CROP BREEDING, GENETICS & CYTOLOGY

SSR and Pedigree Analyses of Genetic Diversity among CIMMYT Wheat Lines Targeted to Different Megaenvironments

S. Dreisigacker, P. Zhang, M. L. Warburton, M. Van Ginkel, D. Hoisington, M. Bohn, and A. E. Melchinger*

ABSTRACT

Improved bread wheat (*Triticum aestivum* L.) cultivars for diverse agroecological environments are important for success in the effort to increase food production. In the 1980s, CIMMYT introduced the megaenvironment (ME) concept to breed wheats specifically adapted to different areas. Our objective was to analyze the genetic diversity among 68 advanced CIMMYT wheat lines targeted to different MEs by using 99 simple sequence repeats (SSRs) and the coefficient of parentage (COP). The average number of alleles detected was higher for the 47 genomic SSRs (5.4) than for the 52 SSRs derived from expressed sequence tags (EST) (3.3), but gene diversity between MEs was similar for both types of markers. No significant differences among the five MEs were observed for the means of SSR-based genetic similarities (GS), calculated as 1 − Rogers' distance, and COP values. Both measures showed a low correlation (r = 0.43). High levels of genetic diversity were found within the germplasm targeted to each ME. However, principle coordinate analysis based on modified Rogers' distances did not separate the genotypes according to their targeted MEs. We conclude that presence of a single core germplasm can reflect large phenotypic differences. A sufficient number of diverse breeding lines for each ME is required because MEs generally combine various production areas. SSRs represent a powerful tool to quantify genetic diversity in wheat, but genotypic differentiation for adaptation to specific MEs in the CIMMYT program could not be proven.

Wheat, together with maize (*Zea mays* L.) and rice (*Oryza sativa* L.), is one of the three major food crops in the world. It is grown in a variety of environments, ranging from fully irrigated (e.g., northern India, Egypt), to high rainfall (e.g., northwestern Europe, eastern Africa, southern zone of Latin America), and drought-prone regions (e.g., U.S. Great Plains, most of Australia, parts of Argentina). In these areas wheat production experiences a range of biotic and abiotic stresses and crop improvement requires precise focusing on the needs of the crop in each area, the producers, the processing industry, and the consumers (Lantican et al., 2002).

More than one half of the wheat production environments are located in developing countries, which fall within the mandate of CIMMYT. In the 1980s, CIMMYT introduced the concept of breeding for different MEs. A ME is defined as a large, not necessarily contiguous area, which usually encompasses more than one country and is frequently transcontinental. It is characterized by similar biotic and abiotic stress conditions, cropping systems, and consumer demands (Rajaram et al., 1994). Twelve MEs have been classified, six of which are focused on efficient selection of better-adapted spring bread wheat, the dominant type of wheat in developing countries. The concept has permitted expanding breeding efforts relevant within each ME. In breeding for enhanced adaptation, adequate genetic diversity is a prerequisite for any crop improvement program. The genetic progress through selection is directly related to the variability present in the gene pool, and the quality of the genes contributed by the parents.

The COP is an indirect measure of genetic diversity among genotypes based on the probability that alleles at a certain locus are identical by descent. Calculation of COP values rests on simplifying assumptions regarding the relatedness of ancestors, parental contribution to the offspring, and absence of selection and genetic drift, which are not met under breeding conditions (Cox et al., 1985; Cowen and Frey, 1987). In contrast, molecular markers measure diversity directly at the DNA level. In studies of autogamous crops with low levels of apparent genetic variability such as wheat, soybean [*Glycine max* (L.) Merr.], and rice, SSRs proved to be a suitable marker system. They are generally genome specific, abundant, codominant in nature, and show a fairly uniform distribution over the genome. SSRs have been applied in many aspects of genetic diversity analyses such as genetic differentiation caused by selection (Stachel et al., 2000), fingerprinting of genotypes to analyze the structure of germplasm collections (Parker et al., 2002; Huang et al., 2002), and the analysis of temporal changes in diversity (Donini et al., 2000; Christiansen et al., 2002).

Traditional methods to develop SSRs are based on isolating and sequencing genomic libraries, which contain putative SSR tracts (Adams et al., 1992). A novel source for generating SSRs is provided by screening EST databases available online (Kota et al., 2001). This recent approach allows researchers to shift from the use of anonymous markers with unknown effect on the phenotype to markers physically associated with coding

S. Dreisigacker and A.E. Melchinger, Inst. of Plant Breeding, Seed Science, and Population Genetics, Univ. of Hohenheim, 70593 Stuttgart, Germany; M. Bohn, Dep. of Crop Science, Univ. of Illinois, Urbana, IL 61801; P. Zhang, M. van Ginkel, M.L. Warburton, and D. Hoisington, CIMMYT, Mexico D.F., Mexico. Received 17 March 2003. *Corresponding author (melchinger@pz.uni-hohenheim.de).

Published in Crop Sci. 44:381–388 (2004).

677 S. Segoe Rd., Madison, WI 53711 USA

Abbreviations: AMOVA, analysis of molecular variance; CIMMYT, International Maize and Wheat Improvement Center; COP, coefficient of parentage; EST, expressed sequence tag; GS, genetic similarity; ME, megaenvironment; SSR, simple sequence repeat.

regions, which may more accurately reflect the effects of selection, both natural and artificial.

The objectives of the present study were to (i) evaluate the use of genomic and EST-derived SSRs for determining the genetic diversity among advanced spring bread wheat lines from the CIMMYT breeding program, (ii) compare genetic distances based on SSRs with the COP estimates of these wheat lines, and (iii) determine the diversity for SSRs within and among sets of lines targeted to different MEs.

