

UC Davis

UC Davis Previously Published Works

Title

Evaluation of genetic diversity and genome-wide linkage disequilibrium among U.S. wheat (*Triticum aestivum* L.) germplasm representing different market classes

Permalink

<https://escholarship.org/uc/item/96q060xs>

Journal

Crop Science, 47(3)

ISSN

0011-183X

Authors

Chao, S

Zhang, W

Dubcovsky, J

et al.

Publication Date

2007-05-01

DOI

10.2135/cropsci2006.06.0434

Peer reviewed

RESEARCH

Evaluation of Genetic Diversity and Genome-wide Linkage Disequilibrium among U.S. Wheat (*Triticum aestivum* L.) Germplasm Representing Different Market Classes

Shiaoman Chao,* Wenjun Zhang, Jorge Dubcovsky, and Mark Sorrells

ABSTRACT

Genetic diversity and genome-wide linkage disequilibrium (LD) were investigated among 43 U.S. wheat (*Triticum aestivum* L.) elite cultivars and breeding lines representing seven U.S. wheat market classes using 242 wheat genomic simple sequence repeat (SSR) markers distributed throughout the wheat genome. Genetic diversity among these lines was examined using genetic distance-based and model-based clustering methods, and analysis of molecular variance. Four populations were identified from the model-based analysis, which partitioned each of the spring and winter populations into two subpopulations, corresponding largely to major geographic regions of wheat production in the United States. This suggests that the genetic diversity existing among these U.S. wheat germplasm was influenced more by regional adaptation than by market class, and that the individuals clustered in the same model-based population likely shared related ancestral lines in their breeding history. For this germplasm collection, genome-wide LD estimates were generally less than 1 cM for the genetically linked loci pairs. This may result from the population stratification and small sample size that reduced statistical power. Most of the LD regions observed were between loci less than 10 cM apart. However, the distribution of LD was not uniform based on linkage distance and was independent of marker density. Consequently, LD is likely to vary widely among wheat populations.

S. Chao, USDA-ARS Biosciences Research Lab., 1605 Albrecht Blvd., Fargo, ND 58105; W. Zhang and J. Dubcovsky, Dep. of Plant Sciences, Univ. of California, Davis, CA 95616; M. Sorrells, Dep. of Plant Breeding and Genetics, Cornell Univ., 240 Emerson Hall, Ithaca, NY 14853. Received 28 June 2006. *Corresponding author (chaos@fargo.ars.usda.gov).

Abbreviations: AMOVA, analysis of molecular variance; CS, Chinese Spring; EST, expressed sequence tag; F_{st} , Wright's fixation index; HRS, hard red spring; HRW, hard red winter; HWS, hard white spring; HWW, hard white winter; LD, linkage disequilibrium; PCR, polymerase chain reaction; PIC, polymorphism information content; QTL, quantitative trait loci; SSR, simple sequence repeat; SRW, soft red winter; SWS, soft white spring; SWW, soft white winter; UPGMA, unweighted pair-group method with arithmetic mean.

WHEAT (*Triticum aestivum* L.) is the principal cereal grain grown in the United States for both domestic consumption and export. Improved cultivars and cultural practices have resulted in a steady increase in the average wheat yield during the past 30 yr, from 2000 kg ha⁻¹ in 1970 to over 2700 kg ha⁻¹ in 2004 (Wheat Yearbook, USDA, <http://usda.mannlib.cornell.edu/MannUsda/viewDocumentInfo.do?documentID=1295>). To maintain a steady rate of wheat improvement, exploring genetic diversity at the molecular levels by means of molecular genetics technologies and integrating such information with conventional breeding methods will be critical.

Genetic diversity plays a vital role in developing improved cultivars. Previously, the level and patterns of genetic diversity among U.S. wheat cultivars have been investigated among soft and hard red winter wheats (Cox et al., 1986), hard red spring wheat (Chen et al., 1994), soft winter wheat from the eastern United States (Kim and

Published in Crop Sci. 47:1018–1030 (2007).

doi: 10.2135/cropsci2006.06.0434

© Crop Science Society of America

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

All rights reserved. No part of this periodical may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Permission for printing and for reprinting the material contained herein has been obtained by the publisher.

Ward, 1997), a sample of spring and winter wheat cultivars from the Pacific Northwest region (Barrett et al., 1998), and hard red winter wheat cultivars from the Northern Great Plains (Fufa et al., 2005). The measures applied in these studies and others can be summarized into four categories: pedigree data, phenotypic traits, storage proteins, and DNA markers. In general, the diversity estimates based on different methods were positively correlated (Parker et al., 2002; Fufa et al., 2005). Among DNA markers used, restriction fragment length polymorphisms (Kim and Ward, 1997; Paull et al., 1998), sequence tagged site (Chen et al., 1994), amplified fragment length polymorphisms (Barrett et al., 1998; Manifesto et al., 2001), and simple sequence repeats (SSRs) (Huang et al., 2002; Fufa et al., 2005) have all been found suitable for diversity studies.

Knowledge of the level of genetic diversity and familiarity with genetic and historical relationships among elite germplasm can further the exploitation of genetic variation in wheat. One genetic analysis that can be used to develop this knowledge is linkage disequilibrium (LD), or nonrandom association of alleles at adjacent loci within a population, which is the basis for association mapping strategies. In association mapping, phenotypic diversity is surveyed among a population of individuals with varying degrees of relationship, followed by identification of marker polymorphisms that correlate with phenotypic variation (Buckler and Thornsberry, 2002). Dissecting complex agronomic traits using LD-based mapping has recently been reported in crop plants such as rice (*Oryza sativa* L.) (Garris et al., 2003), maize (*Zea mays* L.) (Buckler et al., 2006), potato (*Solanum tuberosum* L.) (Simko et al., 2004), barley (*Hordeum vulgare* L.) (Kraakman et al., 2004), and common wheat (Brescghello and Sorrells, 2006). Association mapping depends on the patterns of LD, how far the usable levels of disequilibrium extend in the genome, and how much LD varies from one chromosome region or from one population to another. Factors such as the mating system, the recombination rate, population structure, population history, genetic drift, directional selection, and gene fixation at different rates on different chromosome regions can all affect the patterns of LD (Gaut and Long, 2003). To date, the extent of LD patterns in plants have been examined in *Arabidopsis* (Nordborg et al., 2002), maize (Remington et al., 2001), barley (Kraakman et al., 2004), rice (Garris et al., 2003), sorghum [*Sorghum bicolor* (L.) Moench] (Hamblin et al., 2004), durum wheat (*T. turgidum* L. var. *durum*) (Maccaferri et al., 2005), and loblolly pine (*Pinus taeda* L.) (Brown et al., 2004). These results have indicated that, in general, LD decay with distance occurs at a much slower rate in self-pollinated plants, such as *Arabidopsis*, rice, barley, durum wheat, and sorghum, than in outcrossing species, including maize and loblolly pine. In barley, however, it was further demonstrated that the extent of LD could vary dramatically between populations with different evolutionary histories

(Caldwell et al., 2006). It was found that LD in a population of a wild barley progenitor rapidly decayed at a similar rate as a population of maize inbred lines. In common wheat, the LD patterns have been assessed for chromosomes 2D and part of 5A (Brescghello and Sorrells, 2006), and the distribution of LD was not uniform on these chromosomes.

Currently, data are available online for over 1700 genomic and expressed sequence tag (EST)-derived SSR markers, which have been characterized and genetically and/or physically mapped across all chromosomes for at least 27 wheat accessions (GrainGenes database, <http://wheat.pw.usda.gov>; verified 4 Mar. 2007). Their high information content (Plaschke et al., 1995), uniform distribution throughout the wheat genomes (Somers et al., 2004), and ease of use have made SSR-based markers a valuable tool. In addition, automated protocols for polymerase chain reaction (PCR)-based techniques and data collection have been developed for SSRs (Diwan and Cregan, 1997), which allow high throughput genotyping to be readily applied in wheat. In this report, we analyze the genetic diversity among 43 U.S. wheat cultivars and breeding lines representing seven market classes using a set of 242 SSR markers each mapped to a single chromosome location and distributed over all 21 chromosomes. In addition, the distribution and extent of LD over the portion of the wheat genome covered were estimated to determine the implications of applying association mapping in wheat.

MATERIALS AND METHODS

Plant Materials

Forty-three wheat cultivars, advanced breeding lines, and germplasm representing seven U.S. market classes were used for this study (Table 1). These lines were selected from 18 wheat breeding programs across the country and were used as parents to generate mapping populations segregating for a diverse array of agronomic traits as part of a collaborative project among U.S. public wheat breeding programs (<http://maswheat.ucdavis.edu/>, last update 26 Feb. 2007; verified 4 Mar. 2007). Among the 15 spring wheat genotypes representing three market classes, nine were hard red (HRS), two were hard white (HWS), and three were soft white (SWS). The germplasm IDO556 representing soft white club was grouped with SWS for analysis. Twenty-eight winter wheat genotypes representing four market classes included six hard red (HRW), five hard white (HWW), 10 soft red (SRW), six soft white (SWW), and a facultative SWS line, Q36. Also included in the analysis was cultivar Chinese Spring (CS) as a genotyping control for comparison with SSR marker data from previous studies using different assay methods. The source of CS was the Wheat Genetic Stock Collection located at USDA-ARS, University of Missouri, Columbia.

