bag X1106, Sovenga 0727, South Africa

Present adress:
F. Ordon, Institute of Epidemiology and Resistance,
Federal Centre for Breeding Research on Cultivated Plants,
Theodor-Roemer-Weg 4, 06449 Aschersleben, Germany

Introduction

Sorghum is an important staple food throughout semi-arid Asian and African regions (Ahmed et al. 2000). A large number of different landraces well adapted to low-input conditions as well as to biotic and abiotic stress factors are still cultivated. A low productivity of grown varieties, frequent occurrences of food shortage in sorghum growing areas, and the extension of sorghum cultivation to marginal lands requires extensive breeding programmes followed by the introduction of new varieties fitting small-scale farmers needs (Haussmann et al. 2000). Compared to maize, sorghum breeding has been neglected in recent decades and, furthermore, the availability of high-yielding maize varieties has led to the displacement of sorghum. However, in general, maize is less adapted to

drought conditions and thus of lower yield stability (Wenzel et al. 2001b).

With respect to efficient breeding, the conservation of genetic resources is important, since landraces may bear advantageous genes that are especially useful in resistance breeding and in terms of quality traits (Tanksley and McCouch 1997). Knowledge of genetic diversity has an important impact on the improvement of crop productivity as well as on the conservation of genetic resources (Dean et al. 1999; Simioniuc et al. 2002). In South Africa, sorghum is of regional importance as a main staple food (Wenzel et al. 2001a). A large number of landraces is cultivated; these are well adapted to local conditions and posess superior grain quality but a very poor yielding ability (Wenzel et al. 2001b). An increase in the yield potential of these landraces should be a major objective in sorghum breeding with respect to that region. In the frame of a breeding programme detailed knowledge on the genetic diversity within this genepool would facilitate a more efficient selection of parental genotypes, as knowledge of genetic variation and the genetic relationship between genotypes is an important prerequisite for the efficiency of plant breeding programmes (Russel et al. 1997; Ribaut and Hoisington 1998).

Molecular markers are powerful tools for the assessment of genetic relationships. Polymerase chain reaction (PCR)-based marker systems like amplified fragment length polymorphisms (AFLPs), random amplified polymorphic DNAs (RAPDs) and simple sequence repeats (SSRs) have been widely used in recent years, replacing restriction fragment length polymorphisms (RFLPs) in DNA fingerprinting (Pejic et al. 1998). SSRs and RAPDs have been used to characterize genetic diversity in sorghum (Dean et al. 1999; Djè et al. 1999; Ayana et al. 2000). These methods differ in the number of informative DNA fragments per PCR reaction, in genome coverage and in technical and time demand (for overview, see Karp et al. 1998). RAPDs are relatively less costly and easy to establish, but reproducibility is low. In comparison to RFLPs, AFLPs are advantageous due to the high number of polymorphic bands per assay unit (Bohn et al. 1999). SSRs are highly polymorphic compared to the methods mentioned above. Computations of genetic relatedness are based on different mutation models: the infinite allele model and the stepwise mutation model (Di Rienzo et al. 1994). Goldstein et al. (1995a) established simple stepwise mutation model-based distance computations and compared these to infinite mutation model-based calculations using SSRs. Both models are currently applied in plant science (e.g. Dean et al. 1999; Udupa et al. 1999). Besides this, published genetic similarity data varies widely with respect to the number of informative DNA fragments taken into account. In this respect, the comparability and effectiveness of different marker systems has been discussed – for example, for barley (Russel et al. 1997) and maize (Pejic et al. 1998).

For measuring the reliability of clusters developed on different marker systems, bootstrap-based accuracy measures were developed by Efron and Tibshirani (1986) and applied to RFLP-based distance measures by Tivang et al. (1994). Therefore, the main objectives of this paper are (1) the estimation of genetic relatedness among sorghum cultivars grown in Southern Africa, (2) the comparison of different marker techniques and theoretical mutation models and (3) the estimation of the effect of sample size (i.e. number of markers/bands) on the results obtained.

Materials and methods

Plant material and DNA extraction

The 46 sorghum accessions tested consisted of 23 landraces from Southern Africa, of which 11 landraces were collected in the Northern Province, South Africa and 12 derived from different parts of Southern Africa. Thirteen breeding varieties were derived from the Sorghum and Millet Improvement Project (SMIP) of the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT). Besides this, five newly bred varieties of the Agricultural Research Council – Grain Crops Institute (ARC-GCI), Potchefstroom, South Africa, and five newly bred varieties of the Department of Crop Science, National University of Lesotho, Maseru, Lesotho were included (Table 1). Seeds were sown in the greenhouse and DNA was extracted from leaf tips of 14-day-old plants according to Doyle and Doyle (1990). As this study did not aim at the estimation of the degree of heterozygosity and heterogeneity within landraces and breeding varieties, mixed samples of ten plants from each accession were taken in order to represent the genetic variability possibly present within each sorghum accession.

Amplified fragment length polymorphism (AFLP)

AFLP analysis was performed according to Vos et al. (1995) using a GeneAmp Systeme 9700 (Perkin Elmer, Alameda, Calif.). Genomic DNA was digested using *Eco*RI/*Mse*I and double-stranded adapters were ligated to the restricted DNA fragments. Ligation was followed by two pre-amplifications (+0/+1, Schiemann et al. 1999) prior to an amplification (+3) using selective primer combinations. Fragments were separated by polyacrylamide gel electrophoresis using a Licor 4200 DNA analyzer (for details see Scheurer et al. 2001). In total, 700 primer combinations were screened on the accessions IS 4177 and ZA 4471, which were expected to represent a high level of genetic diversity due to their origin and morphology. The following 28 most informative primer combinations were selected for the analysis of genetic diversity: AAT/AGA, AAT/AGC, ACA/AAC, ACA/AAG, ACA/ACC, AGT/AAG, AGT/ACA, ATA/AAC, ATA/ACC, CAA/CAC, CAA/CAG, CAA/CCA, CAC/CAA, CAC/CAC, CAC/CCA, CAC/CCG, CAG/CAT, CAT/CAC, CCA/CAT, CCA/CCA, CGA/CCA, CGC/CAA, CGC/CAC, CGC/CCA, CGG/CAG, CGT/CAA, CGT/CAG, CTG/CAC, CTT/CCG.