MATERIALS AND METHODS

Plant Materials

A total of 68 CIMMYT advanced spring bread wheat lines from crosses made during 1989 to 1996 were chosen for this study (Table 1). Most of the lines were bred by a "modified bulk" procedure described by Van Ginkel et al. (2002). Seed of outstanding F_7 lines was harvested in bulks for subsequent yield trials. These yield trials were grown in replicated and latinized α-lattice designs at Cd. Obregon (Sonora, Mexico) or Toluca (Mexico State, Mexico) in 2000 and 2001 under conditions simulating the different MEs for which they are being bred (e.g., full irrigation, reduced irrigation, drought, and heat stress, etc.). On the basis of their performance, 8 to 15 advanced lines were selected from yield trials representative of the first five spring bread wheat MEs (Table 2). The lines were chosen as candidates for further evaluation at international testing sites (Van Ginkel et al., 2002). Progenies from three crosses (Alucan/Duluca, PF869107/CEP8825//Milan and Babax/Amadina//Babax) were identified and selected for more than one ME.

SSR Analyses

DNA extraction was performed with the CTAB method of Saghai-Maroof et al. (1984) modified according to CIMMYT Applied Biotechnology Center's Manual of Laboratory Protocols (Hoisington et al., 1994). Twenty seeds per advanced line were grown in the greenhouse and after 2 wk young leaves were harvested from 5 to 10 plants per line. Leaves were bulked for DNA extraction to assess the genetic variability within each line as described by Gilbert et al. (1999). Quality and quantity of the isolated DNA was determined on 1% (w/v) agarose gels by comparing bands to known concentrations of λ DNA.

SSR information was obtained from two different sources: 46 SSRs were collected from a conventional genomic library (genomic SSRs) developed at IPK Gatersleben by Röder et al. (1998 and unpublished data) and 51 SSRs derived from ESTs (EST-SSR) with the prefix "*DuPw*" were kindly provided by DuPont, Wilmington, DE (Dupont, unpublished data; Eujayl et al., 2002). The SSRs from both sources were distributed equally over the genome. In addition, the 1BS/1R translocation EST-SSR marker *Taglgap* (Devos et al., 1995) and the SSR marker *WMC56* developed by the Wheat Microsatellite Consortium (Agrogene, France) were used. Details for each of the 99 SSRs can be found online (http://www.cimmyt.org/english/webp/support/publications/support_materials/ssr_mw1.htm; verified 23 September 2003).

PCR reactions were performed in a model PTC225 thermocycler (MJ Research, Inc., Waltham, MA). Each 20-μL reaction mixture contained 25 ng template DNA, 150 n*M* of each primer, 250 μ*M* dNTPs, 200 μ*M* $MgCl_2$, 1× PCR buffer and 2.5 U of *Taq*-polymerase. Forward primers were labeled at the 5′ end with either one of three phosphoramidite fluorescent dyes 6-carboxyfluoresein, tetrachloro-6-carboxyfluoresein, or hexachloro-6- carboxyfluoresein. PCR was performed with the following standard temperature profile: 29 cycles with a 1 min denaturing step at 94°C, 2 min annealing temperatures between 50 and 64°C depending on the different primer combinations, and 2 min extension at 72°C. The 1-min time spread of the standard profile cycle was modified in some cases to fully optimize amplification conditions.

Amplification products were separated on an ABI™377 Sequencer (Perkin Elmer/Applied Biosystems, Foster City, CA) using 4.5% (w/v) polyacrylamide denaturing gels (acrylamide:bisacrylamide 29:1). Running conditions were 2400 V, 40 mA, 120 W electrophoresis power and 40 mW laser power. Products from up to five SSRs could be distinguished simultaneously because of the three different fluorescent dyes and migration distance differences. Fragment sizes were calculated semiautomatically by the computer software GeneScan 3.1 (Perkin Elmer/Applied Biosystems) by comparing fragments with an internal size standard (GeneScan 350 or 500) labeled with *N,N,N,N*,-tetramethyl-6-carboxyrhodamine. GeneScan fragments were assigned to alleles by the category function of the software Genotyper 2.1 (Perkin Elmer/Applied Biosystems). Sixty-four genotypes were run on each gel plus two wheat lines, Opata and Synthetic, as controls.

We could not optimize the amplification profile of nine SSRs for the scoring with Genotyper. These markers were optimized to run on small (16 by 20 cm) 6% (w/v), 19:1 acrylamide:bis-acrylamide denaturing gels (*ATTO*[8] AE-6220). The gels were run for 2 h at about 350 V, with a 100-bp ladder as a standard. For fragment visualization, silver staining was applied according to Applied Biotechnology Center's Manual of Laboratory Protocols. The fragment length of each SSR was determined with the scientific image system Kodak ID 2.02 (Kodak, New Haven, CT).

Statistical Analyses

Reproducibility of SSR amplification and scoring was determined on the basis of the percentage of disagreements in the fragment size of the two standard lines, Opata and Synthetic, for gels loaded with the same markers. Allele frequencies at the 99 loci, total gene diversity (H_T), gene diversity within MEs (H_S), and the proportion of diversity resulting from gene differentiation between MEs (G_{ST}) were calculated according to Nei (1987). The measures were considered separately for the two SSR sources to examine the influence of the genome location of the markers (genomic and EST-SSRs).

The COP for all pairwise combinations of wheat lines was calculated on the basis of fully expanded genealogical information extracted from CIMMYT's International Wheat Information System (Payne et al., 2002). Calculations of COP were based on the assumptions described by St. Martin (1982), except for sister lines, which were assigned a COP of 0.56 instead of 1.0 following Cox et al. (1985). Genetic similarity (GS) assigned as 1 − Rogers' distance (Rogers, 1972) was estimated to compare SSR-based and COP estimates. Pearson's correlation coefficient (r) between GS and COP values was calculated for related pairs of lines (COP ≥ 0.05). Standard errors of genetic similarity estimates were obtained by a bootstrap procedure with resampling over markers (Weir, 1996). Furthermore, modified Rogers' distance (Wright, 1978) was calculated among all possible pairs of lines as a basis for the application of multivariate methods, because it represents a Euclidean distance.

An analysis of molecular variance (AMOVA) on the basis of SSR data was computed to test the differentiation of the 68 genotypes according to the five MEs. The hierarchical analysis

Table 1. Pedigrees of the 68 CIMMYT spring bread wheat lines classified by megaenvironments (ME).