SSR Marker Genotyping

Genomic DNA was extracted from individuals using seeds collected from one plant at the University of California, Davis. Nuclear DNAs were extracted from precipitated nuclei using the large-scale extraction method previously described by Dvorak

Table 1. Description of 43 wheat accessions included in this study.

Germplasm	Market class [†]	Origin	Year of release	Pedigree [‡]
Thatcher	HRS	Minnesota	1934	Marquis/lumillo//Marquis/Kanred
McNeal	HRS	Montana	1994	PI 125000/Centana//PK176/Fronteira/3/Glenman
ND735	HRS	North Dakota		ND 2907/3/Grandin*3//Ramsey/ND 622/4/ND 2809
Reeder/Bw-277 "R"	HRS	North Dakota		Reeder = IAS20*4/HH567.71//Stoa/3/ND674
Reeder/Bw-277 "S"	HRS	North Dakota		
Steele-ND	HRS	North Dakota	2004	Parshall/5/Grandin/3/IAS20 *4/H567.71//Amidon/4/Grandin *2/Glupro
Jupateco 73S	HRS	CIMMYT, Mexico	1973	II12300//LERMA ROJO 64A/8156/3/NORTE-O M 67
Weebill 1	HRS	CIMMYT, Mexico	1999	BABAX/AMADINA//BABAX
PI610750	HRS	CIMMYT, Mexico		Crocl/Ae. <i>tauschii</i> (205)//Kauz
UC1110	HWS	California		CKR'S*x[YD'S*x[BBxCHA]]
Grandin*5/ND614-A	HWS	North Dakota		Grandin = Len//Butte *2/ND507/3/ND593
IDO556	SWS (Club)	Idaho		Tincurrin/Centennial
Louise	SWS	Washington	2004	Wakanz/Wawawai
Penawawa	SWS	Washington	1985	Potam 70/Fielder
Zak	SWS	Washington	2000	Pavon 'S'/5/PI167822/C113438 113-6//Idaed/Marfed 68-5/4/Lemhi 66/3/Yak-tana 54A *4/Norin 10/Brevor/6/Walladay/7/PI506355/8/Treasure
IDO444	HRW	Idaho	1994	Utah 216c-12- 10/Cheyenne/5/PI 476212/4/Burt/3/Rio/Rex//Nebred//6//Utah 216c- 12-10/Cheyenne/5/PI 476212/4/Burt/3/Rio/Rex//Nebred
Jagger	HRW	Kansas	1994	KS82W418/Stephens
Harry	HRW	Nebraska	2002	NE90614/NE87612 (NE90614 = Brule/4/Parker *4/Agent//Beloterkovskaia 198/Lancer/3/Newton/Brule, NE87612 = Newton//Warrior *5/Agent/3/Agate sib)
Wesley	HRW	Nebraska	1998	KS831936-3/NE86501 = Sumner sib (Plainsman V/Odesskaya 51)//Colt/Cody
2174	HRW	Oklahoma	1998	IL71-5662/PL145//216
TAM 105	HRW	Texas	1979	'short wheat' /Sturdy composite bulk selection
CO940610	HWW	Kansas		KS87H22/MW09 (KS87H22 = H15A13333/5*Larned//Eagle/Sage/3/TAM 105, MW09 = Clark's Cream/5*KS75216(Newton Sib))
Heyne	HWW	Kansas	1998	KS82W422/SWM754308/KS831182/KS82W422
KS01HW163-4	HWW	Kansas		Trego (= KS87H325/Rio Blanco)/Betty 'S'
Platte	HWW	AgriPro		N84-1104/Abilene (OK11252/W76-1226)
Rio Blanco	HWW	Kansas	1989	OK11252A/W76-1226
Jaypee	SRW	Arkansas	1995	Arthur 6/AR 39-3 (AR 39-3 = Doublecrop//Forlani/Garibaldi)
AGS 2000	SRW	Georgia	1999	Pioneer 2555/PF84301//Florida 302
P91193	SRW	Indiana		Benhur//Arthur/Knox62/3/Arthur/NY5715AB/4/Hart/Beau//Arthur/Abe/5/Auburn/Coker 8427/3/OH256/Scotty//jClark.
P92201	SRW	Indiana		Tyler//Caldwell*2/Pioneer S76/3/Clark/CI5549/4/Roazon/Caldwell/5/Glory
Foster	SRW	Kentucky	1996	Coker 65-20/Arthur/4/Chul * 8CC//VA 68-22-7/Abe/3/VA 72-54- 14/Tyler//Suwon 92/Arthur//Arthur/VA 70-52-2
SS 550	SRW	Virginia	2000	Coker 9803/Freedom
McCormick	SRW	Virginia	2002	VA92-51-39/AL870365
USG 3209	SRW	Virginia	1999	Saluda/4/Massey*2/3/Massey*3/Balkan//Saluda
25R26	SRW	Pioneer Hi-Bred		2548 sib/Pioneer line W9057C//Pioneer line W9018A/2555 sib
26R46	SRW	Pioneer Hi-Bred		FL7927-G14//2555*3/Coker 80-28
Q36 (OR9900553)	SWW (SWS)	Oregon		Armindia/3/VPM/MOS951//2*Hill/5/ID#870337
Stephens	SWW	Oregon	1977	Nord Desprez/Pullman Sel. 101
Eltan	SWW	Washington	1990	Luke//BR-70443-4 (PI 167822)/Sel. 101
Finch	SWW	Washington	2001	Dusty//WA7164 (VPM1/Moisson 951//Yamhill/Hyslop)/Dusty
Caledonia	SWW	New York	1998	a selection out of Geneva (Genesee/CI 12658//Burt/3/HeinesVII/4/Genesee)
Cayuga	SWW	New York	1993	reselection of NY262-37-10W (Geneva/Clark's Cream//Geneva)
NY18/Clark's Cream 40-1	SWW	New York		Clark's Cream = composite bulk of KanKing/Golden 50 or Kanking/Golden Chief selections

[†]HRS, hard red spring; HWS, hard white spring; SWS, soft white spring; HRW, hard red winter; HWW, hard white winter; SRW, soft red winter; SWW, soft white winter.

[‡]The pedigree data were obtained from the breeders and the GRIN database (<http://www.ars-grin.gov/npgs/index.html>).

et al. (1988). Altogether 1619 genomic and EST-derived SSR primer pairs were assayed in this study (Table 2). The marker identifiers and sources of the primer sequences for genomic SSR markers were gwm (Röder et al., 1998); gdm (Pestsova et al., 2000); barc (Song et al., 2002, 2005); cfa and cfd (Sourdille et al., 2001; Guyomarc'h et al., 2002); wmc (Gupta et al., 2002; <http://wheat.pw.usda.gov/ggpages/SSR/WMC/>; Daryl Somers, pers. comm.); and gpw (Sourdille et al., 2004). The EST-derived SSR markers, cnl and ksm, were described in Yu et al. (2004) and <http://wheat.pw.usda.gov/ITMI/EST-SSR/>. Polymerase chain reaction amplifications followed the M13-tailed primer PCR method (Schuelke, 2000) with the M13 oligonucleotide labeled with one of the four fluorescent dyes, 6-FAM, VIC, NED, and PET, added in the reaction mix. After multiplexing PCR products labeled with four different fluorescent dyes, the electrophoresis was performed on the Applied Biosystems (Foster City, CA) 3130xl Genetic Analyzer. GeneMapper software v3.7 (Applied Biosystems) was used for fragment analysis and allele calling. The detailed PCR amplification conditions and genotyping process were as described by Somers et al. (2004).

Data Analysis

Among the 1619 SSRs assayed in this study, a subset of 242 SSR markers that were previously mapped to a unique and single chromosome location was selected for data analysis. Most (178/242) of the marker positions on each chromosome were based on the Ta-SSR-2004 consensus map (Somers et al., 2004). A CMap Matrix tool (<http://www.genica.net.au/index.php?title=CMap>) was then used to position the remaining 64 markers by aligning and comparing the Ta-SSR-2004 map with either the wheat composite 2004 map or the Ta-Synthetic × Opata-SSR map. Markers with known physical bin map locations were also used to assist in positioning markers where their genetic map locations were not previously determined. Both the genetic and physical maps were accessed from the GrainGenes database. The fragment sizes for CS generated from the primers used in this study were compared with previously published sizes for CS, if available. The complete information for the 242 markers selected and their genetic and physical bin locations are available in the supplemental Table S1.

Gene diversity, defined as the probability that two randomly chosen alleles from the population are different (Weir, 1996); polymorphism information content (PIC) values, defined by Botstein et al. (1980); and total number of alleles at each SSR locus were calculated using the PowerMarker software (Liu and Muse, 2005). Gene diversity and number of alleles were determined for the entire set of samples as well as for each market class separately. Number of alleles unique to each market class was calculated using the CONVERT software v1.31 (Glaubitz, 2004). The methods implemented in the PowerMarker software were also applied to calculate the genetic distance among 43 samples based on Rogers' distance, the scaled Euclidean distance (Rogers, 1972), and to reconstruct a dendrogram using UPGMA. The Rogers' similarity coefficient factored in allele frequencies calculated from the codominant marker data, assuming no knowledge of evolutionary pressure on divergence of the lines under consideration, and is suitable to evaluate genetic similarity among germplasm derived from breeding programs (Reif et al., 2005). To test the reliability of the relationships among samples suggested by the clusters, a bootstrap analysis with 1000 replications was performed using the

Table 2. Source and total number of simple sequence repeat (SSR) markers assayed, level of marker polymorphism, and number of SSR markers selected for this study.