Random amplified polymorphic DNA (RAPD)

RAPD analysis was carried out as described by Ordon et al. (1995). Reaction products were separated electrophoretically on a 2% agarose gel (Gibco Ultra pure or Sigma 3:1) and subsequently stained with ethidium bromide. Ninety 10-mer primers (Operon Technologies) were screened on the ICRISAT accession and on the landrace mentioned above. A total of 25 primers producing informative and distinct bands was chosen for DNA fingerprinting (Table 2).

Table 1 List of sorghum accessions, their origin, breeding institute and name

Accession[a]	Origin	Breeder	Name
Landraces (LR)			
ZA 0498	South Africa		
RB 1304	Botswana		
ZA1542	North West, South Africa		
ZA1595	Gauteng, South Africa		Radar
ZA 1605	Gauteng, South Africa		SAR-29
MW 1987	Malawi		E102B
RB 2153	Botswana		Segaolane
ZA 2172	Eastern Cape, South Africa		Transkei Red
ZA 2330	Mpumalanga, South Africa		Breytenbach Red
SA 2893[b]	Africa		WSV 397
LS 3986	Lesotho		Tenant White
ZA 3874	KwaZulu Natal, South Africa		
NP 1807	Northern Province, South Africa		
NP 1851	Northern Province, South Africa		
NP 1859	Northern Province, South Africa		
NP 4438	Northern Province, South Africa		
NP 4448	Northern Province, South Africa		
NP 4453	Northern Province, South Africa		
NP 4456	Northern Province, South Africa		
NP 4461	Northern Province, South Africa		
NP 4463	Northern Province, South Africa		
NP 4468	Northern Province, South Africa		
NP 4471	Northern Province, South Africa		
Breeding varieties (BV)			
LS 3716	Lesotho	University of Lesotho	A29
LS 3728	Lesotho	University of Lesotho	ICV 219
LS 3737	Lesotho	University of Lesotho	C15
LS 3744	Lesotho	University of Lesotho	C22
LS 3761	Lesotho	University of Lesotho	5024
ZA 3105	South Africa	Grain Crops Institute	K901
ZA 3984	South Africa	Grain Crops Institute	K222
ZA 4125	South Africa	Grain Crops Institute	A2281
ZA 4258	South Africa	Grain Crops Institute	K023
ZA 4322	South Africa	Grain Crops Institute	Sel 967
IS 4162	Zimbabwe	ICRISAT/SMIP	SV-1
IS 4163	Zimbabwe	ICRISAT/SMIP	SV-2
IS 4166		ICRISAT/SMIP	Larsvyt 19
IS 4170		ICRISAT/SMIP	SDSL 89473
IS 4172		ICRISAT/SMIP	SDSL 89566
IS 4173		ICRISAT/SMIP	SDSL 87046
IS 4175		ICRISAT/SMIP	SDSL 87049T
IS 4177		ICRISAT/SMIP	SDSL 89426
IS 4179		ICRISAT/SMIP	Larsvyt 46-85
IS 4181		ICRISAT/SMIP	85-1L-208
IS 4201		ICRISAT/SMIP	MR-812
IS 4206		ICRISAT/SMIP	R-8602
IS 4394	Zambia	ICRISAT/SMIP	Sima

[a] RB, Botswana; LS, Lesotho; MW, Malawi; ZA, South Africa; IS, ICRISAT; NP, Northern Province, South Africa

[b] According to our information SA 2893 LR is a landrace of unknown African origin, received from the University of Purdue, USA

Simple sequence repeats (SSRs)

Out of 30 tested SSR loci 25 polymorphic primer pairs published by Brown et al. (1996), Taramino et al. (1997), and Kong et al. (2000) were used for the estimation of genetic relatedness (Table 3). DNA amplifications were carried out according to Brown et al. (1996). Amplification products were resolved on polyacrylamide gels using a Licor 4200 DNA analyser. Fragments were stained with IRD 700 or IRD 800, respectively.

Data analysis

Presence or absence of bands was determined using the software package RFLP-SCAN 2.1. The resulting 0/1 matrix was used as input to a bootstrap procedure (Efron and Tibshirani 1986) using the programme package PHYLIP 3.5 (http://evolution.genetics.washington.edu/phylip.html). Each of the 1,000 bootstrap resamples is created by randomly taken single bands or loci from the original data set with replacement. The resulting matrix has exactly the same size as the original data set. Scored AFLP and RAPD fragments were treated as independently evolved, and bootstrapping was carried out on the basis of single bands, while SSRs were block-bootstrapped, with each SSR-locus representing one block. Genetic similarity was estimated according to Nei and Li

Table 2 Sequences of 10-mer RAPD primers used for estimation of genetic diversity

Name	Sequence	Name	Sequence	Name	Sequence
OP A08	5'-GTGACGTAGG-3'	OP E15	5'-ACGCACAACC-3'	OP L06	5'-GAGGGAAGAG-3'
OP A15	5'-TTCCGAACCC-3'	OP K02	5'-GTCTCCGCAA-3'	OP L16	5'-AGGTTGCAGG-3'
OP A17	5'-GACCGCTTGT-3'	OP K13	5'-GGTTGTACCC-3'	OP L18	5'-ACCACCCACC-3'
OP A19	5'-CAAACGTCGG-3'	OP K14	5'-CCCGCTACAC-3'	OP L19	5'-GAGTGGTGAC-3'
OP D02	5'-GGACCCAACC-3'	OP K17	5'-CCCAGCTGTG-3'	OP N06	5'-CCACGGGAAG-3'
OP D15	5'-CATCCGTGCT-3'	OP K19	5'-CACAGGCGGA-3'	OP N08	5'-ACCTCAGCTC-3'
OP D16	5'-AGGGCGTAAG-3'	OP L02	5'-TGGGCGTCAA-3'	OP N14	5'-TCGTGCGGGT-3'
OP E02	5'-GGTGCGGGAA-3'	OP L03	5'-CCAGCAGCTT-3'		
OP E10	5'-CACCAGGTGA-3'	OP L05	5'-ACGCAGGCAC-3'		