No.	Pedigree†
ME1IR (irrigated zones)‡	
1	KAUZ//ALTAR 84/AOS/3/KAUZ
2	ATTILA/3/HUI/CARC//CHEN/CHTO/4/ATTILA
3	PRINIA/WEAVER//STAR/3/WEAVER
4	OASIS/4*BORL95
5	WEAVER/WL3926//SW89.3064
6	RABE/2*MO88
7	CNDO/R143//ENTE/MEXI_2/3/AEGILOPS SQUARROSA (TAUS)/4/WEAVER/5/2*KAUZ
8	CHEN/AEGILOPS SQUARROSA (TAUS)//FCT/3/2*WEAVER
9	CHEN/AEGILOPS SQUARROSA (TAUS)//FCT/3/STAR
10	CHUM18/5*BCN
11	P1.861/RDWG
12	CMH80A.542/CNO79
13	SUPER SERI #2
14	PVN//CAR422/ANA/5/BOW/CROW//BUC/PVN/3/YR/4/TRAP#1
15	BABAX/AMADINA//BABAX (WEEBILL1)
ME1HT (irrigated hot zones)	
16	VEE/PJN//2*TUI
17	PFAU/WEAVER
18	CAZO/KAUZ//KAUZ
19	MNCH/3*BCN
20	W462//VEE/KOEL/3/PEG//MRL/BUC
21	XIANG82.2661/2*KAUZ
22	SW89-5124*2/FASAN
23	KEA/TAN/4/TSH/3/KAL/BB//TQFN/5/PAVON/6/SW89.3064
24	LAJ3302/2*MO88
25	CROC_1/AE.SQUARROSA (205)//2*BCN
26	PICUS/4/CS5A/5RL-1//BUC/BJY/3/ALD/PVN/5/LAJ3302
27	PFAU/MILAN
28	TAM200/TUI
29	SABUF/7/ALTAR 84/AE.SQUARROSA (224)//YACO/6/CROC_1/AE.SQUARROSA (205)/5/BR12*3/4/IAS55*4/CI14123/3/IAS55*4/EG,AUS//IAS55*4/ALD
30	[KASYON]//PVN/SPRW
ME2 (high rainfall zones)	
31	ALUCAN/DUCULA
32	IAS58/4/KAL/BB//CJ71/3/ALD/5/CNR/6/THB/CEP7780
33	R37/GHL121//KAL/BB/3/JUP/MUS/4/2*YMI #6/5/CBRD
34	TNMU/6/CEP80111/CEP81165/5/MRNG/4/YKT406/3/AG/ASN//ATR
35	TNMU/MILAN
36	TNMU/6/PEL74144/4/KVZ//ANE/MY64/3/PF70354/5/BR14/7/BR35
37	DUCULA/TNMU
38	TRAP#1/BOW//VEE#5/SARA/3/ZHE JIANG 4/4/DUCULA
39	PASTOR//MUNIA/ALTAR 84
40	OR791432/VEE#3.2//MILAN
41	MUNIA/ALTAR 84//AMSEL
42	TNMU/MUNIA
43	TNMU/ATTILA
44	HXL8088/DUCULA
45	PF869107/CEP8825//MILAN
ME3 (high rainfall, acid soil zones)	
46	ALUCAN/DUCULA
47	TNMU/TUI
48	TNMU/OCEP17
49	OCEP15/KAUZ//TNMU
50	BR14*2/SUM3//TNMU
51	TNMU/BR35//THB/CEP7780
52	KVZ/3/TOB/CTFN//BB/4/BLO/5/TAN/6/PRL/7/MILAN
53	PF869107/CEP8825//MILAN
ME4 (semiarid zones)	
54	DUCULA//VEE/MYNA
55	SRMA/TUI
56	CROC_1/AE.SQUARROSA (224)//OPATA
57	PIOS/DUCULA
58	LAJ3302/3/GZ156/NAC//PSN/URES/4/WEAVER
59	3VASKAR/G303.1M.1.3.2.2.2//KAUZ/3/SKAUZ/4/KAUZ
60	PASA/SAET
61	NL456/VEE#5//CHIL/3/MUNIA
62	TZPP*2/ANE//INIA/3/CNO67/JAR//KVZ/4/MN72252/5/SHI#4414/CROW
63	TSI/VEE#5//KAUZ
64	KAUZ/5/PAT10/ALD//PAT72300/3/PVN/4/BOW
65	KA/NAC
66	ALTAR 84/AE.SQ//2*OPATA
67	FRET2
68	BABAX/AMADINA//BABAX (WEEBILL1)

† Nomenclature according to Purdy et al. (1968): The initial cross is indicated by a single slash (e.g. A/B), the second cross by a double slash (e.g. A/B//C), and subsequent crosses in numerical order by flanked single slashes (e.g. A/B//C/3/D). Backcrosses are designated with an asterisk (*) and a number indicating the dosage of the recurrent parent.

‡ Refers to the five MEs described in Table 2.

Table 2. Characterization of important spring bread wheat megaenvironments (ME) defined by CIMMYT (Rajaram et al., 1994).

ME	Moisture regime	Temperature regime	Breeding objectives in addition to yield†	Year‡	Yield trials	Contribution of wide crosses§
ME1IR	Low rainfall, irrigated	Temperate	Resistance to lodging, SR, YR and LR, end-use quality	1945	Obregon: 700 mm by irrigations	China, India, Synthetic, Durum wheats
ME1HT	Low rainfall, irrigated	Hot	As for ME1IR plus tolerance to heat	1945	Obregon: 700 mm by irrigations, late planting	China, Argentina, India, Synthetic, Durum wheats
ME2	High rainfall	Temperate	Resistance to SR, YR, LR, *Septoria* spp., FHB, BYDV, waterlogging, pre-harvest sprouting. end-use quality	1972	Toluca: high rainfall (800 mm)	China, Brazil, Durum wheats
ME3	High rainfall	Temperate	As for ME2 plus Al and Mn tolerance, P-use efficiency	1974	Toluca: high rainfall (800 mm), seedling test for low pH and Al toxicity	China, Brazil
ME4	Low rainfall	Temperate or hot	Resistance to SR, YR, LR plus tolerance to drought, end-use quality	1970	Obregon: one pre-seeding irrigation, 300 mm available	Argentina, Nepal, Synthetic wheats

† SY = Stem rust, YR = Yellow rust, LR = Leaf rust, FHB = Fusarium head blight, BYDV = Barley Yellow Dwarf Virus.
‡ Refers to the year in which breeding for the respective ME began at CIMMYT.
§ Refers to the 68 lines included in this study.

divides the total variance into variance components due to intra- or inter-ME differences and tests their significance. Principal coordinate analysis was performed on the basis of the modified Rogers' distances to visualize the dispersion of the genotypes (Gower, 1966). The K-means clustering algorithm was used to identify groups of similar lines, on the basis of a least-squares partitioning method, which divides a collection of objects into *k* clusters depending on minimum distances to the centers of the clusters (MacQueen, 1967). COP values were calculated to the six progenitors most frequently used in the crosses and averaged within each ME and K-means cluster.