SSR source [†]	Total no. used	No. of polymorphic markers (%)	No. of markers selected
BARC	592	239 (40)	73
CFA	54	29 (54)	13
CFD	127	72 (57)	29
GDM	64	24 (38)	7
GWM	239	151 (63)	62
WMC	330	177 (54)	58
GPW	73	19 (26)	0
EST-SSR	140	31 (22)	0
Total	1619	742 (46)	242

[†]BARC: USDA-ARS Beltsville Agricultural Research Center; GDM, Gatersleben D-genome Microsatellite; GWM, Gatersleben Wheat Microsatellite; WMC; Wheat Microsatellite Consortium. CFA, CFD, GPW are lab designators for P. Sourdille at Clermont-Ferrand, France.

PowerMarker software. The consensus UPGMA tree with bootstrap values was reconstructed by the consensus program of Phylip v3.63 (J. Felsenstein, University of Washington, Seattle, WA) and displayed using the TreeView software (Page, 1996).

The genetic variation among and within populations of wheat lines was tested using analysis of molecular variance (AMOVA) implemented in Arlequin v3.01 (Excoffier et al., 2005). This approach considers the number of genotype differences and estimates the variance among and within the defined genetic structure. The variance within populations was expressed as Wright's fixation index (F_{st}) and the statistical significance of F_{st} was evaluated by permuting genotypes among or within all populations 1000 times. Pairwise F_{st} comparisons based on genotype differences were used to evaluate genetic diversity among populations.

Population structure was investigated using a Bayesian clustering approach to infer the number of clusters (populations) with the software *structure* v.2 (Pritchard et al., 2000). The algorithm attempts to identify genetically distinct subpopulations based on the patterns of SSR allele frequencies. Sixty-seven loosely linked SSR markers (>40 cM) with three to four markers distributed on 21 chromosomes were selected for *structure* analysis to minimize detecting background LD caused by tightly linked markers (Falush et al., 2003). No prior information was used to define the clusters, and the number of subpopulations (K) was set from two to seven. Runs with K = 3 to 5 were repeated at least three times. For each run, the burn-in period and the simulation run length were both set at 100000 with admixture model and correlated allele frequency (Falush et al., 2003). This method estimated the proportion of the genomes of each individual derived from the different clusters and assigned individuals to subpopulations based on membership probability. We used the run that assigned all the lines to a single cluster at a probability >0.50.

To evaluate LD, all accessions except one sister line, Reeder/BW-277S, were analyzed. Rare alleles with an allele frequency <5% along with null alleles and residual heterozygosity were treated as missing data. This resulted in eliminating data for 45 SSR markers out of the 242 SSR markers selected with >20% missing data. The final data set used to perform the LD analysis consisted of 197 SSR loci and 42 lines. The LD parameter, r^2 , and significance of each pair of SSR loci were estimated using the program TASSEL (<http://>

Table 3. Simple sequence repeat (SSR) marker distribution, total number of alleles detected, and polymorphism information content (PIC) values by chromosome.

Chromosome	No. of markers	No. of alleles	Mean allele number (range)	Mean PIC value (range)
1A	12	95	7.92 (2–24)	0.64 (0.18–0.93)
2A	13	98	7.54 (4–15)	0.68 (0.36–0.89)
3A	12	90	7.50 (5–12)	0.66 (0.30–0.84)
4A	11	82	7.45 (2–17)	0.56 (0.15–0.92)
5A	13	82	6.31 (3–14)	0.60 (0.29–0.87)
6A	8	47	5.88 (2–9)	0.63 (0.29–0.86)
7A	12	80	6.67 (2–22)	0.52 (0.13–0.93)
1B	8	65	8.13 (3–14)	0.66 (0.38–0.86)
2B	11	76	6.91 (3–19)	0.58 (0.26–0.86)
3B	12	74	6.17 (3–10)	0.63 (0.31–0.80)
4B	11	81	7.36 (2–16)	0.63 (0.12–0.88)
5B	13	98	7.54 (2–14)	0.62 (0.26–0.83)
6B	10	79	7.90 (2–12)	0.69 (0.26–0.86)
7B	12	96	8.00 (3–17)	0.66 (0.27–0.91)
1D	11	60	5.45 (3–12)	0.54 (0.26–0.82)
2D	10	104	10.40 (4–18)	0.73 (0.20–0.91)
3D	12	102	8.50 (5–14)	0.72 (0.46–0.88)
4D	11	73	6.64 (3–12)	0.64 (0.29–0.86)
5D	15	120	8.00 (3–20)	0.66 (0.34–0.90)
6D	13	62	4.77 (2–9)	0.50 (0.12–0.82)
7D	12	79	6.58 (3–14)	0.65 (0.43–0.87)
Homoeologous group				
1	31	220	7.10 (2–24)	0.61 (0.18–0.93)
2	34	278	8.18 (3–19)	0.66 (0.20–0.91)
3	36	266	7.39 (3–14)	0.67 (0.30–0.88)
4	33	236	7.15 (2–17)	0.61 (0.12–0.88)
5	41	300	7.32 (2–20)	0.63 (0.26–0.90)
6	31	188	6.06 (2–12)	0.60 (0.12–0.86)
7	36	255	7.08 (2–22)	0.61 (0.13–0.93)
Genome				
A	81	574	7.09 (2–24)	0.61 (0.13–0.93)
B	77	569	7.39 (2–19)	0.64 (0.12–0.91)
D	84	600	7.14 (2–20)	0.63 (0.12–0.91)
Total	242	1743		
Overall mean			7.20 (2–24)	0.62 (0.12–0.93)

www.maizegenetics.net; verified 10 Mar. 2007). The comparison-wise significance was computed using 1000 permutations and the proportion of permuted gamete distributions greater than observed gamete distribution served as an empirical *P* value (Weir, 1996). To obtain a critical value of r^2 for significant linkage, unlinked estimates of r^2 were square root transformed and the parametric 95th percentile of that distribution was taken as a population-specific critical value of r^2 (Brescghello and Sorrells, 2006).

RESULTS

SSR Diversity

In this study, a total of 1619 SSR markers developed in wheat were used to genotype 43 accessions and 742 (45.8%)

of them detected at least one polymorphism (Table 2). In contrast to the EST-derived and *gpw* SSR markers, most of the genomic SSR markers showed high levels of polymorphism (Table 2). To evaluate genetic diversity among these 43 accessions, a subset of 242 single-locus genomic SSR markers was selected. They are widely distributed on both short and long arms of 21 chromosomes with an average of 10 per chromosome. These primers were also selected because they produced a high rate of amplification with <10% missing data including nulls. Null alleles were considered as missing data because of the difficulty in differentiating them from PCR failure.

Although genomic DNA was extracted from a single plant, a low level of heterozygosity with an overall mean of 0.9% was observed. The residual heterozygosity was kept in the data set used for genetic diversity analysis because of its minimal impact on the outcome of the analysis (data not shown). The number of alleles detected by each marker varied greatly ranging from 2 to 24 with a total of 1743 detected and a mean allele number of 7.2 per marker (Table 3). Polymorphism information content values ranged from 0.12 to 0.93 with a mean of 0.62 for all markers. Overall, markers in the B genome gave slightly higher PIC values compared with markers in A or D genomes (Table 3). Number of alleles, gene diversity and PIC values calculated for each SSR marker are shown in Table S1.

To explore genetic diversity among genotypes within different market classes, the estimates of gene diversity and mean number of alleles were also calculated for each market class (Table 4). Both HRS and SRW lines were very diverse with mean allele numbers of 3.74 and 3.67, respectively, per locus, and gene diversity of 0.56 and 0.57, respectively. The highest number of alleles unique to the market class was found in the SRW class (19.2%). Despite a smaller sample size for SWW, the number of alleles unique to that market class was similar (15.9%) to HRS (16.1%). Nonetheless, the higher genetic diversity levels observed in HRS and SRW reflect the larger sample size included for analysis in this study.

Genetic Relationships among U.S. Wheat Germplasm

To compare genetic relationships among the 43 accessions surveyed, a majority-rule consensus tree was built using a genetic-distance matrix based measure and UPGMA methods (Fig. 1). The soft red winter lines (SRW) adapted to the East and Southeast regions of the United States formed a clearly separated group supported by a high bootstrap value (90%). Within the second major cluster, also supported by a high bootstrap value (77%), the spring germplasm was grouped separately from the winter germplasm, but these two groups were supported with low bootstrap values (<50%), suggesting cross-breeding between spring and winter germplasm was involved in the ancestral lines of these samples. The winter lines that grouped in the second major

cluster were further divided into three subgroups. One included all the hard winter genotypes (HRW and HWW) from Idaho and the Midwest production area (bootstrap <50%), whereas the other two were formed exclusively by soft cultivars (SWW), one from the Northeast (bootstrap 63%) and the other from the Pacific Northwest (bootstrap 76%) region of the United States. The mixture of HRW and HWW lines within the same cluster reflects the frequent intercrossing between these two gene pools, also suggested by the pedigree data. Within the spring cluster, one of two subgroups consisted of those originating from the Northern Great Plains region, including six HRS lines and one HWS line. The other contained mostly lines from the Pacific Northwest region, including all SWS lines, one HWS line, and two HRS lines developed at CIMMYT, Jupateco 73S and Weebill 1. The pedigree data (Table 1) indicated that most of the Pacific Northwest lines analyzed in this study have CIMMYT-derived lines present in their lineages, and thus, were consistent with the relationships based on markers. The HRS line PI610750 was also within this group but was distantly related to all other cultivars as expected for a synthetic derived cultivar.

