Genetic structures of the CIMMYT international yield trial targeted to irrigated environments

Susanne Dreisigacker · Hailemichael Shewayrga ·
Jose Crossa · Vivi N. Arief · Ian H. DeLacy ·
Ravi P. Singh · Mark J. Dieters · Hans-Joachim Braun

Abstract International yield trials are assembled by CIMMYT to disseminate promising wheat breeding materials worldwide. To assess the genomic structure and linkage disequilibrium (LD) within this germplasm, wheat lines disseminated during 25 years of the Elite Spring Wheat Yield Trial (ESWYT) targeted for irrigated environments of the world were genotyped with the high-throughput Diversity Arrays Technology (DArT) marker system. Analyses of population structure assigned the ESWYT germplasm into five major sub-populations that are shaped by prominent CIMMYT wheat lines and their descendants. Based on genetic distance, we concluded that a constant level of genetic diversity was maintained over the years of ESWYT dissemination. Genetic distance between the individual ESWYT lines significantly increased when the ESWYT were grouped according to the differences in years of ESWYT dissemination, suggesting a systematic change in allele frequencies over time, most probably due to breeding and directional selection. By means of multiple regression analyses, 78 markers displaying a significant change in allele frequency across years were identified and interpreted as an indicator for constant selection. The markers identified were partly associated with grain yield, leaf, stem, and yellow rust and point to key genomic regions for further investigation. Large numbers of adjacent DArT marker pairs showed significant LD across the ESWYT population and within each of the five sub-populations identified. Sub-population differentiation measured by the fixation index and average genetic distance were highly correlated with LD levels, suggesting that the sub-populations themselves explain much of the LD.

Keywords International yield trials · Population structure · Temporal genetic diversity · Linkage disequilibrium

S. Dreisigacker (✉) · J. Crossa · R. P. Singh ·
H.-J. Braun
International Maize and Wheat Improvement Center
(CIMMYT), Apdo. Postal 6-641, 06600 México, DF,
México
e-mail: s.dreisigacker@cgiar.org

H. Shewayrga · V. N. Arief · I. H. DeLacy · M. J. Dieters
School of Agriculture and Food Sciences, University of
Queensland, Brisbane, QLD, Australia

Introduction

For international agricultural research centers, dissemination of international nurseries plays an important role for germplasm exchange and fostering cooperation with a global network of crop scientists. The basic goals of the international nurseries are to (i) evaluate promising breeding materials on a worldwide basis (ii) provide information about the adaptability of varieties to a wide range of environments,

yield potential, disease and pest incidence (iii) indicate if varieties might serve as immediate introductions into potentially high wheat production areas, and (iv) serve as parental materials for accelerating the breeding programs (Rajaram et al. 2002).

A looming global food crisis beginning in the 1950s led the Office for Special Studies, the predecessor of the International Centre for Maize and Wheat Improvement (CIMMYT), to formalize the first international stem rust nursery in wheat with a shared objective of increasing wheat production in the developing countries (Byerlee and Dubin 2010). In 1960, the first international spring wheat yield nursery was distributed to 12 countries from Canada to Argentina. Over time, the scope and coverage of the international nurseries have grown and become more specialized. During the 1980s, wheat-growing regions of the world were classified into mega-environments and specific yield trials established to target some of the major mega-environments have been disseminated by CIMMYT in a system that is known as the international wheat improvement network (IWIN). The basic goals and methodology of dissemination of these nurseries have remained largely unchanged until today. Improved germplasm is dispatched through screening nurseries and yield trials to a global network of wheat researchers. Data from these trials are then returned to CIMMYT, catalogued, analyzed, and made available to the global wheat improvement community (Rajaram et al. 2002).

The mega-environment-targeted yield trials have been analyzed in detail across the many locations and for a number of years, providing diverse perspectives of international wheat breeding (Byth et al. 1976; DeLacy et al. 1993, Trethowan and Crossa 2007). Initial analyses involved the screening ability of individual test locations and their associations to identify key locations for any germplasm type (Braun et al. 1992, Trethowan et al. 2003, Lillermo et al. 2005). Heisey et al. (2002) and Trethowan et al. (2002, 2003) analyzed yield gains from the release of new wheat lines included in the international trials over time and observed significant progress in the development of wheat germplasm adapted to both low- and high-yielding conditions.

While phenotypic analyses of wheat lines included in the international yield trials have been extensive, only a few genomic studies of this germplasm have been performed (Crossa et al. 2007, 2010; Arief et al. 2010). Whole-genome marker polymorphisms encompass a range of applications. They are used to estimate the genetic structure of germplasm by inferring the individual identity or relatedness between pairs of individuals or by examining the current differentiation within and between populations. Whole-genome marker profiles are also useful to estimate linkage disequilibrium (LD) in targeted germplasm sets, defined as a non-random association of alleles at two or more loci (e.g. Somers et al. 2007). In combination with phenotypic information, whole-genome association mapping allows the identification of quantitative trait loci (QTL) underlying the target traits (e.g. Crossa et al. 2007; Neumann et al. 2010), or genomic selection the prediction of agronomic performance of untested individuals (Meuwissen et al. 2001; Crossa et al. 2010).

The overall objective of the research reported here was to perform a genomic study of an historical set of wheat lines that have been disseminated via the CIMMYT international Elite Spring Wheat Yield Trial (ESWYT) that is targeted to irrigated environments. Individual objectives were to (i) determine the general and temporal genetic structures of the selected wheat lines with the high-density Diversity Arrays Technology (DArT; Akbari et al. 2006) (ii) examine the allele spectra of marker loci across years of trial distribution, and (iii) investigate the extent and genomic distribution of LD between pairs of marker loci as indicators of selection.

Materials and methods

Plant materials and DArT procedure

The study included 606 advanced lines derived from 23 ESWYT (ESWYT 2 to ESWYT 9, ESWYT 11 to ESWYT 25) disseminated in the 25 year period from 1979/80 to 2004/05. Detailed information on all ESWYT lines included in the study is published in Arief et al. (2010). Seed of all lines was obtained from the CIMMYT Genetic Resource Center and genomic DNA was extracted from bulked leaves of ten two-week-old seedlings using a CTAB procedure (Saghai-Maroof et al. 1984) modified based on CIMMYT protocols (http://www.cimmyt.cgiar.org/ABC/Protocols/manualABC.html). For each ESWYT, the

number of lines genotyped varied and ranged from 4 (ESWYT 6) to 44 (ESWYT 25). Fifty-six lines were represented in more than one ESWYT.

