Theor Appl Genet (2009) 119:1523–1537
DOI 10.1007/s00122-009-1153-8

ORIGINAL PAPER

Analysis of diversity and linkage disequilibrium along chromosome 3B of bread wheat (*Triticum aestivum* L.)

Aniko Horvath · Audrey Didier · Jean Koenig · Florence Exbrayat · Gilles Charmet · François Balfourier

Received: 3 April 2009 / Accepted: 30 August 2009 / Published online: 16 September 2009

Abstract A highly polymorphic core collection of bread wheat and a more narrow-based breeding material, gathered from pedigrees of seven modern cultivars, was analysed in order to compare genetic diversity indices and linkage disequilibrium (LD) patterns along the chromosome 3B with microsatellite (SSR) and Diversity Arrays Technology markers. Five ancestral gene pools could be identified within the core collection, indicating a strong geographical structure (Northwest Europe, Southeast Europe, CIMMYT–ICARDA group, Asia, Nepal). The breeding material showed a temporal structure, corresponding to different periods of breeding programmes [old varieties (from old landraces to 1919), semi-modern varieties (1920–1959), modern varieties (1960–2006)]. Basic statistics showed a higher genetic diversity in the core collection than in the breeding material, indicating a stronger selection pressure in this latter material. More generally, the chromosome 3B had a lower diversity than the whole B-genome. LD was weak in all studied materials. Amongst geographical groups, the CIMMYT–ICARDA pool presented the longest ranged LD in contrast to Asian accessions. In the breeding material, LD increased from old cultivars to modern varieties. Genitors of seven modern cultivars were found to be different; most marker pairs in significant LD were observed amongst genitors of Alexandre and Koreli varieties, indicating an important inbreeding effect. At low genetic distances (0–5 cM), the breeding material had higher LD than the core collection, but globally the two materials had similar values in all classes. Marker pairs in significant LD are generally observed around the centromere in both arms and at distal position on the short arm of the chromosome 3B.

Communicated by H. H. Geiger.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-009-1153-8) contains supplementary material, which is available to authorized users.

A. Horvath · A. Didier · J. Koenig · F. Exbrayat · G. Charmet · F. Balfourier (✉)
INRA, UMR1095 Genetics,
Diversity and Ecophysiology of Cereals,
63100 Clermont-Ferrand, France
e-mail: balfour@clermont.inra.fr

Present Address:
A. Horvath
INRA, UR419 Fruit Species,
33883 Villenave d'Ornon, France

Introduction

Bread wheat (*Triticum aestivum* L.) is one of the most cultivated crops in the world. It covers the largest surface amongst cereals (216 million ha) worldwide. The production takes the third place, after maize and rice, with 605 million tons (FAOSTAT 2008). Bread wheat belongs to the *Triticum* genus of the large *Poaceae* family, which gathers about 3,500 species (Bonnier and Douin 1990). Bread wheat is an autogamous, allohexaploid species ($2n = 6x = 42$), with seven groups of chromosomes. Each group is constituted by three homoeologous chromosomes belonging to the A, B and D genomes (AABBDD) (Gill et al. 2004).

Numerous molecular markers (isozymes, AFLP, RFLP, RAPD) have been successfully used to assess genetic diversity amongst cereal crops (Liu et al. 2002; Ozkan et al. 2002; Tanyolac et al. 2003). Recently, simple sequence repeat (SSR) or microsatellites were applied to assess diversity in French and European wheat germplasms (Roussel et al. 2004, 2005) and Diversity Arrays

Technology (DArT) to describe diversity of barley (Wenzl et al. 2004) and of bread wheat cultivars (White et al. 2008).

The commonly used molecular techniques have been presented and compared recently in a review (Able et al. 2007). These previous studies have demonstrated a high level of diversity in different cereal crops. However, this overall genetic diversity is structured, depending on the plant material studied. For example, the diversity of European bread wheat varieties is influenced both by temporal and geographical variations (Roussel et al. 2005). Linkage disequilibrium (LD) studies on plant genomes are important in order to identify genetic regions associated with agronomical traits (Rafalski and Morgante 2004). Most of the demographic events observed in population genetics affect LD pattern, since it is influenced by allele frequency, recombination rate between sites, mating system and admixture level of investigated populations. Thus, LD can be increased in populations undergoing a bottleneck event, followed by a genetic drift (Flint-Garcia et al. 2003). This scenario is often observed in cultivated crops as, e.g., in maize, where selection for adaptive traits may have created LD in regions covering major genes for these traits (Remington et al. 2001).

Recently, genome-wide LD analysis has been performed on different cereal crops, showing that the LD is depended on the population structure (Chao et al. 2007; Maccaferri et al. 2005; Malysheva-Otto et al. 2006; Somers et al. 2007). Similarly, patterns of intragenic LD revealed differences in a worldwide set of cultivars in barley, depending on the geographical origin of variety subsets (Malysheva-Otto and Roder 2006). When comparing barley and wheat accessions for genes coding for the transcription factor GAMYB, barley was found to have a moderate level of LD whilst wheat was shown to have an absolute LD between polymorphic sites, decreasing within the gene (Haseneyer et al. 2008). The usefulness of LD studies has been recently demonstrated in a paper describing the association mapping of kernel size and milling quality in wheat cultivars (Breseghello and Sorrels 2006).

Here, we focused on chromosome 3B which is the largest amongst bread wheat chromosomes with 995 Mb (Lee et al. 2004; Paux et al. 2008b), and contains numerous repetitive elements (Paux et al. 2006). Chromosome 3B can be easily isolated by flow cytometry (Vrana et al. 2000). Amongst the homoeologous group 3 chromosomes of wheat genome, the most expressed sequence tags were reported to be located onto chromosome 3B (Qi et al. 2004) and it was estimated that it contains 6,000 genes based upon the analysis of 11 Mb of BAC-end sequences (Paux et al. 2006). Numerous quantitative trait loci (QTLs) have been already identified, notably concerning resistance to pre-harvest sprouting (Groos et al. 2002), to *Fusarium* head blight (Ma et al. 2006; Shen et al. 2006), to kernel shattering (Zhang and Mergoum 2007) and QTLs implicated in tolerance to nitrogen deficiency (Laperche et al. 2007). A cytological map of chromosome 3B has been established thanks to deletion lines (Endo and Gill 1996) and a physical map has been published (Paux et al. 2008a, b).

Studies of wheat genetic diversity preserved in germplasm collections are necessary to highlight combinations of new alleles which can be useful for plant breeding activity. However, intensive breeding practices using related genetic material may lead over the time to an important loss in diversity and modify both the level and the pattern of LD (Flint-Garcia et al. 2003).

With the help of microsatellite (SSR) and DArT markers, the objectives of the present paper are to investigate the genetic diversity and to analyse the LD patterns along the chromosome 3B in two different, contrasted genetic pools of bread wheat: (1) a highly polymorphic worldwide core collection which maximises overall genome diversity and is geographically structured and (2) a breeding material containing several modern inbred varieties and their genitors which are related accessions mainly originated from Europe.

Materials and methods

Plant material

Two contrasting plant materials were sampled for the present study.

The bread wheat core collection used in the present study is a sample representative of worldwide wheat diversity. The methodology used to build the core collection and its composition was described in Balfourier et al. (2007). Briefly, this collection is made up of 372 accessions which capture more than 98% of the whole allelic diversity observed at 38 SSR polymorphic loci in a sample of 4,000 accessions; the core includes wheat accessions from 70 different countries. Accessions are either landraces from the nineteenth century or cultivars from throughout the twentieth century.

The breeding material was composed of seven modern French cultivars of bread wheat (cv. Alexandre, Apache, Isengrain, Koreli, Oracle, Ornicar, Taldor) and their genitors were also included. Genealogical trees of modern cultivars were determined with the help of the online Wheat Pedigree Database (http://genbank.vurv.cz/wheat/pedigree/) maintained by the Crop Research Institute VURV (Praha, Czech Republic) and with the ERGE Database of the Cereal Genetic Resources Collection of

INRA (http://www1.clermont.inra.fr/umr1095). In total, 359 individuals participated in the breeding of these seven modern varieties. Amongst the 359 individuals, seeds from 235 genitors were used, i.e. 65.45% of the total material. The reasons for the missing plant material were as follows: (1) intermediate inbred lines were not conserved in breeding programmes and (2) the names of the genitors in the databases were uncertain, mainly for landraces and wild genitors. Modern cultivars, registration year, breeder's name, number of genitors and available material are shown in Table 1. The available 242 individuals (i.e. 235 genitors and 7 elite cultivars) have been conserved or introduced into the Clermont-Ferrand Cereals Genetic Resources Centre (http://www1.clermont.inra.fr/umr1095) and they are part of the cereal crop germplasm collection of INRA. They form the second sample used and named "breeding material" in the present study.

Genitors have been distributed into seven groups, according to their registration or cultivation period. Temporal groups and accession number are given in Table 2.

DNA extraction

Seeds used for DNA extraction were collected from self-pollinated ears. Leaves of five plants per accession were pooled and genomic DNA was extracted from 100 mg frozen leaves with the kit Biosprint 96 using Biosprint workstation (Qiagen). DNA quality was checked by electrophoresis onto 0.8% agarose gel and DNA concentration was determined with a NanoDrop® ND-1000 UV-Vis Spectrophotometer (NanoDrop Technologies).