Genetic Diversity and Population Structure among U.S. Wheat Germplasm

We used AMOVA to assess genetic diversity within and among market classes by dividing the population based on growth habit and market class. The results indicated that 81.2% of the genetic variation ($P < 0.0001$) resided within market classes and 16.6% ($P < 0.0001$) resided among market classes. Only 2.2% of the total variation was explained by spring and winter growth habit classes and was not statistically significant ($P > 0.05$). Genetic variation between market classes was tested using the F_{st} statistic estimated from pairwise comparisons as a measure for genetic distance between market classes. Pairwise comparisons showed considerable variation in the F_{st} values ranging from 0.087 to 0.290 (Table 5). A low level of population differentiation was present between HRS and HWS ($F_{st} = 0.087$), and HRW and HWW ($F_{st} = 0.096$). The differences between HWS and two other market classes, SWS and SWW, were not statistically significant ($P > 0.05$) (even though SWS and SWW were significantly different), suggesting the presence of some level of heterogeneity within HWS. This concurs with the results from the distance-based analysis, where the two HWS lines were clustered in two separate subgroups within the spring wheat cluster. For the remaining pairwise comparisons, a more substantial population differentiation was observed. The overall F_{st} value estimated within market classes was 0.188 indicating moderate population structure. The AMOVA test showed that the division of wheat accessions based on market classes resulted in the partitioning of

Table 4. Summary of simple sequence repeat (SSR) marker diversity estimates among wheat market classes.

Market class [†]	Sample size	No. of alleles	Mean allele number (range)	Mean gene diversity (range)	No. (%) of allele unique to market class
HRS	9	905	3.74 (1–9)	0.56 (0–0.88)	146 (16.1)
HWS	2	378	1.56 (1–3)	0.29 (0–0.62)	17 (4.5)
SWS	4	544	2.25 (1–4)	0.42 (0–0.75)	45 (8.3)
HRW	6	726	3.00 (1–6)	0.51 (0–0.83)	74 (10.2)
HWW	5	622	2.57 (1–5)	0.47 (0–0.80)	48 (7.7)
SRW	10	889	3.67 (1–10)	0.57 (0–0.90)	171 (19.2)
SWW	7	757	3.13 (1–7)	0.52 (0–0.86)	120 (15.9)
Overall	43	1743	7.20 (2–24)	0.66 (0.13–0.94)	621 (35.6)

[†]HRS, hard red spring; HWS, hard white spring; SWS, soft white spring; HRW, hard red winter; HWW, hard white winter; SRW, soft red winter; SWW, soft white winter.

the majority of the genetic variance to within each market class, but no significant difference was found between spring and winter growth habits. The low proportion of variation explained by growth habit is likely the result of divergent subgroups within each growth habit class. This is supported by the fact that all 12 pairwise comparisons between spring and winter subgroups were significantly different.

We further explored the genetic structure among the samples using a model-based method. A model-based method is a cluster analysis that evaluates genetic similarity among genotypes without using prior information about the growth habit and the market class. This analysis suggested several theoretical subpopulation sizes with high significance; however, four subpopulations were optimal for assigning all except five lines into one of the four clusters at a posteriori probability >0.80. The five genotypes assigned to individual clusters with a posteriori probability >0.50 were McNeal, Thatcher, Wesley, Heyne, and the breeding line NY18/Clark's Cream 40–1. The proportion of membership clustered in each of the four subpopulations for each of the 43 accessions is shown in Table S2. These four model-based populations, indicated by different symbols in Fig. 1, corresponded well with their growth habit and market classification. The model-based analysis split the spring wheat from the winter wheat, which is consistent with breeding populations based on growth habit. The analysis further partitioned each of the spring and winter populations into two subpopulations and generally concurred with the genetic relationships based on the distance measure. However, the model-based clustering suggested evidence of substructure within HRS and SWW wheat that the genetic distance-based method did not detect. Splitting the SWW lines resulted in dividing all winter wheat lines into two major populations adapted to wheat production regions east and west of the Mississippi River. McNeal, a HRS cultivar developed in Montana, was more similar to the spring lines adapted to the Pacific Northwest than to the lines from the Northern Great Plains region. The F_{st} values for the four populations identified by the model-based

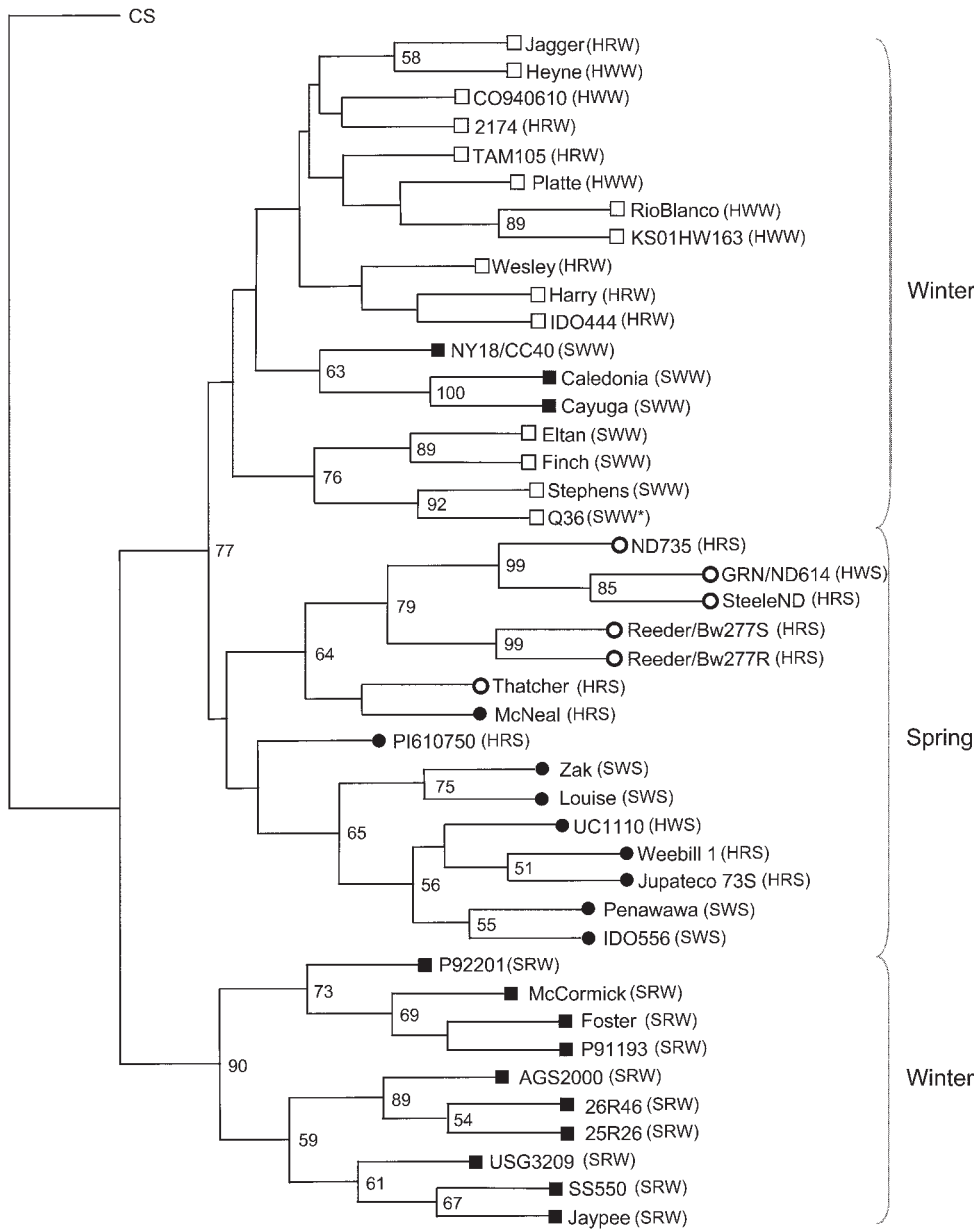


Figure 1. A consensus UPGMA dendrogram of 43 wheat accessions constructed using the genetic distance-based method based on 1000 bootstrap replications. The bootstrap values >50% are shown in the tree. The market class each sample represents is indicated in the parentheses. The four model-based populations were depicted as ■ = Winter–East, □ = Winter–West, ○ = Spring–Northern Plains, and ● = Spring–Pacific Northwest groups. *Q36, a facultative SWS line, was treated as SWW in this study.

analysis were 0.39 for Spring–Northern Plains group, 0.22 for Spring–Pacific Northwest group, 0.14 for Winter–East group, and 0.12 for Winter–West group.