Table 3 SSRs used for the studies on genetic relatedness: their PCR product range, number of alleles per locus and diversity index as observed on 46 sorghum accessions

Primer	Author	LG[a]	Type	Observed PCR product range	Number of alleles	$1 - \sum x_i^2$
Xtxp 1	Kong et al. (2000)	B	$(AG)_{34}$	176–224	11	0.84
Xtxp 6	Kong et al. (2000)	I	$(CT)_{33}$	105–134	14	0.87
Xtxp 10	Kong et al. (2000)	F	$(CT)_{14}$	171–177	8	0.76
Xtxp 20	Kong et al. (2000)	C	$(AG)_{21}$	199–282	9	0.74
Xtxp 30	Kong et al. (2000)	J	$(AAT)_{25}$	270–278	8	0.73
Xtxp 33	Kong et al. (2000)	C	$(TC)_{20}C(TG)_5 + (CT)_9CC(TG)_7$	206–246	16	0.87
Xtxp 34	Kong et al. (2000)	C	$(CT)_{29}$	340–368	8	0.76
Xtxp 40	Kong et al. (2000)	D	$(GGA)_7$	134–144	3	0.08
Xtxp 43	Kong et al. (2000)	A	$(CT)_{28}$	164–183	10	0.74
SbAGA01	Taramino et al. (1997)	H	$(AG)_{33}$	96–118	4	0.36
SbAGB02	Taramino et al. (1997)	A	$(AG)_{35}$	171–193	8	0.82
SbAGB03	Taramino et al. (1997)	F	$(AG)_{41}$	133–155	6	0.63
SbAGH04	Taramino et al. (1997)	F	$(AG)_{39}$	112–182	20	0.93
SbKAFGK1	Taramino et al. (1997)	J	$(ACA)_9$	98–186	11	0.88
Sb1-1	Brown et al. (1996)	H	$(AG)_{16}$	241–263	6	0.73
Sb1-10	Brown et al. (1996)	D	$(AG)_{27}$	394–460	5	0.58
Sb4-15	Brown et al. (1996)	E	$(AG)_{16}$	120–132	6	0.38
Sb4-32	Brown et al. (1996)	E	$(AG)_{15}$	177–191	7	0.76
Sb4-121	Brown et al. (1996)	D	$(AC)_{14}$	215–227	10	0.84
Sb5-206	Brown et al. (1996)	E	$(AC)_{13}/(AG)_{20}$	126–148	8	0.80
Sb5-236	Brown et al. (1996)	G	$(AG)_{20}$	193–217	6	0.73
Sb6-34	Brown et al. (1996)	I	$(AC)/(CG)_{15}$	158–188	5	0.34
Sb6-36	Brown et al. (1996)	C	$(AG)_{19}$	174–196	7	0.67
Sb6-57	Brown et al. (1996)	C	$(AG)_{18}$	307–318	6	0.63
Sb6-342	Brown et al. (1996)	A	$(AC)_{25}$	200–286	15	0.86

[a] LG, Linkage group according to Kong et al. (2000), Taramino et al. (1997) or Dean et al. (1999), respectively

(1979) using a modified GENDIST version of the software package PHYLIP 3.5 (Le Thierry d'Ennequin et al. 2000) for estimating genetic similarity:

$$GS_{ij} = \frac{2N_{ij}}{(N_i + N_j)},$$

where N_i and N_j are the total number of bands of accessions i or j, respectively, and N_{ij} is the total number of shared bands in the accessions i and j. In addition to the Nei and Li (1979) estimates for genetic similarity, genetic relatedness based on SSR loci was estimated following the stepwise mutation model of Goldstein et al. (1995b). In contrast to the Nei and Li (1979) index, which is the quotient of shared alleles and total number of alleles, the $\Delta\mu$ distance (Goldstein et al. 1995b) is the squared average allelic difference in terms of tandem repeats over all SSR loci as follows:

$$\Delta\mu = \left(\frac{\sum (M_{ix} - M_{jx})}{n}\right)^2,$$

where M_{ix} and M_{jx} are the number of tandem repeats of allele x in accession i or j, respectively.

Bootstraps of tandem repeat data and $\Delta\mu$ distances were computed using MICROSAT (http://hpgl.stanford.edu/projects/microsat/). UPGMA-clusters were calculated for each of the 1,000 bootstrap re-samples and a consensus tree was drawn using PHYLIP 3.5.