DArT genotyping

The 372 accessions of the core collection were genotyped by DArT markers. DArT fingerprinting was performed at the Diversity Arrays Technology Pty Limited (Canberra, Australia; http://triticarte.com.au) with the panel Wheat *Pst*I(*Taq*I) v2.3. Polymorphism scoring was done as described in Wenzl et al. (2004).

Table 2 Temporal groups and number of accessions per group

Group	Period	Number of accessions
1	Landraces to 1919	74
2	1920–1944	52
3	1945–1959	41
4	1960–1969	34
5	1970–1979	16
6	1980–1989	14
7	1990–2006	11

Microsatellite analysis

All the accessions (372 + 242) were genotyped with 62 SSR markers, mapped on the chromosome 3B (http://wheat.pw.usda.gov/GG2/index.shtml) of wheat. Six BARC (Gupta et al. 2002), 1 CFA (Sourdille et al. 2001), 7 CFE (Zhang et al. 2005), 12 CFB, 13 GPW, 6 GWM (Roder et al. 1998), 1 NW, 10 WMC (Somers et al. 2004) and 6 WMM SSR markers were tested. Further information concerning wheat microsatellite markers is available in the GrainGenes 2.0 database (http://wheat.usda.gov/GG2/index.shtml).

PCR reactions were realised as follows: 1× PCR buffer with $MgCl_2$ (Qiagen), 207.5 mM of each dNTP, 50 nM of M13-marked forward primer, 500 nM of reverse primer, 523 nM fluorochrom, 0.031 U of Taq Polymerase (Qiagen) and 3.8 ng of genomic DNA in a 6.5 µl final volume.

PCR amplification conditions were similar as described in Roussel et al. (2004, 2005), with some modifications: (1) initial denaturation (5 min at 95°C), 30 cycles of 30 s at 95°C for denaturation, 30 s at 60°C for annealing and 30 s at 72°C for extension, followed by 8 cycles of 30 s at 95°C, 30 s at 56°C and 30 s at 72°C, then final extension (5 min at 72°C); (2) initial denaturation (5 min at 95°C) followed by 7 cycles of 30 s at 95°C, 30 s at 62°C with touchdown −1°C in each cycle and 30 s at 72°C, then 20 cycles of 30 s at 95°C, 30 s at 55°C, 30 s at 72°C, followed by 8 cycles of 30 s at 95°C, 30 s at 56°C, 30 s at 72°C, then a

Table 1 List of studied bread wheat modern cultivars, genitors' number and inbreeding coefficients

Modern cultivar	Registration year	Breeder	Genitors total number	Available genitors	IBC
Apache	1997	Nickerson	110	71 (64.55%)	0.017
Alexandre	1997	Nickerson	188	127 (67.55%)	0.024
Isengrain	1996	Desprez	122	86 (70.49%)	0.012
Koreli	2006	INRA	194	133 (68.56%)	0.033
Oracle	1994	Benoist	159	111 (69.81%)	0.022
Ornicar	1997	Benoist	131	82 (62.60%)	0.015
Taldor	1997	GAE	98	70 (71.43%)	0.003

IBC inbreeding coefficient

final extension step for 5 min at 72°C; (3) the same as (2) but 30 cycles instead of 20.

Fragment analysis was performed using an ABI PRISM 3100 Genetic Analyser (Applied Biosystems) and GENESCAN/GENOTYPER software as described in Roussel et al. (2004, 2005).

Data analysis

Genetic structure of the worldwide core collection was analysed based on DArT genotyping, with the help of the STRUCTURE software (Pritchard et al. 2000). The procedure places individuals into K clusters. For each run, we set a 10^5 iteration burn-in period and a 10^5 iteration sampling period and we perform five independent runs of structure for a number of cluster K varying from 2 to 10, leading to 45 structure outputs. Then, we estimated the number of subpopulations and the best output on the basis of Evanno criterion (Evanno et al. 2005) and, to a limited extent, on the basis of consistency of group composition based on a priori knowledge of genetic and geographic origins.

A principal coordinates analysis (PCoA) was also performed on the core collection, based on Sokal and Michener dissimilarity index (= simple matching coefficient) computed from DArT markers in DARwin software (Perrier and Jacquemoud-Collet 2006).

Basic statistics of genetic diversity, i.e. total number of alleles and allele frequencies per SSR locus, were calculated using the GENETIX software (Belkhir et al. 2004). Polymorphic information content (PIC) was calculated according to the formula $H = 1 - \sum p_i^2$ (Nei 1973), where p_i means the frequency of the ith allele. To compare groups with different sample sizes, the allele rarefaction method is recommended in order to standardise allelic richness of samples (Petit et al. 1998).

Temporal structure of breeding material was analysed amongst the seven temporal groups. Unbiased genetic distance (Nei 1978) was calculated with GENETIX software (Belkhir et al. 2004), then UPGMA (Unweighted Pair Group Method with Arithmetic mean) dendrogram was constructed in the PHYLIP program package (Felsenstein 2004), visualised by TREEVIEW (Page 1996). In order to compare allelic richness of seven temporal groups, allele rarefaction method (Petit et al. 1998) was used owing to the different sample sizes. Inbreeding coefficient (IBC) within pedigree structures of seven modern varieties was calculated using the circular edge walking method in Peditree software (van Berloo and Hutten 2005).

LD between markers was calculated by TASSEL 2.0.1 software (Zhang et al. 2006). As SSRs are multiallelic markers, a weighted average is suggested to estimate D' and r^2 values (Farnir et al. 2000). In our analyses, the r^2 correlation index between alleles at loci was chosen, as the index most frequently used in the literature.

Data file was filtered for rare alleles (a percentage less than 5% in the whole population), thus LD values cannot be biased by low allele frequencies. 100,000 permutations have been realised in order to determine the significance with the two-sided Fisher's exact test. p values have not been computed with the "rapid permutations" option, since this latter slightly reduces them (Zhang et al. 2006).

For every comparison between the large core collection and groups having a smaller sample size, we processed a random re-sampling of the core collection, i.e. 242 and 85 individuals, when comparing it to the breeding material and to geographical structures, respectively. The random re-sampling was realised 19 times with the sample algorithm of the base package in R language and environment for statistical computing (R Development Core Team 2006), then manual verification was realised on the 19 random samples. Thus, the highest (lowest) value out of the 19 obtained for a given parameter gives a rough estimate of its 95% confidence highest (lowest bound). For practical reasons, it was not possible to carry out hundreds of simulations runs which would have given more accurate results.

In order to compare the number of marker pairs in significant LD in different classes of genetic distance (0–5 cM), a genetic distance matrix between marker pairs was established using information on consensus map (Paux et al. 2008b) and the maximum distance algorithm of the stats package of R language and environment for statistical computing (R Development Core Team 2006).

Results

Genetic diversity

Global diversity

The 62 SSRs gave comparable results with few missing data in both sets of plant materials. Allele number per marker and PIC are presented in Table 3. A total of 595 polymorphic alleles were detected. The total number of alleles ranged from 2 to 35 for the core collection and from 2 to 29 for the breeding material. PIC values ranged from 0.010 to 0.936 in the core collection and from 0.017 to 0.897 in the breeding material. Based on PIC values of individual markers, the less informative was the marker cfe129 both in the core collection and in the breeding material. The more informative markers were not the same for the core collection and the breeding material: wmm1758 and wmm1966, respectively. Mean PIC values were similar in both material (0.553 and 0.521). Using the

Table 3 Microsatellite loci in the order of their position on the chromosome 3B, allele number and polymorphic information content (PIC) values in the core collection and breeding material