Whole-genome marker polymorphisms were generated via DArT profiling by Triticarte Pty. Ltd. (Canberra, Australia; http://www.triticarte.com.au) using the composite array v2.3 and following the standard DArT procedure described in Akbari et al. (2006). A total of 1,447 DArT markers polymorphic across all lines were detected. From the total number of DArT markers, 435 were located on the integrated map published by Crossa et al. (2007). The absence of a DArT marker score was considered as a missing value.

Statistical analyses

To assess how informative each DArT marker is, the polymorphic information content (PIC) for each locus was calculated following Anderson et al. (1993). Population differentiation was investigated with a subset of 83 mapped DArT markers equally distributed in the genome using the software STRUCTURE version 2.2 (Pritchard et al. 2000). Repeated lines were excluded from this analysis. Due to the biallelic nature of DArT, we followed the recommended procedure for dominant markers in STRUCTURE. Models with a putative number of sub-populations (K) from 1 to 20 with admixture and correlated allele frequencies were considered. Five runs per K, with a burn-in period of 50,000 and data collection of 250,000 iterations, were implemented. The K with the minimum stable estimate of $\Pr(X/K)$ predicted was chosen as optimal K. In addition, delta$K$ as described in Evanno et al. (2005) was calculated to confirm the optimum K. The neighbor joining clustering algorithm was performed for the lines within each sub-population, using Rogers' genetic distance (RD; Rogers 1972) as proximity measure and based on the same 83 polymorphic DArT markers. This step was undertaken to refine the results of the population structure obtained from STRUCTURE.

Average genetic distance between lines was also used as an index of genetic diversity. All polymorphic DArT markers were used to estimate RD among pairs of lines within and between individual ESWYT, and the entire ESWYT dataset. The trend of genetic diversity over the 25 year period of ESWYT dissemination was examined by performing multiple regression analyses. ESWYT 6 and ESWYT 11 were excluded from this analysis because of the low number of lines in these trials (four and nine lines in ESWYT 6 and 11, respectively).

Genetic distance estimates between all pairs of ESWYT (e.g. ESWYT 2 and 3, ESWYT 2 and 5, ESWYT 3 and 5, etc.) were further grouped based on the number of years separating the dissemination of each ESWYT pair. The average genetic distance between all ESWYT pairs in each group was then calculated; e.g., group one represented the average genetic distance between ESWYT that were disseminated in subsequent years, group two the average genetic distance of individual ESWYT pairs that were disseminated 2 years apart, etc. This was performed to determine changes of allele frequencies based on the ongoing breeding process. To further investigate in which parts of the genome allelic changes occurred over time, the frequency spectra of all 435 mapped DArT markers were determined across the 23 ESWYT. The DArT markers extract only biallelic information on the bases of hybridization presence or absence, and therefore the spectra of only the least frequent allele across all lines was considered. Significant changes of allele frequencies within the observed allele spectra were examined using linear regression analyses.

As an additional measure of genetic diversity, pairwise linkage disequilibrium (LD) and its significance was calculated by means of the square of the correlation coefficient of allele pairs between two loci (r^2) according to Weir (1996). LD was examined for all adjacent loci pairs in the genome as well as for all pairwise genome-wide comparisons. Linkage disequilibrium was evaluated based on all of the mapped DArT markers across the entire ESWYT population and within each of the sub-populations determined via population structure analyses. Rare alleles (DArT markers with a frequency <5%) that may inflate LD measurements were not observed and therefore none of the 435 mapped DArT markers were excluded from the analyses. Significance (P values) of LD for marker pairs was determined with Fischer's Exact Test (Fisher 1922) with 1,000 permutations.

Genetic distance and LD was calculated using PowerMarker software (Lui and Muse 2005). The software package R (www.r-project.org) was used to perform multiple regressions.

Results

Population structure

The entire ESWYT population of 606 wheat lines generated a total of 1447 DArT markers; 1431 were taken into account, while 16 with a high proportion of missing data (>10%) were eliminated. A summary of the marker information is given in Table 1. The PIC value of the markers ranged from 0.02 to 0.38 and averaged 0.27. The average genetic distance across all ESWYT lines was 0.335.

Model-based clustering was implemented with a subset of 83 equally distributed DArT markers to infer population structure. The optimal number of populations that describes a given set of germplasm is determined by STRUCTURE by comparing goodness-of-fit values for each run under an a priori number of populations K. According to the STRUCTURE documentation (http://pritch.bsd.uchicago.edu/software/readme_2_1/readme), goodness-of-fit is expected to show a quick increase until it reaches an optimal group number and then presents a "plateau phase" characterized by constant value or very low increase. The optimal number of sub-populations in the ESWYT population was somewhat unstable; the average goodness-of-fit values showed a constant increase with increasing K without reaching a clear plateau (data not shown). Only by calculating delta$K$ could the germplasm be assigned into five sub-populations (P1–P5), shaped by prominent CIMMYT wheat lines and their descendants (Fig. 1). Prominent wheat lines included the lines Bobwhite (P1), Bluejay (P2), Veery wheat (P3), Pavon F76 (P4) and Kauz (P5). The size of the sub-populations ranged between 68 and 113 lines and 127 lines could not be assigned to any of the groups with a probability equal to or

higher than 0.5. The fixation index (F_{ST}) calculated by the STRUCTURE software varied widely amongst the five sub-populations. F_{ST} is defined as the correlation of gametes within sub-populations relative to gametes drawn at random from the entire population and is therefore a measure of population differentiation. Estimated F_{ST} values were 0.26, 0.11, 0.45, 0.39 and 0.73 for sub-populations P1, P2, P3, P4 and P5, respectively. Average genetic distances of the lines within each sub-population varied equally and ranged from 0.121 for sub-population P5 to 0.323 for sub-population P2 (Table 3). Neighbor joining clustering for each of the five sub-populations revealed different levels of additional sub-structures (Fig. 1). An additional three to four clusters in each sub-population were observed. The larger main clusters contained predominantly the lines per se, sister-lines and crosses with the lines that described the respective sub-population. Smaller discrete clusters mainly included crosses with non-CIMMYT germplasm sources or CIMMYT synthetic derivatives.

Genetic diversity and allele frequency changes over time

The results of the analysis of diversity change over time are presented in Fig. 2. Over the time period of 25 years of ESWYT dissemination, the genetic diversity in the ESWYT has remained relatively constant but some fluctuations were evident. The highest level of genetic diversity was observed in ESWYT 13 and the lowest level in ESWYT 20. The average genetic distance between ESWYT increased with increasing time between the year each ESWYT was disseminated. This result suggests a genetic shift from early to recent ESWYT examined based on a systematic change in allele frequencies and is most likely a consequence of breeding and selection during this period (Fig. 3).

