Mol Breeding (2011) 27:37–58
DOI 10.1007/s11032-010-9411-7

Genome-wide association mapping: a case study in bread wheat (*Triticum aestivum* L.)

K. Neumann · B. Kobiljski · S. Denčić · R. K. Varshney · A. Börner

Received: 30 September 2009 / Accepted: 9 February 2010 / Published online: 6 March 2010

Abstract Association-based trait mapping is an innovative methodology based on linkage disequilibrium. Studies in plants, especially in cereals, are rare. A genome-wide association study of wheat is reported, in which a large number of diversity array technology markers was used to genotype a winter wheat core collection of 96 accessions. The germplasm was structured into two sub-populations. Twenty agronomic traits were measured in field trials conducted over up to eight growing seasons. Association analysis was performed with two different approaches, the general linear model incorporating the Q-matrix only and the mixed linear model including also the kinship-matrix. In total, 385 marker–trait associations significant in both models were detected. The intrachromosomal location of many of these coincided with those of known major genes or quantitative trait loci, but others were detected in regions where no known genes have been located to date. These latter presumptive loci provide opportunities for further wheat improvement, based on a marker approach.

Keywords Agronomic traits · General linear model · Genetic mapping · Linkage disequilibrium · Marker–trait associations · Mixed linear model · *Triticum aestivum*

Electronic supplementary material The online version of this article (doi:10.1007/s11032-010-9411-7) contains supplementary material, which is available to authorized users.

K. Neumann · A. Börner (✉)
Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK), Corrensstraße 3, 06466 Gatersleben, Germany
e-mail: boerner@ipk-gatersleben.de

B. Kobiljski · S. Denčić
Institute of Field and Vegetable Crops, M. Gorkog 30, 21000 Novi Sad, Serbia

R. K. Varshney
International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, Greater Hyderabad 502 324, India

Introduction

Most agronomically important traits in cereals are quantitatively inherited, making the genes underlying variation for these traits hard to detect. Based on mapping in classical mapping populations developed from a cross of two parents (recombinant inbred lines, doubled haploid lines, etc.), many quantitative trait loci (QTL) have been identified over the last decade. Examples for wheat are given by Börner et al. (2002), Quarrie et al. (2005), Huang et al. (2006) and Kumar et al. (2007). An alternative means of detecting QTL is based on correlating genotype with phenotype in germplasm collections or natural populations, and this is commonly referred to as association mapping. The underlying principle of this approach is that linkage disequilibrium (LD) tends to be maintained over many generations between loci

which are genetically linked to one another. High LD is expected to be observed between loci in tight linkage, as recombination events since the mutation should have eliminated LD between loci that are not in close distance (Breseghello and Sorrells 2006b).

The two major advantages of association mapping over standard genetic mapping based on populations of bi-parental crosses are, first, that a much larger and more representative genepool can be surveyed, and, second, that it bypasses the expense and time of mapping studies through making the crossing cycles in population development unnecessary and enabling the mapping of many traits in one set of genotypes. Another advantage is the much finer mapping resolution, resulting in small confidence intervals of the detected loci compared to classical mapping, where the identified loci need to become fine-mapped (Remington et al. 2001). However, the statistical tools required to perform the analysis are more complex, since false-positive associations between a marker and a trait can be caused by the existence of a population structure causing a seeming LD between loci that are not linked or only in loose linkage. For dealing with such a hidden population structure, statistical options have been developed (Falush et al. 2003). This situation can occur through selection, which can be artificial or natural, genetic drift and species-dependent characters such as the mating system (Flint-Garcia et al. 2003). In our crop plants, the domestication and breeding process may have caused such LD. Methods to deal with population structure range from the quantitative transmission disequilibrium test (QTDT) through genomic control (GC) to structured association (SA) (Yu and Buckler 2006). In the latter, a set of random markers is used to estimate population structure (Q), which is incorporated in a general linear model for testing associations. This has been widened to a mixed linear model, including also the kinship relations of the samples, and offers improved control of both type I and type II error rates as described by Yu et al. (2006). Little is known about a direct comparison of these approaches in identifying loci with real empirical data.

Pioneering association mapping studies in plants began only a few years ago (Ivandic et al. 2002; Skøt et al. 2004; Aranzana et al. 2005). For some traits, the approach can be simplified by searching for LD between a number of candidate genes and the trait (the "candidate gene approach"), but where no such prior information is available, the whole genome needs to be scanned in a search for LD between a framework of mapped marker loci and the target trait ("genome-wide association analysis"). The candidate gene approach has been the more widely applied (e.g., Tommasini et al. 2007; Matthies et al. 2008), while genome-wide association mapping in cereals has not often been attempted to date (Breseghello and Sorrells 2006a; Kraakman et al. 2006; Roy et al. 2006), largely because of the impracticality of genotyping large numbers of entries at the required number of marker loci. However, the development of high-throughput systems such as diversity array technology (DArT) (Jaccoud et al. 2001) has overcome this difficulty, since it provides a highly multiplexed platform, which allows for rapid and cost-effective genome-wide genotyping (Wenzl et al. 2004). DArT markers are bi-allelic dominant anonymous markers obtained by cloning random fragments of genomic representations. A first analysis of sequencing these markers showed that most of them are derived from the genespace (Wenzl et al. 2006).

A first genome-wide association mapping study in wheat exploiting DArT marker technology was published by Crossa et al. (2007). The authors analysed historical multi-locational field trials for both grain yield and resistance to various foliar diseases. Numerous marker–trait associations (MTAs) were detected. Here, we report a DArT-based association study of a breeder's core collection of 96 winter wheat accessions of diverse origin, evaluated for a set of 20 agronomic characters over a number of seasons. MTAs were identified using both the general linear model (GLM) and the mixed linear model (MLM). Association analysis based on elite lines and breeding material has the power also to detect loci for traits with low heritability such as yield and its components (Breseghello and Sorrells 2006b). In breeding programs, lines are evaluated over a broad range of locations and years or both, which gives a useful database for the analysis.

Materials and methods

Plant materials and phenotypic evaluation

The test set of 96 winter wheats comprised entries from 21 countries across five continents (Table 1),

Table 1 The identity (name and origin) of the 96 entries in the germplasm set analysed, along with their sub-group Q as defined by STRUCTURE analysis (assigned to a sub-group if probability >0.5); country code from the UN website

Name	Q	Origin	Name	Q	Origin	Name	Q	Origin
Magnif 41	1	ARG	Ai-bian	1	JPN	WWMCB 2	1	USA
Cook	1	AUS	Norin 10	1	JPN	Ivanka	1	SER
Kite	1	AUS	Saitama 27	1	JPN	Mina	1	SER
Min. Dwarf	1	AUS	Tr.Compactum	1	LV	NS 22/92	1	SER
Timson	1	AUS	Inia 66	1	MEX	NS 559	1	SER
Triple Dirk B	1	AUS	Mex. 120	1	MEX	NS 602	1	SER
Triple Dirk S	1	AUS	Mex.17 bb	1	MEX	NS 63-24	1	SER
Rusalka	1	BUL	Mex.3	1	MEX	NS 66/92	1	SER
Lambriego Inia	1	CHL	S. Cerros	1	MEX	NS 79/90	1	SER
Al-Kan-Tzao	1	CHN	Vireo"S"	1	MEX	Renesansa	1	SER
Ching-Chang 6	1	CHN	F 4 4687	1	ROM	Sava	1	SER
Peking 11	1	CHN	Donska polupat.	1	RUS	Slavija	1	SER
Ana	1	CRO	Tibet Dwarf	1	TIB	Gala	2	ARG
ZG 1011	1	CRO	Tom Thumb	1	TIB	Triple Dirk B (bulk)	2	AUS
ZG 987/3	1	CRO	Mironovska 808	1	UKR	BCD 1302/83	2	MDA
ZG K 3/82	1	CRO	Benni multifloret	1	USA	Cajeme 71	2	MEX
ZG K 238/82	1	CRO	Florida	1	USA	Bezostaja 1	2	RUS
ZG K T 159/82	1	CRO	Hays 2	1	USA	Centurk	2	USA
Capelle Desprez	1	FRA	Helios	1	USA	Lr 12	2	USA
Durin	1	FRA	Holly E	1	USA	Tr.Sphaerococcum	2	USA
Avalon	1	GBR	Hope	1	USA	L 1/91	2	SER
Brigant	1	GBR	INTRO 615	1	USA	L 1A/91	2	SER
Highbury	1	GBR	Lr 10	1	USA	Nizija	2	SER
TJB 990-15	1	GBR	Norin 10/Brev.14	1	USA	Nov.Crvena	2	SER
Bankut 1205	1	HUN	Phoenix	1	USA	Nova banatka	2	SER
L-1	1	HUN	Purd./Loras	1	USA	NS 33/90	2	SER
Szegedi 768	1	HUN	Purd.39120	1	USA	NS 46/90	2	SER
Hira	1	IND	Purd.5392	1	USA	NS 55-25	2	SER
Sonalika	1	IND	Red Coat	1	USA	NS 74/95	2	SER
Suwon 92	1	IND	Semillia Eligulata	1	USA	PKB Krupna	2	SER
UPI 301	1	IND	UC 65680	1	USA	Pobeda	2	SER
Acciaio	1	ITA	Vel	1	USA	Sofija	2	SER

and was assembled from a larger core collection created at the Institute of Field and Vegetable Crops (Novi Sad, Serbia), based on contrasting phenotypic expression for the breeding traits evaluated in this study (Kobiljski et al. 2002; Quarrie et al. 2003). The entries were evaluated for 20 morphological and agronomic important characters over a number of (up to eight) seasons.