All analyses were performed with the Plabsim software (Frisch et al., 2000), which is implemented as an extension of the statistical software R (Ihaka and Gentleman, 1996). The AMOVA was performed by the software package Arlequin 2.0 (Schneider et al., 2000).

RESULTS

For the 68 CIMMYT advanced lines analyzed with 99 SSRs, a total of 425 alleles was detected with an average of 4.3 alleles per locus. The average number of alleles was considerably lower in EST-SSRs (3.3) than in genomic SSRs (5.4), with seven out of the 52 EST-SSRs being monomorphic (Table 3). However, an important feature of EST-SSRs was the high-quality fragment patterns obtained, which were devoid of stutter bands, resulting in a higher reproducibility (98.8%) and lower residual heterozygosity (1.4%) than with genomic SSRs (89.5 and 3.7%, respectively).

The total gene diversity (H_T) varied widely among loci from 0.01 at *DuPw138* to 0.83 at *Xgwm437*, with an average of 0.47 (Table 3). Considering the two different SSR sources, the average H_T and H_S values were lower for EST-SSRs than for genomic SSRs. The corresponding G_{ST} value for all loci was 0.09 and for EST-SSRs (0.10) just slightly higher than for genomic SSRs (0.09).

The mean COP value over the 68 wheat genotypes was 0.14 and ranged from 0.01 to 0.87 for closely related pairs (Table 4). The mean COP values within MEs did not substantially differ between the five MEs. Thirty-nine percent of the COP values were smaller than 0.10, indicating that theoretically less than 10% of the genetic material segregating in ancestral populations was identical by descent in any two cultivars.

GS for all pairs of lines ranged from 0.39 to 0.91 with an average of 0.59 for all genotypes (Table 4). Similar to the COP values, MEs were not significantly different in their mean GS, but with an equal range of values. In the specific cases in which two progenies of the crosses Alucan/Duluca, PF869107/CEP8825//Milan and Babax/Amadina//Babax were selected for different MEs, GS values were high (0.88, 0.86, and 0.72, respectively), as expected.

The mean COP and GS values between MEs were of similar size as the means within MEs (Table 5). ME1HT was most distant to ME2 and ME3 on the basis of COP, and ME1IR most distant to ME3 on the basis of GS values. The AMOVA confirmed these results in that 92% of the total variation was found within MEs and just 8% between MEs (data not shown). The correlation between GS and COP values was $r = 0.43$.

The principal coordinate analysis based on modified Rogers' distances did not separate the genotypes according to their targeted MEs (Fig. 1). Fourteen of the chosen genotypes cluster somewhat together because of their resistance to acid soil. The K-means cluster algorithm identified more than one solution, the most frequent (90%, 1000 repetitions) comprising three defi-

Table 3. Average number of alleles per locus, residual heterozygosity, reproducibility and gene diversity estimated over the two different sources of markers used in this study.

SSR source	No. of SSR	Avg. no. of alleles/locus	Heterozygosity	Reproducibility	Gene diversity†		
					H_T	H_S	G_{ST}
Genomic SSRs	47	5.4	3.7	89.5	0.57	0.52	0.09
EST derived SSRs	52	3.3	1.4	98.8	0.37	0.33	0.10
Total	99	4.3	2.5	95.7	0.47	0.43	0.09

† H_T = total gene diversity, H_S = diversity within megaenvironments (MEs), G_{ST} = diversity between MEs.

Table 4. Total and unique number of alleles, number of monomorphic loci, mean genetic similarities (GS), and mean coefficient of parentage (COP) within each megaenvironment (ME).

ME	No. of lines	No. of alleles	No. of unique alleles†	No. of monomorphic loci	GS			COP		
					Mean	Min.	Max.	Mean	Min.	Max.
ME1IR	15	267	20	21	0.59 ± 0.06	0.44	0.85	0.18 ± 0.11‡	0.02	0.59
ME1HT	15	289	33	17	0.59 ± 0.07	0.45	0.91	0.14 ± 0.11	0.01	0.48
ME2	15	253	12	22	0.64 ± 0.06	0.52	0.80	0.11 ± 0.08	0.02	0.37
ME3	8	232	18	24	0.60 ± 0.06	0.47	0.73	0.13 ± 0.06	0.04	0.24
ME4	15	272	23	17	0.63 ± 0.06	0.48	0.85	0.18 ± 0.08	0.04	0.50
Total	68	425	–	7	0.59 ± 0.06	0.39	0.91	0.14 ± 0.09	0.01	0.87

† Alleles occurring only in one ME.
‡ Standard deviation.

nite centers. K-means tended to form the clusters on the basis of common progenitors used in the crosses made during 1989 and 1996. Seven of the 11 lines with Kauz in the pedigree were included in cluster K1 (Table 6). Four lines containing Kauz did not group into this cluster. These lines had Chinese wheats in their pedigree or were selected in later segregating generations under different environmental conditions by the International Center of Agricultural Research in Dry Areas in Syria. On the basis of COP, Weaver holds the highest parental contribution to cluster K2 and Milan to cluster K3. Progenitor Tinamou contributed about equally to clusters K2 and K3. Twenty of the 68 genotypes had durum (*T. durum* Desf.) wheat, Chinese wheat, or synthetic hexaploid wheat in their pedigree. These lines were scattered all over the principal coordinate analysis plot because of different sources of Chinese lines and *Aegilops squarrosa* L. in the synthetic hexaploid wheats used as crossing parents.