Results from AMOVA for four model-based populations revealed that within-population differences among individuals accounted for 83.8% of genetic variation ($P < 0.0001$), 15.5% of the variation was attributed to differences among populations ($P < 0.0001$), and differences between winter and spring growth habits constituted 0.7% ($P < 0.05$) (data not shown). Overall F_{st} within populations was 0.1623. When F_{st} values were computed for pairwise comparisons of four model-based populations, highly significant popula-

tion differentiation ($P < 0.001$) was detected among all pairwise combinations (data not shown).

Taken together, our analyses suggest that to better evaluate the extent and sources of genetic diversity among wheat accessions, it is necessary to assess and account for the presence of population substructure among samples. This is particularly important for samples with high levels of genetic diversity as reflected by a large proportion of genetic variance within populations from the AMOVA test.

Genome-wide LD Analysis and Distribution of LD on a Chromosomal Scale

The extent of genome-wide LD among the entire set of samples was evaluated through pairwise comparisons among 197 SSR loci (67, 59, and 71 on genomes A, B, and D, respectively) yielding 19306 estimates. Among them, 889 (4.6%) showed significant association at a comparison-wise 0.01 level. A large proportion of loci pairs (86%, 766/889) in LD were from different chromosomes. Out of 123 significant loci pairs within the same chromosomes, 70 of them were linked at <10 cM. Figure 2 shows the distribution of r^2 values as a function of genetic distance in centimorgans for loci pairs located on the same chromosome. The r^2 values declined rapidly to 0.2 within 10 cM and to 0.1 within 20 cM for linked loci in wheat.

The patterns of LD on a chromosomal scale were further evaluated for all chromosomes (Fig. 3). The r^2 value corresponding to the 95th percentile of the distribution of those estimates was 0.056 and this value was used as a population-specific threshold for r^2 as evidence of genetic linkage. Chromosome arms were populated with well-distributed markers except for 1BL, 1DS, 2BL, 2DL, 4AS, 4BS, 5AS, 5BS, 6AS, 6BS, 7AL, and 7DL. Regions of significant LD that involved loci 10 cM or less apart were present on most of the chromosomes, except 1B, 3D, and 7D. High levels of LD were detected among closely spaced loci physically encompassing the centromeric region of 2A, 2B, 3B, 6B, and

7B. Notable exceptions, however, were observed on 3DL, 4DL, and 6AL (two regions) where large genetic distances (>30 cM) remained in LD in this group of genotypes. Significant LD estimates among unlinked loci were detected in 15 regions on chromosomes 1A, 2A, 3B, 4D, 5A, 5B, 6A, 6D, and 7D. In contrast, several regions with multiple markers spaced less than 4 cM apart did not exhibit LD such as 3DL, 5BL, and 5DL. The pairwise r^2 estimates among 197 SSR loci within each chromosome ranged from 0 to 0.547, with a mean of 0.041 and a median of 0.030 (data not shown). The percentage of loci pairs in LD on individual chromosomes ranged from 0% on 1B to 30% on 2D and 7B. The B genome showed the highest proportion of significant LD even though it had the fewest markers.

The estimated locations of several genes controlling agronomic traits were compared to the significant LD regions identified in this study (Fig. 3). Surprisingly, none of the LD regions seemed to be associated with these traits, except the 3DL region that includes genes for kernel color and leaf and stem rust (caused by *Puccinia triticina* Eriks. and *P. graminis* Pers.:Pers. f. sp. *tritici* Eriks. and E. Henn, respectively) resistance. Although the genes associated with seed protein, market class, and disease resistance have been subjected to selection throughout the breeding process, recombination has apparently reduced the LD.

DISCUSSION

SSR Diversity

SSR marker analyses revealed a gene diversity of 0.66 and a mean allele number of 7.2 over 242 loci for this cross-section of U.S. wheat germplasm. These values are relatively higher than previously reported estimates of 6.2 alleles among 40 elite European cultivars over 23 loci (Plaschke et al., 1995), 5.4 alleles and a gene diversity of 0.57 among 68 advanced CIMMYT lines over 47 genomic SSR markers (Dreisigacker et al., 2004), 4.8 alleles over 93 loci among 95 eastern U.S. soft winter wheat lines and cultivars (Bressegello and Sorrells, 2006), and 6.9 alleles and a gene diversity of 0.55 over 70 loci detected among 134 durum accessions representing the elite germplasm from wide geographic regions (Maccaferri et al., 2005). In contrast, Roussel et al. (2004) reported an allele number of 14.5 and a gene diversity of 0.662 among 559 French wheat accessions over 42 polymorphic loci. Using wheat accessions from the IPK gene bank, Huang et al. (2002) reported a gene diversity of 0.77 and 18.1 alleles detected over 26 loci among 998 accessions from 68 countries.

Table 5. Pairwise Wright's fixation index (Fst) statistics among wheat market classes.† Upper diagonal is Fst value, lower diagonal is P value.

	HRS	HWS	SWS	HRW	HWW	SRW	SWW
HRS	–	0.087	0.170	0.144	0.180	0.156	0.164
HWS	0.924	–	0.290	0.256	0.282	0.232	0.254
SWS	0.010	0.066	–	0.219	0.244	0.202	0.207
HRW	0.004	0.041	0.002	–	0.096	0.169	0.147
HWW	0.002	0.046	0.021	0.534	–	0.220	0.184
SRW	0.000	0.021	0.000	0.000	0.001	–	0.171
SWW	0.000	0.060	0.001	0.001	0.000	0.000	–

†HRS, hard red spring; HWS, hard white spring; SWS, soft white spring; HRW, hard red winter; HWW, hard white winter; SRW, soft red winter; SWW, soft white winter.

The much higher levels of SSR marker diversity found in these studies can be attributed to the use of landraces and diverse germplasm maintained in the gene bank. In our study, the higher gene diversity of this set of cultivars and experimental lines likely reflects the representation of seven market classes in the United States, thus encompassing a wider gene pool, and the use of a large number of SSR markers to detect genetic diversity over the entire wheat genome.

Genetic Relationships and Genetic Structure among U.S. Wheat Germplasm

While known pedigree information can provide a breeding history of germplasm under consideration, errors, missing information, and unknown effects of selection and drift can affect the accuracy of interpretations using pedigree data alone to assess genetic relationships among wheat cultivars (van Beuningen and Busch, 1997; Kim and Ward, 1997; Fufa et al., 2005). The genetic distance-based clustering method separated winter wheat into two clusters but kept all the

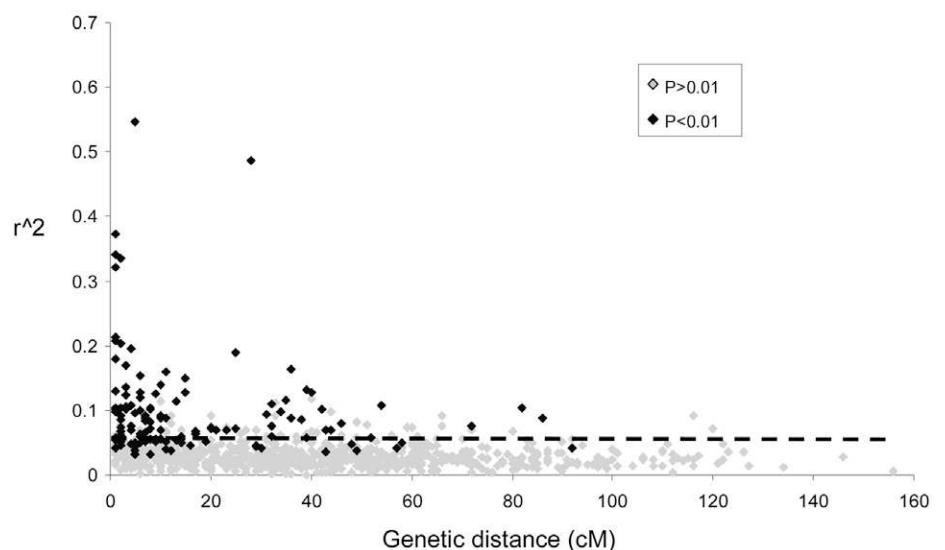


Figure 2. Genome-wide LD decay among SSR marker pairs as a function of genetic distance (cM). The P values were determined using 1000 permutations. The line corresponds to the population-specific threshold as the 95th percentile of the distribution of r^2 as evidence of genetic linkage.