The investigated traits were heading date (Hd), flowering time (Fl), plant height (Hg), peduncle length (PL), biomass (Bm), grain yield (GY), grain number per spike (GN), grain weight per spike (GW), thousand kernel weight (TKW), harvest index (HI), spike length (SL), spike number per m^2 (SN), spike weight (SW), spike index (SI), the number of spikelets per spike (SPS), the number of sterile spikelets per spike (St), resistance against leaf rust (LR) and powdery mildew (PM), protein content (Pr) and sedimentation value (Sd). The field trials were conducted in Novi Sad (Serbia) and each entry was represented by three plots per season. The years of investigation for each trait are summarized in

Table 2 Overview of the years in which the traits were investigated (indicated by x) and the criterion for a significant marker–trait association (MTA) for the calculation with the two models (GLM and MLM) in the last column

	1993	1994	1995	1996	1997	1998	1999	2000	2001	Years	No. of years with $P < 0.05$ in the two models
Hg		x	x	x	x	x	x	x	x	8	6 in both
Hd		x	x	x	x	x	x	x	x	8	6 in both
TKW	x	x	x	x	x	x	x			7	5 in both
Fl		x	x	x	x	x	x			6	4 in both
Bm		x	x	x	x	x	x			6	4 in both
GW		x	x	x	x	x	x			6	4 in both
GN		x	x	x	x	x	x			6	4 in both
HI		x	x	x	x	x	x			6	4 in both
SW		x	x	x	x	x	x			6	4 in both
SL		x	x	x	x	x	x			6	4 in both
SI		x	x	x	x	x	x			6	4 in both
SPS		x	x	x	x	x	x			6	4 in both
St		x	x	x	x	x	x			6	4 in both
GY			x	x	x	x	x			5	3 in both
Pr			x	x	x	x	x			5	3 in both
Sd			x	x	x	x	x			5	3 in both
LR					x	x	x	x	x	5	3 in both
PM					x	x	x	x	x	5	3 in both
SN				x	x	x	x			4	3 in both
PL					x	x	x			3	3 in one model, 2 in the second

Table 2. The plot size was 1.2 m^2 and each plot consisted of six rows, separated from one another by 20 cm. Traits Hg, Hd, Fl, HI, Bm, SN, LR and PM were measured on a per plot basis, and the remaining traits were evaluated from five spikes sampled from five plants per plot. Hd and Fl were assigned when 50% of the spikes within a given plot had reached the relevant stage (respectively, emergence of the spike from the flag leaf sheath and anthesis). Harvest index represents the ratio between grain yield and biomass at harvest, and spike index the ratio between grain weight and spike weight. Biomass was calculated from the above ground biomass present in a 50 × 50 cm quadrat sampled from the centre of each plot. Infection with LR and PM were scored after flowering, at a time when disease reactions were fully developed, and were expressed as the percentage of leaf area infected. Protein content of whole meal flour was estimated by the Kjeldahl method (Tecator, Höganäs, Sweden), according to ICC standard 105/2, and the determination of sedimentation value followed the method of Zeleny according to ICC standard 116/1 (ICC 2008).

DArT assay, population structure and linkage disequilibrium

Diversity array technology profiling was performed by Triticarte Pty. Ltd (Canberra, Australia; http://www.triticarte.com.au/). Each of the 874 informative markers was designated by the prefix '*wPt*'. Alleles occurring at a low frequency ($f < 0.05$) were excluded from all analysis, resulting in the removal of 39 DArT markers. Assignment of 525 of the markers to linkage groups was based on Crossa et al. (2007). All chromosomes except 6D are covered by markers but the D-genome in general only poorly (the whole genetic map is shown in Electronic Supplementary Material Fig. 1). Most markers map to only one position in the genome but 24 markers map to more than one and are

considered as multilocus markers. The chromosomal location of at least 177 of the 315 unmapped markers was provided by Triticarte.

To determine the population structure of the test material, a subset of the genotypic data (219 markers) was processed by the software program STRUCTURE (Pritchard et al. 2000), applying the admixture model, a burn-in of 10,000 iterations and a 10,000 MCMC duration to test for a K value in the range 1–12. The likely number of sub-populations present was estimated following Evanno et al. (2005), in which the number of sub-groups (ΔK) is maximized.

Linkage disequilibrium was identified with the software program TASSEL 2.01 (Bradbury et al. 2007). Allele frequency correlations (r^2) according to Weir (1996) were calculated using the LD function. The significance of pair-wise LD (P-values) was computed using 1,000 permutations. LD was calculated separately for loci on the same chromosome (intrachromosomal pairs) and unlinked loci (interchromosomal pairs). From the latter a critical value of r^2 was estimated following Breseghello and Sorrells (2006a) in root transforming the r^2-values and taking the 95% percentile of this distribution as the threshold, beyond which LD is likely caused by real physical linkage. The intrachromosomal r^2-values were plotted against the genetic distance and a LOESS curve was fitted to the plot to check at which distance it intercepts the line of the critical r^2 in order to see how fast the LD decay occurs. The LD analysis for the mapped markers (without multilocus markers) was performed for the whole wheat collection and separately for the Q-groups from STRUCTURE. LD was also estimated among the unmapped markers, and between the unmapped and mapped markers (including the multilocus markers but only considering one position) but here only for the whole population and not for the subgroups.

Phenotype–genotype association analysis

The software program TASSEL 2.01 was used to calculate associations between the markers and each trait in turn, employing the general linear model (GLM) based on the chosen Q-matrix derived from STRUCTURE. With the newer version TASSEL 2.1 the mixed linear model (MLM) suggested by Yu et al. (2006) was additionally implemented using Q-Matrix and the kinship-Matrix, which was also calculated with the version 2.1 in TASSEL considering all mapped markers. The Kinship-Matrix is generated by TASSEL through converting the distance matrix calculated from TASSEL's Cladogram function to a similarity matrix. In the MLM approach the option to calculate the heritability separately for each marker was used. The EMMA method (Kang et al. 2008) was chosen and the MLM parameters were left at the default settings from TASSEL. Each trait was represented by its mean of the three plots in each season and the years were separately analysed. Depending on the final number of investigated years, different criteria for a final significant marker–trait association (MTA) were defined (Table 2). Statistical calculations as correlations between years and between traits were managed through SPSS v16.0.

Results

Population structure and trait correlations

According to the method of Evanno et al. (2005), ΔK was plotted against the number of sub-groups K. The maximum value of ΔK occurred at $K = 2$, so each entry was assigned to one of two sub-populations. The majority of entries (76) belonged to sub-group 1 (Q1) whereas the smaller group 2 (Q2) consisted of the remaining 20 genotypes (Table 1). This structure is consistent with the known origin and pedigrees of the material. Thus, for example, 'Nova Banatka' and 'Nizija' are both descended from 'Bezostaja 1', and all three of these cultivars belonged to sub-group Q2, which largely consists of European (mainly Serbian) elite lines and cultivars. Most of the non-Serbian materials belonged to the bigger group Q1.

The two sub-groups differed with respect to grain yield, thousand kernel weight, spike index, leaf rust resistance and sedimentation value. Q2 entries had significantly higher GY, TKW, SI and Sd, and were more sensitive to LR (ANOVA, $P < 0.01$ in all cases).

The wheat collection offered partially large phenotypic differences (Table 3). Across years, with one exception all traits had significant correlations; the highest Pearson correlation coefficient appeared for height (Min = 0.91, Max = 0.97), followed by heading (Min = 0.88, Max = 0.96), flowering (Min = 0.81, Max = 0.95), spike length (Min = 0.82, Max =

Table 3 Descriptive statistics for the 20 agronomic traits

Trait	Minimum	Maximum	Mean	Std. deviation	Variance
Fl	128.7	150.2	138.7	5.3	27.6
Hd	121.3	146.4	133.2	5.8	33.9
Hg	23.2	122.8	80.4	19.7	386.1
PL	12.7	59.1	33.1	7.6	58.3
Bm	1.03	2.25	1.68	0.26	0.07
GY	2.66	9.18	6.12	1.45	2.09
GN	20.3	77.6	44.1	9.8	95.9
GW	0.49	2.95	1.51	0.38	0.14
TKW	23.0	54.0	36.1	6.3	39.6
HI	0.180	0.480	0.366	0.062	0.004
SL	4.07	16.90	9.77	1.76	3.09
SN	279.0	1,086.0	608.4	131.2	17,206.7
SW	0.96	4.50	2.02	0.51	0.26
SI	0.59	0.85	0.75	0.05	0.00
SPS	15.5	52.2	20.8	5.5	30.2
St	0.62	9.87	2.28	1.36	1.86
LR	1.4	88.0	39.5	21.4	460.0
PM	1.4	53.0	17.3	10.4	108.7
Pr	10.32	15.94	12.36	1.15	1.32
Sd	20.6	53.4	34.5	6.8	46.9

0.93) and peduncle length (Min = 0.80, Max = 0.86). The Pearson correlations for TKW, grain number, spike weight and spikelets per spike ranged between 0.60 and 0.86 (mean = 0.73). Moderate correlations were observed for grain weight, harvest index, spike number, sterile spikelets per spike, protein content and leaf rust scores (Min = 0.42, Max = 0.88, mean = 0.65). The mean correlation for the traits biomass, spike index, grain yield and sedimentation value was only 0.49 (Min = 0.25, Max = 0.77). The incidence of powdery mildew was low in 2000, so there was no significant correlation between PM in 2000 and the other years. Even without the 2000 data the correlation between years were the lowest for all traits (mean for all correlations only 0.36).

Running partial correlations and taking the factor year into account, several significant correlations were found (Table 4). The highest Pearson correlation coefficients were observed for the correlation between grain weight and spike weight (0.88) and for plant height and peduncle length (0.84). In addition, moderately high correlations (0.66–0.72) exist between spike weight and grain number, grain weight and grain number as well as for the number of spikelets and the number of sterile spikelets. Moderate correlations were found between spike length and spike weight, grain number and number of spikelets and for grain weight and spike length ($\sim$0.5). Another moderate correlation was found for heading and harvest index but here the relationship is negative. The correlation between heading and flowering, which is very high (0.976) considering only the mean values of all years, is only low (0.36) taking the factor year into account, which indicates a high influence of the environment for this interaction. Most correlations for Fl/Hd to the yield-related traits are negative and of only low strength (except Hd and HI with higher correlation). Also the (low) correlations between LR resistance and Pr to yield characters were negative (note that for LR negative correlations should be interpreted as positive, because a higher disease score indicates greater susceptibility, not a higher level of resistance).

Linkage disequilibrium

In the entire collection under investigation 1,549 (14.9%) of the 10,412 intrachromosomal marker pairs showed a significant level of LD ($P < 0.01$). The average of r^2 for all pairs was 0.069. Although significant decay of LD with the increase of genetic distance was observed in our population, the correlation between the LD *P*-value and genetic distance in cM was relatively low (0.186, $P < 0.0001$). According to Maccaferri et al. (2005), four classes of marker pairs were defined: class 1—tight linkage (distance < 10 cM); class 2—moderate linkage (10–20 cM); class 3—loosely linked (20–50 cM); and class 4—independent pairs (>50 cM). The amount of significant LD and the r^2 values in the entire collection and in the two subgroups are listed in Table 5. In all classes the values for the group Q1 are very close to the values of the whole collection, in the amount of pairs in LD as well as in the mean r^2 values. The group Q2 shows much fewer pairs in significant LD but the mean r^2 values are much higher in all genetic classes. Furthermore the number of pairs in total LD is much higher in this group (416 pairs compared to 205 in the whole collection and 210 in Q1). The pairs in total LD have a mean distance of 11.6 cM in the Q2, whereas in the Q1 and the entire collection it spans only over a mean distance of 3.6 cM. Also the critical r^2 for the Q2 is, with 0.535, higher compared to the whole collection (0.263) and the Q1

Table 4 Significant phenotypic trait correlations (partial correlations controlling for year) with mean values for each genotype for each single year