DISCUSSION

Use of Genomic and EST-SSRs in Breeding Programs

Genomic SSR markers have been intensively used to detect the variability between bread wheat genotypes, but the large genome size of wheat is a challenge in identifying sufficiently robust and informative SSRs for fingerprinting. EST-SSRs present a novel source of SSRs and have some intrinsic advantages over genomic SSRs. They can be developed from available EST databases and their frequency is abundantly high in transcribed regions (Morgante et al., 2002). A concern is that the coding character of EST-SSRs limits their level of polymorphism.

Our results agree with other studies in rice (Cho et al., 2000), grape (*Vitis* ssp., Scott et al., 2000), and wheat (Eujayl et al., 2002) in that the overall level of polymorphism for genomic SSRs was higher than for EST-SSRs. However, compared with the latter study with 64 durum wheat lines, where on average 5.5 alleles per locus for genomic SSRs and 4.1 for EST-SSRs were found, we detected slightly lower average numbers of alleles for both genomic (5.2) and EST-SSRs (3.2). The somewhat higher level of diversity reported by Eujayl et al. (2002) may be attributable to the more variable material in their study, which comprised a sample of various lines and landraces with different genetic backgrounds. Here, we studied advanced breeding lines ready for international dissemination to developing country breeding programs.

Higher-order repeat motifs were applied in our sample of EST-SSRs. In comparison to dinucleotide motifs, they are generally less polymorphic and insensitive to single-nucleotide polymorphisms in the flanking regions of the SSRs, which facilitate the designation of allele sizes (Chakraborty et al., 1997; Song et al., 2002; Mogg et al., 2002). This explains the higher reproducibility and the lower degree of heterozygosity or heterogeneity for EST-SSRs compared with genomic SSRs in our study.

EST-SSRs are assumed to reflect more accurately the effects of selection under which the germplasm has been developed. However, only a slightly better differentiation of the allele frequencies between MEs was observed for EST-SSRs than for genomic SSRs. Wheat is highly autogamous and, therefore, differential selection of fitness-related target loci will also affect genomic SSRs linked to them. Further evidence shows a functional importance of genomic SSR structures, which may cause some form of balancing selection (Innan et al., 1997; Li et al., 2000; Li et al., 2002).

Future opportunities to combine markers and phenotypic data in association studies may improve the application of EST-SSRs in the evaluation of germplasm, as exemplified in maize by Thornsberry et al. (2001). A specific marker could then be used to examine the functional diversity at a certain locus. We speculate that EST-SSRs from genes contributing to specific ME adap-

Table 5. Mean coefficient of parentage (above diagonal) and genetic similarity estimates (below diagonal) between five CIMMYT megaenvironments (ME).

	ME1IR	ME1HT	ME2	ME3	ME4
ME1IRR		0.16 ± 0.12†	0.11 ± 0.06	0.11 ± 0.05	0.18 ± 0.10
ME1HT	0.57 ± 0.05		0.10 ± 0.05	0.10 ± 0.07	0.16 ± 0.10
ME2	0.62 ± 0.06	0.59 ± 0.05		0.14 ± 0.11	0.12 ± 0.04
ME3	0.56 ± 0.05	0.58 ± 0.08	0.59 ± 0.06		0.12 ± 0.06
ME4	0.58 ± 0.06	0.59 ± 0.07	0.62 ± 0.07	0.60 ± 0.07	

† Standard deviation.