Spearman's rank correlation coefficients were calculated in order to compare indices on genetic relatedness generated by different types of molecular markers. The correlation between the original similarity indices and cophenetic values, expressed on each node of a dendrogram, was computed using the Mantel test procedure (Mantel 1967) of the software package NTSYS-PC (Rohlf 2000). With the goal to detect the minimum number of bands required for a reliable estimation of genetic relatedness, we created sub-samples by the random deletion of scored bands for AFLP and RAPD data. SSR sub-samples were taken by random deletion of loci. Bootstraps and pairwise genetic relatedness were computed as described above. Coefficients of variation (CV) of the mean genetic similarity computed from bootstrap re-samples were calculated from each sub-set as a measurement for statistical accuracy as described by Tivang et al. (1994). Genetic diversity (H) on the basis of AFLP and RAPD data was computed according to Lynch and Milligan (1994) using AFLP-SURV (http://www.ulb.ac.be/sciences/

Table 4 Number of assay units, total number of bands and maximum, minimum and mean genetic relatedness according to Nei and Li (1979) or Goldstein et al. (1995b)

	Number of			Genetic relatedness		
	Assay units	Bands	Polymorphic bands	Minimum	Maximum	Mean
AFLP[a]	28	1,135	701	0.80	0.97	0.88
RAPD[a]	25	378	305	0.77	0.99	0.85
SSR[a]	25	217	217	0.00	0.84	0.31
$\Delta\mu$-SSR[b]	25	217	217	2.1	169.4	57.0

[a] Genetic similarity estimated according to Nei and Li (1979)
[b] Genetic distance estimated according to Goldstein et al. (1995b)

Table 5 Genetic diversity among Sorghum accessions from Southern Africa

	N[a]	AFLP H[b]	RAPD H[b]	SSR	
				DI[c]	Alleles/locus
Landraces	23	0.136	0.160	0.597	6.56
Southern Africa	12	0.169	0.187	0.628	5.36
Northern Province	11	0.073	0.110	0.451	3.72
Breeding varieties	23	0.140	0.151	0.580	6.12
South Africa/Lesotho	10	0.141	0.147	0.517	4.20
ICRISAT/SMIP	13	0.136	0.147	0.563	4.61
All accessions	46	0.167	0.183	0.665	8.68

[a] N, Number of accessions within each group or subgroup
[b] H, Genetic diversity according to Lynch and Milligan (1994)
[c] DI, Genetic diversity according to Nei (1973)

lagev/software.html; Vekemans et al. 2002). Genetic diversity of SSR data was estimated by the number of alleles per locus and the mean diversity index (DI) over all loci was estimated according to Nei (1973):

$$DI = \frac{1}{n}\sum_{j}\left(1 - \sum_{i} {x_{ij}}^2\right),$$

where x_{ij} is the frequency of the i th allele of locus j and n the number of loci. For estimating genetic diversity the 46 sorghum accessions were divided into two groups (landraces and breeding varieties), each consisting of two subgroups (Northern Province landraces versus Southern African landraces and breeding varieties from South Africa and Lesotho versus ICRISAT/SMIP varieties, respectively).

Results

The total number of scored bands on the set of 46 genotypes was 1,135 based on 28 AFLP primer combinations, 378 bands using 25 RAPD primers and 217 alleles detected on 25 SSR loci. All 25 SSR loci and 61.76% of the AFLP and 80.69% of the RAPD bands were polymorphic (Table 4). The average number of scored bands was 39.3 for each AFLP primer combination and 15.1 for the RAPD primers. The mean number of alleles per SSR locus was 8.68, ranging from 3 to 20 (Table 3). Based on these data total genetic diversity was estimated at H = 0.167 on AFLPs, H = 0.183 on RAPDs and DI = 0.665 on SSRs. Within the subset of landraces and breeding varieties genetic diversity was calculated at a similar level, i.e. H = 0.136 (AFLPs), H = 0.160 (RAPDs), DI = 0.597 (SSRs) for landraces and H = 0.140 (AFLPs), H = 0.151 (RAPDs), DI = 0.580 (SSRs) for breeding varieties (Table 5). However, within the subset of landraces lower genetic diversity was found in those accessions derived from the Northern Province, i.e. H = 0.073 (AFLPs), H = 0.110 (RAPDs) and DI = 0.451 (SSRs, Table 5).

Table 6 Spearman's rank correlation coefficients among diversity indices obtained using AFLP, RAPD and SSR markers

	AFLP[a]	RAPD[a]	SSR[a]	$\Delta\mu$-SSR[b]
AFLP[a]	1.00			
RAPD[a]	0.81	1.00		
SSR[a]	0.57	0.51	1.00	
$\Delta\mu$-SSR[b]	0.66	0.67	0.52	1.00

[a] Genetic similarity data estimated according to Nei and Li (1979)
[b] Genetic distance data estimated according to Goldstein et al. (1995b)

The genetic similarity (GS) data are summarized in Table 4. Mean, minimum and maximum GS of AFLP- and RAPD-based data are similar in size. Mean GS of SSR-based data according to Nei and Li (1979) was significantly lower, and the range of genetic similarity was higher. Since distance estimates according to Goldstein's stepwise mutation model ($\Delta\mu$, Goldstein et al. 1995a, b) are expressed as the squared mean of the differences in tandem repeats over all SSR loci taken into account, absolute $\Delta\mu$ values are not comparable to Nei and Li (1979) based estimates. Spearman's rank correlation coefficient (Table 6) is highest between GS values based on AFLP and RAPD data, and the lowest correlation coefficient was estimated between GS values based on SSR and RAPD data. A similarly low correlation was estimated between SSR and $\Delta\mu$-SSR data. The correlation coefficient between the GS of AFLP and the GS of SSR was lower than the coefficient between AFLP and $\Delta\mu$-SSR estimates. The same holds true for RAPD and SSR

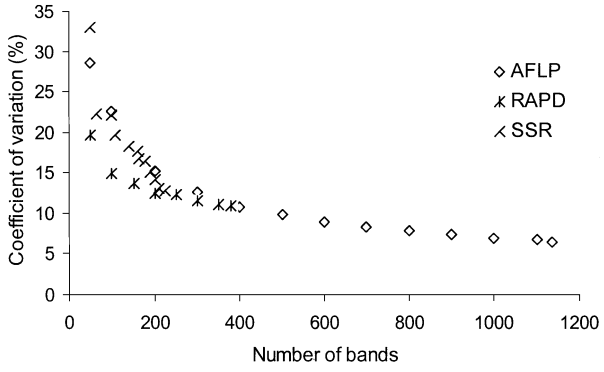

Fig. 1 Coefficients of variation estimated by bootstrap analysis for subsamples with an increasing number of bands. For SSRs each data point refers to the number of alleles corresponding to an increasing number of loci

data. The Mantel test (Mantel 1967) resulted in a very good (AFLPs: $r = 0.92$) to good (RAPDs: $r = 0.88$, SSRs: $r = 0.87$, $\Delta\mu$-SSRs: $r = 0.85$) fit of the cophenetic values to the original data set.