To further investigate in which parts of the genome allelic changes occurred, the allele frequency spectra of the 435 mapped markers was examined. By means of linear regression analyses, 17.9% of the mapped markers revealed a significant change of allele frequency ($P < 0.01$) over time. Allele frequency changes occurred in 14 of the 21 wheat chromosomes (Table 2). Not all markers pointed to a single locus; some markers represented a cluster of markers at the same locus. DArT markers on the

Table 1 Distribution of DArT markers polymorphic in 23 ESWYT lines

Distribution of DArT markers	
Total available	1,431
A genome	483
B genome	787
D genome	124
Unassigned markers	37
Mapped in Crossa et al. (2007)	435

Fig. 1 Population structure of ESWYT lines. **a** Membership coefficient (Q values) revealed using STRUCTURE software. Each *horizontal line* represents one wheat line, which is partitioned into five sub-populations. **b** Neighbor joining trees for each sub-population
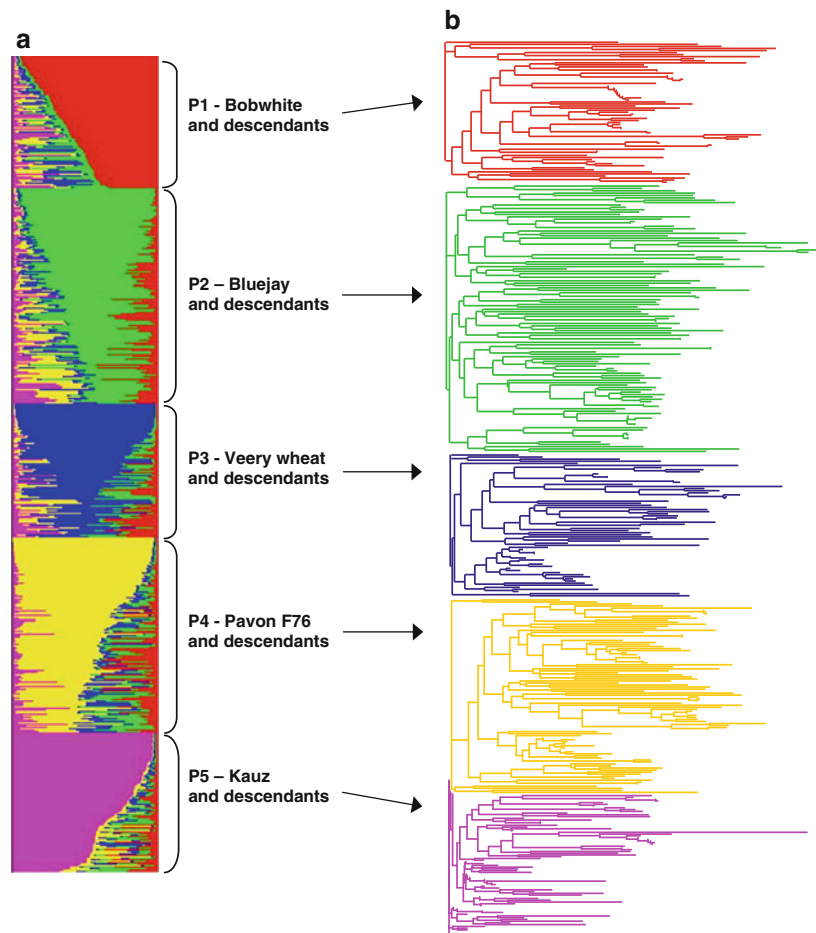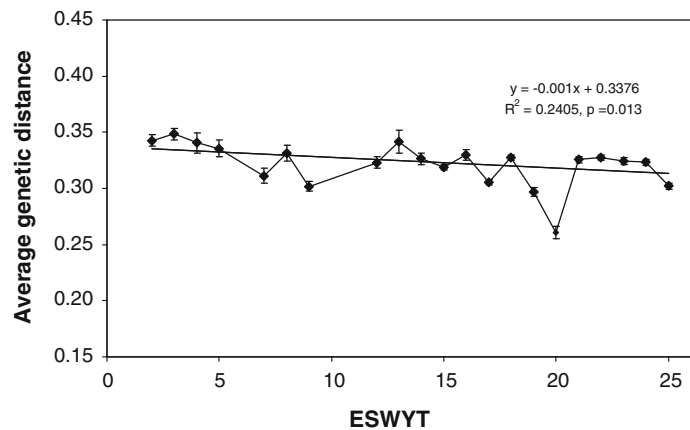

P1 - Bobwhite and descendants

P2 – Bluejay and descendants

P3 - Veery wheat and descendants

P4 - Pavon F76 and descendants

P5 – Kauz and descendants

Fig. 2 Fluctuations in genetic diversity measured on the basis of average Rogers' distances within each ESWYT line over time

$y = -0.001x + 0.3376$
$R^2 = 0.2405$, p =0.013

same or a subset of ESWYT lines have been previously considered in the studies of Crossa et al. (2007, 2010) and Arief et al. (2010) to determine marker–trait associations via whole-genome association mapping and genomic prediction. We surveyed the synergies between the observed marker–trait associations and our results. Out of the 78 observed markers with a significant change in allele

Fig. 3 Average genetic
distance between ESWYT
pairs grouped according to
the number of years
separating the
dissemination of the two
ESWYT lines in each pair
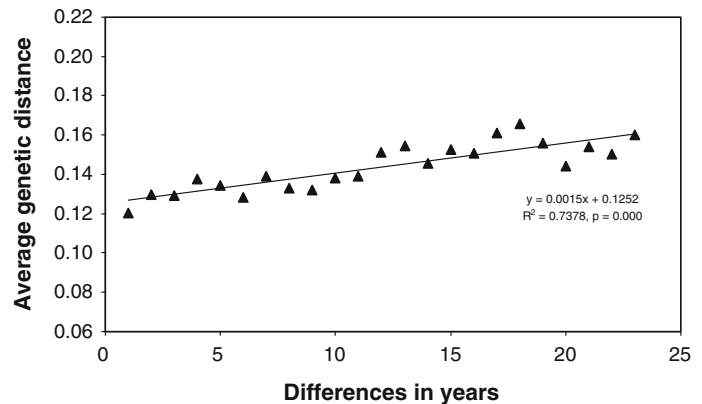
frequency, we found 35 markers that were previously
identified to be trait-associated. Thirteen markers
were associated with grain yield, located at loci on
chromosomes 2A, 2D, 3A, 3B, 5B and 7B. Three
markers associated with yellow rust mapped on
chromosome 2B and a total of 24 markers associated
with two or three traits, including grain yield, leaf,
yellow or stem rust, mapped on chromosomes 1A,
1B, 3A, 3B, 4A, 5B and 7B (Table 2).