Locus	Allele number		PIC	
	Core collection	Breeding material	Core collection	Breeding material
gwm389	22	15	0.891	0.868
barc075	9	5	0.581	0.546
gpw7774	2	3	0.372	0.384
nw1644	11	9	0.798	0.765
cfb3530	6	5	0.492	0.511
cfb3417	5	5	0.624	0.549
barc147	13	9	0.522	0.241
gwm493	24	19	0.730	0.570
gpw7031	2	2	0.099	0.044
gpw3248	20	14	0.904	0.894
barc092	5	3	0.238	0.198
gpw4146	8	6	0.505	0.602
wmc078	15	15	0.816	0.735
wmc43	7	7	0.336	0.278
wmc231	12	6	0.786	0.647
wmc418(1)	4	3	0.361	0.096
gwm566	14	8	0.831	0.778
cfb3260	8	7	0.668	0.508
gpw7452	11	8	0.657	0.630
wmc540	7	3	0.271	0.195
gwm284	7	7	0.764	0.746
cfb3481	3	2	0.494	0.373
cfb3266	8	5	0.651	0.625
cfb3328	3	3	0.137	0.239
gpw4207	9	7	0.424	0.450
gpw3134	11	9	0.761	0.720
gpw1145	7	4	0.182	0.121
wmc527	17	10	0.766	0.691
gpw1107	10	7	0.726	0.536
cfe278	4	6	0.406	0.530
cfb3330	13	13	0.819	0.678
barc164	22	16	0.827	0.811
wmm1002	6	5	0.673	0.677
wmc418(2)	8	8	0.696	0.685
gpw4034	12	14	0.613	0.541
cfe326	2	2	0.497	0.451
cfe294	2	2	0.035	0.154
wmm1703	7	7	0.720	0.673
cfb053	3	3	0.494	0.503
cfb3063	4	4	0.518	0.562
gpw8064	2	2	0.409	0.466
cfb3059	9	7	0.577	0.544
gpw8144	2	3	0.011	0.079
wmc291	17	13	0.751	0.806
wmm280	20	17	0.823	0.820
cfb3109	3	3	0.053	0.125
cfe129	3	3	0.010	0.017
cfe107	2	2	0.041	0.043
cfe364	2	2	0.038	0.133
cfa2170	7	6	0.680	0.654
wmm638	7	6	0.610	0.607
barc206	3	3	0.509	0.490
cfe365	2	2	0.479	0.390
wmc326	24	19	0.878	0.803
barc077	17	16	0.883	0.853
wmc322	10	5	0.768	0.681
wmm1966	23	21	0.914	0.897
wmm1758	29	35	0.936	0.890
gwm299	13	10	0.825	0.755
gpw5007	4	3	0.343	0.387
cfb3366	4	2	0.167	0.212
gwm340	23	15	0.894	0.818
Mean	9.59 ± 7.32	7.66 ± 5.79	0.553 ± 0.272	0.521 ± 0.253
Rarefied mean	8.88	7.59		

rarefaction method, allele numbers of the core collection and breeding material were expected to be 8.88 and 7.55, respectively. Consequently, a decrease of 15% of allelic richness could be observed between the core collection and the genitors of elite inbred lines.

Table 1 shows IBCs calculated within pedigrees of seven modern varieties. Amongst them, Koreli had the highest number of genitors which appeared several times in its genealogical tree (IBC = 0.033). Varieties could be ordered from the more to the less consanguineous as follows: Koreli, Alexandre, Oracle, Apache, Isengrain, Ornicar, Taldor.

Genetic structure in the core collection

The analysis of population structure inferred with the STRUCTURE software, using 578 DArT markers, led to five groups ($K = 5$) for the core collection. Figure 1 presents a visualisation of the structure results in five groups. The major part of accessions from France and Northwest European countries (AUT, BEL, CHE, DEU, GBR, NLD SWE) formed a first group of accessions (Western Europe group), whilst a majority of East European and Southeast European accessions (BGR, FIN, HUN, POL, ROM, RUS, UKR, YUG) together with North American accessions (CAN, USA) were clustered in a second group (Eastern Europe group). A third group included accessions scattered in very diverse regions such as Central and South America (ARG, BRA COL, MEX, URY), Africa (EGY, KEN, TUN, ZAF), the Mediterranean area (ESP, PRT, GRC, ISR, SYR), Australia (AUS) and even North America (CAN,

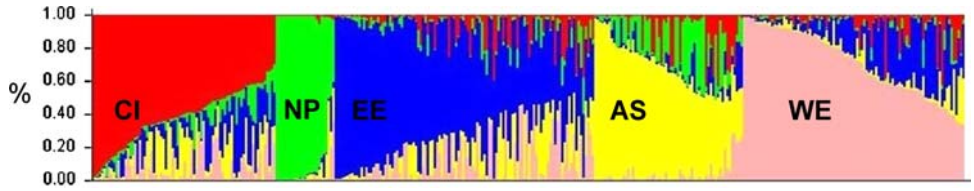

Fig. 1 Estimated population structure. Each individual is represented by a *thin vertical line*, which is partitioned into $K = 5$ segments that represent the individual's estimated membership fractions (%) in K clusters: *CI* CIMMYT–ICARDA cluster, *NP* Nepalese cluster, *EE* Eastern European cluster, *AS* Asian cluster, *WE* Western European cluster

USA): this third group may be considered as the International Maize and Wheat Improvement Centre (CIMMYT) and the International Centre for Agricultural Research in the Dry Areas (ICARDA) group since accessions may be related to the breeding programme of these two institutes. Most part of Caucasian, Middle East and Central Asian accessions (AFG, ARM, AZE, GEO, TJK, TKM TUR) were clustered in a fourth group, together with southeastern ones (CHN, IND, JPN, PAK) and formed the Asian group. Finally, every Nepalese accession was gathered in a last small group. Figure 2 clearly shows the discrimination between these five groups on the plan of the two first axes of a PCoA based on the Sokal and Michener dissimilarity matrix calculated with 578 DArTs markers and highlights the strong geographical structure existing within the core collection.

Genetic structure of the breeding material

The seven temporal groups of the breeding material were compared for their allele number with the rarefaction method described in Petit et al. (1998) in order to minimise the differences due to the heterogenous sample sizes. Rarefied allele number decreased from landraces to recent varieties, in the order of the temporal groups: 10.00, 9.96, 9.83, 9.60, 8.36, 7.69, and 6.47, respectively. A more important decrease in rare alleles could be observed from the early 1970s.

UPGMA dendrogram based on the unbiased genetic distance (Nei 1978) of seven temporal groups is given in Fig. 3. The dendrogram is made of three major clades. Landraces and ancient varieties (period 1) were clearly separated from the rest. The remainder periods constituted

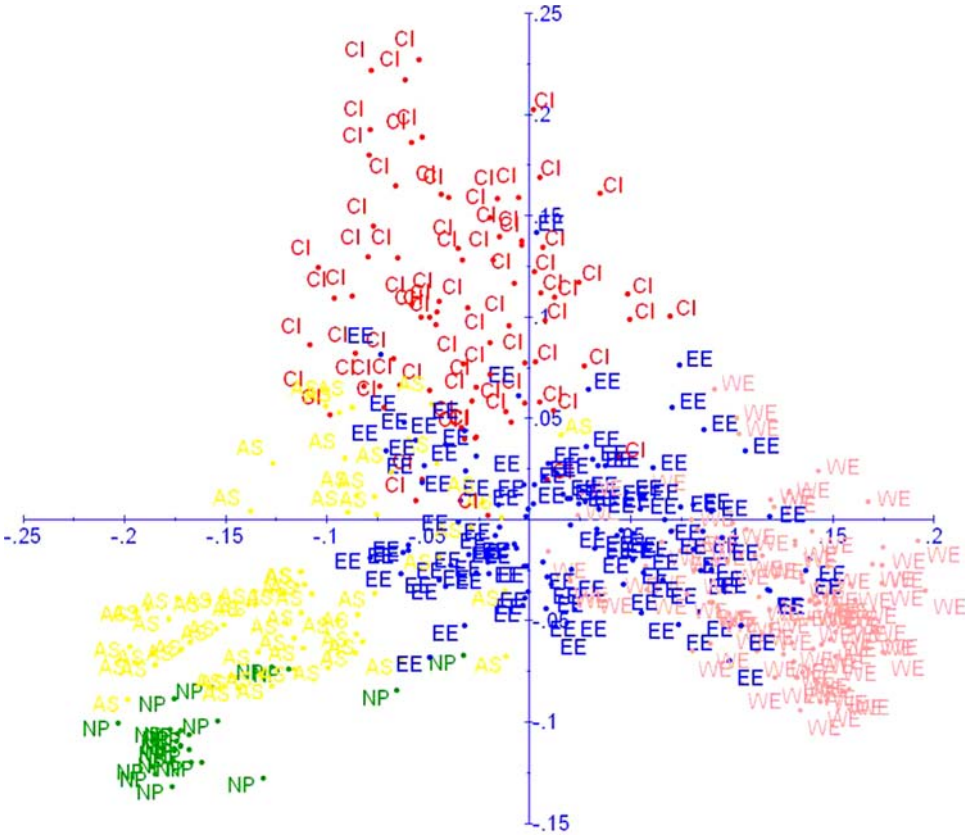

Fig. 2 Projection of 372 accessions of the core collection on the two first axes of principal coordinates analysis. Individuals are clustered in five groups according to population structure analysis and represented by abbreviations used in Fig. 1

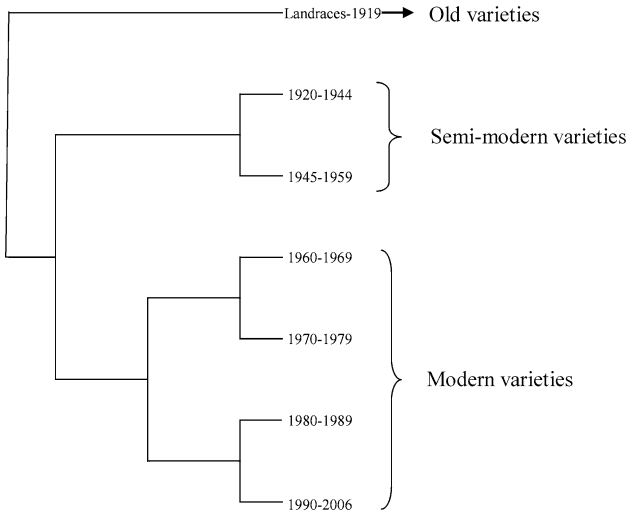

Fig. 3 UPGMA dendrogram based on Nei genetic distance (1978). Three major clades can be distinguished, constituting temporal groups as follows: (1) old varieties (landraces to 1919); (2) semi-modern varieties (1920–1959); (3) modern varieties (1960–2006)

two clades, with periods 2–3 (years 1920–1944 and 1945–1959) on one hand, and periods 4–7 (years 1960–1969, 1970–1979, 1980–1989, 1990–2006, respectively) on the other hand. Thus, temporal structure meets diversity relationships amongst temporal groups.