In order to specify the effect of the sample size on the accuracy of the estimated genetic similarity, CV of the mean genetic similarity estimated on bootstrap re-samples were computed using an increasing number of bands (AFLP and RAPD) and SSR loci (Fig. 1). With an increasing number of bands CV decreased and was estimated at CV = 6.25% for AFLPs, CV = 10.91% (RAPDs) and CV = 11.86 (SSRs) taking into account all of the bands generated. In this respect it is interesting to note that the CV for AFLPs became already smaller than 10% by 500 bands. The highest mean bootstrap probability (56%) was obtained with 28 AFLP primer combinations. RAPD- and SSR-based mean bootstrap probabilities were low; the mean value adapted from 25 assay units was about 30% (RAPDs) 29% (SSRs) and 30% ($\Delta\mu$-SSRs).

On all marker systems two main clusters were built by UPGMA-analysis. One highly reliable branch with 98.9% bootstrap probability was built out of 18 landraces and one ICRISAT accession using AFLPs (Fig. 2). A similar cluster was formed on the basis of RAPDs, but bootstrap probability was low (10.2%), and two newly developed South African varieties (ZA 3984 and ZA 4322) are part of this cluster. When both of these varieties were excluded, the bootstrap probability increased to 52.7%. Even using SSRs, similar clusters consisting mainly of landraces were received. However, bootstrap probabilities were low (25.2%) and intermediate using $\Delta\mu$ (60.4%). Taking a closer look at the branch comprising mainly landraces, it turned out that two landraces and the ICRISAT accession IS 4170 are not part of the dendrogram based on SSR data in contrast to the other dendrograms; IS 4166 was re-ordered into the 'landrace' cluster. Using $\Delta\mu$, three newly developed South African varieties cluster with the landraces. In the AFLP-based UPGMA-cluster analysis it turned out that the landrace NP 4453 is separated from the main 'landrace-cluster' with a 85.7% bootstrap probability and that the landrace RB 1304 from Botswana is separated from the residual cluster with a 99.3% reliability. These results are confirmed by SSR and RAPD data. Four out of the ten landraces collected in the Northern Province cluster with a 96.6% bootstrap probability based on AFLP data. A similar cluster with lower bootstrap probabilities was found on RAPD (41.6%) and SSR data (12.0% bootstrap probability). The Northern Province landraces NP 1851 and NP 1807 are closely related according to the AFLP data (99.9% bootstrap probability).

Having a closer look at the breeding varieties, a close relation between ZA 4258 and LS 3744 is evident and highly reliable taking into account the bootstrap probabilities of the GS estimated according to Nei and Li (1979) of all three marker systems, i.e. 100% (AFLPs), 89.9% (RAPDs) and 73.4% (SSRs). Another highly significant cluster consists of the ICRISAT accession IS 4173 and IS 4175, which is validated by high bootstrap probabilities on the basis of AFLP and SSR estimates (100% and 92.2%, respectively) and by an intermediate bootstrap probability using RAPDs (64.1%). However, these genotypes are more distantly related applying the stepwise mutation model. Similar results are obtained between ZA 3874 and IS 4162. Three accessions considered to be landraces (SA 2893, ZA 3874 and MW 1987) are more closely related to breeding varieties than to landraces.

Discussion

With all marker techniques, the 46 sorghum accessions analysed were unequivocally fingerprinted. The range and mean of the estimated genetic similarity were similar for AFLPs and RAPDs but a much wider range was found for SSRs, leading to a lower mean GS. Similar results have been reported by Russel et al. (1997) for barley and Pejic et al. (1998) for maize. In the present study, estimated mean genetic similarities amounted to 0.88 and 0.85 based on AFLP and RAPD data, respectively, and to 0.31 for SSRs. A comparison of genetic similarity data using the same techniques but different plant material is difficult, since most estimates are based on randomly detected polymorphic bands rather than on all randomly detected bands – i.e. excluding monomorphic fragments leads to lower genetic similarities. Besides this, similarity indices based on investigations on closely related accessions in general over-estimate polymorphisms in comparison to similarity measures based on distantly related genotypes.

Correlation coefficients between RAPDs and AFLPs were higher than those between SSRs and AFLPs. In contrast to this, Russel et al. (1997) and Pejic et al. (1998) found overall low correlation coefficients comparing RAPD data with SSR, AFLP and RFLP data, while data based on SSRs and AFLPs were more closely correlated. Pejic et al. (1998) assumed that the high rate of mismatch annealing of RAPDs may be the main reason for this.