Level of LD

The percentage of loci pairs in significant LD
($r^2 > 0.1$, $P < 0.001$) and the average LD values
obtained across the entire ESWYT population and
within each of the five sub-populations are reported in
Table 3. Two percent of independent loci pairs in
significant LD were observed in the entire data set
and ranged from a minimum of 2.1% in sub-
population P5 to a maximum of 5.2% in sub-
population P3. The percentage of adjacent loci pairs
in significant LD within sub-populations was in all
five cases lower than the 25.0% observed in the entire
data set and ranged from 17.2% in sub-population P1
to 21.8% in sub-population P4. The average r^2 values
for both independent and adjacent loci pairs varied
similarly. All five sub-populations showed higher r^2
values in comparison to the entire ESWYT popula-
tion. The highest values r^2 were observed in sub-
population P5 (0.327, 0.689) and the lowest values in
sub-population P2 (0.164, 0.508). The r^2 values
declined within 30 cM to the 0.1 threshold (data not
shown).

According to Maccaferri et al. (2005), the loci
pairs were additionally subdivided into four classes

on the basis of their inter-marker genetic distance:
loci pairs tightly to moderately linked (0–10 and
11–20 cM apart) and loci pairs loosely linked (21–50
and >50 cM). The average inter-marker distance (the
average distance in cM between adjacent loci pairs in
significant LD) was similar in the entire data set and
in the five sub-populations (Table 3). In all cases the
majority of loci pairs in significant LD were repre-
sented in the classes of tightly to moderately linked
markers. Sub-populations P2 and P5 revealed the
lowest and highest r^2 values in each distance class.
Correspondingly, the rate of LD decay in these two
sub-populations evolved faster and slower (respec-
tively) with increasing distance than in the other sub-
populations.

Discussion

Population structure

The CIMMYT wheat program develops germplasm
for regions of the world where spring wheat is grown,
with particular emphasis on the developing world.
This breeding effort is collaborative and many
international cooperators share their best lines and
return yield, disease and agronomic data collected in
CIMMYT's international trials (Rajaram et al. 2002).
For the strategic planning of a breeding program, the
understanding of accessible germplasm, exploitation
of genetic variation and knowledge of population
structure is important. The population structure of
diverse bread wheat germplasm has been previously
explored at the DNA level mainly using simple
sequence repeat (SSR) markers (e.g., Balfourier et al.

Table 2 Mapped DArT markers with allele frequencies that significantly changed across 23 ESWYT lines

Marker	Chr.[a]	Pos.	P	Marker–trait association[b]	Marker	Chr.[a]	Pos.	P	Marker–trait association[b]
wPt4765	1AS	6.3	0.006	GY, YR	wPt1940	3BL	58.2	0.000	
wPt9317	1AS	12.4	0.009		wPt0280	3BL	63.3	0.003	GY
wPt2527	1AS	23.5	0.003		wPt5825	4AL	143.7	0.003	
wPt7030	1AS	23.6	0.008	GY, YR	wPt3349	4AL	160.9	0.004	
wPt9402	1AS	47.9	0.001	YR, SR	wPt2084	4AL	161.1	0.000	GY, YR, SR
wPt3477	1BS	8.2	0.005		wPt3795	4AL	161.4	0.000	GY, YR, SR
wPt8949	1BS	11.4	0.000	GY, YR, SR	wPt7807	4AL	161.7	0.000	GY, YR, SR
wPt8267	1BS	11.6	0.008		wPt1155	4AL	163.3	0.007	
wPt5647	2AL	76.7	0.000	GY	wPt5172	4AL	167.9	0.002	
wPt7024	2AL	84.6	0.004	GY	wPt1362	4AL	171.3	0.000	
wPt1489	2BS	57.8	0.001	YR	wPt3491	4AL	174.1	0.009	
wPt8072	2BS	57.8	0.000		wPt4424	4AL	177.0	0.007	
wPt9350	2BL	91.9	0.008	YR	wPt4620	4AL	177.6	0.000	GY, LR
wPt0047	2BL	106.3	0.003	YR	wPt2151	4AL	183.6	0.000	
wPt2644	2DS	3.3	0.005	GY	wPt1931	4BS	40.9	0.003	
wPt4144	2DS	5.3	0.001	GY	wPt6136	5BS	16.0	0.001	GY
wPt3728	2DS	9.5	0.000		wPt1420	5BS	41.1	0.007	GY
wPt7608	3AS	9.9	0.005		wPt9666	5BS	41.1	0.010	
wPt6854	3AS	11.7	0.004		wPt5346	5BS	62.5	0.005	
wPt9049	3AS	15.2	0.007		wPt5914	5BS	62.5	0.008	GY
wPt1681	3AS	15.6	0.008	GY, YR	wPt5175	5BS	62.7	0.006	GY, YR
wPt4131	3AS	15.6	0.000		wPt9814	5BS	69.7	0.006	GY
wPt7756	3AS	15.7	0.008	GY, YR	wPt3457	5BL	92.3	0.006	
wPt2698	3AS	15.9	0.005	GY	wPt1482	5BL	130.2	0.000	
wPt2755	3AS	16.6	0.007	GY	wPt9103	5BL	130.2	0.000	
wPt7992	3AS	17.0	0.003	GY, YR	wPt9598	5BL	130.2	0.000	GY, LR
wPt4725	3AS	21.0	0.003	GY, YR	wPt4858	6BS	66.0	0.008	
wPt9268	3AS	23.1	0.007	GY	wPt3309	6BS	70.6	0.002	
wPt0365	3BS	44.0	0.002	GY, LR, SR	wPt5333	6BS	70.6	0.002	
wPt0995	3BS	44.0	0.002	GY, YR, SR	wPt7627	6BL	102.5	0.003	
wPt1822	3BS	44.0	0.001		wPt9270	6BL	115.1	0.005	
wPt4370	3BS	44.0	0.002		wPt1048	6BL	118.1	0.000	
wPt4576	3BS	44.0	0.002		wPt1149	7BL	64.7	0.000	YR, SR
wPt7266	3BS	44.0	0.002		wPt2305	7BL	66.0	0.006	
wPt7956	3BS	44.0	0.002		wPt8069	7BL	67.2	0.006	
wPt8983	3BS	44.0	0.002	GY	wPt3939	7BL	117.8	0.000	GY
wPt1159	3BS	44.4	0.006		wPt0884	7BL	156.3	0.006	
wPt6047	3BS	52.7	0.007	GY, LR, SR	wPt4555	7DL	97.1	0.000	
wPt0405	3BS	52.8	0.001						
wPt5906	3BS	55.3	0.007						

Marker–trait associations reported previously in a subset of ESWYT germplasm by Crossa et al. (2007, 2010) are indicated

[a] Chromosome assignments and positions are given according to the integrated map published in Crossa et al. (2007)