	Fl	Hd	Hg	PL	Bm	GY	GN	GW	TKW	HI	SL
Fl	1										
Hd	0.356***	1									
Hg			1								
PL			0.841***	1							
Bm			0.275***	0.230***	1						
GY		−0.216***	0.148*	0.143*	0.374***	1					
GN							1				
GW		−0.247***			0.155**	0.383***	0.661***	1			
TKW	−0.145*	−0.471***	0.166**	0.201**	0.264***	0.439***	−0.254***	0.409***	1		
HI		−0.536***	−0.225***	−0.146*		0.428***		0.358***	0.443***	1	
SL		0.134*	0.134*	0.191*			0.472***	0.506***			1
SN			0.122*		0.132*		−0.398***	−0.397***			−0.405***
SW					0.119*	0.284***	0.724***	0.881***	0.308***	0.260***	0.563***
SI	−0.121*	−0.408***	0.275***	0.250***	0.205**	0.443***		0.336***	0.455***	0.383***	−0.125*
St		0.281***	0.213***	0.137*		−0.251***	0.149*		−0.271***	−0.363***	0.320***
SPS		0.329***	0.158**	0.133*	0.121*		0.530***	0.276***	−0.208***	−0.214***	0.471***
LR	0.191**					−0.120*	−0.151*	−0.206***			−0.235***
PM										0.185**	−0.186**
Pr		0.226***				−0.310***	−0.213***	−0.252***		−0.395***	
Sd		−0.157**	0.117*	0.176**	0.181**	0.157**	−0.228***		0.219***		

	SN	SW	SI	St	SPS	LR	PM	Pr	Sd
Fl									
Hd									
Hg									
PL									
Bm									
GY									
GN									
GW									
TKW									
HI									
SL									

Table 4 continued

	SN	SW	SI	St	SPS	LR	PM	Pr	Sd
SN	1								
SW	−0.354***	1							
SI			1						
St			−0.171**	1					
SPS	−0.287***	0.391***	−0.169**	0.686***	1				
LR		−0.222***		−0.303***	−0.286***	1			
PM				−0.321***	−0.144*	0.286***	1		
Pr		−0.210***	−0.159**	0.131*				1	
Sd	0.204**		0.161**	−0.123*	−0.172**			0.134*	1

Significance level indicated with asterisks as follows: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$

(0.279). Beyond this critical r^2 value, LD is likely to be caused by genetic linkage. For the entire collection and the two subgroups LOESS curves were fitted to the LD plots, which do not intercept the line of the critical r^2 (Fig. 1a). This reflects that the overall LD decayed fast even in the Q2 group. The fast decay of LD for the groups and the whole collection is better visible in Fig. 1b. The Q2 follows the same pattern in decay of LD as the Q1 and the total collection but from a higher level of r^2 values. In the complete collection and the Q1, mean LD decays permanently below 0.15 within 4 cM; the Q2 falls below the 0.15 threshold only from a distance over 18 cM.

In the complete collection only 614 of the pairs are in LD ($P < 0.01$) because of physical linkage, as their r^2 values are higher than the critical value of 0.263. The frequency of physically linked pairs versus non-physically linked pairs according to higher genetic distance follows a logarithmic function (Fig. 2). The mean r^2 value for the physically linked pairs is 0.695, mainly due to a high number of pairs in complete LD (205); without these pairs the mean is 0.542. In 152 of the totally linked pairs, the two marker loci are separated from one another by less than 5 cM. The largest separation between markers in complete LD is 45.4 cM, on chromosome 7A. A further 19 pairs in complete LD have a distance between 10 and 30 cM. Though high LD over long distances exists, r^2 decreases fast for many close marker pairs and so the mean r^2 for all pairs with a distance of 0 cM is only 0.502.

An overview of the interchromosomal LD is given in Table 6. The mean r^2 is much lower than for the intrachromosomal LD. Once more the Q2 shows higher r^2-values and a lower number of significant pairs. The Q1 is only slightly higher in mean r^2 and slightly lower in the number of significant pairs. For the entire collection significant LD was detected for 3.2% of the 115,215 possible marker pairs. Of these, 24 were in complete LD. Much of this unexpected total LD resulted from the interaction of two markers (one on chromosome arm 3AS, the other on 7DS) with eight other markers mapping within a 17-cM interval of chromosome arm 1BS (17 from the 24 pairs). Another four pairs resulted from an interaction between a single marker on 6BL with four mapping close to the centromere of chromosome 3B. In the Q1 as in the whole collection, the same 24 pairs are in total LD. In the Q2 the number of totally linked pairs is remarkably high; here 66 pairs are in complete LD.

Table 5 Overview of LD in the intrachromosomal pairs in the whole population (Pop) and in the two Q-groups

	Class	Pairs total	Mean r^2 of all pairs	No. significant pairs	% Significant pairs	No. physically linked pairs	% Physically linked pairs	Mean r^2 for physically linked	No. pairs in complete LD
Pop	Class 1	2,405	0.189	922	38.3	518	56.2	0.699	185
Pop	Class 2	1,530	0.049	208	13.6	40	19.2	0.657	14
Pop	Class 3	3,326	0.037	290	8.7	55	19.0	0.692	6
Pop	Class 4	3,151	0.020	129	4.1	1	0.8	0.378	0
Pop	Total	10,412	0.069	1,549	14.9	614	39.6	0.695	205
Q1	Class 1	2,405	0.187	791	32.9	501	63.3	0.704	190
Q1	Class 2	1,530	0.049	126	8.2	35	27.8	0.686	14
Q1	Class 3	3,326	0.038	206	6.2	59	28.6	0.640	6
Q1	Class 4	3,151	0.023	90	2.9	1	1.1	0.359	0
Q1	Total	10,412	0.069	1,213	11.7	596	49.1	0.696	210
Q2	Class 1	1,862	0.294	243	13.1	220	90.5	0.920	315
Q2	Class 2	1,094	0.138	47	4.3	35	74.5	0.805	34
Q2	Class 3	2,363	0.101	216	9.1	46	21.3	0.890	49
Q2	Class 4	2,342	0.076	16	0.7	9	56.3	0.910	18
Q2	Total	7,661	0.146	522	6.8	310	59.4	0.902	416

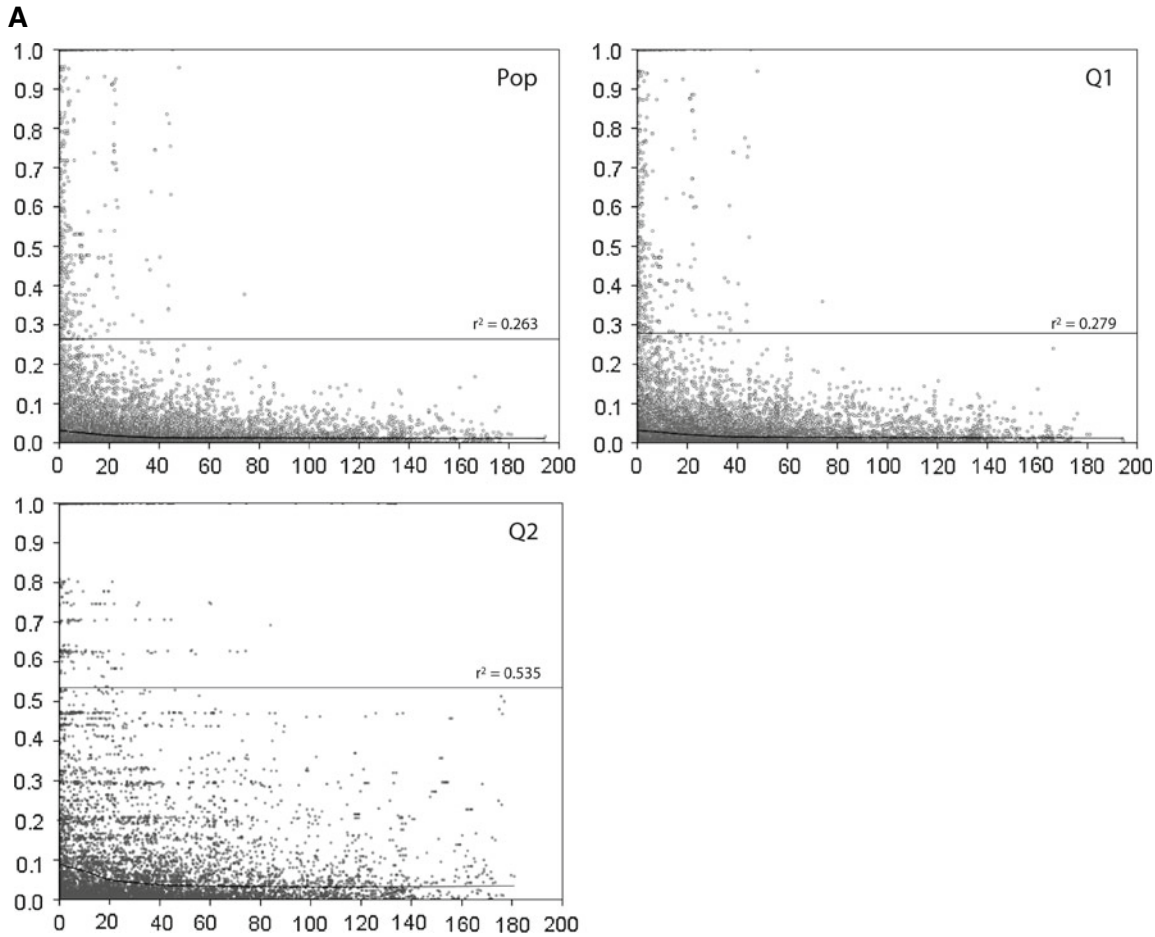
Class 1—genetic distance <10 cM, class 2—distance >10, <20 cM, class 3—distance >20, <50 cM, class 4—>50 cM; mean allele frequency correlations (r^2) for all pairs, number (no.) of pairs and percentage (%) significant in LD ($P < 0.01$), no. and % of physically linked pairs ($r^2 >$ critical r^2, $P < 0.01$), no. of pairs in complete LD ($r^2 = 1$)

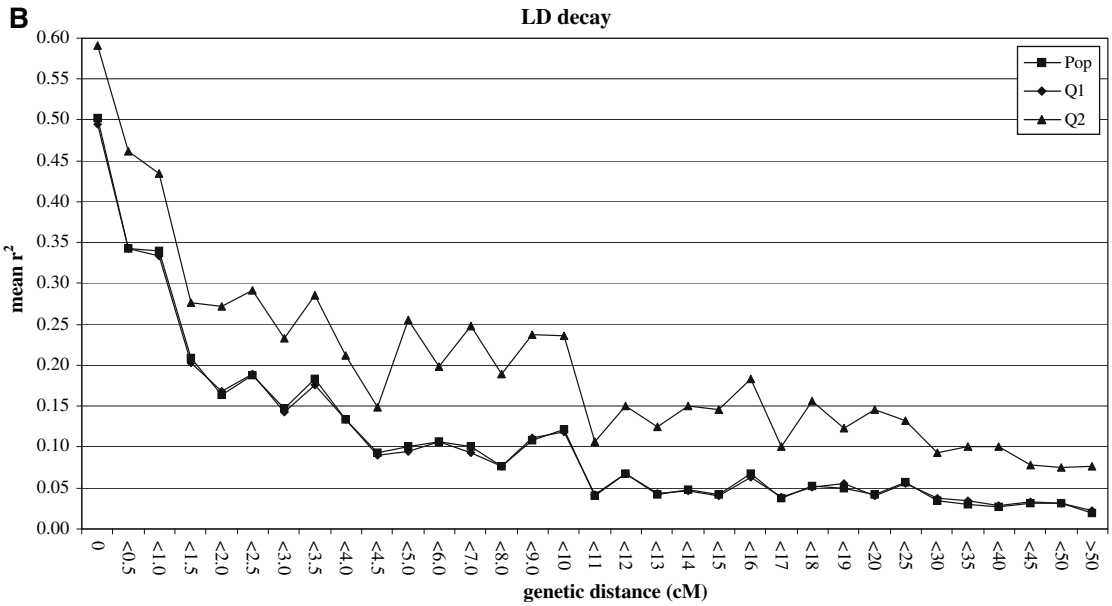
Out of these, 17 pairs result from an interaction between 3 markers on 2B (position at ~58 cM) and 6 markers on 6B (position ~40–46 cM). The other 49 pairs in complete LD are spread throughout the chromosomes and cannot be traced back to a peculiar interaction. The 17 marker pairs in total LD (3A and 7D with 1B) described for the collection and the Q1 are here not in significant LD ($P \sim 0.05$) though their r^2 values are also 1. The same is true for the interaction of 6BL with 3B markers ($P \sim 0.06$) indicating that the small population size makes it difficult to detect significant LD through a permutation test even if pairs are totally linked.