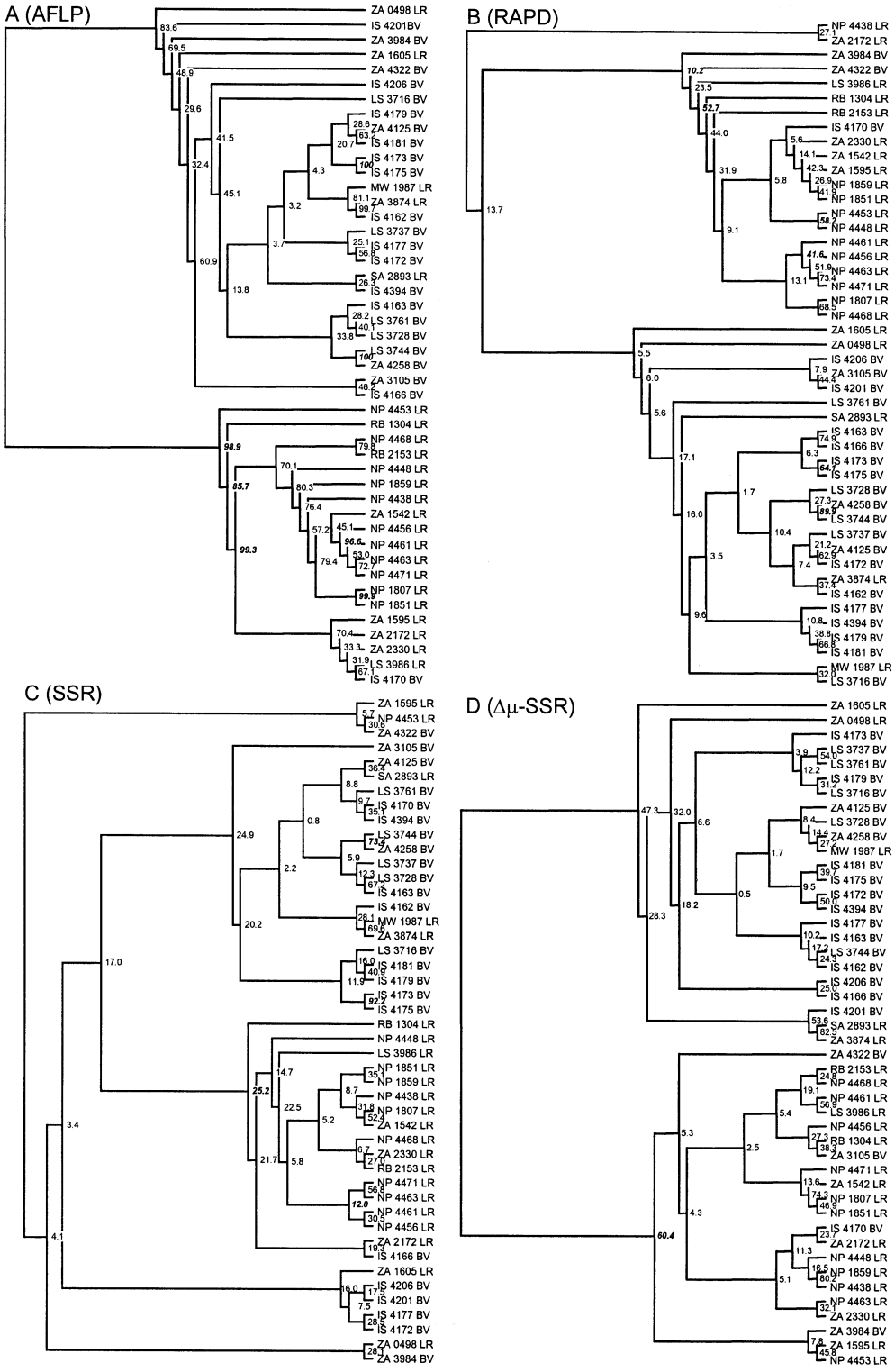

Fig. 2 AFLP (**A**), RAPD (**B**), SSR (**C**) and Δμ-SSR (**D**) UPGMA-clusters of 46 sorghum accessions derived from southern Africa. Bootstrap probabilities mentioned in the text are printed in *bold* and *italic*. *LR* Landrace, *BV* breeding variety, *RB* Botswana, *LS* Lesotho, *MW* Malawi, *ZA* South Africa, *IS* ICRISAT, *NP* Northern Province, South Africa

However, in our studies the lowest correlation coefficients were found between Δμ-SSR data and RAPDs. A similar low correlation coefficient was revealed between Δμ and Nei and Li-based SSR data. Bohn et al. (1999) reported low correlations between AFLP and SSR data among winter wheat cultivars and speculated that this may be due to a low linkage between marker loci from different marker systems resulting in sampling of different genome parts.

A CV of 10% was considered to be the threshold for reliable genetic similarity estimates by Tivang et al. (1994). Based on AFLP data 500 bands were necessary

for obtaining a CV smaller than 10%, using SSRs and RAPDs all 25 assay units (217 or 378 bands, respectively) were required for reliable similarity estimates. CVs of all three marker systems corresponding to the same number of bands were on a similar level (Fig. 1). Tivang et al. (1994) used the CV as an accuracy measure for the genetic diversity estimation in maize based on RFLPs and found that 375 polymorphic bands were required to obtain a CV lower than 10%. Using bootstrap-based standard errors as accuracy measure, Pejic et al. (1998) assumed that 150 bands would be sufficient for genetic similarity estimates among maize inbred lines no matter which marker technique was used. Lima et al. (2002) considered a CV of 5% to obtain a good precision in analysing genetic diversity and found 800 polymorphic bands to be necessary.

Different computations are established using SSR markers for analysing genetic relatedness. Nei's distance and the Nei and Li index are based on the infinite allele model. Di Rienzo et al. (1994) reported that, in human microsatellite evolutionary processes, mutations occur 'stepwise' and the most probable case is a change in one or two tandem repeats. This means that, in general, larger differences in fragment lengths are based on a higher incidence of mutations and, thus, on a lower genetic similarity. This contradicts the infinite mutation model (Feldmann et al. 1997) and implies that SSR loci contain more information than a simple allele sharing on which Nei and Li estimates are based. In fact, Slatkin (1995) was able to show that squared differences of repeat units in SSR loci are meaningful estimators for the time of separation of populations; Goldstein et al. (1995a) and Pollock et al. (1998) found that the average squared distance of microsatellites had in contrast to simple allele sharing an almost linear relation to time. Our studies revealed a stronger correlation using the squared average ($\Delta\mu$) as the distance measure between SSR-based data and AFLPs as well as between SSRs and RAPDs than using the Nei and Li index for SSR similarity estimates. The differentiation between landraces and breeding varieties was stronger using $\Delta\mu$. Thus, the stepwise mutation model may be more suitable for analysing SSRs. This holds true especially for more distantly related genotypes (Goldstein et al. 1995a). However, limitations of genetic similarity estimates, based on a strict stepwise mutation model, are obvious whenever there is no knowledge of the incidence and size of mutations, which do not occur in single steps. On the other hand, distance measures based on allele sharing are superior with respect to variance (Goldstein et al. 1995a).