[b] *GY* Grain Yield, *YR* Yellow rust, *SR* Stem rust, *LR* Leaf rust

Table 3 Fixation index (F_{ST}), average genetic distance, percentage and average LD ($r^2 > 0.1$, $P < 0.001$) between independent and adjacent DArT markers as well as average inter-marker distance between adjacent DArT markers. LD analysis was performed on the entire ESWYT population and on the five sub-populations identified by STRUCTURE analysis (coded as P1, P2, P3, P4, P5)

	Entire ESWYT	P1	P2	P3	P4	P5
No. of lines	563	69	113	68	96	90
F_{ST}	–	0.26	0.11	0.39	0.45	0.73
Average genetic distance	0.335	0.279	0.325	0.204	0.257	0.121
% LD (genome-wide markers)	2.0	2.6	2.6	5.2	4.5	2.1
% LD (adjacent markers)	25.0	17.2	21.3	21.3	21.8	19.1
Average LD (genome-wide markers)	0.155	0.288	0.164	0.256	0.210	0.327
Average LD (adjacent markers)	0.487	0.660	0.508	0.614	0.585	0.689
Average inter-marker distance	12.1	14.5	12.5	12.6	17.0	14.6

2007; Chao et al. 2007; Somers et al. 2007) and more recently with DArT (Stodart et al. 2007; White et al. 2008). The focus in most published studies was to determine the genetic structure and diversity in different national and global collections of wheat varieties, while in this study elite germplasm of an individual breeding program was employed.

Our results indicate that CIMMYT's breeding program targeted to irrigated environments has developed gene pools in partial isolation from each other derived from different cornerstone breeding parents and hallmark germplasm. Five sub-populations were identified in the ESWYT germplasm. The series of Bobwhite sister lines and their descendents predominant in sub-population P1 are wheat lines developed by the CIMMYT wheat program in the early 1970s. The lines were generated from a top cross with the pedigree Aurora//Kalyan/Bluebird/3/ Woodpecker and entered the international nurseries in the late 1970s. Between 1984 and 1991, 16 of the Bobwhite sister lines were released as commercial cultivars in Africa, South America and the Middle East (Rajaram et al. 1983).

Sub-population P2 includes primarily lines which share the variety Bluejay in their pedigree. The line is early maturing but susceptible to leaf rust. Bluejay is (amongst others) a parent of Opata M85 and WH 542. The Veery wheat cultivars developed in CIMMYT and predominant in sub-population P3 were derived from a cross between a spring wheat Mexican semidwarf and the winter wheat cultivar Kavkaz containing the 1RS chromosome from Petkus rye (Rajaram et al. 1983). Veery wheat (e.g. the lines

Glennson, Ures, or Seri) are popular under drought or heat-stressed conditions and may on average produce a 7% grain yield advantage (Rajaram et al. 1993). One hundred and sixty-five cultivars with Veery wheat in their pedigree have been released in 35 countries, and up to 50% of all wheat cultivars in China may carry the 'Veery' 1RS translocation (Zhou et al. 2003). In the sub-population P4, the Mexican cultivar Pavon F76 released in 1976 and its descendents were very frequent. The cultivar is well known to display moderate levels of adult-plant resistance to leaf and yellow rust that have remained effective in Mexico and other parts of the world where it has been released and grown. Singh et al. (1998) located the slow rusting resistance gene Lr46 in chromosome 1B of Pavon F76. The cultivar Kauz (also named Bacanora), observed as a cornerstone parent in sub-population P5, is one of several Veery wheat progenies that have demonstrated a superior level of abiotic tolerance to a number of stresses (e.g. drought, heat, etc.) and improved nutrient efficiencies (e.g. N- and P-efficiency). Kauz is especially known to contribute to higher heat tolerance (Rajaram et al. 2002).

Although it was possible to identify major gene pools that are present in this breeding population, analyses of the population structure of the ESWYT germplasm showed a complex pattern of genetic relationships. Camus-Kulandaivelu et al. (2007) suggested that the STRUCTURE algorithm does not converge to an optimal K when complex genetic structures exist, such as strong relatedness within some germplasm families. The ESWYT lines, with an

intensive breeding history, include a large number of sister and closely related lines along with introgressions of unrelated materials, which may explain the lack of stability in the structure analyses. For association mapping, the determination of population structure avoids identifying spurious associations between phenotype and genotype (Pritchard et al. 2000). Subsequent implementation of whole-genome association mapping will need to take into account the strong relatedness of lines within the sub-populations via kinship or pedigree estimation, as suggested by Yu et al. (2006).

Temporal genetic diversity

To date, only a few studies have applied DArT markers to the study of genetic diversity in wheat. White et al. (2008) previously specified the limitations of potential statistical analyses based on the biallelic nature of DArT, which, however, are compensated by the large number of markers that allow precise average values to be obtained. Equivalent to the research investigations of White et al. (2008) and Stodart et al. (2007), the DArT markers provided a suitable genotyping platform for determining genetic structures in wheat in this study.

The overall level of genetic diversity in the ESWYT was reasonably high and within the range of reported UK, US, and Australian wheat cultivars released from 1930 to 2005 (White et al. 2008). Average genetic distance in the ESWYT was higher (0.335) than in the 94 UK cultivars (0.298), but lower than in the 96 US (0.349) and 30 Australian cultivars (0.405). Unsurprising, the level of genetic diversity in the ESWYT was lower than that observed in 44 landrace accessions derived from 25 different countries and held in the Vavilov Center of Diversity (Stodart et al. 2007). Genetic diversity could be compared via the average PIC values of DArT markers, yielding values of 0.270 in the ESWYT and 0.420 in the landrace germplasm respectively.

The level of genetic diversity in CIMMYT germplasm has been determined in earlier studies (Dreisigacker et al. 2004; Reif et al. 2005). CIMMYT's wheat-breeding program traditionally aimed to increase the genetic diversity by taking into account the necessity for diversification, environmental protection and durable resistance, while building on the wide geographic adaptation of the germplasm achieved in previous decades (Reeves et al. 1999). To reach this goal, spring and winter wheat germplasm from different regions, exotic germplasm such as Chinese or Brazilian wheat cultivars, and landraces from different regions, as well as wild relatives such as *Agropyrum* derivatives, were introgressed into the breeding pool.