Based on this result, temporal groups could be brought together according to the clades of dendrogram. Consequently, regrouping periods led to balance sample sizes by forming three groups: (1) old varieties (landraces to 1919), $N = 74$; (2) semi-modern varieties (1920–1959), $N = 93$; (3) modern varieties (1960–2006), $N = 75$. These categories better reflect the history of the breeding material, its temporal structure from landraces to modern cultivars. Further analyses were realised on these periods, thus old and modern cultivars could be compared directly, without any influence of sample size.

Linkage disequilibrium

The dense microsatellite genotyping made possible to realise a fine LD analysis along the chromosome 3B and compare LD patterns to the structure of the chromosome.

Core collection

After filtering out rare alleles, 1,329 pairs of markers were observed amongst the theoretically possible 1,891, because alleles with too low frequency are excluded from the analysis. A total of 357 pairs revealed significant LD at $p < 0.001$ (18.87% calculated for the possible pair number of 1,891). Amongst them, four pairs (0.21%) showed an r^2 value higher than 0.2 (Table 4).

Comparison of the core collection and geographically structured groups

In order to compare directly the whole core collection and geographical groups identified with the help of STRUCTURE software, we considered the mean sample size value (85) calculated on the four first groups. Nepalese group was removed from this comparison, since it is an isolate group made of a small number of individuals. Then, we processed a random re-sampling of 85 individuals in the core collection. The amount of significant marker pairs was revealed to be the highest in the third group (130; 6.87% at $p < 0.001$), which gathers mainly line issues from CIMMYT and ICARDA breeding programmes (Table 4). Western European group was comparable to the core collection (3.80 and 2.64%, respectively). Asian accessions showed a small quantity (1.26%) of significant LD. Eastern European group was found intermediate (2.06%) between Asian and Western European samples.

In order to assess LD decrease in function of the genetic distance between markers, percentage of the number of marker pairs in significant ($p < 0.001$) LD was compared in distance classes by step of 5 cM, between 0 and 50 cM. The same classes were used in further analysis to avoid confusion caused by classes' differences.

Figure 4 shows r^2 decrease profile of geographical groups. The most marker pairs in highly significant LD were observed in the CIMMYT–ICARDA gene pool, then in Western European group. At 0–5 cM distance, the amount of significant r^2 was the highest in the CIMMYT–ICARDA group (2.37% calculated from the 1,891 pairs possible), then decreased as follows: Western Europe, Eastern Europe, core collection, Asia (1.58, 1.16, 1.11, and 0.42%, respectively). At 5–20 cM distance, important differences were identified between every group. From a distance of 25–30 cM, geographical groups presented a similar profile, except CIMMYT–ICARDA pool which stayed higher than others.

Breeding material

Pair-wise LD analysis of markers revealed 1,530 observed marker pairs from the theoretically possible 1891 (Table 4). Amongst them, 254 pairs (13.43% calculated for the possible pair number of 1891) revealed a significant ($p < 0.001$) LD and 4 pairs (0.21%) showed an r^2 value higher than the conventional threshold $r^2 = 0.2$ (Table 4).

Comparison of old and modern varieties

Globally, LD increased from old cultivars to modern varieties, when considering the number of significant marker pairs and $r^2 > 0.2$ values. In the pair-wise analysis

Table 4 Linkage disequilibrium patterns of geographical structures, temporal groups and genealogical trees

Material	*N*	Possible[a]	Observed[b]	Significant[c]	%[d]	$r^2 > 0.2$[e]	%[f]
Core collection	372	1,891	1,329	357	18.87	4	0.21
Core re-sampling	85	1,891	1,396	50	2.64	7	0.37
Western Europe	95	1,891	1,354	72	3.80	9	0.47
Eastern Europe	110	1,891	1,286	39	2.06	7	0.37
CIMMYT, ICARDA	74	1,891	1,528	130	6.87	13	0.68
Asia	63	1,891	1,208	24	1.26	6	0.31
Old varieties	74	1,891	1,378	48	2.53	4	0.21
Semi-modern varieties	93	1,891	1,165	67	3.54	3	0.15
Modern varieties	75	1,891	1,431	78	4.12	10	0.52
Genitors Alexandre	127	1,891	1,510	117	6.18	5	0.26
Genitors Apache	71	1,891	1,321	63	3.33	12	0.63
Genitors Isengrain	86	1,891	1,415	71	3.75	11	0.58
Genitors Koreli	133	1,891	1,479	141	7.45	8	0.42
Genitors Oracle	111	1,891	1,417	103	5.44	9	0.47
Genitors Ornicar	82	1,891	1,326	52	2.74	5	0.26
Genitors Taldor	70	1,891	1,278	55	2.90	9	0.47
Core re-sampling	242	1,891	1,447	254	13.43	2	0.10
Breeding material	242	1,891	1,530	254	13.43	4	0.20

[a] Number of possible marker pairs

[b] Number of observed marker pairs

[c] Number of marker pairs presenting a significant LD at $p < 0.001$

[d] Percentage of marker pairs presenting a significant LD at $p < 0.001$

[e] Number of marker pairs presenting a significant LD > 0.2 at $p < 0.001$

[f] Percentage of marker pairs presenting a significant LD > 0.2 at $p < 0.001$

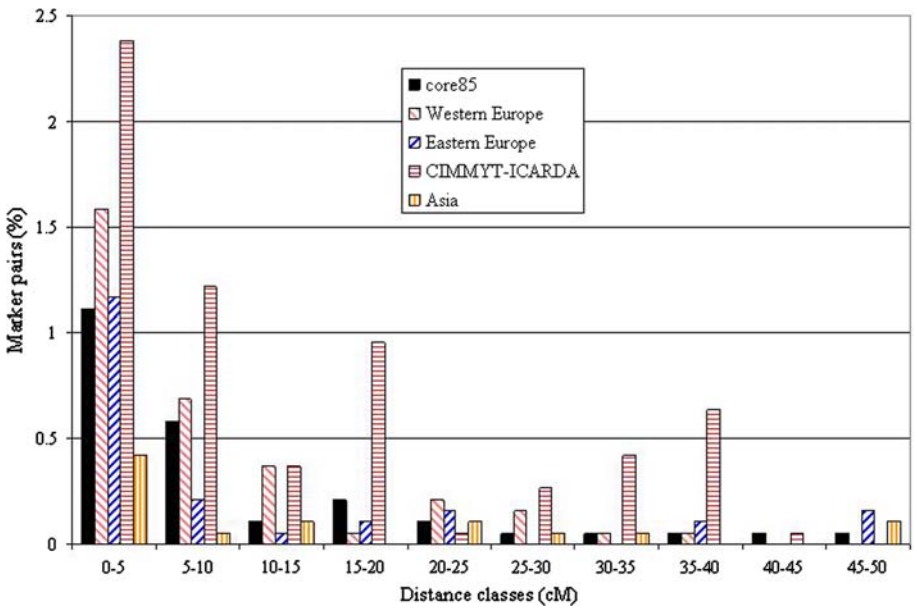

Fig. 4 Percentage of marker pairs in significant ($p > 0.001$) LD related to marker distance classes, amongst the core collection and different geographical structures

of LD (Table 4), in the old category, 1,378 pairs were observed for the possible 1,891 markers pairs, amongst them 48 were significant at $p < 0.001$ (2.53% calculated for the possible pair number of 1,891). In the modern group, 1,431 marker pairs were described, with 78 significant marker pairs (4.12%). Amongst r^2 values, 4 (0.21%) reached the threshold value 0.2 in the old category, and 10 (0.52%) in the modern group.

Figure 5 presents the percentage of the number of marker pairs in significant ($p < 0.001$) LD in ancient and modern cultivars in distance classes. At a distance of 0–5 cM, modern cultivars had twofold higher LD values compared to the older ones (2.22 and 0.89%, respectively). Semi-modern cultivars revealed intermediate with 1.21%. A similar decrease was observed in the three materials from about 20 cM distance.

Comparison of genitors of seven modern varieties per genealogical tree

Genitors of seven modern varieties showed important differences in LD patterns (Table 4). The most marker pairs in significant LD were observed amongst genitors of Koreli and Alexandre (7.45 and 6.18%, respectively). Genitors of Ornicar and Taldor presented only 2.74 and 2.90%, respectively, of significant LD at $p < 0.001$. Concerning r^2 values reaching the conventional threshold $r^2 = 0.2$, the highest percentage was observed amongst genitors of Apache (0.63%). Koreli had a smaller amount of $r^2 > 0.2$ (0.42%), similar to those of Oracle and Taldor (both 0.47%).