With all marker systems two major clusters were detected. The first cluster mainly consists of landraces and the second one of newly developed varieties. Further sub-clusters were only partially verified by the different marker systems (Fig. 2). The newly developed varieties, including semi-dwarfs as well as varieties adapted to low-input agriculture, were not clearly separated – for example, newly developed low-input varieties derived from ICRISAT/SMIP and those bred at the Grain Crops Institute, Potchefstroom, South Africa. In addition, the sorghum accessions were not clearly separated according to their origin as has been demonstrated for barley (Ordon et al. 1997). Sub-clusters of breeding varieties contain cultivars from South Africa, Lesotho and ICRISAT, and those consisting mainly of Northern Province landraces include landraces of distant geographic origin. Ayana et al. (2000) also reported a weak differentiation of Ethiopian and Eritrean sorghum accessions according to both agro-ecological adaptation zones and regions of origin and supposed out-crossing and seed movement to be the reason for this. Most of the accessions of the present study can be assigned to the same agro-ecological zone, but the cultivation of varieties for different purposes (such as porridge and beer preparation) may blur regional differentiations in part.

With all of the marker techniques used the genetic diversity of the 23 landraces and 23 breeding varieties was estimated on the same level, except for the landraces derived from the Northern Province, which revealed a lower genetic diversity (Table 5). This may be due to the fact that these landraces have been sampled in a restricted area in comparison to those derived from different parts of Southern Africa. However, Djè et al. (1999) estimated a higher genetic diversity when analysing 25 sorghum landraces derived from a restricted area of North Western Morocco with three SSRs. The same holds true for the genetic diversity estimated on all accessions (Table 5) when compared to results of Djè et al. (2000) based on five SSRs and to studies of Grenier et al. (2000). However, in these studies sorghum accessions derived from different parts of the world were included and the SSRs used were more informative, for example, the five SSRs analysed by Djè et al. (2000) generated between 14 and 24 different alleles – i.e. 19.2 on average on 25 accessions – while in our studies on average 8.68 alleles per SSR loci were detected.

Our results on genetic relatedness and genetic diversity within sorghum accessions grown in Southern Africa revealed a clear separation between landraces and breeding varieties (Fig. 2) but a similar level of genetic diversity within both subgroups (Table 5). This information in connection with knowledge on agronomic traits (Wenzel et al. 2001a, b) may have an impact on sorghum breeding in this region. One breeding strategy may be to choose well-adapted parents that posses many random genetic differences in the hope of an increased number of transgressive recombinants (Tinker et al. 1993; Graner et al. 1994). In comparison to pedigree data, detailed genetic similarity estimates based on molecular markers are better suited for this selection of parental genotypes since additional information about the number of segregating loci is provided. Besides this, these molecular markers facilitate an unequivocal identification of respective germplasms for conservation purposes (Karp et al. 1998).

Acknowledgements The author would like to thank Natalia Lest, Franziska Müller and Roland Kürschner for their excellent

technical assistance. Financial support of the Bundesministerium für Wirtschaftliche Zusammenarbeit (BMZ) and the Deutsche Forschungsgemeinschaft (DFG), Project No. FR 682/9-2 is gratefully acknowledged.

References

Ahmed MM, JH Sanders JH, Nell WT (2000) New sorghum and millet cultivar introduction in Sub-Saharan Africa: impacts and research agenda. Agric Systems 64:55–65

Ayana A, Bryngelsson T, Bekele E (2000) Genetic variation of Ethiopian and Eritrean sorghum (*Sorghum bicolor* (L.) Moench) germplasm assessed by random amplified polymorphic DNA (RAPD). Genet Res Crop Evol 47:471–428

Bohn M, Utz FH, Melchinger AE (1999) Genetic similarities among winter wheat cultivars determined on the basis of RFLPs, AFLPs and SSRs and their use for predicting progeny variance. Crop Sci 39:228–237

Brown SM, Hopkins MS, Mitchell CV, Senior ML, Wang TY, Duncan RR, Gonzalez-Candelas F, Kresovich S (1996) Multiple methods for the identification of polymorphic simple sequence repeats (SSRs) in sorghum (*Sorghum bicolor* (L.) Moench). Theor Appl Genet 93:190–198

Dean RE, Dahlberg JA, Hopkins MS, Mitchell CV, Kresovich S (1999) Genetic redundancy and diversity among 'orange' accessions in the U.S. national sorghum collection as assessed with simple sequence repeat (SSR) markers. Crop Sci 39:1215–1221

Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M, Freimer NB (1994) Mutational processes of simple-sequence repeat loci in human populations. Proc Natl Acad Sci USA 91:3166–3170

Djè Y, Forcioli D, Ater M, Lefèbvre C, Vekemans X (1999) Assessing population genetic structure of sorghum landraces from North-Western Morocco using allozyme and microsatellite markers. Theor Appl Genet 99:157–163

Djè Y, Heurtz M, Lefèbre C, Vekemans X (2000) Assessment of genetic diversity within and among germplasm accessions in cultivated sorghum using microsatellite markers. Theor Appl Genet 100:918–925

Doyle JF, Doyle JL (1990) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Focus 12:13–15

Efron B, Tibshirani T (1986) Bootstrap coefficient of variation, confidence limits and other methods for statistical accuracy measurements. Theor Popul Biol 53:256–271

Feldmann MW, Bergmann A, Pollock DD, Goldstein DB (1997) Microsatellite genetic distances with range constraints: analytical description and problems of estimation. Genetics 145:207–216