It is a public concern that modern plant breeding introduces crop uniformity, thereby jeopardizing future selection gain and increasing the genetic vulnerability to biotic and abiotic stresses (Harlan 1972). Such concerns have a theoretical basis, as modern plant breeding usually introduces intensive selection within a narrow range of plant germplasm with limited allele introgression over time (Allard 1999). The level of genetic diversity across the time period of 25 years of ESWYT dissemination was stable and thus did not support the concern that modern plant breeding leads to a reduction in available diversity (Fig. 2). These results were in agreement with the study of van de Wouw et al. (2010), which demonstrated with meta-analytical techniques that overall, in the long run, no substantial reduction has taken place in the regional diversity of crop varieties released by plant breeders. The candidates for each ESWYT trial are annually nominated on the basis of performance in preliminary yield trials in Mexico. Fluctuations of average genetic distances across the ESWYT were evident. Factors to consider which influenced this diversity flux are (i) the access to new germplasm; (ii) the continuous introgression of new germplasm into the current gene pool and its evolution during the selection process; (iii) changes of personnel and breeding schemes in the program; and (iv) the current generic and local targets of the germplasm. The latter is a crucial factor for CIMMYT breeders, who must determine the optimum distribution of effort between developing germplasm with generically valuable traits for distribution to broad mega-environments, breeding cultivars that also encompass local-adaptive traits, and disseminating sufficiently diverse cultivars such that more agro-ecological niches are reached and genetic vulnerability to disease and pests is reduced. The more comprehensive the feedback from the international testing network, the better targeted new germplasm will become in terms of both generic and local-adaptive traits.

Changes in allele frequencies of marker loci over time

An increasing average genetic distance between ESWYT with increasing time between the year of dissemination suggested a genetic shift from early to recent years of ESWYT based on a systematic change in allele frequencies during the breeding processes (Fig. 3). It can be conjectured that selection is the main factor contributing to the allelic shifts in the ESWYT population, increasing the frequency of favorable alleles and decreasing the frequency of those unfavorable for the traits under selection. Over the 25 year time period of ESWYT dissemination, the allele frequencies of 17.9% of the 435 mapped DArT markers have changed significantly. Interestingly, nearly half of these alleles were associated with grain yield and rust resistance in previous studies of Crossa et al. (2007, 2010), indicating continuous selection pressure towards those traits and warranting further investigation.

Genomic regions associated with grain yield in this and in at least two other cited publications have been identified in chromosomes 2AL, 2DS and 7BL (Table 2). Li et al. (2007) and Kuchel et al. (2007) described minor QTL for spike number and time to ear-emergence in the same region of chromosome 2AL as that examined here. The markers wPt-2644, wPt-4144 and wPt-3728 were moderately linked to the microsatellite Xgwm261 associated with the well known QTL of the photoperiod response gene Ppd-D1 and plant height gene Rht8 located in chromosome 2DS (Sourdille et al. 2000). Ppd-D1 is one of the most potent genes affecting the photoperiod response of wheat. Since the 1970 s, the Ppd-D1a allele is present in nearly all CIMMYT wheat lines, which is why it is rather unlikely that the QTL found with an increasing/decreasing frequency in the ESWYT population is equivalent. A strong QTL effect for grain number per head was observed on chromosome 7BL by Quarrie et al. (2005) in the same region as the DArT marker wPt-3939, indicative of the presence of the same QTL.

Supplementary genomic regions were associated with grain yield and leaf, yellow or stem rust on chromosomes 1AS, 1BS, 2BL, 3AS, 3BS, 4AL, 5BS/L and 7BL (Table 2). No QTL for grain yield and rust have been previously observed on chromosome 1AS, whereas on chromosome 1BS QTL effects are known to be detected on the basis of the 1B/1R translocation. The marker wpt-9350 was closely linked to the two allelic yellow rust resistance genes Yr5 and Yr7 on chromosome 2BL. The gene Yr39 was mapped on chromosome 7BL. Yr5 and Yr39 provide resistance to all known races of yellow rust and are present in the CIMMYT germplasm pool. No yellow rust QTL have yet been reported on chromosomes 3AS, 4AL and 5BS. A QTL for anthesis date was detected on chromosome 3AS, explaining significant variation for kernel number, kernel weight and plant height (Dil-birligi et al. 2006). A QTL for spikelets per spike was mapped by Kumar et al. (2007) on chromosome 4AL in the same region as in this study. Kuchel et al. (2007) and Groos et al. (2003) identified QTL associated with the yield components kernel number, kernel size and heads per plant in the same region of chromosome 5BS as examined here. The QTL observed on chromosome 5BL is possibly linked to the leaf rust gene Lr18, but it is not clear how frequently this gene is present in the ESWYT. It can also be suspected that this QTL represents the vernalization gene Vrn-B1. The presence of the Vrn-B1 gene has been characterized in CIMMYT international nursery germplasm targeted to semi-arid regions in the world. The data showed strong selection pressure towards Vrn-B1 and Vrn-D1 (unpublished data). It is likely that selection in the ESWYT was directed towards the same genes.

Whole-genome LD among different sub-populations

The size of the ESWYT population used for LD analyses in this study was larger than in any previous studies in wheat. The population represents elite lines from a global but single breeding program, and thus LD was expected to be relatively high (Breseghello and Sorrells 2006).

Association results of DArT marker pairs pointed to the presence of extensive LD, both in the entire ESWYT population and within each of the five main sub-populations. Equivalent to earlier studies, our analyses observed decreasing LD with increasing genetic distance, an indication of LD maintained by genetic linkage. The LD decline ($\sim$30 cM) was faster in comparison to the study of Crossa et al. (2007) using a subset of lines of the same ESWYT population, but slower than in previously published whole genome studies in wheat which estimated an
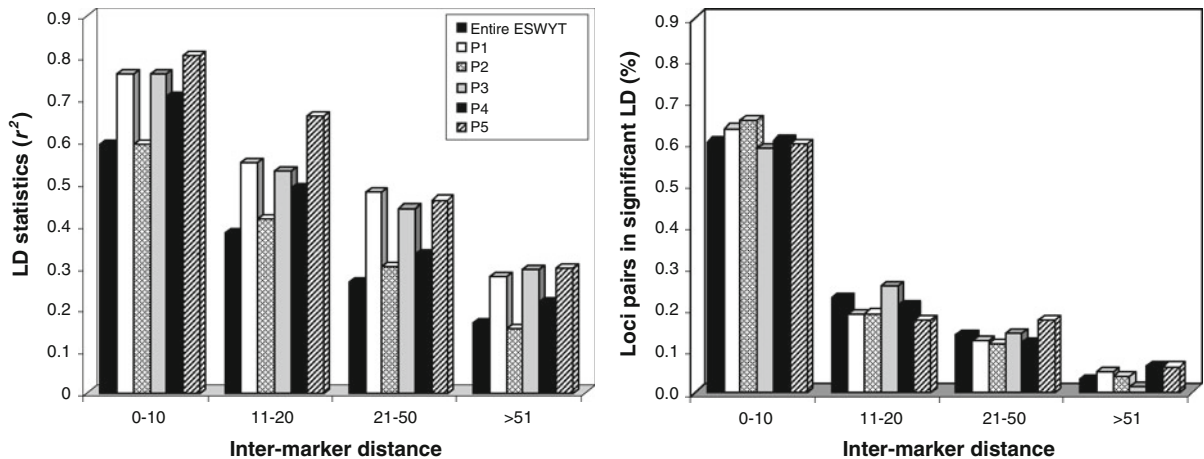
Fig. 4 Percentage and average LD between adjacent DArT markers ($r^2 > 0.1$, $P < 0.001$). Pair-wise DArT markers have been classified based on inter-marker genetic distances revealed using the integrated map in Crossa et al. (2007) as follows: tightly to moderately linked (0–10 and 11–20 cM apart), loosely linked (21–50 cM), and those with a distance >50 cM. LD analysis was performed across the entire ESWYT population and within each of the five sub-populations (coded as P1, P2, P3, P4, P5) identified according to STRUCTURE analyses