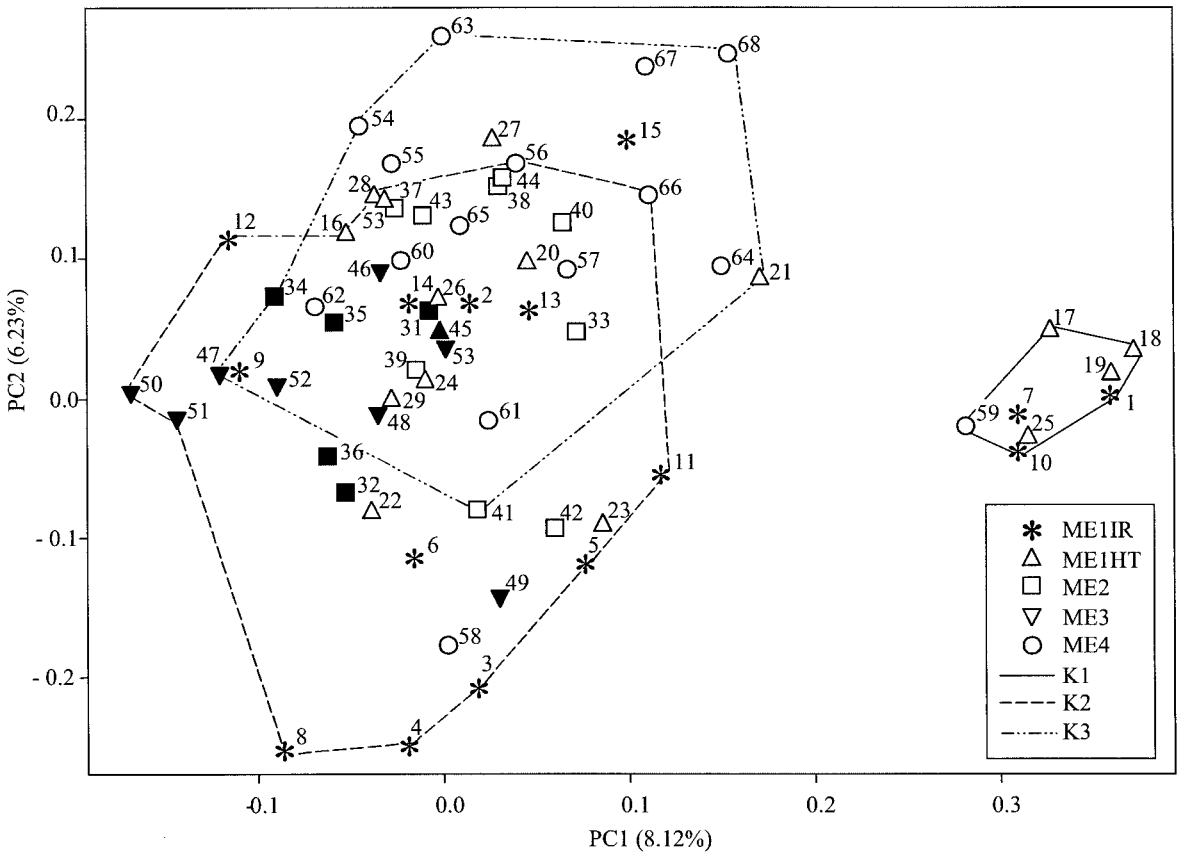

Fig. 1. Associations among 68 CIMMYT wheat lines revealed by principal coordinate analysis performed with modified Rogers' distance estimates calculated from 99 SSR loci. Numbers refer to the 68 wheat lines, megaenvironments (ME) are designated by circles, triangles, and squares. K-mean clusters are indicated by solid and dashed lines, respectively.

tation might have revealed more clearly the effects of selection in our germplasm.

Correlation between Pedigree and SSR-Based Distance Estimates

In agreement with previous studies in wheat, the correlation between GS and COP estimates reported here was low but significant ($r = 0.43$). This low correlation can be explained by the unrealistic assumptions in calculating COP values and the substantial variation in GS estimates of unrelated lines (Graner et al., 1994). With increasing relatedness of lines, the association between COP and their corresponding GS values should become tighter. In fact, a higher correlation ($r = 0.55$) between COP (COP ≥ 0.01) and GS estimates was found in the study of Plaschke et al. (1995), where the average COP among a set of European wheats was 0.29. However, selection and drift are presumably also important factors reducing the correlation between GS and COP due to a shift in the genomic contribution of parental lines, particularly during the early selfing generations.

Variation of Genetic Diversity in Megaenvironments

Provided molecular markers represent an accurate picture for the genetic diversity at functional genes, the

Table 6. Average coefficient of parentage (COP) of predominant progenitors used (i) in crosses of this study within megaenvironments (ME) and (ii) in clusters revealed by the K-means algorithm performed on modified Rogers' distance.

		Predominant progenitor					
Groups	**Advanced lines**	**Kauz**	**Parula**	**Weaver**	**Tinamou**	**Milan**	**Ducula**
ME							
ME1IR	**1-15**	**16.5**	**12.1**	**18.8**	**0.0**	**0.0**	**0.0**
ME1HT	**16-30**	**20.8**	**4.2**	**3.3**	**0.0**	**3.3**	**0.0**
ME2	**31-45**	**0.0**	**0.0**	**0.0**	**18.3**	**16.7**	**13.3**
ME3	**46-53**	**3.1**	**0.0**	**0.0**	**28.1**	**12.5**	**6.3**
ME4	**54-68**	**12.5**	**1.7**	**3.3**	**0.0**	**0.0**	**6.7**
Mean		**11.4**	**4.0**	**5.6**	**7.4**	**3.7**	**5.2**
K-means cluster performed on MRD							
K1	**1, 7, 10, 17, 18, 19, 25, 59**	**71.5**	**4.7**	**9.4**	**0.0**	**0.0**	**0.0**
K2	**3-6, 8, 9, 11, 12, 16, 22, 23, 26, 28, 29, 32, 36, 42, 45, 48-53, 56, 58, 62, 66**	**0.8**	**6.9**	**11.0**	**8.9**	**0.0**	**5.4**
K3	**2, 13-15, 20, 21, 24, 27, 30, 31, 33-35, 37-41, 43, 44, 46, 47, 54, 55, 57, 60, 61, 63-66, 67, 68**	**5.2**	**1.1**	**0.0**	**7.4**	**10.6**	**2.9**

true genetic differences between the germplasm targeted for different MEs are small. Several reasons might explain this observed absence of genetic differentiation: (i) selection based on ME adaptation has not been practiced long enough to differentiate the germplasm, (ii) genes conferring fitness to one ME are not unique to that ME and may confer fitness to several MEs, and (iii) adaptation to MEs is not based on an accretion of random genes but rather a limited set of specific genes.

CIMMYT's concept of breeding for different MEs was implemented in the 1980s. Thus, selection history for ME adaptation has presumably been too short to result in a detectable genetic differentiation. In addition, shuttle breeding between two environmentally contrasting sites in Mexico (Cd. Obregon and Tuluca) during the selfing generations and intermating of germplasm adapted to different ME may have leveled the genetic differentiation between MEs (Rajaram and Van Ginkel, 2001).

The uniform level of relatedness between genotypes targeted to the five MEs (Table 4) and the principal coordinate analysis (Fig. 1) suggest the presence of a single core germplasm providing genes or gene combinations conferring fitness to several MEs. This is not surprising, because some MEs differ only in few of the classification criteria. For example, the high rainfall environment ME3 can be characterized as a specific subenvironment of ME2 and, consequently, selection to both environments depends on the same major abiotic and biotic stresses (Table 2).

Although many genetic diversity studies confirmed the use of SSRs for germplasm identification, it is an open question whether they are able to reveal functional diversity. Adaptation to different ME might be attributable to only a small number of genes regulating the underlying physiological processes, which were not reflected by the applied SSRs (Nevo, 2001). An exception was the cluster of acid soil resistant lines in Fig. 1, but this could also be explained by the fact that 8 of the 14 lines had the highly acid soil resistant line Tinamou as a parent in their pedigree.

High levels of genetic diversity were found within the germplasm targeted to each ME. To warrant diversity in every breeding cycle, on average about 25% of the crossing block entries are replaced with outstanding new introductions (Van Ginkel et al., 2002). In our study too, at least 20 wheat lines were included with ancestors from nonconventional sources such as Chinese, durum, or synthetic hexaploid wheats.