Goldstein DB, Ruíz-Linares A, Cavalli-Sforza LL, Feldmann MW (1995a) An evaluation of genetic distances for use with microsatellite loci. Genetics 139:463–471

Goldstein DB, Ruíz-Linares A, Feldman M, Cavalli-Sforza LL (1995b) Genetic absolute dating based on microsatellites and the origin of modern humans. Proc Natl Acad Sci USA 92:6720–6727

Graner A, Ludwig WF, Melchinger AE (1994) Relationships among European barley germplasm. Crop Sci 34:1199–1205

Grenier C, Deu M, Kresovich S, Bramel-Cox PJ, Hamon P (2000) Assessment of genetic diversity in three subsets constituted from the ICRISAT sorghum collection using random vs non-random sampling procedures B. Using molecular markers. Theor Appl Genet 101:197–202

Haussmann BIG, Obilana AB, Ayiecho PO, Blum A, Schipprack W, Geiger HH (2000) Yield and yield stability of four population types of grain sorghum in a semi arid area of Kenya. Crop Sci 40:319–329

Karp A, Isaac PG, Ingram DS (1998) Molecular tools for screening biodiversity. Chapman & Hall, London

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

Le Thierry d'Ennequin M, Pannaud O, Toupance B, Sarr A (2000) Assessment of genetic relationships between *Setaria italica* and its wild relative *S. viridis* using AFLP markers. Theor Appl Genet 100:1061–1066

Lima MLA, Garcia AAF, Oliveira KM, Matsuoko S, Arizono H, de Souza Jr CL, de Souza AP (2002) Analysis of genetic similarity detected by AFLP and coefficient of parentage among genotypes of sugar cane (*Saccarum* spp). Theor Appl Genet 104:30–38

Lynch M, Milligan BG (1994) Analysis of population genetic structure with RAPD markers. Mol Ecol 3:91–99

Mantel M (1967) The detection of disease clustering and generalized regression approach. Cancer Res 27:209–220

Nei M (1973) Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci USA 70:3321–3323

Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleasis. Proc Natl Acad Sci USA 76:5269–5273

Ordon F, Bauer E, Friedt W, Graner A (1995) Marker-based selection for the ym4 BaMMV-resistance gene in barley using RAPDs. Agronomie 15:481–485

Ordon F, Schimann A, Friedt W (1997) Assessment of the genetic relatedness of barley accessions (*Hordeum vulgare* s.l.) resistant to soil-borne mosaic-inducing viruses (BaMMV, BaYMV, BaYMV-2) using RAPDs. Theor Appl Genet 94:325–330

Pejic I, Ajmone-Marsan P, Morgante M, Kozumplick V, Castiglioni P, Taramino G, Motto M (1998) Comparative analysis of genetic similarity among maize inbred lines detected by RFLPs, RAPDs, SSRs and AFLPs. Theor Appl Genet 97:1248–1255

Pollock DD, Bergmann A, Feldmann MW, Goldstein DB (1998) Microsatellite behaviour with range constraints: Parameter estimation and improved distances for use in phylogenetic reconstruction. Theor Pop Biol 53:256–271

Ribaut JM, Hoisington D (1998) Marker-assisted selection: new tools and strategies. Trends Plant Sci 3:236–239

Rohlf FJ (2000) NTSYS-pc numerical taxonomy and multivariate analysis system, version 2.1. Exeter Publ, New York

Russel JR, Fuller JD, Macaulay M, Hatz BG, Jahoor A, Powell W, Waugh R (1997) Direct comparison of levels of genetic variation among barley accessions detected by RFLPs, AFLPs, SSRs and RAPDs. Theor Appl Genet 95:714–722

Scheurer KS, Friedt W, Huth W, Waugh R, Ordon F (2001) QTL analysis of tolerance to a German strain of BYDV-PAV in barley (*Hordeum vulgare* L.). Theor Appl Genet 103:1074–1083

Schiemann A, Dauck V, Friedt W, Streng S, Graner A, Ordon F (1999) Establishment of a flourescence-based AFLP technique and rapid marker detection for the resistance locus *rym5*. Barley Genet Newslett 29:http://wheat.pw.usda.gov/ggpages/bgn/29/a29-01.html

Simioniuc D, Uptmoor R, Friedt W, Ordon W (2002) Genetic diversity and relationships among pea cultivars (*Pisum sativum* L.) revealed by RAPDs and AFLPs. Plant Breed 121:429–435

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

Tanksley SD, McCouch R (1997) Seed banks and molecular maps: Unlocking genetic potential from the wild. Science 277:1063–1066

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

Tinker NA, Fortin MG, Mather DE (1993) Random amplified polymorphic DNA and pedigree relationships in spring barley. Theor Appl Genet 85:976–984

Tivang JG, Nienhuis J, Smith OS (1994) Estimation of sampling variance of molecular marker data using the bootstrap procedure. Theor Appl Genet 89:259–264

Udupa SM, Robertson LD, Weigand F, Baum M, Kahl G (1999) Allelic variation at $(TAA)_n$ microsatellite loci in a world chickpea collection (*Cicer arietinum* L.) germplasm. Mol Gen Genet 261:354–363

Vekemans X, Beauvens T, Lemaire M, Roldan-Ruiz I (2002) Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. Mol Ecol 11:139–151

Vos P, Hogers R, Bleeker M, Reijans M, Van De Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA Fingerprinting. Nucleic Acids Res 23:4407–4414

Wenzel WG, van Loggerenberg M, Ordon F (2001a) Quick screening methods for sorghum quality traits. J Appl Bot 75:43–45

Wenzel WG, Ayisi KK, Mogashoa A, Donaldson G, Mohammed R, Uptmoor R, Ordon F, Friedt W (2001b) Improved sorghum varieties for smallholder farmers. J Appl Bot 75:207–209