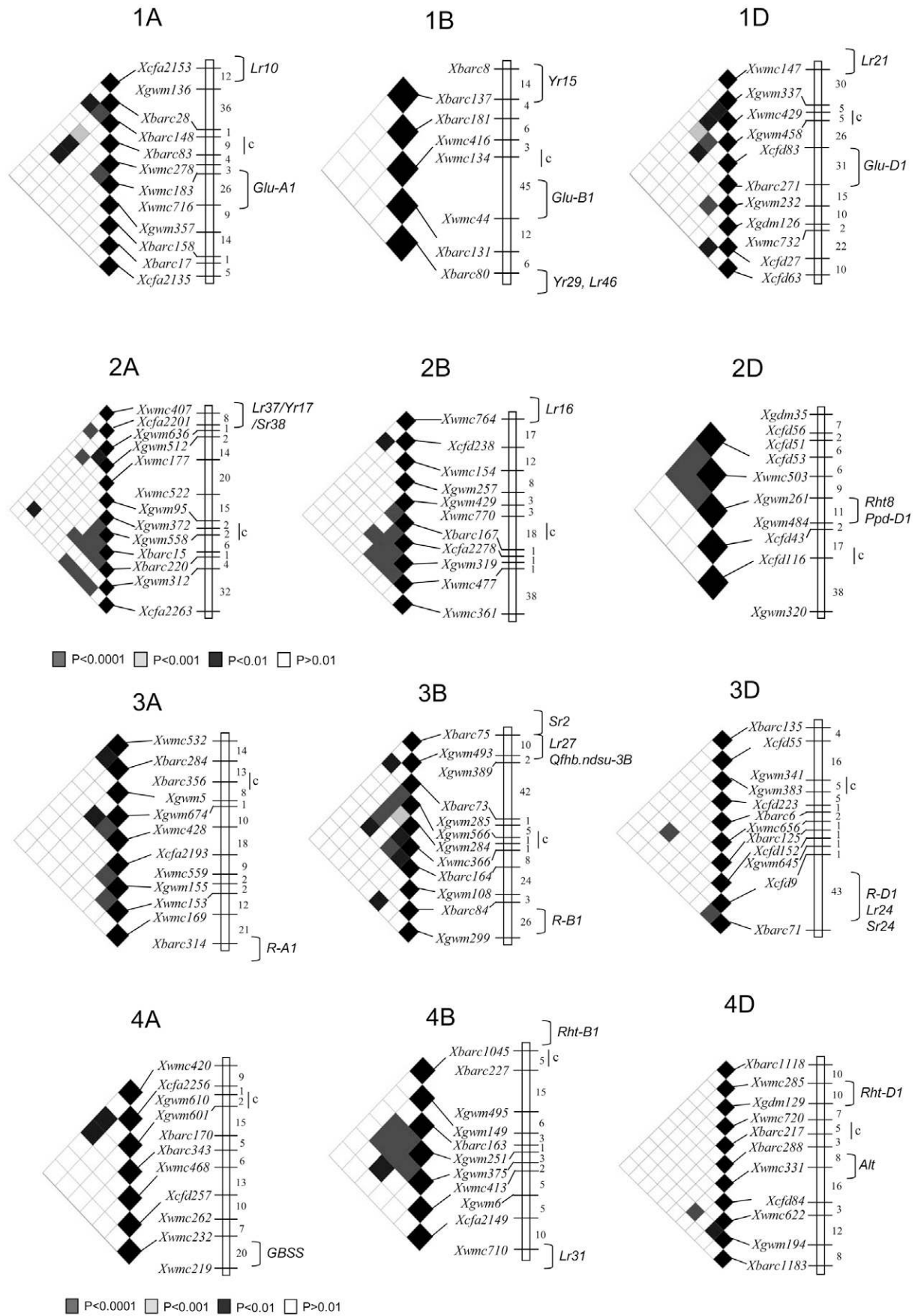


Figure 3. Wheat consensus map and distribution of LD patterns by chromosome. The gray scale shown in each square corresponds to *P* value. The original colored display can be accessed from the supplemental Fig. S1 file. Also included on the map are the chromosome locations of genes controlling selected agronomic traits.

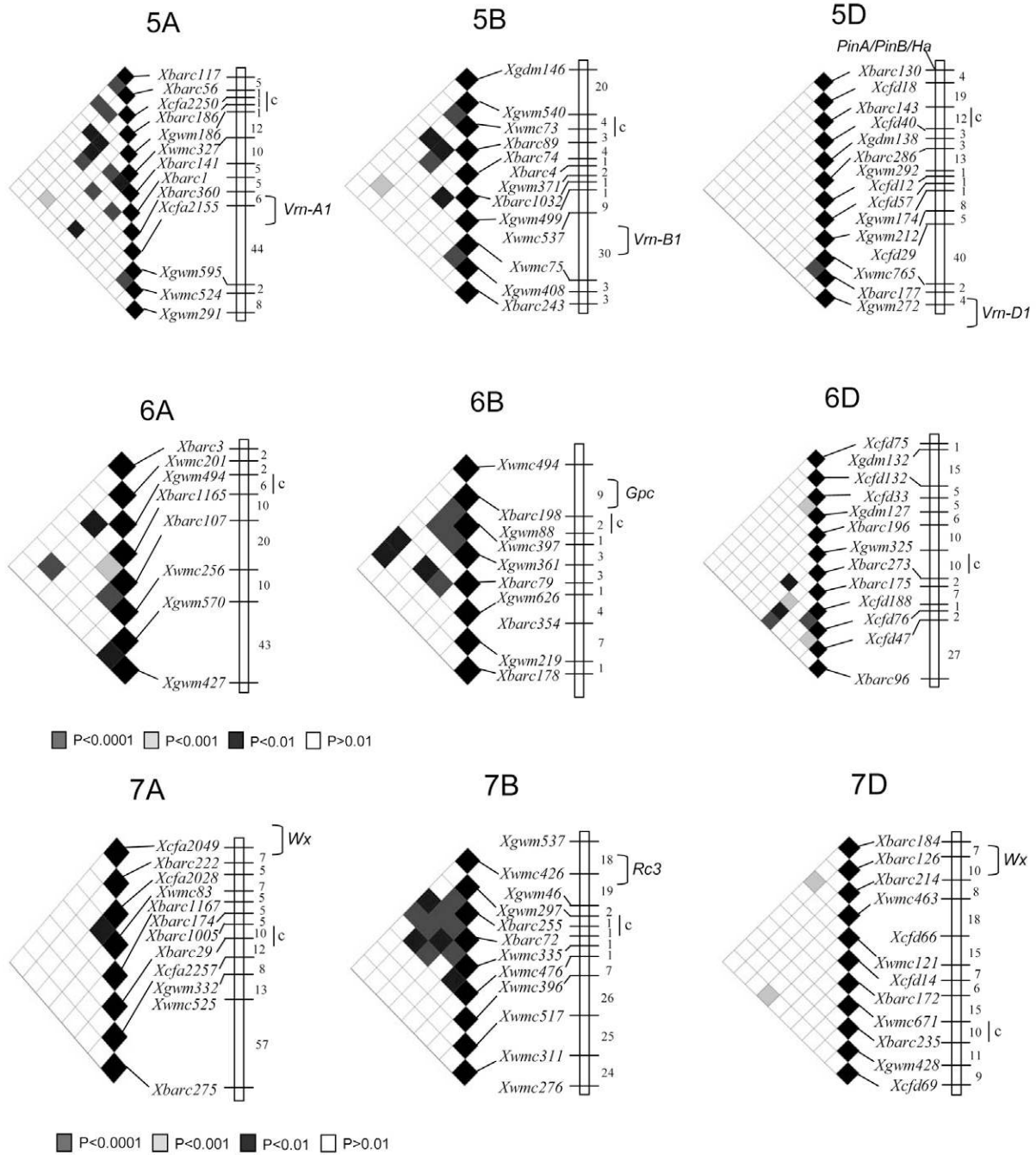


Figure 3. Continued.

spring wheat accessions together, even though the group as a whole was more diverse. Within each of these clusters, except for SRW and SWW, five other market classes, HRS, HWS, HRW, SWS, and HWW, did not form a single cluster. These results contrast with the findings of Barrett and Kidwell (1998) involving a set of wheat lines adapted mainly to the Pacific Northwest region. Our studies indicate that the germplasm surveyed from different wheat production regions across the country were quite diverse. While results from genetic distance-based and model-based approaches generally agreed with each other, some differences were noted. The four populations identified by the model-based analysis corresponded largely to major geographic regions

of wheat production in the United States indicating that the genetic diversity existing among the U.S. wheat germplasm in this study was likely the result of regional adaptation rather than market class difference, and that the individuals clustered in the same population likely shared related ancestral lines in the breeding history. In this study, Jagger was believed to be a direct descendent of Stephens based on pedigree. Structure analysis assigned them to the same population, both with a probability of >0.80 . However, the relationship estimate based on the distance-based method indicated that they have diverged greatly. Also suggested from the pedigree data was the sharing of similar ancestry among HWW lines and SWW–Pacific Northwest lines,

thus agreeing with the model-based results. But such relationships were not supported by the distance-based analysis. Taken together, our analysis indicates that the model-based results may be more consistent with the pedigree data. A similar finding was also reported among 260 maize inbred lines (Liu et al., 2003). However, the distance-based results were found to correspond better to the pedigree data among 115 U.S. rice cultivars (Lu et al., 2005). Our study involved a much smaller sample size (43), and samples for some market classes were underrepresented. Therefore, further study including more lines from different production regions is necessary to verify this finding.

Chromosomal Distribution of Linkage Disequilibrium Patterns in Wheat

Genome-wide r^2 values declined rapidly to 0.2 within 10 cM and to the baseline significance level within 20 cM for linked loci. Similar or larger estimates of LD were reported for *Arabidopsis* (Nordborg et al., 2002) and barley (Kraakman et al., 2004), while for sugar beet (*Beta vulgaris* L.) genotypes, LD was <3 cM (Kraft et al., 2000). In *Lolium perenne* L., LD was <3.4 cM (Skøt et al., 2005), and in maize LD declined over a distance of 2000 bp (Remington et al., 2001).