Of the unmapped markers, 1,809 (3.5%) of the 51,677 pairs were in significant LD with one another, of which 69 were in complete LD. Considering the LD between unmapped and mapped markers, it was found that from the 171,434 pairs in total, 5,991 (3.5%) were in significant LD ($P < 0.01$). 101 mapped–unmapped marker pairs were in total LD.

Comparison of the models

The two model approaches were compared for all traits except PL, because the significance criteria there were not equal for both models. In general the number of significant marker–trait associations (MTAs) with the GLM is much higher than with the MLM. For all markers, mapped and unmapped, numbers of 999 and 405 MTAs (note: one marker might have MTAs to more than one trait) was found with the GLM and the MLM, respectively. Thus, with the MLM the amount of MTAs was reduced by 60%. 352 identical MTAs are shared by the two models and fulfilled the significance criteria in both. 186 of the MTAs, detected as significant with the GLM only, seem to be present in the MLM as well but fulfilled the significance criteria defined for the traits only on the $P < 0.1$ level. Very few cases (27) were observed where a MTA was fully significant in the MLM and in the GLM only on the $P < 0.1$ level. Excluding these almost-identical MTAs, 461 unique for the GLM and 26 for the MLM remained. Overall, the number of these unique MTAs for GLM was 46.1% compared to the MLM where it was only 6.4%. Regarding the traits separately (Table 7), large differences appear for these values. Thus, for traits like SPS, HI, Hg, Fl, Hd and PM, the amount of unique GLM-MTAs ranged from 50 to 86%, in contrast to Sd, TKW, GN, GY and SN where this


A
Pop
r² = 0.263
Q1
r² = 0.279
Q2
r² = 0.535
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0 20 40 60 80 100 120 140 160 180 200


B
LD decay
Pop
Q1
Q2
mean r²
0.60
0.55
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0 <0.5 <1.0 <1.5 <2.0 <2.5 <3.0 <3.5 <4.0 <4.5 <5.0 <6.0 <7.0 <8.0 <9.0 <10 <11 <12 <13 <14 <15 <16 <17 <18 <19 <20 <25 <30 <35 <40 <45 <50 >50
genetic distance (cM)

◀ **Fig. 1** Overview of LD parameter r^2 of the intrachromosomal pairs in the whole population (Pop) and the two Q-groups (Q1 and Q2). **a** Scatterplots showing the distribution of the LD parameter r^2 plotted against the genetic distance in cM. *Horizontal line* indicates the 95% percentile of the distribution of unlinked r^2, which gives the critical value of r^2. Second degree LOESS curve fitted to the plot (*black bottom line*). **b** Decay of LD as mean r^2 values for different centimorgan classes

amount ranged only from 0 to 6%. Fl and Hd had by far the highest amount of unique GLM associations. Unique MLM-MTAs ranged mainly from 0 to 7% for most traits but for Bm, St and Hg the amount was 17–24% and for PM up to 33%. Nevertheless, the absolute number of unique associations was a maximum of 8 for the MLM which is very low compared to the GLM with a maximum of 111. For PM, only two MTAs were significant in both models, which was the lowest number of MTAs of all traits shared between the two models. The number of unique MLM-MTAs was higher than unique GLM-MTAs for only two traits (GY and GN). For another two traits (TKW and Sd) there were no unique MLM- or GLM-MTAs at all, meaning that they were fully significant in both models, or at least in one model and in the other on the $P < 0.1$ level. Comparison between GLM and MLM P values were made for a trait with a high amount of unique GLM-MTAs (Hd), a medium amount (GW) and an amount of zero unique associations (GN). The distribution of the P values for the two models for these traits is shown in Fig. 3. The P values for Hd showed the biggest differences between the two models; the GLM gave smaller P values than the MLM. This difference is reduced for the medium unique GLM-MTAs trait GW and not visible for GN, which had no unique GLM-MTAs. Correlations between P values for GLM and MLM are therefore lowest for Hd (0.53), medium for GW (0.76) and highest for GN (0.86). Thus, overall a reduction of significant associations through implementing the kinship relations took place, but the influence is very variable for single traits and there are also traits where these relations seem to have no great influence.

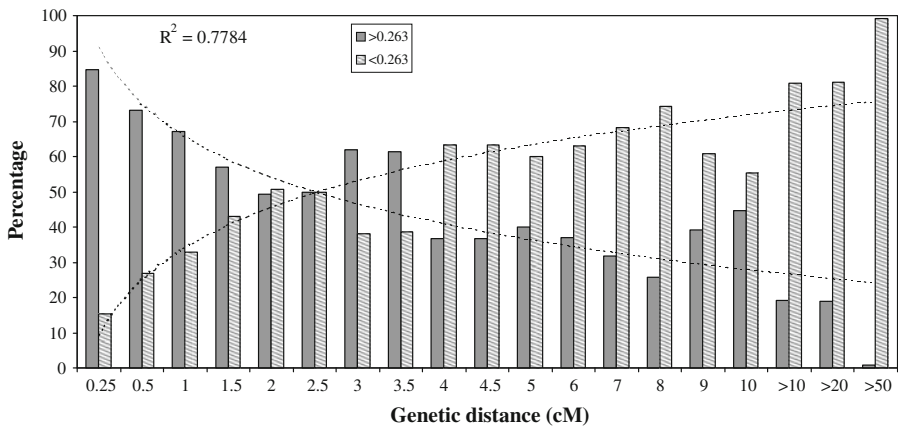


Fig. 2 Percentage of intrachromosomal marker pairs in LD ($P < 0.01$) with r^2 higher and below the critical value of r^2 (0.263) as a logarithmic function of genetic distance (distance in classes) in the entire collection

Table 6 Interchromosomal LD in the whole population (Pop) and in the two Q-groups; mean allele frequency correlations (r^2) for all pairs, number (no.) of pairs and percentage (%) significant in LD ($P < 0.01$), no. of pairs in complete LD ($r^2 = 1$)

	Pairs total	Mean r^2 of all pairs	No. significant pairs	% Significant pairs	No. pairs in total LD
Pop	115,215	0.019	3,646	3.2	24
Q1	115,215	0.021	2,366	2.1	24
Q2	69,936	0.071	268	0.4	66

Table 7 Comparison of the two models for the calculation of associations between markers and traits

Trait	No. GLM sig.	No. MLM sig.	No. sig. in GLM and MLM	No. sig. GLM/ $P < 0.1$ in MLM	No. sig. MLM/ $P < 0.1$ in GLM	No. GLM unique	No. MLM unique	% GLM unique	% MLM unique
Fl	159	33	31	17	0	111	2	69.8	6.1
Hd	142	28	25	11	2	106	1	74.6	3.6
Hg	122	34	23	26	3	73	8	59.8	23.5
Bm	25	18	15	5	0	5	3	20.0	16.7
GY	26	18	16	9	1	1	1	3.8	5.6
GN	37	31	28	9	2	0	1	0.0	3.2
GW	30	14	13	8	1	9	0	30.0	0.0
TKW	13	11	11	2	0	0	0	0.0	0.0
HI	71	12	11	19	1	41	0	57.7	0.0
SL	52	35	29	14	4	9	2	17.3	5.7
SN	33	21	20	11	1	2	0	6.1	0.0
SW	31	15	15	6	0	10	0	32.3	0.0
SI	21	14	13	4	0	4	1	19.0	7.1
SPS	103	39	38	13	1	52	0	50.5	0.0
St	65	35	23	14	6	28	6	43.1	17.1
LR	25	17	16	6	1	3	0	12.0	0.0
PM	8	3	2	2	1	4	0	50.0	0.0
Pr	26	17	14	9	2	3	1	11.5	5.9
Sd	10	10	9	1	1	0	0	0.0	0.0
Total	999	405	352	186	27	461	26	46.1	6.4

GLM = general linear model (Q-matrix), MLM = mixed linear model (Q-matrix + kinship-matrix), sig. = significant with $P < 0.05$, number (no.) of GLM/MLM unique refers to associations only significant in that model and not even significant at the $P < 0.1$ level in the other model; peduncle length was excluded from this comparison

Trait mapping

Regarding significant associations, we present only the MTAs fully significant in both models. Altogether 385 MTAs were identified for 215 of the DArT markers, mapped and unmapped. From that number of markers, 118 are specific for a single trait, the rest consists of associations of up to six traits. SPS was involved in the highest number of MTAs (38), followed by PL and Fl (each 31). The fewest MTAs were associated with PM (2), Sd (8), TKW and HI (each 11). The strong phenotypic correlation between Hg/PL and SPS/St was reflected in the large number of shared MTAs for these traits (16 and 14, respectively). Nevertheless, the highest number of overlapping MTAs shared Fl and Hd (22), though having a lower correlation coefficient when taking the factor year into account. The highest number of trait-specific MTAs were detected for the traits PL (14), Bm (13), GN and SL (each 12). Considering the total number of MTAs found for these traits, Bm has by far the highest amount (87% vs. 41–45%) of trait-specific associations.