Figure 6 shows distance classes amongst markers and the percentage of the number of marker pairs in significant ($p < 0.001$) LD per modern cultivars. Global tendency was a decrease of r^2 around 20 cM distance between markers, but this value changed depending on cultivars. Thus, globally, Koreli and Alexandre presented the highest number of marker pairs in significant LD. A lower percentage of LD was observed amongst genitors of each cultivar until 15–20 cM, except for Koreli and Alexandre, which stayed at this distance (0.68 and 0.58%, respectively), then decayed.

Comparison of breeding material and core collection

The re-sampled ($N = 242$) core collection showed 254 marker pairs in significant LD, which represent 13.43% amongst the theoretically possible 1,891. Interestingly, in the breeding material, the same amount of significant marker pairs was observed (Table 4). At short distances (0–5 cM), the breeding material had higher amount of marker pairs in significant LD: 3.06 versus 2.90% detected in the core collection. Globally, the two materials had similar values in all classes.

Organisation of LD pattern along the chromosome

Globally, marker pairs in significant LD are observed around the centromere in both arms and at distal position on the short arm. When comparing different types of materials, important differences can be described. Figure 7 and supplementary material show p values associated to r^2 values at the chromosomal scale in different materials.

In the analysis amongst the re-sampled ($N = 85$) core collection and structures, practically the same LD blocks were observed in all materials, mainly in deletion bins 3BS-1–3BS-9, 3BL-2–3BL-7 and 3BL-7 (see supplementary material). In the Eastern European group, more blocks were present than in Western European cultivars, but in contrast of this latter case, they did not extend over larger distances. In Asian group, highly significant LD was observed only between a few number of marker pairs. CIMMYT–ICARDA group was characterised by highly significant and longer range LD than other structures.

When comparing old and modern cultivars (Fig. 6), mainly the same blocks were observed along the chromosome 3B, particularly on the short arm. Modern varieties

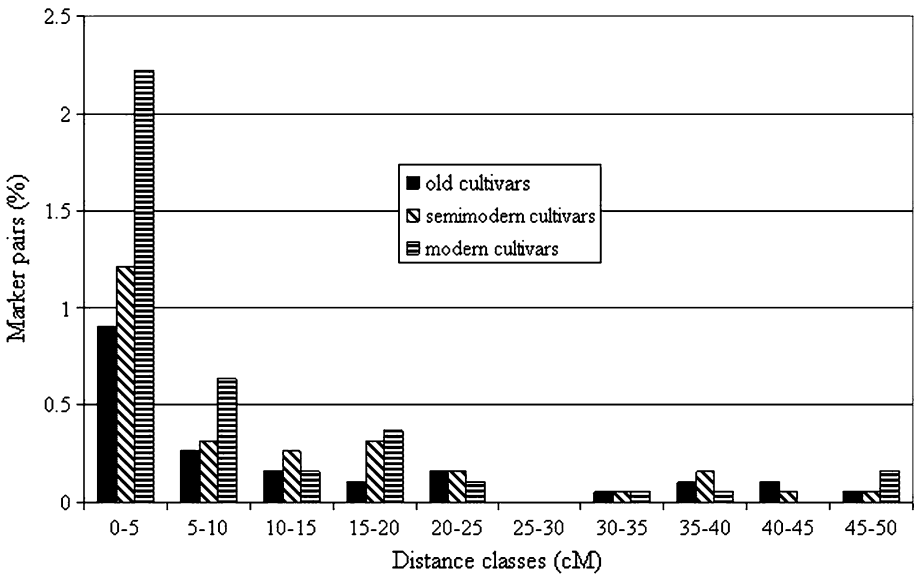

Fig. 5 Percentage of marker pairs in significant ($p > 0.001$) LD related to marker distance classes, amongst, old, semi-modern and modern varieties of breeding materiel

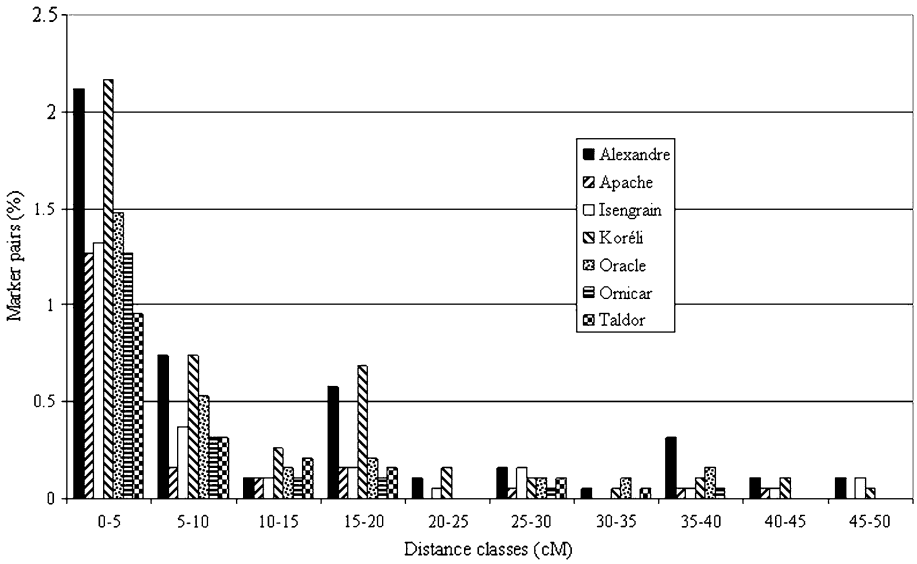

Fig. 6 Percentage of marker pairs in significant ($p > 0.001$) LD related to marker distance classes, amongst genitors of elite bread wheat varieties

showed more extended LD patterns than the older ones, e.g., in bins 3BS-1 and 3BS-9.

In pedigrees of elite varieties, Koreli's genitors showed the most important LD, particularly in deletion bins 3BL-2 and 3BS-1. In addition, highly significant long-range LD was observed, between markers at longer distance. In Taldor, Ornicar and Oracle pedigrees, LD seemed not to be important. Isengrain, Apache and Alexandre showed an intermediate state; this latter was comparable to Koreli in deletion bin 3BS-1.

When comparing the re-sampled core collection ($N = 242$) and the breeding material, there was no important differences between the two types of materials. The core collection presented more important LD at the distal end of the deletion bin 3BL-7. In contrast, the breeding material showed highly significant LD extending on bigger distance than in the core collection.

Discussion

Genetic diversity and structure analysis

The set of 62 SSR markers revealed a total of 595 polymorphic alleles in our material composed by two contrasted sample types. When comparing rarefied allele numbers in two samples, a loss of 15% of allelic richness was observed, agreeing with the domestication and selection effect reported by Rafalski and Morgante (2004). Mean PIC values of markers were comparable in the two types of material and microsatellite markers were successfully informative to describe and to compare diversity patterns of our material, as it has been already reported by numerous studies in bread wheat (Eivazi et al. 2008; Landjeva et al. 2006; Roder et al. 2002; Roussel et al. 2004, 2005; Routray et al. 2007).

Here, we report the results obtained on the 3B chromosome only. Comparison with the whole B-genome should be interesting in order to determine if its diversity characteristics are chromosome specific or if they can be applied generally. For this comparison, we used a data set already published (Balfourier et al. 2007). This latter paper presented the core collection and analysed its diversity with the help of genome-wide distributed genomic and EST-SSR loci. In the present study, we used the same types of markers, on the same plant material, so the results are directly comparable. The mean value of allele number per locus of the whole B-genome (based on 25 SSR markers of the previous study) was 14.12 whilst, on the chromosome 3B, here we found 9.50 with 62 SSRs. After rarefaction method recommended by Petit et al. (1998) for unbalanced sample sizes, the mean allele number of the core collection detected on the whole B-genome (14.12) fell to 8.88 on chromosome 3B, then to 7.55 in the breeding material. Thereby, the chromosome 3B seems to have a lower diversity than the whole B-genome. This difference in diversity can be related to distinct selection pressure amongst chromosomes of the B-genome. Amongst the 6,000 genes estimated on chromosome 3B (Paux et al. 2006), many could have been submitted to various selection effects, since for instance some of them are implicated in disease resistances, as reaction to tan spot (Singh et al. 2006; Tadesse et al. 2008), to *Fusarium* head blight (Gervais et al. 2003; Shen et al. 2006) or to the stem rust (McNeil et al. 2008).