The genome-wide LD analysis results also revealed that overall fewer than 5% of marker loci pairs showed significant LD, and that the majority of them involved two independent loci. This could be due, in part, to the small sample size that reduced statistical power and potentially increased the effects of drift. Also, as indicated from the previous studies in maize, genetic relatedness increases LD (Liu et al., 2003; Stich et al., 2005). The wide genetic diversity found among different wheat market classes and the little or no genetic similarity shared between growth habits may have contributed to lower levels of LD detected as well. Structure analysis indicated a moderate degree of population differentiation among the samples studied and the F_{st} value for one of the four model-based populations, Spring–Northern Plains group, was estimated at 0.39, indicating the presence of population stratification in this group. The identification of substructure among wheat accessions has been used in analyses to reduce the number of spurious associations found in independent loci pairs (Maccaferri et al., 2005; Brescghello and Sorrells, 2006).

For the significant LD detected within chromosomes, consistent with previous findings in both animals and plants, our analysis showed decreasing disequilibrium with increasing distance between markers, an indication of LD maintained by genetic linkage. Most of the LD was observed between loci less than 10 cM apart, and it tended to be sporadically distributed on various chromosomes. However, closely linked markers were frequently not in LD, and LD could also be seen between distant markers. In wheat, the recombination rates tend to be elevated at

the distal end of the chromosomes (Akhunov et al., 2003). From our analysis, low-recombination regions near centromeres that were in LD were found on only 5 of the 21 chromosomes. In contrast, significant LD was also found in markers physically located in the distal chromosome bins of 3DL (3DL3–0.81–1.00), 5DL (5DL5–0.76–1.00), and 6AL (6AL8–0.90–1.00). The long-range LD on 3DL and 6AL involved markers more than 40 cM apart, and higher marker density is needed to better define the LD blocks in these regions. Nonetheless, the nonuniform chromosome distribution of LD observed in this study is comparable to the findings in humans (Huttley et al., 1999; Ardlie et al., 2002), maize (Stich et al., 2005), and a previous study in wheat (Brescghello and Sorrells, 2006).

Our analysis indicated that there is extensive variation in the extent of LD throughout the wheat genome, and this nonuniform pattern of LD in the genome is likely the result of selection and genetic drift. We hypothesize that there might be regions of high LD surrounding genes that differentiate the U.S. wheat market classes. However, there was little evidence of LD among this set of genotypes in the regions around the kernel hardness locus on 5DS or the red kernel color genes on the long arms of chromosomes 3A and 3B, with the exception of *R-D1* on 3DL. The *VRN-A1* gene on 5AL has been reported to be located in a region of reduced recombination (Yan et al., 2003), but LD was not significant in the flanking regions of genes determining the spring and winter growth habit. This might be related to the fact that spring growth habit can be conferred independently by mutations in any of the three copies of the *Vrn-1* gene (*Vrn-A1*, *Vrn-B1*, or *Vrn-D1*; Yan et al., 2004; Fu et al., 2005) relaxing the selection pressure. The genes controlling market classes may have been fixed in the ancestral lines many generations ago. Consequently, these genes and their surrounding regions have been subjected to little or no selection pressure during the breeding process and recombination has eroded LD. The LD surrounding *R-D1* on 3DL may result from recent intermating of red and white wheat cultivars by breeders which is much more likely than hybridization of hard and soft or spring and winter cultivars. However, we cannot rule out the possibility that insufficient genome coverage and density of markers in some regions contributed to our inability to detect LD in those regions.

Association mapping strategies would benefit from the knowledge of LD patterns. In this study, r^2 values decreased with increasing distance between markers, particularly below 10 cM, suggesting that the mapping resolution using this set of genotypes would generally be well below 10 cM for most of the genome. A comparison of the available LD studies in wheat indicates that LD varies widely among different populations, both within and among species (Maccaferri et al., 2005; Brescghello and Sorrells, 2006). Because the extent of LD is likely to be population

dependent, thus reducing the predictability in estimating the marker density required for association studies, caution must be used in designing association studies in wheat. Consequently, most association mapping studies will need to focus on either previously identified quantitative trait loci (QTL) intervals and saturate those regions with markers or on candidate genes for traits of interest. The use of populations with more related breeding lines, typically those adapted to the same wheat production region, may increase the levels of LD and facilitate the detection of associations between markers and QTL using fewer markers.

Acknowledgments

The authors thank Jamie Rust for excellent technical assistance, and Dave Matthews for assistance in querying data from the GrainGenes database. Thanks also go to the wheat breeders participating in the wheat CAP grant who made the wheat cultivars and unreleased germplasm available for this study. This research was supported in part by the funds from the USDA, Cooperative State Research, Education and Extension Service, Coordinated Agricultural Project grant number 2006-55606-16629.

References

- Akhunov, E., A.W. Goodyear, S. Geng, L. Qi, B. Echaliier, B.S. Gill, Miftahudin, J.P. Gustafson, G. Lazo, S. Chao, O.D. Anderson, A.M. Linkiewicz, J. Dubcovsky, M. La Rota, M.E. Sorrells, D. Zhang, H. Nguyen, V. Kalavacharla, K. Hossain, S.F. Kianian, J. Peng, N.L.V. Lapitan, J.L. Gonzalez-Hernandez, J.A. Anderson, D. Choi, T.J. Close, M. Dilbirligi, K.S. Gill, M.K. Walker-Simmons, C. Steber, P.E. McGuire, C.O. Qualset, and J. Dvorak. 2003. The organization and rate of evolution of the wheat genomes are correlated with recombination rates along chromosome arms. *Genome Res.* 13:753–763.
- Ardlie, K.G., L. Kruglyak, and M. Seielstad. 2002. Patterns of linkage disequilibrium in the human genome. *Nat. Rev. Genet.* 3:299–309.
- Barrett, B.A., and K.K. Kidwell. 1998. AFLP-based genetic diversity assessment among wheat cultivars from the Pacific Northwest. *Crop Sci.* 38:1261–1271.
- Barrett, B.A., K.K. Kidwell, and P.N. Fox. 1998. Comparison of AFLP and pedigree-based genetic diversity assessment methods using wheat cultivars from the Pacific Northwest. *Crop Sci.* 38:1271–1278.
- Botstein, D., R.L. White, M. Skolnick, and R.W. Davis. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* 32:314–331.
- Breseghele, F., and M.E. Sorrells. 2006. Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. *Genetics* 172:1165–1177.
- Brown, G.R., G.P. Gill, R.J. Kuntz, C.H. Langley, and D.B. Neale. 2004. Nucleotide diversity and linkage disequilibrium in loblolly pine. *Proc. Natl. Acad. Sci. USA* 101:15255–15260.
- Buckler, E.S., B.S. Gaut, and M.D. McMullen. 2006. Molecular and functional diversity of maize. *Curr. Opin. Plant Biol.* 9:172–176.
- Buckler, E.S., IV, and J.M. Thornsberry. 2002. Plant molecular diversity and applications of genomics. *Curr. Opin. Plant Biol.* 5:107–111.
- Caldwell, K.S., J. Russell, P. Langridge, and W. Powell. 2006. Extreme population-dependent linkage disequilibrium detected in an inbreeding plant species, *Hordeum vulgare*. *Genetics* 172:557–567.
- Chen, H.B., J.M. Martin, M. Lavin, and L.E. Talbert. 1994. Genetic diversity in hard red spring wheat based on sequence-tagged-site PCR markers. *Crop Sci.* 34:1628–1632.
- Cox, T.S., J.P. Murphy, and D.M. Rodgers. 1986. Changes in genetic diversity in the red winter wheat regions of the United States. *Proc. Natl. Acad. Sci. USA* 83:5583–5586.
- Diwan, N., and P.B. Cregan. 1997. Automated sizing of fluorescent-labeled simple sequence repeat (SSR) markers to assay genetic variation in soybean. *Theor. Appl. Genet.* 95:723–733.
- Dreisigacker, S., P. Zhang, M.L. Warburton, M. Van Ginkel, D. Hoisington, and A.E. Melchinger. 2004. SSR and pedigree analyses of genetic diversity among CIMMYT wheat lines targeted to different megaenvironments. *Crop Sci.* 44:381–388.
- Dvorak, J., P.E. McGuire, and B. Cassidy. 1988. Apparent sources of the A genomes of wheats inferred from the polymorphism in abundance and restriction fragment length of repeated nucleotide sequences. *Genome* 30:680–689.
- Excoffier, L., G. Laval, and S. Schneider. 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evol. Bioinformatics Online* 1:4750.
- Falush, D., M. Stephens, and J.K. Pritchard. 2003. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* 164:1567–1587.
- Fu, D., P. Szucs, L. Yan, M. Helguera, J. Skinner, P. Hayes, and J. Dubcovsky. 2005. Large deletions in the first intron of the *Vrn-1* gene are associated with spring growth habit in barley and polyploidy wheat. *Mol. Genet. Gen.* 273:54–65.
- Fufa, H., P.S. Baenziger, B.S. Beecher, I. Dweikat, R.A. Graybosch, and K.M. Eskridge. 2005. Comparison of phenotypic and molecular marker-based classification of hard red winter wheat cultivars. *Euphytica* 145:133–146.
- Garris, A.J., S.R. McCouch, and S. Kresovich. 2003. Population structure and its effect on haplotype diversity and linkage disequilibrium surrounding the *xa5* locus of rice (*Oryza sativa* L.). *Genetics* 165:759–769.
- Gaut, B.S., and A.D. Long. 2003. The lowdown on linkage disequilibrium. *Plant Cell* 15:1502–1506.
- Glaubitz, J.C. 2004. CONVERT: A user-friendly program to reformat diploid genotypic data for commonly used population genetic software packages. *Mol. Ecol. Notes* 4:309–310.
- Gupta, P.K., H.S. Balyan, K.J. Edwards, P. Isaac, V. Korzun, M. Roder, M.-F. Gautier, P. Joudrier, A.R. Schlatter, J. Dubcovsky, R.C. de la Pena, M. Khairallah, G. Penner, M.J. Hayden, P. Sharp, B. Keller, R.C.C. Wang, J.P. Hardouin, P. Jack, and P. Leroy. 2002. Genetic mapping of 66 new microsatellite (SSR) loci in bread wheat. *Theor. Appl. Genet.* 105:413–422.
- Guyomarç'h, H., P. Sourdille, G. Charmet, K.J. Edwards, and M. Bernard. 2002. Characterization of polymorphic microsatellite markers from *Aegilops tauschii* and transferability to the D-genome of bread wheat. *Theor. Appl. Genet.* 104:1164–1172.
- Hamblin, M.T., S.E. Mitchell, G.M. White, J. Gallego, R. Kukatla, R.A. Wing, A.H. Peterson, and S. Kresovich. 2004. Comparative population genetics of the Panicoid grasses: Sequence polymorphism, linkage disequilibrium and selection in a diverse sample of *Sorghum bicolor*. *Genetics* 167:471–483.
- Huang, X.Q., A. Borner, M.S. Roder, and M.W. Ganai. 2002. Assessing genetic diversity of wheat (*Triticum aestivum* L.) germplasm using microsatellite markers. *Theor. Appl. Genet.* 105:699–707.