Mapped markers

The 115 associated mapped markers are included in the genetic map shown in Fig. 4. Each trait was associated with at least one (PM) and maximum ten (Fl) different chromosomes. Eighteen of the chromosomes (all except 3D, 4D and 6D; the last is not covered by the map) were involved in MTAs. Of a total of 115 associated markers, 71 were associated with only one trait and therefore can be called trait-specific MTAs; the other 44 were associated with up to six traits and are further referred as multi-trait MTAs. Excluding the multilocus markers, the highest number of associated markers were found on chromosome 7B (11), followed by 5B and 7A (10), with the

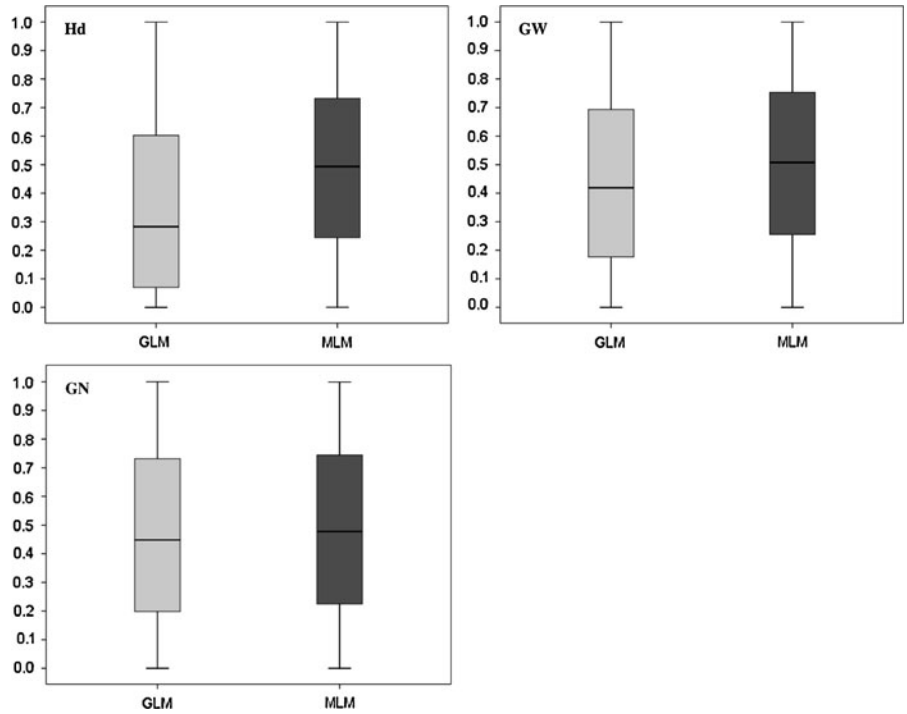


Fig. 3 Boxplots of *P* values for the general linear model (GLM) and the mixed linear model (MLM) for the traits heading (Hd), grain weight (GW) and grain number (GN)

lowest on 5D and 7D (only 1). Considering the homoeologous groups, group 7 contained the highest number (22) followed by group 1 with 17 associated markers; the lowest number was found for the groups 4, 5 and 6 (14 each). Altogether the B-genome offered the highest number of markers involved in MTAs (57), followed by the A- (44) and D-genomes (10).

In the following, the identified MTAs are described for each trait, and cM-positions for the trait-specific MTAs are given in brackets. For *flowering time* thirteen markers were identified as significant; three of them are only associated with this trait, while all other markers carry other trait-associations too, mainly for heading. Altogether 10 different chromosomes were involved: 1B, 1D (2 markers), 2B, 2D, 4B, 5B, 5D, 6A (2 markers), 6B and 7A (2 close markers). The trait-specific markers are located on 1D (82.2), 5B (77.9) and 6A (3.9). *Heading date* was associated with eleven markers; nine coincide with Fl-MTAs, the other two with different traits. These eleven multi-trait MTAs are spread over the ten chromosomes 1B, 1D, 2D, 4A, 4B, 5D, 6A (2 markers), 6B and 7A (2 close markers). For Hd no trait-specific MTA was found. Nevertheless, six MTAs specific only for Fl and Hd were found on 1B (45.6), 5D (17.6), 6A (20.6), 6B (47.6) and on 7A (88.8, 89.6).

Plant height and *peduncle length* shared six MTAs which were specific for these two traits. They are located on 1A (83.3), 2B (68.4), 4A (183.6) and 7B (138.0, 144.7, 156.1), and all MTAs except the 1A locus were significant in all investigated years. The three loci on 7B are all in high LD with each other (>0.9). Five markers were associated only with Hg; they are located on 1B (44.7), 4A (174.6), 6B (181.0) and 7A (69.6 and 70.3, both markers in complete LD). Only one MTA for Hg coincides with a trait other than PL; this MTA for Hg and Hd was found on 6A (77.7). Another seven markers are specific only for PL and were identified on 1A (98.9), 2B (56.1, 57.8), 5A (34.7), 5B (160.5), 7A (129.3) and for a multilocus marker *wPt5707* (1A/2B, 41.5/57.4).

The twelve MTAs identified for *biomass* are all trait-specific and are spread over the chromosomes 1D (49.9), 3B (87.8, 104.4), 4B (111.9), 5B (82.9, 130.9), 6A (26.5, 37.4), 6B (65.5) and 7B (43.3, 101.5) as well as the multilocus marker *wPt7599* mapping to 6A/6B (56.5/7.2). The 6A marker at 37.4 cM is in high LD (>0.9) with the two loci on 5B, which are in high LD with each other as well.

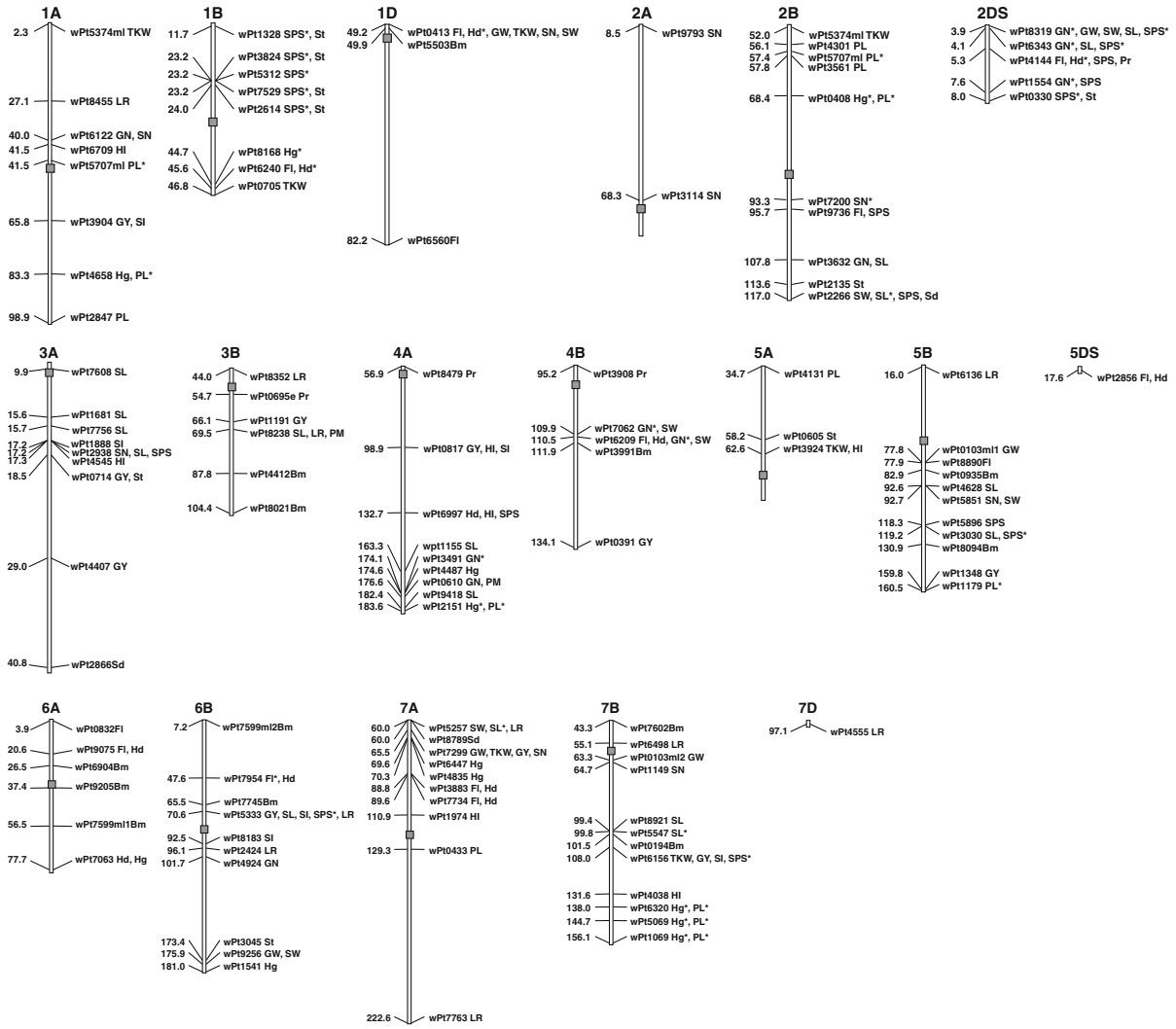


Fig. 4 Genetic map including only the markers with marker–trait associations (MTAs). Approximate position of centromeres indicated by *grey squares*. *Asterisks* indicate MTAs significant in all years investigated

The trait *grain yield* was involved in six multi-trait MTAs, which are spread over 1A, 3A, 4A, 6B, 7A and 7B. Four GY-specific MTAs were found on 3A (29.0), 3B (66.1), 4B (134.1) and on 5B (159.8). For *grain number* two of the ten MTAs are trait-specific and are located on 4A (174.1) and 6B (101.7); the 4A locus is significant in all years. The eight multi-trait MTAs involving GN were found on 1A, 2B, 2D (3 markers), 4A and 4B (2 close markers). *Grain weight* MTAs were identified on 1D, 2D, 6B and 7A and for the multilocus marker *wPt0103* (5B/7B, 77.8/63.3), which is the only trait-specific MTA for GW. For the *thousand kernel weight* four multi-trait MTAs were identified on 1D, 5A, 7A and 7B. Two TKW-specific MTAs were found for the two multilocus markers *wPt0705* (1B/1B, 46.7/47.4) and *wPt5374* (1A/2B, 2.3/52.0). Specific MTAs for *harvest index* are located on 1A (41.5), 3A (17.3), 7A (110.9) and 7B (131.6). Three more multi-trait MTAs for HI were found on 4A (2 markers) and 5A. Several multi-trait MTAs where *spike length* is involved were identified on 2B (2 close markers), 2D (2 close markers), 3A, 3B, 5B, 6B and 7A. SL-specific MTAs are located on 3A (9.9, 15.6, 15.7), 4A (163.3, 182.4), 5B (92.6) and 7B (99.4, 99.8). All three 3A markers are in complete LD as well as the two markers on 4A. One of the

markers on 7B (99.8) is significant in all years. For *spike number* the five multi-trait MTAs are located on 1A, 1D, 3A, 5B and 7A. Specific SN MTAs were identified on 2A (8.5, 68.3), 2B (93.3) and 7B (64.7); the 2B locus is significant in all years. *Spike weight* was associated only with multi-trait MTAs. These eight are located on 1D, 2B, 2D, 4B (2 close markers), 5B, 6B and 7A. Four non-trait-specific MTAs on 1A, 4A, 6B and 7B were identified for *spike index*. The two SI-specific MTAs are located on 3A (17.2) and 6B (92.5).