LD decline within a average distance of 10–20 cM (Chao et al. 2007; Somers et al. 2007). However, the majority of loci pairs in LD were observed within 10 cM, which suggests that association mapping with DArT can be an effective method for QTL identification and validation in these high-level LD regions. Low LD observed among loci pairs in this study may also be partly due to less information available on map positions of DarT markers, the map used being one of the first consensus maps for DArT available (Crossa et al. 2007). Further, actual LD may differ among populations and may need to be evaluated for each population on a case-by-case basis (Breseghello and Sorrells 2006).

Subdividing the germplasm into sub-populations has been applied in LD analyses to reduce the number of spurious associations found in independent loci pairs. A decrease of loci pairs in LD accompanying the subdivision of the entire population is therefore expected. Somers et al. (2007) observed a 16-fold decrease of loci pairs in LD forming sub-populations of 16–44 members, with elevated r^2 values in each sub-population. In our study the number of loci pairs in significant LD was highly variable among sub-populations, but not substantially reduced. Remington et al. (2001) argued that the observed drop in LD significance by forming sub-populations could be partly due to a reduced power of the LD test in the smaller populations. Our study supports this argument. The largest number of lines in the sub-populations was 113 and was thus comparably high, precluding extensive LD decline due to small population size. Elevated LD in sub-populations can also occur with narrow genetic differentiation. By means of F_{ST} values and average genetic distance, genetic differentiation of our sub-populations showed good correlation with the average values of LD, suggesting that the sub-populations themselves explained much of the LD. The highest F_{ST} as well as average r^2 value was observed in sub-population P5, which includes closely related lines derived from the parent Kauz. Large chromosome segments may be shared between the line, resulting in a non-random distribution of haplotypes and leading to elevated LD values. The lack of reduction of loci pairs in significant LD might thus be due to the residual substructure within each sub-population.

References

Akbari M, Wenzl P, Vanessa C, Carling J, Xia L, Yang S, Uszynski G, Mohler V, Lehmensiek A, Kuchel H, Hayden MJ, Howes N, Sharp P, Rathmell B, Vaughan P, Huttner

E, Kilian A (2006) Diversity arrays technology (DArT) for high-throughput profiling of the hexaploid wheat genome. Theor Appl Genet 113:1409–1420

Allard RW (1999) History of plant population genetics. Annu Rev Genet 33:1–37

Anderson JA, Churchill GA, Sutrique JE, Tanksley SD, Sorrells ME (1993) Optimizing parental selection for genetic linkage maps. Genome 36:181–186

Arief VN, Kroonenberg PM, DeLacy IH, Dieters MJ, Crossa J, Dreisigacker S, Braun H-J, Basford KE (2010) Construction and three-way ordination of the Wheat Phenome Atlas. J Appl Prob Stat 5:95–118

Balfourier F, Roussel V, Strelchenko P, Exbrayat-Vinson F, Sourdille P, Boutet G, Koenig J, Ravel C, Mitrofanova O, Beckert M, Charmet G (2007) A worldwide bread wheat core collection arrayed in a 384-well plate. Theor Appl Genet 114:1265–1275

Braun H-J, Pfeiffer WH, Pollmer WG (1992) Environments for selecting widely adapted spring wheat. Crop Sci 32: 1420–1427

Breseghello F, Sorrells ME (2006) Association analysis as a strategy for improvement of quantitative traits in plants. Crop Sci 46:1323–1330

Byerlee D, Dubin HJ (2010) Crop improvement in the CGIAR as a global success story of open access and international collaboration. Int J Commons 4(1):452–480

Byth DE, Eisemann RL, DeLacy IH (1976) Two-way pattern analysis of a large data set to evaluate genotypic adaptation. Heredity 37:215–230

Camus-Kulandaivelu L, Veyieras JB, Gouesnard B, Charcosset A, Manicacci D (2007) Evaluating the reliability of structure outputs in case of relatedness between individuals. Crop Sci 47:887–892

Chao S, Zhang W, Dubcovsky J, Sorrells ME (2007) Evaluation of genetic diversity and genome-wide linkage disequilibrium among U.S. wheat (Triticum aestivum L.) germplasm representing different market classes. Crop Sci 47:1018–1030

Crossa J, Burgueno J, Dreisigacker S, Vargas M, Herrera-Foessel SA, Lillemo M, Singh RP, Trethowan RM, Warburton ML, Franco J, Reynolds M, Couch JH, Ortiz R (2007) Association analysis of historical bread wheat germplasm using additive genetic covariance of relatives and population structure. Genetics 177:1889–1913

Crossa J, de Los Campos G, Perez P, Pianola D, Bargueño J, Araus JL, Makumbi D, Dreisigacker S, Yan J, Arief VN, Banziger M, Braun H-J (2010) Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. Genetics 186:713–724

DeLacy IH, Fox PN, Corbett JD, Crossa J, Rajaram S, Fischer RA, van Ginkel M (1993) Long-term association of locations for testing spring bread wheat. Euphytica 72: 95–106

Dilbirligi M, Erayman M, Campbell BT, Randhawa HS, Baenzinger PS (2006) High density mapping and comparative analysis of agronomically important traits on wheat chromosome 3A. Genomics 88:74–87

Dreisigacker S, Zhang P, Warburton ML, Van Ginkel M, Hoisington D, Bohn M, Melchinger AE (2004) SSR and pedigree analysis of genetic diversity among CIMMYT

wheat lines targeted to different mega environments. Crop Sci 44:381–388

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620

Fisher RA (1922) On the interpretation of χ^2 from contingency tables, and the calculation of P. J R Stat Soc 85:87–94