- Huttley, G.A., M.W. Smith, M. Carrington, and S.J. O'Brien. 1999. A scan of linkage disequilibrium across the human genome. *Genetics* 152:1711–1722.
- Kim, H.S., and R.W. Ward. 1997. Genetic diversity in eastern US soft winter wheat (*Triticum aestivum* L. Em. Thell.) based on RFLPs and coefficients of parentage. *Theor. Appl. Genet.* 94:472–479.
- Kraakman, A.T.W., R.E. Niks, P.M.M.M. Van den Berg, P. Stam, and F.A. Van Eeuwijk. 2004. Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. *Genetics* 168:435–446.
- Kraft, T., M. Hansen, and N.-O. Nilsson. 2000. Linkage disequilibrium and fingerprinting in sugar beet. *Theor. Appl. Genet.* 101:323–326.
- Liu, K., M. Goodman, S. Muse, J.S. Smith, E. Buckler, and J. Doebley. 2003. Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellite. *Genetics* 165:2117–2128.
- Liu, K., and S.V. Muse. 2005. PowerMarker: An integrated analysis environment for genetic marker analysis. *Bioinformatics* 21:2128–2129.
- Lu, H., M.A. Redus, J.R. Coburn, J.N. Rutger, S.R. McCouch, and T.H. Tai. 2005. Population structure and breeding patterns of 145 U.S. rice cultivars based on SSR marker analysis. *Crop Sci.* 45:66–76.
- Maccaferri, M., M.C. Sanguineti, E. Noli, and R. Tuberosa. 2005. Population structure and long-range linkage disequilibrium in a durum wheat elite collection. *Mol. Breed.* 15:271–289.
- Manifeto, M.M., A.R. Schlatter, H.E. Hopp, E.Y. Suarez, and J. Dubcovsky. 2001. Quantitative evaluation of genetic diversity in wheat germplasm using molecular markers. *Crop Sci.* 41:682–690.
- Nordborg, M., J.O. Borevitz, J. Bergelson, C.C. Berry, J. Chory, J. Hagenblad, M. Kreitman, J.N. Maloof, T. Noyes, P.J. Oefner, E.A. Stahl, and D. Wiegand. 2002. The extent of linkage disequilibrium in *Arabidopsis thaliana*. *Nat. Genet.* 30:190–193.
- Page, R.D.M. 1996. TREEVIEW: An application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.* 12:357–358.
- 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.
- Paull, J.G., K.J. Chalmers, A. Karakousis, J.M. Kretschmer, S. Manning, and P. Langridge. 1998. Genetic diversity in Australian wheat varieties and breeding material based on RFLP data. *Theor. Appl. Genet.* 96:435–446.
- Pestsova, E., M.W. Ganal, and M.S. Roder. 2000. Isolation and mapping of microsatellite markers specific for the D genome of bread wheat. *Genome* 43:689–697.
- Plaschke, J., M. Ganal, and M. Roder. 1995. Detection of genetic diversity in closely related bread wheat using microsatellite markers. *Theor. Appl. Genet.* 91:1001–1007.
- Pritchard, J.K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945–959.
- Reif, J.C., A.E. Melchinger, and M. Frisch. 2005. Genetical and mathematical properties of similarity and dissimilarity coefficients applied in plant breeding and seed bank management. *Crop Sci.* 45:1–7.
- Remington, D.L., J.M. Thornsberry, Y. Matsuoka, L.M. Wilson, S.R. Whitt, J. Doebley, S. Kresovich, M.M. Goodman, and E.S. Buckler, IV. 2001. Structure of linkage disequilibrium and phenotypic association in the maize genome. *Proc. Natl. Acad. Sci. USA* 98:11479–11484.
- 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 genetic similarity and genetic distance. p. 145–153. *In* Studies in genetics VII. University of Texas Publication 7213. Univ. of Texas, Austin.
- Roussel, V., J. Koenig, M. Beckert, and F. Balfourier. 2004. Molecular diversity in French bread wheat accessions related to temporal trends and breeding programmes. *Theor. Appl. Genet.* 108:920–930.
- Schuelke, M. 2000. An economic method for the fluorescent labeling of PCR fragments. *Nat. Biotechnol.* 18:233–234.
- Simko, I., S. Costanzo, K.G. Haynes, B.J. Chris, and R.W. Jones. 2004. Linkage disequilibrium mapping of a *Verticillium dahliae* resistance quantitative trait locus in tetraploid potato (*Solanum tuberosum*) through a candidate gene approach. *Theor. Appl. Genet.* 108:217–224.
- Sköt, L., M.O. Humphreys, I. Armstead, S. Heywood, K.P. Sköt, R. Sanderson, I.D. Thomas, K.H. Chorlton, and N.R.S. Hamilton. 2005. An association mapping approach to identify flowering time genes in natural populations of *Lolium perenne* (L.). *Mol. Breed.* 15:233–245.
- Somers, D.J., P. Isaac, and K. Edwards. 2004. A high-density microsatellite consensus map for bread wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 109:1105–1114.
- Song, Q.J., E.W. Fickus, and P.B. Cregan. 2002. Characterization of trinucleotide SSR motifs in wheat. *Theor. Appl. Genet.* 104:286–293.
- Song, Q.J., J.R. Shi, S. Singh, E.W. Fickus, J.M. Costa, J. Lewis, B.S. Gill, R. Ward, and P.B. Cregan. 2005. Development and mapping of microsatellite (SSR) markers in wheat. *Theor. Appl. Genet.* 110:550–560.
- Sourdille, P., S. Singh, T. Cadalen, G.L. Brown-Guedira, G. Gay, L. Qi, B.S. Gill, P. Dufour, A. Murigneux, and M. Bernard. 2004. Microsatellite-based deletion bin system for the establishment of genetic-physical map relationships in wheat (*Triticum aestivum* L.). *Funct. Integr. Genomics* 4:12–25.
- Sourdille, P., M. Tavaud, G. Charmet, and M. Bernard. 2001. Transferability of wheat microsatellite to diploid Triticeae species carrying the A, B and D genomes. *Theor. Appl. Genet.* 103:346–352.
- Stich, B., A.E. Melchinger, M. Frisch, H.P. Maurer, M. Heckenberger, and J.C. Reif. 2005. Linkage disequilibrium in European elite maize germplasm investigated with SSRs. *Theor. Appl. Genet.* 111:723–730.
- van Beuningen, L.T., and R.H. Busch. 1997. Genetic diversity among North American spring wheat cultivars: I. Analysis of the coefficient of parentage matrix. *Crop Sci.* 37:570–579.
- Weir, B.S. 1996. Genetic data analysis II. Sinauer Associated, Inc., Sunderland, MA.
- Yan, L., M. Helguera, K. Kato, S. Fukuyama, J. Sherman, and J. Dubcovsky. 2004. Allelic variation at the *Vm-1* promoter region in polyploidy wheat. *Theor. Appl. Genet.* 109:1677–1686.
- Yan, L., A. Loukoianov, G. Tranquilli, M. Helguera, T. Fahima, and J. Dubcovsky. 2003. Positional cloning of wheat vernalization gene *Vrn1*. *Proc. Natl. Acad. Sci. USA* 100:6263–6268.
- Yu, J.K., M. La Rota, R.V. Kantety, and M.E. Sorrells. 2004. EST derived SSR markers for comparative mapping in wheat and rice. *Mol. Genet. Gen.* 271:742–751.