For the *number of spikelets per spike* and *sterile spikelets per spike* we identified five associated markers on 1B which are all in complete LD with a position of 11.7 and ~23 cM. One of these markers is only associated with SPS but only marginally misses the significance for St. All the 1B markers for SPS are highly significant in all investigated years ($P < 0.01$). One more SPS-St MTA is located on 2D (at 8 cM). One marker associated only with SPS was found on 5B (118.3). Multi-trait MTAs involving SPS are located on 2B (2 markers), 2D (4 markers), 3A, 4A, 5B, 6B and 7B. Besides the SPS-St MTAs already mentioned, only one more multi-trait MTA for St on 3A (St and GY) was found. Three St-specific markers were identified on 2B (113.6), 5A (58.2) and 6B (173.4).

Most of the MTAs where *leaf rust resistance* is involved are trait-specific. They are spread over 1A (27.1), 3B (44.0), 5B (16.0), 6B (96.1), 7A (222.6), 7B (55.1) and 7D (97.1). The three multi-trait MTAs for LR are located on 3B, 6B and 7A. For *powdery mildew resistance* only two MTAs were identified, both multi-trait MTAs located on 3B (69.5) and 4A (176.6). The *protein content* was involved in one multi-trait MTA on 2D. The three Pr-specific MTAs were found on 3B (54.7), 4A (56.9) and 4B (95.2). The *sedimentation value* was involved in one multi-trait MTA on 2B. Two Sd-specific markers were identified on 3A (40.8) and 7A (60.0).

Unmapped markers

Of the 315 unmapped markers, 100 were associated with at least one single trait; 53 of them were involved in multi-trait MTAs with at least two associated traits (Electronic Supplementary Material Table 1). No associated marker was found for PM; the other traits were involved in at least three (Bm) up to 20 MTAs (SPS). For some of these associated markers a strong or complete LD was detected to mapped markers, mostly also involved in the same MTAs and can therefore be considered likely as one locus.

From the 18 with *flowering* associated markers, 13 offer at least a chromosome location. They are spread over the following chromosomes (number of markers in brackets if more than one marker): 1B (2), 2B, 2D (2), 5D (3), 7A and 7D (4). Two of the 5D MTAs associated wit Fl, Hd and Pr (*wPt0400* and *wPt9788*) are in complete LD. The unassigned 7A marker (*wPt9796*) is in complete LD with the mapped Fl-Hd MTA on 7A at 89.6 cM (*wPt7734*). Three of the four unmapped Fl-Hd MTAs on 7D are in high LD (>0.9) with each other and may indicate one locus (*wPt8084* is not in LD with them). Except one, all 14 MTAs for *heading* are shared with Fl. The one exception is a trait-specific Hd MTA to a marker with unknown position in the genome. One multi-trait MTA associated with Fl, Hd and HI to a chromosome 1B marker *wPt7273* with unknown position was in high LD ($r^2 > 0.7$) with a mapped marker carrying a Fl-Hd MTA on 1B (*wPt6240* at 45.6 cM). One multi-trait MTA for Fl, Hd and SPS on 1B (*wPt3566*) is highly significant for Hd in all eight investigated years ($P < 0.001$) and may indicate a strong locus with high influence. It is not in significant LD to the mapped 1B locus and may be located somewhere else on 1B.

With one exception, the 11 MTAs for *height* are MTAs for PL as well. Also here, the one exception is a trait-specific MTA to a marker with unknown position in the genome. The Hg-PL MTAs with known chromosome refer to the chromosomes 2B, 3D, 4D, 5A and 7B. For *peduncle length* another nine MTAs, from which eight are specific and one is a PL-LR MTA, were found. The markers where chromosomes are known are spread over the chromosomes 1B, 3D, 5B (2), 6D and 7A.

Three MTAs for *biomass* were found: two are multi-trait MTAs and one is specific for Bm. Only one marker from these three has a known chromosome location, a multilocus marker (2A/2D) associated with Bm and SI. One unmapped marker (*wPt5320*) is in complete LD with the mapped Bm locus on 1D at 49.9 cM (*wPt5503*) and may belong to this locus position.

Seven MTAs were found for *grain yield*; one is specific, the other six are multi-trait MTAs. Except for one multi-trait MTA on 7A, the chromosomes are not known. The GY-specific unlocalized marker *wPt8449*

is in total LD with the mapped GY locus on 5B (159.8 cM). Nine specific MTAs for *grain number* were identified; from seven the chromosome is known: 2A, 3A, 5B (2), 6B (2) and 7D. The two 6B MTAs are in complete LD. Additionally, ten multi-trait MTAs where GN was involved were detected. Here, for six markers a chromosome location is known: 2D (4), 6A and 7A. Three of the 2D multi-trait MTAs (*wPt1301*, *wPt3757* and *wPt9848*) are in complete LD with one another and with a mapped locus for GN and other traits on 2D at ~4 cM (*wPt6343* and *wPt8319*). These five markers may represent a strong set for the associated traits. The fourth 2D MTA (*wPt9997*) is not in LD with all these markers.

Eight multi-trait MTAs for *grain weight* were identified, from five of which the chromosome is known: 2D (2), 6A and 7A (2); four are MTAs for GN as well. The 2D MTAs fall into the above-mentioned locus for GN. Only five MTAs for *thousand kernel weight* were detected, from which two are specific but with unknown chromosome position. One of the multi-trait MTAs maps to 2D (*wPt997*, not in LD with the mapped 2D locus), the rest are in unknown positions. One unmapped multi-trait MTA (*wPt8938*) for TKW, SI and SPS is in complete LD with the mapped locus for TKW, GY, SI and SPS on 7B at 108.0 cM (*wPt6156*). *Harvest index* was involved in one trait-specific (unknown chromosome) and three multi-traits MTAs (1B, 7D, unknown). For *spike length* four specific (one on 3B, rest unknown) and eight multi-trait MTAs were identified. Two of them belong to the 2D locus for GN and SW, two others are located on 6A and 7A, the rest remains unmapped. For *spike number* eleven MTAs were found, three of them specific (one on 5A, two with unknown chromosomes). The eight multi-trait SN MTAs are located on 6A and 7A (3), the rest is completely unmapped. *Spike weight* was involved in seven multi-trait-MTAs; the ones with known chromosomes were found on 2D (3), 6A and 7A. The three 2D MTAs are identical with the locus mentioned for GN. Trait-specific MTAs for SW were not identified. One trait-specific (unknown position) and six multi-trait MTAs were found for *spike index*. One of these refers to a multi-locus marker mapping to 2A and 2D, a second one to 7D, but the rest has no information on the chromosome location.

Far more MTAs were identified for the *number of spikelets per spike*: sixteen multi-trait and four specific MTAs. Some of the unmapped markers are highly significantly associated ($P < 0.01$) in all of the investigated years (*wPt0170*, *wPt3677*, *wPt3757*, *wPt8658*, *wPt9963*, *wPt9997*). All four multi-traits MTAs for 2D are identical with the ones identified for GN as well; other such MTAs were found on 1B, 5D and 6B. The SPS-specific MTAs refer to markers on 3A (2) and 4A and to a marker with unknown location. Nine multi-trait MTAs for the *number of sterile spikelets* are shared with SPS; only one of them (*wPt3757*) belongs to the known 2D-locus. Two more MTAs, associated with other traits than SPS, were identified for St, one on 7D, the second with no location. Additionally three St-specific MTAs were found, one on 6B, the other two with unknown position. Some SPS-St MTAs could be connected to mapped loci as they are in LD to these. The unassigned marker *wPt0170* associated with SPS and St is in complete LD with the mapped SPS-St loci on 1B located at 22.0–24.0 cM (*wPt2614*, *wPt3824*, *wPt5312* and *wPt7529*) and at 11.7 cM (*wPt1328*). Another unassigned SPS-St MTA (*wPt8713*) is in partial LD ($r^2 \sim 0.4$) with four markers in close positions to the above-mentioned 2D locus at ~4 cM. The SPS-St MTA to the unassigned marker *wPt9963* could also be connected to this locus but it is only in some LD ($r^2 \sim 0.4$) to one of these mapped markers.

Six MTAs for *leaf rust resistance* could be identified. Three LR-specific MTAs are spread over 5B, 5D and 7A. One of the multi-trait MTAs for LR is also located on 7A; the other two are completely unmapped. The *protein content* was involved in seven multi-trait and three Pr-specific MTAs (2A, 2B, unknown). One of the multi-trait MTAs is the one marker on 2D (*wPt9997*), which is not in LD with the known mapped 2D locus. Other such MTAs with a known chromosome location were found on 2B and 5D (2). Two of the Pr-specific MTAs are located on 2A and 2B. For the *sedimentation value* two specific (3A, unknown) and three multi-trait MTAs were found (4D, 5A, unknown).

Discussion

Linkage disequilibrium

Typically, LD is not constant, either across the genome as a whole, or along single chromosomes. LD can occur over large distances but can also

decrease very quickly for nearby loci (Comadran et al. 2009). This fact was also observed in this population as there is on the one hand complete LD over a long range (up to 45 cM on 7A) but, on the other hand, there were many closely linked markers which did not show any significant LD. In rice, LD decayed over 20–30 cM. However, in some cases, it was still detectable over much longer distances, while in others the decay occurred over a rather short distance (Agrama et al. 2007). A similar situation clearly obtains in wheat, which is not unexpected given that rice and wheat are related species and both are self-pollinators. In a study of elite European barley germplasm, Malysheva-Otto et al. (2006) identified genomic regions where LD extended up to 50 cM and in the study with wheat of Crossa et al. (2007) LD blocks extended even up to 87 cM. The studies of Kraakman et al. (2004) and Rostoks et al. (2005) in barley also showed such long-range LD. The extent of LD in these and our studies probably reflects the fact that the test populations comprised cultivars, for which selection has tended to assemble blocks of chromosome containing genes for agronomic fitness.