The results concerning population structure in the core collection are quite consistent with those obtained by Balfourier et al. (2007) analysing a larger sample of

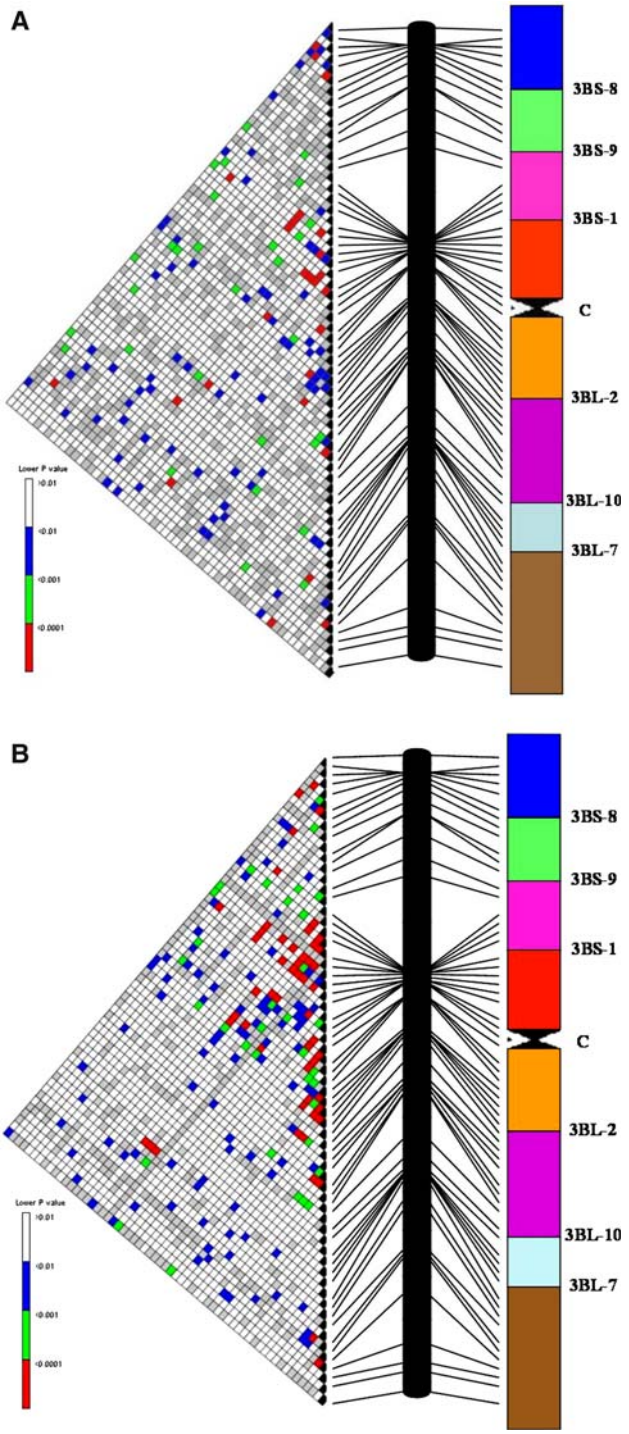

Fig. 7 Wheat cytogenetic and consensus map of 3B chromosome with distribution of LD patterns between pairs of markers in old (**a**) and modern (**b**) cultivars of breeding material. The *grey scale* shown in each *square* corresponds to *p* values

accessions (744 accessions, containing the present 372 core collection) with a set of 73 polymorphic SSR markers covering the whole genome map. When considering a neighbour-joining tree based on a Cavalli-Sforza and Edwards distance matrix between 45 geographical origins of bread wheat germplasm, these authors showed a clear clustering in six main groups: (1) Northwest European cluster, (2) Southeast European + North American cluster, (3) Mediterranean cluster, (4) Central Asian cluster, (5) Eastern Asian cluster, and (6) South American and African cluster. In the present study, only five groups were observed with STRUCTURE software: the two European groups were similar to the previous study. The South American + African accessions were grouped with the Mediterranean accessions to form the CIMMYT + ICARDA group, whilst the remaining accessions from Middle East and Eastern Asia were gathered in a single cluster, except for Nepalese accessions which were apart. Explanations and arguments to interpret such structure were widely developed by Balfourier et al. (2007): the Asian and the two European clusters are probably the result of both the historical process of the spread of initial wheat germplasm from the Fertile Crescent through Asia and Europe, then the biological process of adaptation and selection to specific environmental conditions in Northwest and Southeast Europe, respectively (Roussel et al. 2005). In contrast, the South American and African group of accessions together with some accessions of Mediterranean area might reflect recent breeding practices developed by CIMMYT and ICARDA during the 1960s; according to the historical events, this group is probably the more recent one in term of evolution, with the highest level of admixture, whilst Asian group would be the oldest. Finally, the small group of Nepalese accessions (only 25 accessions) is more difficult to interpret in terms of population genetics: it may be an artefact group due to our core sample size or to the specific genetic structure of these accessions which are all landraces. They can also represent a separated cluster of differentiated lines after isolation by distance.

Allele number of breeding material decreased with the time, as already reported by Roussel et al. (2005), who showed that allele number was stable until the 1960s when he started to decrease regularly. The same situation was reported in a study of 1,680 Chinese wheat varieties (Hao et al. 2006).

The diversity loss during plant breeding activity has been already discussed by several authors. Comparing bread wheat cultivars originated from UK, US and Australia with DArT markers, White et al. (2008) observed a fluctuation of diversity over the time, but they concluded that modern plant breeding did not lead to an important diversity reduction. Similarly, in a study performed on 91 Bulgarian bread wheat cultivars, genetic diversity characteristics seemed to be preserved by wheat breeders after 1960 and there was no declining trend detected caused by modern breeding process (Landjeva et al. 2006). In our study, the dense genotyping performed on one chromosome made it possible to obtain a high resolution of the

analysis and the detection of diversity loss around selected genes positioned in the proximity of neutral microsatellite markers. This result correlates with Roussel et al. (2004), describing chronological trends of diversity in 559 French wheat varieties cultivated between 1800 and 2000.

UPGMA dendrogram of temporal groups distinguished three major clusters, corresponding to the breeding history of accessions. The first group gathered landraces and the oldest varieties, mostly coming from the mass selection or from the first breeding programmes using hybridisation techniques. The second group corresponded to the period 1920–1959 when modern selection has started and accelerated, thanks to introduction of foreign and exotic material in crosses and to technological development which made it possible to reach a considerable gain in productivity, in the adaptation to the environment constraints and in baking quality (Bonjean et al. 2001). The third group is clearly separated from the previous ones, probably due to the phenomenon called "green revolution" in developing countries, characterised by the intensive development of genetically improved cultivars with strongly increased yields, first created at the CIMMYT in Mexico (Khush 2001). The "green revolution" led in particular to improve semi-dwarf varieties possessing different dwarfing genes which are today always present in the French wheat germplasm (Baron et al. 2006).

Linkage disequilibrium analysis

Linkage disequilibrium patterns were analysed by comparing the whole core collection to geographically structured groups, old and modern varieties, genitors of seven modern varieties according to their genealogical tree, finally breeding material and the core collection. Generally, the observed LD was low in both plant material types, only 0.10% of marker pairs reached the threshold of 0.2 r^2 value in the re-sampled ($N = 242$) core collection and 0.20% in the breeding material, in accordance with a study describing LD patterns of the B-genome ancestral species, the tetraploid emmer wheat (Seguin et al. 2006). This finding is quite surprising because it is known that plant mating systems affect LD and generally self-fertilisation leads to a more extensive LD due to the severely reduced effective recombination level (Flint-Garcia et al. 2003). In the 0–5 cM class, the percentage of significant r^2 values is in concordance with the amount of global distribution of high r^2 values, indicating that the most marker pairs in important LD are positioned in this distance class.

After comparison of different population structures and the core collection, a classification can be established amongst them according to the LD level. The highest r^2 values were detected in the CIMMYT–ICARDA group, then in the Western European pool, followed by the whole core collection, Eastern European group and Asian pool. In theory, LD of a population in admixture, like the core collection, is expected to be higher due to the introduction of chromosomes having different origin and allele frequencies. The five structures reported here were not totally independent, since they were results of a geographical separation which occurred progressively and was linked to the history of agriculture. As the core collection is representative for the worldwide diversity of bread wheat, geographical structures can reflect history of bread wheat dissemination around the world. Asian group is considered to be the original gene pool, with many accessions coming from Caucasus, Central Asia or from the Middle East which is not too far from the Fertile Crescent situated in the Near East, considered as a region where agriculture and cereal domestication developed together (Salamini et al. 2002). Low level of LD can be explained by the effective population size having remained high since the time of wheat domestication and beginning of agriculture being dated at 8000 BP. The European gene pool was divided into two distinct groups with a clear separation between Western and Eastern subpopulations. The higher LD of Western European structure can be explained either by an historical bottleneck during agriculture spread from the Middle East during the Neolithic or by the more intensive modern breeding activity. Eastern European group can be dated at 7–6000 BP, when wheat has been disseminated in Central Europe through Anatolia and Greece, then across the Balkans, the Transcaucasia, the Caucasus and Southern Russia. Then, from the Danube and Rhine Valley, bread wheat spread to Western and Northern Europe by 5000 BP. During the 10,000 years of cultivation, genetic variability of bread wheat increased due to the spontaneous hybridisations and mutations (Bonjean et al. 2001), likely creating balanced LD patterns. At the end of the nineteenth century, modern intensive selection started in Western Europe two decades before Eastern Europe thanks to hybridisation techniques developed first in the middle of the nineteenth century by English breeders. Thereby, LD patterns were more influenced in Western Europe by the artificial bottleneck caused by selection pressure, and consequently this structure presents higher r^2 values. Concerning accessions belonging to the CIMMYT–ICARDA group, their highest LD values, extending to a 20 cM distance, are not surprising, because these cultivars are originated from breeding programmes of the two institutions created in 1959 and 1977, respectively. Consequently, modern varieties have been subjected to high selection pressure, perhaps from an originally narrow genetic basis, and these breeding activities resulted in higher LD values.