The high genetic diversity observed within each ME is desired in the CIMMYT wheat program for two reasons. First, the wheat producing areas combined in each ME are fairly diverse and often transcontinental. Diseases and especially races of fungal pests may vary considerably between regions as well as quality demands due to different wheat processing techniques and utilization of the end product. Furthermore, environmental conditions in the target areas are fluctuating based on highly variable seasons (Trethowan et al., 2002). Newly developed CIMMYT lines should reach the cooperating wheat research programs in the respective MEs with still sufficient inherent genetic variation remaining among them for many traits. This enables cooperators to reselect and release cultivars adapted to their own local needs. Therefore, a sufficient number of diverse CIMMYT breeding lines for each ME reduces the risk of genetic vulnerability.

Second, breeding conditions at the test sites in Mexico and the targeted ME are not always closely associated. Trethowan et al. (2001) examined the relationship among various locations using yield data from CIMMYT's Semi-Arid Wheat Yield Trials corresponding to ME4. They found that the yield performance under residual moisture stress at the test location in Cd. Obregon (Mexico) was a good predictor of yield performance at other locations experiencing equivalent late drought stress, but this was not true when drought patterns were different, e.g., when early drought stress occurred.

The common genetic base of our germplasm targeted to different ME did not reflect CIMMYT's efforts to breed for a large number of environments. We speculate that assembling more diversified germplasm pools for different MEs and yield evaluation at ME-specific key sites could achieve higher genetic differentiation. However, increased genetic diversity does not necessarily lead to higher productivity or adaptation. While SSRs are a powerful tool for genotype identification, their usefulness for revealing genotype differentiation regarding specific traits such as adaptation to certain MEs could not be proven in the present study and warrants further research. In particular, the augmented use of EST-SSRs from genes with known functions should be very powerful for this purpose and, thus, could give further directions in the management of wheat breeding programs aiming to optimally serve their clients.

ACKNOWLEDGMENTS

This paper is dedicated to Prof. Dr. Dr. h.c. F.W. Schnell on the occasion of his 90th anniversary. The authors are indebted to Richard Trethowan, CIMMYT, who provided the germplasm of ME4; to Emiliano Villordo and Leticia Diaz for their technical assistance, and to the Vater & Sohn Eiselenstiftung, Ulm, and the Federal Ministry for Economic Cooperation and Development (BMZ) for their financial support and collaboration under the project "Efficient management of genetic diversity in wheat: DNA marker for use in wheat breeding programs and gene banks."

REFERENCES

Adams, M.D., M. Dubnick, A.R. Kerlavage, R. Moreno, J.M. Kelley, T.R. Utterback, J.W. Nagle, C. Fields, and J.C. Venter. 1992. Sequence identification of 2375 human brain genes. Nature 355: 632–634.

Chakraborty, R., M. Kimmel, D.N. Stivers, L.J. Davison, and R. Deka. 1997. Relative mutation at di-, tri-, and tetranucleotide microsatellite loci. Proc. Natl. Acad. Sci. (USA) 94:1041–1046.

Cho, Y.G., T. Ishii, S. Temnykh, X. Chen, L. Lipovich, S.R. McCouch, W.D. Park, N. Ayres, and S. Cartinhour. 2000. Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). Theor. Appl. Genet. 100:713–722.

Christiansen, M.J., S.B. Anderson, and R. Ortiz. 2002. Diversity changes in an intensively bread wheat germplasm during the 20th century. Mol. Breed. 9:1–11.

Cox, T.S., Y.T. Hkiang, M.B. Gorman, and D.M. Rogers. 1985. Relationships between coefficient of parentage and genetic indices in the soybean. Crop Sci. 25:529–532.