Despite long-range LD, only roughly one third of close marker pairs (<10 cM distance) showed significant LD. Nevertheless, a highly significant (although weak) correlation exists between the probability of LD (*P*-value) and genetic distance, as noted also for a population of elite maize lines (Stich et al. 2005), who reported a very similar correlation coefficient to that we obtained. Breseghello and Sorrells (2006a) observed an overall LD decay within 5 cM for LD caused by genetic linkage on chromosome 5A and a decay below <1 cM on 2D in a population of 95 soft winter wheat lines. The LD decay in the present population is also very fast (LOESS curve does not intercept the critical r^2) and different for single chromosomes (data not shown). Here, for most of the chromosomes the decay is in the same range as in the overall picture (considering only chromosomes with sufficient numbers of markers). Exceptions are chromosomes 1D and 7D, where the LOESS curve intercepts the critical r^2 at $\sim$8 cM and at $\sim$2 cM, respectively. The pattern of LD decay is also in accordance with the 100 lines of winter barley from Comadran et al. 2009. The number of all intrachromosomal pairs in significant LD is at 14.9% rather lower than for the bread wheat collection described by Somers et al (2007), who found 56.7% intrachromosomal marker pairs in LD on the same chromosome. They also reported huge differences between single chromosomes. Crossa et al. (2007) reported $\sim$26% though the population is nearly of the same size. The reason might be the larger numbers of markers analysed in the present study compared to Somers' and Crossa's studies (370 and 242, respectively), which yields a breakdown of larger LD blocks. The fact that LD decayed in many cases very quickly is reflected by the low proportion of pairs in LD (38.3%) for the tight linkage class 1, so even the majority of close markers is not in LD here. This proportion also varies strongly across the chromosomes, ranging from 0 to 73% (data not shown). These facts are important for the resolution of genome-wide association studies as they indicate that even a higher number of markers than used here is needed for detecting all possible loci. The proportion of LD caused by population structure increases with higher genetic distance and in the entire wheat collection follows a logarithmic function described by Comadran et al. (2009) for barley. The critical value of r^2 for our collection is nearly double those described by Crossa et al. (2007) and Comadran et al. (2009) which probably results from more pairs in higher LD and not the overall amount of LD. The high number in total LD in the subgroup Q2 resulted in a much higher critical r^2 compared to the whole collection and the subgroup Q1. The differences between the two subgroups in amount and extent of LD are similar to those reported by Comadran et al. (2009) in a barley collection with five subgroups using DArT markers. Their group of north Mediterranean two-row barley (52 lines) is similar to the whole barley collection, as it is here also for the Q1, and their smaller groups of Turkish (21) and eastern Mediterranean (16) accessions show very high levels of intrachromosomal LD over long distances, even more than found here in the Q2 group.

The mean r^2 of all interchromosomal pairs is at 0.019 close to the median of 0.022 described by Breseghello and Sorrells (2006a) in a population of the same size as investigated here, even though they had only 630 SSR marker pairs. Though this value is similar, the critical r^2 value is much higher here. This may result from the much higher marker density which gives more opportunities to detect more cases of high interchromosomal LD but also from the fact

that another marker type was used which produces in general lower LD. Thus, the values in Breseghello and Sorrells (2006a) ranged from 0 to maximally only 0.133 compared to the DArT marker set in the present study, where even cases of total LD between markers on different chromosomes were obtained. This is supported by Somers et al. (2007) who used microsatellite markers and had maximum r^2 values of only 0.68 even for intrachromosomal LD, while the DArT marker set of Crossa et al. (2007) revealed several LD blocks where the pairs of adjacent loci were in complete LD, which was also obtained in the present study but on a lower distance. The total amount of interchromosomal pairs in significant LD (3.2%) is comparable with the results of the wheat lines of Crossa et al. (2007) and Tommasini et al. (2007), who reported 5.0% and 2.9%, respectively. Crossa et al. reported a marginally higher average of r^2 for those pairs than that we obtained. Only the number of interchromosomal pairs in total LD was unexpectedly high, as other studies did not report on such cases. The reason might be an effect of the small population consisting of breeding material.

Evaluating LD between mapped and unmapped markers can help in assigning them to approximate map locations, when r^2 is high (Comadran et al. 2009), and help to interpret detected marker–trait associations. Rostoks et al. (2006) remapped loci of known map positions with different r^2-value cut-offs and found an r^2 threshold >0.5 sufficient to map them with an accuracy >50% with ~5% false-positive calls. Therefore we could consider the unmapped associated markers in high LD with mapped associated markers as one locus for the specific trait. However, because high LD can also occur through population structure, one has to handle this with care.

Genotype–phenotype associations

In general, we found the more complex (polygenic) and/or environmentally dependent the investigated trait, the fewer loci and/or only minor loci were detectable. So, for traits with low correlations between the years, only a few associations were found: two for powdery mildew and nine for sedimentation value, but more than 30 for flowering and the number of spikelets, traits showing high correlations between the investigated years.

MTAs for heading and flowering

Hd and Fl are determined by genes affecting the vernalization response, the photoperiod response and earliness per se. Because all the entries in the germplasm panel were pure or facultative winter wheats, no allelic variation at the *Vrn-1* genes was expected, and indeed none of the Hd or Fl MTAs mapped to the regions on the homoeologous group 5 chromosomes where these genes reside (McIntosh et al. 2008). One MTA was identified on the long arm of chromosome 5B, but in a position only ~10 cM away from the centromere (whereas *Vrn-B1* is genetically independent of the centromere: Leonova et al. 2003). The photoperiod response genes (*Ppd-1*) map to the short arms of the homoeologous group 2 chromosomes (McIntosh et al. 2008) but a Hd/Fl MTA was detected only on chromosome arm 2DS. Considering the GLM only, there are MTAs on the short arms of 2A and 2B but with MLM they do not fulfill the significance criteria (Electronic Supplementary Material Table 2). Although the germplasm entries were all assumed to be photoperiod-insensitive, some of these MTAs may reflect allelic variation at *Ppd-1* genes. Another QTL associated with photoperiod response and described by Kuchel et al. (2006) is located in the centromere region of chromosome 7AS. This locus may correspond to the major locus detected on 7AS in the present study.

Flowering/heading time MTAs mapped close to the centromeres of chromosomes 1B and 1D, and may therefore reflect variation at *Eps* genes, since such a gene has been identified both on chromosome 1BL (Tóth et al. 2003) as well as in a syntenic region of barley (Laurie et al. 1995). A further *Eps* locus (and a coincident MTA) maps to chromosome arm 3AL (Miura et al. 1999; Börner et al. 2002) as well as to barley chromosome arm 3HL (Laurie et al. 1995), but we found a significant association on 3AL only with the GLM (*wPt3697* at 27.9 cM); with the MLM this marker is significant in four out of 6 years only at the $P < 0.1$ level. Flowering/heading time MTAs not coinciding with any known mapped genes were detected on chromosomes 2BL, 6AS, 6AL and 6BS.

MTAs for plant height and peduncle length

Height is under the control of genes distributed across at least 17 of the 21 wheat chromosomes (Snape et al.

1977). Apart from the numerous semi-dwarfing *Rht* genes, many have in the meantime been detected by QTL mapping. Taking both models into account together, no Hg MTAs were detected in the regions of chromosomes 2D, 4B or 4D harbouring the widely-used dwarfing genes *Rht8*, *Rht-B1* and *Rht-D1*. Because most of the germplasm entries are known to carry *Rht8*, allelic variation at these loci was not expected to be present here. For *Rht-B1* and *Rht-D1* the situation is different. Some of the germplasm is known to carry either *Rht-B1* or *Rht-D1* or even both. Considering *Rht-B1*, a significant MTA was detected in a highly comparable position in the centromeric region of 4BS applying MLM but did not fulfill the significance criteria (but in 5 years $P < 0.1$) with GLM. The lack of MTAs on chromosome 4D is most probably due to the very low marker number/density on that particular chromosome (only four markers spanning 20.5 cM in total). MTAs detected on chromosomes 2B, 4A and 7A may be comparable to loci *QHt.inra-2B*, *QHt.inra-4A* and *QHt.inra* 7A, respectively, described by Gervais et al. (2003). Of special interest seems to be the locus on chromosome 2BS which may be homoeoallelic to *Rht8* on 2DS. A 20-cM long interval on chromosome arm 7BL carries three Hg/PL MTAs, a region that, however, is distinct from that carrying the QTL *QHt.crc-7B* and *QHt.fra-7B* (McCartney et al. 2005; Cadalen et al. 1998) whereas another locus detected in the present study only with the GLM (*wPt7602* at 43.3 cM) on chromosome 7BS may be comparable to *QHt.fra-7B*. It may be concluded that the loci on chromosomes 1AL, 6AL, 6BL and 7BL detected here are new ones.

Less attention has been paid in the past to mapping variation for peduncle length, although this trait is of some importance in the context of avoiding ear diseases. One QTL (*QPdl.ipk-6A*) described by Börner et al. (2002) could coincide with a locus which is significant for Hg and PL in all years with the GLM (*wPt7330* on 6A at 20.6 cM) but not significant at all with the MLM. A number of PL MTAs coincided with Hg MTAs but in genomic regions not yet described.

MTAs associated with grain yield, grain quality and disease resistance

With respect to both grain yield and disease resistance, it was possible to compare the present outcomes with those derived by Crossa et al. (2007). Identical GY MTAs were associated with the markers *wPt3904* (1AL, 65.8 cM), *wPt4407* (3AL, 29.0 cM), and *wPt6156* (7BL, 108.0 cM). In addition, we identified GY MTAs associated with mapped markers not included in Crossa's study, *wPt1191* (3B, 66.1 cM) and *wPt7299* (7A, 65.5 cM), which however were in close distances to loci detected by Crossa et al. as well as the two loci *wPt1974* (7A, 110.9 cM) and *wPt4038* (7B, 131.6 cM) which are fulfil the GLM significance criteria but the MLM criteria only on the $P < 0.1$ level. This large-scale coincidence was somewhat unexpected, given that the two studies had involved very different sets of germplasm. GY has also been targeted by a number of conventional QTL mapping experiments, allowing for a comparison between MTA and known QTL locations. Thus, the location of *QYld.crc-4A* (4AL) (McCartney et al. 2005) coincides with the GY MTA detected here, whereas others do not. One MTA (*wPt6148* at 48.6 cM) significant for the GLM on the $P < 0.05$ level but only on the $P < 0.1$ level in the MLM coincides with *QYld.crc-2A* (2AS) (McCartney et al. 2005). QTL for GN, GW and SL have been described in the literature (collated by McIntosh et al. 2008), but the location of these does not coincide with any of the MTAs identified here. In contrast, MTAs were detected here for the traits Bm, HI, SPS and St but no QTL for these traits have been mapped to date. Of some future interest are the MTAs associated with the unmapped markers *wPt3677*, *wPt8658* and *wPt9963* which were associated with SPS and other traits (mainly St) and were highly significant in all years for SPS. The location of one of the MTAs for protein on chromosome 4B is comparable to that of *Qpro.mbg-4B* (Blanco et al. 1996). Three QTL for sedimentation value have been identified, on chromosome 2BL, 3AL and 7AS (Blanco et al. 1998). The position of *Qsev.mbg-7A* is similar to the Sd MTA detected here.

The association mapping study of Crossa et al. (2007) also targeted the response to powdery mildew and leaf rust infection. The two MTAs for PM located on chromosomes 3BL and 4BL, respectively, do not match any other known locus. Loci for LR coinciding with Crossa et al. (2007) are the MTAs on 1AS (different marker but only less than 2 cM away from Crossa's location), 3BS (different marker, same position), 3BL (<1 cM away) and 7AS (different marker, same position). Regions on chromosomes

1AS, 3BS and 7AS may correspond to the known major genes *Lr10*, *Lr27* and *Lr47*, respectively. This is a fairly high number of coinciding LR MTAs despite the fact that Crossa et al. (2007) had a wide geographical spread of the trials (data were collected from a series of internationally-based trials), whereas the present experiments were conducted at a single location.