Some important differences were observed amongst genitors of the seven modern varieties, when LD was analysed according to genealogical trees. The seven

modern varieties can be considered rather related thanks to breeding activity, where the same genitor was crossed repeatedly in different breeding programmes. When comparing IBCs and the number of marker pairs in significant ($p < 0.001$) LD, varieties with the highest IBC in their genealogical trees presented the highest amount of significant r^2 values. This finding is logical, thus the repetition of the same genitors in a pedigree results in fixation of several chromosomal regions, LD increases.

Despite this consanguinity, some important differences were identified amongst varieties, suggesting that LD could be more efficiently studied on a pedigree-related material than on a temporally structured gene pool.

The re-sampled ($N = 242$) core collection and the breeding material were not different for LD values. In contrast to our hypothesis, the highly polymorphic core collection had the same LD than the selected breeding material: 13.43% of marker pairs showed a significant r^2 value in both materials.

The distribution of LD blocks on the whole chromosome was similar in each type of materials, differences were described in the extent of LD. Highly significant r^2 values were mainly concentrated at the distal end of the short arm of 3B chromosome and in deletion bins around the centromere (3BS-1 and 3BL-2). As crossing-over weakens intrachromosomal LD (Flint-Garcia et al. 2003), recombination rates along the chromosome 3B are likely to shape the distribution of marker pairs with significant r^2. Recently, a gradient of recombination was reported on the short arm, from the centromere to the telomere, with an important decrease at the telomeric end in the last deletion bin (Saintenac et al. 2008).

In conclusion, we reported here the genetic diversity and LD patterns in two contrasted materials. To our knowledge, this is the first study dissecting these characteristics for two differently structured gene pools with the same methods and at fine scale, along one chromosome. Diversity and LD analyses performed on the core collection made of different structures, reinforced available information concerning this worldwide genetic material. These results will be useful to further association genetic studies possibly performed on different agronomical traits, as flowering date, plant height, resistance to diseases and pre-harvest sprouting, on grain quality characteristics, as kernel weight, grain protein content, and grain hardness, and on flour rheology. These characteristics have been recently evaluated in the worldwide core collection and this is an important basis for further studies on wheat quality (Bordes et al. 2008). Power and accuracy of methods using the two types of association panels (core collection vs. breeding lines) will be compared in a next paper.

In addition, our results concerning the strongly related breeding material can be useful to initiate a study to assess signature of selection in pedigrees of modern bread wheat cultivars, in order to identify the selection pressure and effects during the 180 years of modern wheat breeding.

Acknowledgments The authors are grateful to P. Sourdille for providing CFA, CFB, CFE microsatellite primer pairs and thank Genoplante Consortium, RAGT and Syngenta firms for GPW, WMM and NW microsatellite primers, respectively. This work is part of the SDD project (Semences De Demain) which has been financed by the "Industry competitiveness Fund" of the French Ministry of Industry (http://www.competitivite.gouv.fr) and labelled by the pole "Céréales vallée" (http://cereales-vallee.org/). It is also a part of the EXEGESE project founded by NRA (French National Research Agency).

References

Able JA, Langridge P, Milligan AS (2007) Capturing diversity in the cereals: many options but little promiscuity. Trends Plant Sci 12(2):71–79

Balfourier F, Roussel V, Strelchenko P, Exbrayat-Winson 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

Baron M, Bastergue P, Branlard G, Cauderon Y, Guiard J, Jahier J, Lonnet P, Quandalle C, Ronsin T, Rousset M, Triverio F, Trottet M (2006) Le blé tendre. In: Doré C, Varoquaux F (eds) Histoire et amélioration de cinquante plantes cultivées. INRA Editions, Paris, France, pp 137–161

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2004) Genetix 4.05.2, Logiciel sous Windows^{TM} pour la Génétique des Populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier, France

Bonjean AP, Doussinault G, Stragliati J (2001) French wheat pool. In: Bonjean AP, Angus WP (eds) The world wheat book. A history of wheat breeding. Intercept Ltd., Editions Tec & Doc, Lavoisier Publishing Inc., Londres, Paris, New York, pp 127–165

Bonnier G, Douin R (1990) La grande flore en couleurs de Gaston Bonnier, France, Suisse, Belgique et pays voisins. Belin Editions, Paris, France, pp 1–1401

Bordes J, Branlard G, Oury F-X, Charmet G, Balfourier F (2008) Agronomic characteristics, grain quality and flour rheology of 372 bread wheats in a worldwide core collection. J Cereal Sci. doi:10.1016/j.jcs.2008.05.005

Breseghello F, Sorrels ME (2006) Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. Genetics 172:1165–1177

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

Eivazi AR, Naghavi MR, Hajheidari M, Pirseyedi SM, Ghaffari MR, Mohammadi SA, Majidi I, Salekdeh GH, Mardi M (2008) Assessing wheat (*Triticum aestivum* L.) genetic diversity using quality traits, amplified fragment length polymorphisms, simple sequence repeats and proteome analysis. Ann Appl Biol 152:81–91

Endo TR, Gill BS (1996) The deletion stocks of common wheat. J Hered 87:295–307

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

FAOSTAT (2008) http://faostat.fao.org/ (last accessed July 2008)

Farnir F, Coppieters W, Arranz J-J, Berzi P, Cambisano N, Grisart B, Karim L, Marcq F, Moreau L, Mni M, Nezer C, Simon P, Vanmanshoven P, Wagenaar D, Georges M (2000) Extensive genome-wide linkage disequilibrium in cattle. Genome Res 10:220–227

Felsenstein J (2004) PHYLIP (Phylogeny Inference Package) version 3.6. Department of Genome Sciences, University of Washington, Seattle, USA

Flint-Garcia SA, Thornsberry JM, Buckler ES IV (2003) Structure of linkage disequilibrium in plants. Annu Rev Plant Biol 54:357–374

Gervais L, Derdryver F, Morlais J-Y, Bodusseau V, Negre S, Bilous M, Groos C, Trottet M (2003) Mapping of quantitative trait loci for field resistance to *Fusarium* head blight in an European winter wheat. Theor Appl Genet 106:961–970

Gill BS, Appels R, Botha-Oberholster A-M, Buell CR, Bennetzen JL, Chalhoub B, Chumley F, Dvorak J, Iwanaga M, Keller B, Li W, McCombie WR, Ogihara Y, Quetier F, Sasaki T (2004) A workshop report on wheat genome sequencing: international genome research on wheat consortium. Genetics 168:1087–1096

Groos C, Gay G, Perretant M-R, Gervais L, Bernard M, Derdryver F, Charmet G (2002) Study of the relationship between pre-harvest sprouting and grain color by quantitative trait loci analysis in a white × red grain bread-wheat cross. Theor Appl Genet 104:39–47

Gupta PK, Balyan HS, Edwards KJ, Isaac P, Korzun V, Roder MS, Gautier M-F, Joudrier P, Schlatter AR, Dubcovsky J, De la Pena RC, Khairallah M, Penner G, Hayden MJ, Sharp P, Keller B, Wang RCC, Hardouin JP, Jack P, Leroy P (2002) Genetic mapping of 66 new microsatellite (SSR) loci in bread wheat. Theor Appl Genet 105:413–422

Hao C, Lanfen W, Xueyong Z, Guangxia Y, Yushen D, Jizeng J, Xu L, Xunwu S, Sancai L, Yongsheng C (2006) Genetic diversity in Chinese modern wheat varieties revealed by microsatellite markers. Sci China C Life Sci 49:218–226

Haseneyer G, Ravel C, Dardevet M, Balfourier F, Sourdille P, Charmet G, Brunel D, Sauer S, Geiger HH, Graner A, Stracke S (2008) High level of conservation between genes coding for the GAMYB transcription factor in barley (*Hordeum vulgare* L.) and bread wheat (*Triticum aestivum* L.) collections. Theor Appl Genet 117:321–331

Khush GS (2001) Green revolution: the way forward. Nat Rev Genet 2:815–822

Landjeva S, Korzun V, Ganeva G (2006) Evaluation of genetic diversity among Bulgarian winter wheat (*Triticum aestivum* L.) varieties during the period 1925–2003 using microsatellites. Genet Res Crop Evol 53:1605–1614

Laperche A, Brancourt-Hulmel M, Heumez E, Gardet O, Hanocq E, Devienne-Barret F, Le Gouis J (2007) Using genotype × nitrogen interaction variables to evaluate the QTL involved in wheat tolerance to nitrogen constraints. Theor Appl Genet 115:399–415

Lee J-H, Ma Y, Wako T, Li LC, Kim K-Y, Park S-W, Uchiyama S, Fukui K (2004) Flow karyotypes and chromosomal DNA contents of genus *Triticum* species and rye (*Secale cereale*). Chromosome Res 12:93–102

Liu F, Sun G-L, Salomon B, Von Bothmer R (2002) Characterization of genetic diversity in core collection accessions of wild barley, *Hordeum vulgare* ssp. *spontaneum*. Hereditas 136:67–73