Cowen, N.M., and K.J. Frey. 1987. Relationships between genealogical

distance and breeding behavior in oat (*Avena sativa* L.). Euphytica 36:413–424.
Devos, K.M., G.J. Bryan, A.J. Collins, P. Stephenson, and M.D. Gale. 1995. Application of two microsatellite sequences in wheat storage proteins as molecular markers. Theor. Appl. Genet. 90:247–252.
Donini, P., P. Stephenson, G.J. Bryan, and R.M.D. Koebner. 2000. The potential of microsatellites for high throughput genetic diversity assessment in wheat and barley. Genet. Resour. Crop Evol. 45: 415–421.
Eujayl, I., M.E. Sorrells, M. Braun, P. Wolters, and W. Powell. 2002. Isolation of EST-derived microsatellite markers for genotyping the A and B genomes of wheat. Theor. Appl. Genet. 104:399–407.
Frisch, M., M. Bohn, and A.E. Melchinger. 2000. Plabsim: Software for simulation of marker-assisted backcrossing. J. Hered. 91:86–87.
Gilbert, J.E., R.V. Lewis, M.J. Wilkinson, and P.D.S. Caligari. 1999. Developing an appropriate strategy to assess genetic variability in plant germplasm collections. Theor. Appl. Genet. 98:1125–1131.
Gower, J.C. 1966. Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika 53:325–388.
Graner, A., W.F. Ludwig, and A.E. Melchinger. 1994. Relationships among European barley germplasm: II. Comparison of RFLP and pedigree data. Crop Sci. 34:1199–1205.
Hoisington, D., M. Khairallah, and D. Gonzalez-de-Leon. 1994. Laboratory Protocols: CIMMYT Applied Molecular Genetics Laboratory. 2nd ed. (Available online with updates at (http://www.cimmyt.cgiar.org/ABC/Protocols/manualABC.html; verified 22 September 2003). CIMMYT, Mexico, D.F.
Huang, X.Q., A. Börner, M.S. Röder, and M.W. Ganal. 2002. Assessing genetic diversity of wheat (*Triticum aestivum* L.) germplasm using microsatellite markers. Theor. Appl. Genet. 105:699–707.
Ihaka, R., and R. Gentleman. 1996. R: A Language for data analysis and graphics. J. Comput. Graph. Stat. 5:299–314.
Innan, H., R. Terauchi, and T. Miyashita. 1997. Microsatellite polymorphism in natural populations of the wild plant *Arabidopsis thaliana.* Genetics 146:1441–1452.
Kota, R., R.K. Varshney, T. Thiel, K.J. Dehmer, and K.J. Graner. 2001. Generation and comparison of EST-derived SSRs and SNPs in barley (*Hordeum vulgare* L.). Hereditas 135:145–151.
Lantican, M.A., P.L. Pingali, and S. Rajaram. 2002. *In* J. Ekboir (ed.) World wheat overview and outlook 2000–2001: Developing no-till packages for small scale farmers. CIMMYT, Mexico, D.F.
Li, Y.-C., T. Fahima, J.H. Peng, M.S. Röder, V.M. Kirzhner, A. Beiles, A.B. Korol, and E. Nevo. 2000. Edaphitic microsatellite DNA divergence in wild emmer wheat, *Triticum dicoccoides*, at a microsite: Tabigha, Israel. Theor. Appl. Genet. 101:1029–1038.
Li, Y.-C., A.B. Korol, T. Fahima, A. Beiles, and E. Nevo. 2002. Mircosatellites: Genomic distribution, putative functions and mutational mechanisms: A review. Mol. Ecol. 11:2453–2465.
MacQueen, J. 1967. Some methods for classification and analysis of multivariate observations. p. 281–297. *In* Proc. of the Berkeley Symposium on Mathematical Statistics and Probability, 5th. Vol. 1. University of California Press, Berkeley, CA.
Mogg, R., J. Batley, S. Hanley, D. Edwards, H. O'Sullivan, and K.J. Edwards. 2002. Characterization of the flanking regions of *Zea mays* microsatellites reveals a large number of useful sequence polymorphisms. Theor. Appl. Genet. 105:532–543.
Morgante, M., M. Hanafey, and W. Powell. 2002. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. Nat. Genet. 30:194–200.
Nei, M. 1987. Molecular evolutionary genetics. Columbia University press, New York.
Nevo, E. 2001. Evolution of genome-phenome diversity under environmental stresses. Proc. Natl. Acad. Sci. (USA) 98:6233–6240.
Parker, G.D., P.N. Fox, P. Langridge, K. Chalmers, B. Whan, and P.F. Ganter. 2002. Genetic diversity within Australian wheat breeding programs based on molecular and pedigree data. Euphytica 124: 293–306.
Payne, T.S., B. Skovmand, C.G. Lopez, E. Brandon, and A. McNab. (ed.) 2002. The international wheat information system (IWIS™), Version 4. 2001. On compact disk. CIMMYT, Mexico, D.F.
Plaschke, J., M.W. Ganal, and M.S. Röder. 1995. Detection of genetic diversity in closely related brad wheat using microsatellite markers. Theor. Appl. Genet. 91:1001–1007.
Purdy, L.H., W.Q. Loegering, C.F. Konzak, C.J. Perterson, and R.E. Allan. 1968. A proposed standard method for illustrating pedigrees of small grain varieties. Crop Sci. 8:405–406.
Rajaram, S., and M. Van Ginkel. 2001. Mexico, 50 years of International Wheat Breeding. p. 579–608. *In* A.P. Bonjean and W.J. Angus (ed.) The world wheat book, a history of wheat breeding. Lavoisier Publishing, Paris.
Rajaram, S., M. Van Ginkel, and R.A. Fischer. 1994. CIMMYT's wheat breeding mega-envionments (ME). p. 1101–1106. *In* Z.S. Li and Z.Y. Xin (ed.) Proc. of the International Wheat Genetics Syposium, 8th, Beijing. 19–24 July. Agricultural Scientech Press, Beijing.
Röder, M.S., V. Korzun, K. Wendehake, J. Plaschke, M.-H. Tixier, P. Leroy, and M.W. Ganal. 1998. A microsatellite map of wheat. Genetics 149:2007–2023.
Rogers, J.S. 1972. Measures of similarities and genetic distances. p.145–153. *In* M.R. Wheeler (ed.) Studies in genetics VII. University of Texas, Austin, TX.
Saghai-Maroof, M.A., K. Soliman, R.A. Jorgensen, and R.W. Allard. 1984. Ribosomal DNA spacer length polymorphism in barley: Mendelian inheritance, chromosomal location and population dynamics. Proc. Natl. Acad. Sci. (USA) 81:8014–8018.
Schneider, S., D. Roessli, and L. Excoffier. 2000. Arlequin version 2.000: A software for population genetics data analysis. Genetics and biometry laboratory, University of Geneva, Switzerland (see also http://lgb.unige.ch/arlequin/; verified 22 September 2003).
Scott, K.D., P. Eggler, G. Seaton, M. Rossetto, E.M. Ablett, L.S. Lee, and R.J. Henry. 2000. Analysis of SSRs derived from grape ESTs. Theor. Appl. Genet. 100:723–726.
Song, Q.J., E.W. Fickus, and P.B. Cregan. 2002. Characterization of trinucleotide SSR motifs wheat. Theor. Appl. Genet. 104:286–293.
Stachel, M., T. Lelley, H. Grausgruber, and J. Vollmann. 2000. Application of microsatellites in wheat (*Triticum aestivum* L.) for studying genetic differentiation caused by selection for adaptation and use. Theor. Appl. Genet. 100:242–248.
St. Martin, S.K. 1982. Effective population size for the soybean improvement program in maturity groups 00 to IV. Crop Sci. 22: 151–152.
Thornsberry, J.M., M.M. Goodman, J. Kresovich, S. Nielsen, and D. Buckler. 2001. Dwarf8 polymorphisms associate with variation in flowering time. Nat. Genet. 28:286–289.
Trethowan, R.M., J. Crossa, M. van Ginkel, and S. Rajaram. 2001. Relationships among bread wheat international yield testing location in dry areas. Crop Sci. 41:1461–1469.
Trethowan, R.M., M. van Ginkel, and S. Rajaram. 2002. Progress in breeding wheat for yield and adaptation in global drought affected environments. Crop Sci. 42:1441–1446.
Van Ginkel, M., R.M. Trethowan, K. Ammar, J. Wang, and M. Lillemo. 2002. A guide to bread wheat breeding at CIMMYT. Wheat special report No. 5 (Rev. ed.). CIMMYT, Mexico, D.F.
Weir, B.S. 1996. Genetic data analysis II. Sinauer Associates, Sunderland, MA.
Wright, S. 1978. Evolution and genetics of populations. University Chicago Press, Chicago.