Power of detecting associations with GLM and MLM

The approach of Yu et al. (2006) including the kinship relations in a mixed model proved to be powerful enough to detect significant associations, even in a relative small sample set, despite the fact that the use of Q-Matrix and Kinship-Matrix leads to a loss of degrees of freedom, because it reduces the power of detecting associations in small populations (Stich et al. 2008). Nevertheless, it can be useful to use both models, the GLM and the MLM, to compare the results with already known loci. In our study some of these known loci could only be detected with the GLM approach (see flowering, height, peduncle) which is in accordance with Yu et al. (2009) who stated that the GLM approach with the Q-matrix is likely more sensitive than the kinship approach in some cases. New loci should be verified definitively by the MLM approach instead of with the Q-matrix in the GLM alone; too many associations appear that cannot be detected with the MLM and might be false positives. The difference between the two models is trait-dependent. For traits like flowering and heading far more associations were detected with the GLM alone, but there were also traits such as grain number or sedimentation value with only small differences in significant associations between the two approaches. Again this is in accordance with Yu et al. (2006) who noted differences in their model approaches for different traits. More comparisons of different models in genome-wide association mapping scans should be examined to study this phenomenon in depth.

Concluding remarks

Genome-wide association mapping is a new approach in cereal genetics, and particularly in wheat. In contrast to conventional bi-parental segregation-based mapping, which can only analyse allelic differences between the two parental lines, association mapping attempts to scan genetic variation across a wide spectrum of material. The study underlines the value of historical field trail data from breeding material in elucidating the genetic basis of important agronomic traits even with a relatively small collection of genotypes. A substantial number of MTAs for a whole set of agronomically important traits was detected in only one study. With a mix of the kinship and the Q-matrix approach, many loci were detected that coincide with known major genes or QTL, indicating the power of association mapping. Additionally, potential novel loci were identified that may help to better understand the architecture of complex genetic traits. The novel loci provide opportunities for further improvement of wheat, based on a marker approach.

Acknowledgments The authors thank Susanne Dreisigacker, CIMMT, Mexico for providing the mapping data as well as Inge Matthies, IPK Gatersleben, Germany for her constructive discussions during the course of study.

References

Agrama HA, Eizenga GC, Yan W (2007) Association mapping of yield and its components in rice cultivars. Mol Breed 19:341–356

Aranzana MJ, Kim S, Zhao K, Bakker E, Horton M et al (2005) Genome-wide association mapping in Arabidopsis identifies previously known flowering time and pathogen resistance genes. PLoS Genet 1(5):e60

Blanco A, De Giovanni C, Laddomada B, Sciancalepore A, Simeone R, Devos KM, Gale MD (1996) Quantitative trait loci influencing grain protein content in tetraploid wheats. Plant Breed 115:310–316

Blanco A, Bellomo MP, Lotti C, Manglio T, Pasqualone A, Simeone R, Troccoli A, Di Fonzo N (1998) Genetic mapping of sedimentation volume across environments using recombinant inbred lines of durum wheat. Plant Breed 117:413–417

Börner A, Schumann E, Fürste A, Cöster H, Leithold B, Röder MS, Weber WE (2002) Mapping of quantitative trait loci determining agronomic important characters in hexaploid wheat (*Triticum aestivum* L.). Theor Appl Genet 105: 921–936

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635

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

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

Cadalen T, Sourdille P, Charmet G, Tixier MH, Gay G, Boeuf C, Bernard S, Leroy P, Bernard M (1998) Molecular markers linked to genes affecting plant height in wheat using a doubled-haploid population. Theor Appl Genet 96:933–940

Comadran J, Thomas WTB, van Eeuwijk FA, Ceccarelli S, Grando S, Stanca AM, Pecchioni N, Akar T, Al-Yassin A, Benbelkacem A, Ouabbou H, Bort J, Romagosa I, Hackett CA, Russell JR (2009) Patterns of genetic diversity and linkage disequilibrium in a highly structured *Hordeum vulgare* association-mapping population for the Mediterranean basin. Theor Appl Genet 119:175–187

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

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

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567–1587

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

Gervais L, Dedryver 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

Huang XQ, Cloutier S, Lycar L, Radovanovic N, Humphreys DG, Noll JS, Somers DJ, Brown PD (2006) Molecular detection of QTLs for agronomic and quality traits in a doubled haploid population derived from two Canadian wheats (*Triticum aestivum* L.). Theor Appl Genet 113:753–766

ICC (2008) http://www.icc.or.at/methods3.php

Ivandic V, Hackett CA, Nevo E, Keith R, Thomas WTB, Forster BP (2002) Analysis of simple sequence repeats (SSRs) in wild barley from the fertile crescent: associations with ecology, geography and flowering time. Plant Mol Biol 48:511–527

Jaccoud D, Peng K, Feinstein D, Kilian A (2001) Diversity arrays: a solid state technology for sequence information independent genotyping. Nucleic Acids Res 29:e25

Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckermann D, Daly MJ, Eskin E (2008) Efficient control of population structure in model organism association mapping. Genetics 178:1709–1723

Kobiljski B, Quarrie S, Denčić S, Kirby J, Ivegěs M (2002) Genetic diversity of the Novi Sad Wheat collection revealed by microsatellites. Cell Mol Biol Lett 7:685–694

Kraakman ATW, Niks RE, Van den Berg PM, Stam P, Van Eeuwijk FA (2004) Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. Genetics 168:435–446

Kraakman ATW, Martinez F, Mussiraliev B, van Eeuwijk FA, Niks RE (2006) Linkage disequilibrium mapping of morphological, resistance, and other agronomically relevant traits in modern spring barley cultivars. Mol Breed 17:41–58

Kuchel H, Hollambi G, Langridge P, Williams K, Jefferies SP (2006) Identification of genetic loci associated with ear emergence in bread wheat. Theor Appl Genet 113:1103–1112

Kumar N, Kulwal PL, Balyan HS, Gupta PK (2007) QTL mapping for yield and yield contributing traits in two mapping populations of bread wheat. Mol Breed 19(2):63–177

Laurie DA, Pratchett N, Bezant J, Snape JW (1995) RFLP mapping of five major genes and eight quantitative trait loci controlling flowering time in a winter × spring barley (*Hordeum vulgare* L.) cross. Genome 38:575–585

Leonova I, Pestsova E, Salina E, Efremova T, Röder M, Börner A (2003) Mapping of the Vrn-B1 gene in *Triticum aestivum* using microsatellite markers. Plant Breed 122:209–212

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, Ganal MW, Röder 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

Matthies IE, Weise S, Röder MS (2008) Association of haplotype diversity in the a-amylase gene amy1 with malting quality parameters in barley. Mol Breed 23(1):139–152

McCartney CA, Somers DJ, Humphreys DG, Lukow O, Ames N, Noll J, Cloutier S, McCallum BD (2005) Mapping quantitative trait loci controlling agronomic traits in the spring wheat cross RL4452 × 'AC Domain'. Genome 48:870–883

McIntosh RA, Yamazaki Y, Dubcovsky J, Rogers J, Morris C, Somers DJ, Apples R, Devos KM (2008) Catalogue of gene symbols for wheat. http://www.shigen.nig.ac.jp/wheat/komugi/genes/symbolClassList.jsp

Miura H, Nakagawa M, Worland AJ (1999) Control of ear emergence time by chromosome 3A of wheat. Plant Breed 118:85–87

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

Quarrie SA, Dodig D, Pekiç S, Kirby J, Kobiljski B (2003) Prospects for marker-assisted selection of improved drought responses in wheat. Bulg J Plant Physiol (special issue):83–95

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, Yessimbekova M, Turuspekov Y, 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

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

Rostoks N, Mudie S, Cardle L, Russell J, Ramsay L, Booth A, Svensson JT, Wanamaker SI, Walia H, Rodriguez EM, Hedley PE, Liu H, Morris J, Close TJ, Marshall DF, Waugh R (2005) Genome-wide SNP discovery and linkage analysis in barley based on genes responsive to abiotic stress. Mol Genet Genom 274:515–527

Rostoks N, Ramsay L, MacKenzie K, Cardle L, Bhat PR, Roose ML, Svensson JT, Stein N, Varshney RK, Marshall DF, Graner A, Close TJ, Waugh R (2006) Recent history of artificial outcrossing facilitates whole-genome association mapping in elite inbred crop varieties. Proc Natl Acad Sci USA 103:18656–18661

Roy JK, Bandopadhyay R, Rustgi S, Balyan HS, Gupta PK (2006) Association analysis of agronomically important traits using SSR, SAMPL and AFLP markers in bread wheat. Curr Sci 90:683–689

Skøt L, Humphreys MO, Armstead I, Heywood S, Skøt KP, Sanderson R, Thomas ID, Chorlton KH, Hamilton NRS (2004) An association mapping approach to identify flowering time genes in natural populations of *Lolium perenne* (L.). Mol Breed 15:233–245

Snape JW, Law CN, Worland AJ (1977) Whole chromosome analysis of height in wheat. Heredity 38:25–26

Somers D, 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(6):557–567. doi:10.1139/G07-031

Stich B, Melchinger AE, Frisch M, Maurer HP, Heckenberger M, Reif J (2005) Linkage disequilibrium in European elite maize germplasm investigated with SSRs. Theor Appl Genet 111:723–730

Stich B, Möhring J, Piepho H-P, Heckenberger M, Buckler ES, Melchinger AE (2008) Comparison of mixed-model approaches for association mapping. Genetics 178:1745–1754

Tommasini L, Schnurbusch T, Fossati D, Mascher F, Keller B (2007) Association mapping of Stagonospora nodorum blotch resistance in modern European winter wheat varieties. Theor Appl Genet 115:697–708

Tóth B, Galiba G, Fehér E, Sutka J, Snape JW (2003) Mapping genes affecting flowering time and frost resistance on chromosome 5B of wheat. Theor Appl Genet 107:509–514

UN Country Code (2009) http://unstats.un.org/unsd/methods/m49/m49alpha.htm

Weir BS (1996) Genetic data analysis II: methods for discrete population genetic data. Sinauer Associates Inc., Sunderland

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. Proc Natl Acad Sci USA 101(26):9915–9920

Wenzl P, Li H, Carling J, Zhou M, Raman H, Paul E, Hearnden P, Maier C, Xia L, Caig V, Ovesná J, Cakir M, Poulsen D, Wang J, Raman R, Smith KP, Muehlbauer GJ, Chalmers KJ, Kleinhofs A, Huttner E, Kilian A (2006) A high-density consensus map of barley linking DArT markers to SSR, RFLP and STS loci and agricultural traits. BMC Genomics 7:206. doi:10.1186/1471-2164-7-206

Yu J, Buckler ES (2006) Genetic association mapping and genome organization of maize. Curr Opin Biotechnol 17:155–160

Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB et al (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203–208

Yu J, Zhang Z, Zhu C, Tabanao DA, Pressoir G, Tuinstra MR, Kresovich S, Todhunter RJ, Buckler ES (2009) Simulation appraisal of the adequacy of number of background markers for relationship estimation in association mapping. Plant Genome 2:63–77. doi:10.3835/plantgenome2008.09.0009