Ma HX, Zhang KM, Gao L, Bai GH, Chen HG, Cai ZX, Lu WZ (2006) Quantitative trait loci for resistance to fusarium head blight and deoxynivalenol accumulation in Wangshuibai wheat under field conditions. Plant Pathol 55:739–745

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

Malysheva-Otto LV, Roder MS (2006) Haplotype diversity in the endosperm specific β-amylase gene *Bmy1* of cultivated barley (*Hordeum vulgare* L.). Mol Breed 18:143–156

Malysheva-Otto LV, Ganal MW, Roder MS (2006) Analysis of molecular diversity, population structure and linkage disequilibrium in a worldwide survey of cultivated barley germplasm (*Hordeum vulgare* L.). BMC Genet 7(6). doi:10.1186/1471-2156-7-6

McNeil MD, Kota R, Paux E, Dunn D, McLean R, Feuillet C, Li D, Kong X, Lagudah E, Zhang JC, Jia JZ, Spielmeyer W, Bellgard M, Appels R (2008) BAC-derived markers for assaying the stem rust resistance gene, *Sr2*, in wheat breeding programmes. Mol Breed 22:15–24

Nei M (1973) Analysis of gene diversity in subdivised populations. PNAS 70:3321–3323

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89:583–590

Ozkan H, Brandolini A, Schafer-Pregl R, Salamini F (2002) AFLP analysis of a collection of tetraploid wheats indicates the origin of emmer and hard wheat domestication in southeast Turkey. Mol Biol Evol 19(10):1797–1801

Page RDM (1996) TREEVIEW: an application to display phylogenetic trees on personal computers. Comput Appl Biosci 12:357–358

Paux E, Roger D, Badaeva E, Gay G, Bernard M, Sourdille P, Feuillet C (2006) Characterizing the composition and evolution of homoeologous genomes in hexaploid wheat through BAC-end sequencing on chromosome 3B. Plant J 48:463–474

Paux E, Legeai F, Guilhot N, Adam-Blondon A-F, Alaux M, Salse J, Sourdille P, Leroy P, Feuillet C (2008a) Physical mapping in large genomes: accelerating anchoring of BAC contigs to genetic maps through in silico analysis. Funct Integr Genom 8:29–32

Paux E, Sourdille P, Salse J, Saintenac C, Choulet F, Leroy P, Korol A, Michalak M, Kianian SF, Spielmeyer W, Lagudah E, Somers D, Kilian A, Alaux M, Vautrin S, Berges H, Eversole K, Appels R, Safar J, Simkova H, Dolezel J, Bernard M, Feuillet C (2008b) A physical map of the 1-gigabase bread wheat chromosome 3B. Science 322:101–104

Perrier X, Jacquemoud-Collet JP (2006) DARwin software. http://darwin.cirad.fr/darwin

Petit RJ, El Mousadik A, Pons O (1998) Identifying populations for conservation on the basis of genetic markers. Conserv Biol 12(4):844–855

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Qi LL, Echalier B, Chao S, Lazo GR, Butler GE, Anderson OD, Akhunov ED, Dvorak J, Linkiewicz AM, Ratnasiri A, Dubcovsky J, Bermudez-Kandianis CE, Greene RA, Kantety R, La Rota CM, Munkvold JD, Sorrels SF, Sorrels ME, Dilbirligi M, Sidhu D, Erayman M, Randhawa HS, Sandhu D, Bondareva SN, Gill KS, Mahmoud AA, Ma X-F, Miftahudin, Gustafson JP, Conley EJ, Nduati V, Gonzalez-Hernandez JL, Anderson JA, Peng JH, Lapitan NLV, Hossain KG, Kalavacharla V, Kianian SF, Pathan MS, Zhang DS, Nguyen HT, Choi D-W, Fenton RD, Close TJ, McGuire PE, Qualset CO, Gill BS (2004) A chromosome bin map of 16,000 expressed sequence tag loci and distribution of genes among the three genomes of polyploid wheat. Genetics 168:701–712

R Development Core Team (2006) R: a language and environment for statistical computing, version 2.4.0. R Foundation for Statistical Computing, Vienna, Austria

Rafalski A, Morgante M (2004) Corn and humans: recombination and linkage disequilibrium in two genomes of similar size. Trends Genet 20(2):103–111

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

Roder MS, Korzun V, Wendehake K, Plaschke J, Tixier M-H, Leroy P, Ganal MW (1998) A microsatellite map of wheat. Genetics 149:2007–2023

Roder MS, Wendehake K, Korzun V, Bredemeijer G, Laborie D, Bertrand L, Isaac P, Rendell S, Jackson J, Cooke RJ, Vosman B, Ganal MW (2002) Construction and analysis of a microsatellite-based database of European wheat varieties. Theor Appl Genet 106:67–73

Roussel V, Koenig J, Beckert M, Balfourier F (2004) Molecular diversity in French bread wheat accessions related to temporal trends and breeding programmes. Theor Appl Genet 108:920–930

Roussel V, Leisova L, Exbrayat F, Stehno Z, Balfourier F (2005) SSR allelic diversity changes in 480 European bread wheat varieties released from 1840 to 2000. Theor Appl Genet 111:162–170

Routray P, Basha O, Garg M, Singh NK, Dhaliwal HS (2007) Genetic diversity of landraces of wheat (*Triticum aestivum* L.) from hilly areas of Uttaranchal, India. Genet Res Crop Evol 54:1315–1326

Saintenac C, Choulet F, Faure S, Paux E, Feuillet C, Sourdille P (2008) Recombination analysis on bread wheat chromosome 3B. In: Proceedings of the 11th international wheat genetics symposium. Sydney University Press, Brisbane, Australia

Salamini F, Ozkan H, Brandolini A, Schafer-Pregl R, Martin W (2002) Genetics and geography of wild cereal domestication in the Near East. Genetics 3:429–441

Seguin M, Attard A, Bataillon T, Billot C, Cenci A, Chantret N, Courtois B, David J, Deu M, El Azhari N, Glaszmann J-C, Glemin S, Haudry A, Le Guen V, Maynadier M, Pomies V, Ronfort J, Tsitronne A, Weber C (2006) Analyse et prédiction des patrons de déséquilibre de liaison dans les collections de ressources génétiques de plantes pérennes ou annuelles, autogames ou allogames. Les Actes du BRG 6:57–74

Shen X, Francki MG, Ohm HW (2006) A resistance-like gene identified by EST mapping and its association with a QTL controlling *Fusarium* head blight infection on wheat chromosome 3B. Genome 49:631–635

Singh PK, Gonzalez-Hernandez JL, Mergoum M, Ali S, Adhikari TB, Kianian SF, Elias EM, Hughes GR (2006) Identification and molecular mapping of a gene conferring resistance to *Pyrenophora tritici-repentis* race 3 in tetraploid wheat. Phytopathology 96:885–889

Somers DJ, Isaac P, Edwards K (2004) A high-density microsatellite consensus map for bread wheat (*Triticum aestivum* L.). Theor Appl Genet 109:1105–1114

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, Guyomarc'h H, Baron C, Gandon B, Chiquet V, Artiguenave F, Edwards K, Foisset N, Dufour P (2001) Improvement of the genetic maps of wheat using new microsatellite markers. In: Abstracts 167, plant and animal genome IX conference, San Diego, USA

Tadesse W, Hsam SLK, Wenzel G, Zeller FJ (2008) Chromosome location of a gene conferring resistance to *Pyrenophora tritici-repentis* in Ethiopian wheat cultivars. Euphytica 162:423–430

Tanyolac B, Linton E, Ozkan H (2003) Low genetic diversity in wild emmer (*T. turgidum* L. subsp. *dicoccoides* (Korn. ex Asch. et Graebn.) Thell.) from South-eastern Turkey revealed by restriction fragment length polymorphism. Genet Res Crop Evol 50:829–833

van Berloo R, Hutten RC (2005) Peditree: pedigree database analysis and visualisation for breeding and science. J Hered 96(4):465–468

Vrana J, Kubalakova M, Simkova H, Cihalikova J, Lysak MA, Dolezel J (2000) Flow sorting of mitotic chromosomes in common wheat (*Triticum aestivum* L.). Genetics 156:2033–2041

Wenzl P, Carling J, Kudrna D, Jaccoud D, Huttner E, Kleinhofs A, Kilian A (2004) Diversity Arrays Technology (DArT) for whole-genome profiling of barley. PNAS 101:9915–9920

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

Zhang G, Mergoum M (2007) Molecular mapping of kernel shattering and its association with Fusarium head blight resistance in a Sumai3 derived population. Theor Appl Genet 115:757–766

Zhang LY, Bernard M, Leroy P, Feuillet C, Sourdille P (2005) High transferability of bread wheat EST-derived SSRs to other cereals. Theor Appl Genet 111:677–687

Zhang Z, Bradbury PJ, Kroon DE, Casstevens TM, Buckler ES (2006) TASSEL 2.0: a software package for association and diversity analyses in plants and animals (www.maizegenetics.net). In: Plant and animal genomes XIV conference, poster P956/CP012, San Diego, USA