Groos C, Robert N, Bervas E, Charmet G (2003) Genetic analysis of grain-protein content, GY, and thousand-kernel weight in bread wheat. Theor Appl Genet 106:1032–1040

Harlan JR (1972) Genetics of disaster. J Environ Qual 1:212–215

Heisey PW, Lantican MA, Dubin HJ (2002) Impacts of international wheat breeding research in developing countries, 1966–1997. CIMMYT, Mexico, DF

Kuchel H, Williams KJ, Langridge P, Eagles HA, Jefferies SP (2007) Genetic dissection of grain yield in bread wheat. I. QTL analysis. Theor Appl Genet 115:1029–1041

Kumar N, Kulwal PL, Balyan HS, Gupta PK (2007) QTL mapping for yield and yield contribution traits in two mapping populations of bread wheat. Mol Breed 19: 163–177

Li S, Jia J, Wei X, Zhang X, Li L, Chen H, Fan Y, Sun H, Zhao X, Lei T, Xu Y, Jiang F, Wang H, Li L (2007) A intervarietal genetic map and QTL anaysis for yield traits in wheat. Mol Breed 20:167–178

Lillermo M, Van Ginkel M, Trethowan RM, Hernandez E, Crossa J (2005) Differential adaptation of CIMMYT bread wheat to global high temperature environments. Crop Sci 45:2443–2453

Lui K, Muse SU (2005) Power marker: integrated analysis environment for genetic marker data. Bioinformatics 21: 2128–2129

Maccaferri M, Sanguineti MC, Noli E, Tuberosa R (2005) Population structure and long-range linkage disequilibrium in a durum wheat elite collection. Mol Breed 15: 271–289

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829

Neumann K, Kobiljski B, Denčić S, Varshney RK, Börner A (2010) Genome-wide association mapping: a case study in bread wheat (Triticum aestivum L.). Mol Breed 27:35–58. doi:10.1007/s11032-010-9411-7

Pritchard JK, Stephen M, Donnely P (2000) Inference on population structure using multilocus genotype data. Genetics 155:945–959

Quarrie SA, Steed A, Calestani C, Semikhodskii A, Lebreton C, Chinoy C, Steele N, Pljevljakusic D, Waterman E, Weyen J. Schondelmaier J, Habash DZ, Farmer P, Saker L, Clarkson DT, Abugalieva A, Yessimekova M, Turuspekov T, Abugalieva S, Tuberosa R, Sanguineti MC, Hollington PA, Aragues R, Royo A, Dodig D (2005) A high density genetic map of hexaploid wheat (Triticum aestivum L.) from the cross Chinese Spring × SQ1 and its use to compare QTLs for grain yield across a range of environments. Theor Appl Genet 110:865–880

Rajaram S, Mann CHE, Ortiz-Ferrara G, Mujeeb-Kazi A (1983) Adaptation, stability, and high yield potential of certain 1B/1R CIMMYT wheats. In: Sakamoto S (eds)

Proceedings of 6th international wheat genetics symposium, Kyoto, Japan. pp 613–621

Rajaram S, Van Ginkel M, Fischer RA (1993) CIMMYT's wheat breeding mega-environments (ME). International wheat genetics symposium 8, Beijing, China, 20–25 July 1993

Rajaram S, Borlaug NE, Van Ginkel M (2002) CIMMYT international wheat breeding. In: Curtis BC, Rajaram S, Gomez Macpherson H (eds) Bread wheat improvement and production. FAO, Rome, pp 103–117

Reeves T, Rajaram S, Van Ginkel M, Trethowan R, Braun H-J, Cassaday K (1999) New wheats for a secure, sustainable future. CIMMT, Mexico, DF

Reif JC, Zhang P, Dreisigacker S, Warburton ML, Van Ginkel M, Hoisington D, Bohn M, Melchinger AE (2005) Wheat genetic diversity trends during domestication and breeding. Theor Appl Genet 110:859–864

Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler ES (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc Natl Acad Sci USA 98:11479–11484

Rogers JS (1972) Studies in genetics V II. University of Texas Publication, Austin, pp 145–153

Saghai-Maroof MA, Soliman K, Jorgensen RA, Allard RW (1984) Ribosomal DNA spacer length polymorphism in barley: Mendelian inheritance, chromosomal location and population dynamics. Proc Natl Acad Sci USA 81: 8014–8018

Singh RP, Huerta-Espino J, Rajaram S, Crossa J (1998) Agronomic effects from chromosome translocations 7DL.7Ag and 1Bl.1RS in spring wheat. Crop Sci 38: 27–33

Somers DJ, Banks T, DePauw R, Fox S, Clarke J, Pozniak C, McCartney C (2007) Genome-wide linkage disequilibrium analysis in bread wheat and durum wheat. Genome 50:557–567

Sourdille P, Snape JW, Cadalen T, Charmet G, Nakata N, Bernard S, Bernard M (2000) Detection of QTLs for heading time and photoperiod response in wheat using a doubled-haploid population. Genome 43:487–494

Stodart BJ, Mackay MC, Raman H (2007) Assessment of molecular diversity in landraces of bread wheat (*Triticum aestivum* L.) held in an ex-situ collection with Diversity Arrays Technology (DArT^{TM}). Austr J Agron Res 58:1174–1182

Trethowan RM, Crossa J (2007) Lessons learnt from forty years of international spring bread wheat trials. Euphytica 157:385–390

Trethowan RM, Van Ginkel M, Rajaram S (2002) Progress in breeding wheat for yield and adaptation in global drought affected environments. Crop Sci 42:1441–1446

Trethowan RM, Van Ginkel M, Ammar K, Crossa J, Payne PS, Cucadar B, Rajaram S, Hernandez E (2003) Associations among twenty years of bread wheat evaluation environments. Crop Sci 43:1698–1711

Van de Wouw M, Van Hintum T, Kik C, Van Treuren R, Visser B (2010) Genetic diversity trends in twentieth century crop cultivars: a meta analysis. Theor Appl Genet 120:1241–1252

Weir BS (1996) Genetic data analysis II. Sinauer Associated, Inc

White J, Law JR, MacKay I, Chalmers KJ, Smith JSC, Kilian A, Powell W (2008) The genetic diversity of UK, USA and Australian cultivars of *Triticum aestivum* measured by DArT markers and considered by genome. Theor App Genet 116:439–453

Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2006) A unified mixed-model method for association mapping that counts for multiple level of relatedness. Nat Genet 28:203–208

Zhou Y, He ZH, Liu JJ, Liu L (2003) Distribution of 1BL/1RS translocation in Chinese winter wheat and its effect on noodle quality. In: Pogna NE, Romano M, Pogna E, Galterio G (eds) Proceedings 10th international wheat genetics symposium, Paestum Italy, vol 3, pp 1419–1421