GEFÖRDERT VOM

Bundesministerium
für Bildung
und Forschung

Veröffentlichung der Ergebnisse von Forschungsvorhaben im BMBF-Programm

B I O L O G I E

"Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 0315067B gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt beim Autor".

Forschungsvorhaben:

GABI-FUTURE: Verbundvorhaben

GABI-WHEAT: Etablierung einer Plattform zur genomweiten Assoziationskartierung in Weizen
Teilprojekt B

Förderkennzeichen:

0315067B

Zuwendungsempfänger:

Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK)

Corrensstr. 3

06466 Stadt Seeland OT Gatersleben

Projektleitung:

Dr. Marion Röder

Laufzeit:

01.01.2008 bis 30.06.2011

Inhaltsverzeichnis

I. Kurze Darstellung

1 Aufgabenstellung

In diesem Verbundprojekt sollten die Grundlagen für eine Assoziationskartierung in europäischem Winterweizen gelegt werden und Erfahrungen bezüglich der Datenauswertung gewonnen werden. Dazu wurde die Populationsstruktur in europäischem Elite-Material (hauptsächlich aus Deutschland, Frankreich und Großbritannien) mittels einer großen Anzahl von Mikrosatellitenmarkern und einer beschränkten Anzahl von SNPs aus bekannten Kandidatengenen untersucht, sowie das Ausmaß von Gametenphasenungleichgewicht (Linkage Disequilibrium = LD) bestimmt. Des weiteren sollten Merkmale für Ertrag und Ertragskomponenten, Backqualität, phenologische Merkmale und Krankheitsresistenzen mit den molekularen Markerdaten assoziiert werden und auf diesem Wege die genetische Basis für agronomisch wichtige Merkmale bei Winterweizen untersucht werden. Das Projekt wurde von den vier Projektpartnern gemeinsam realisiert, wobei Genotypisierung, Phänotypisierung und Datenauswertung zusammen durchgeführt wurden. Aus diesem Grund werden die Ergebnisse dieses Projektes zum großen Teil gemeinsam und in einheitlicher Form im Abschlussbericht präsentiert.

2 Voraussetzungen, unter denen das Vorhaben durchgeführt wurde

QTL-Kartierung und die Kartierung von Kandidatengenen wurden bislang hauptsächlich in eigens für diesen Zweck generierten biparentalen Kartierungspopulationen durchgeführt. Obwohl diese Methode eine Anzahl Erfolge aufzuweisen hat, sind auch Nachteile mit ihr verbunden. Ein Nachteil sind die hohen Kosten und der zusätzliche Aufwand für die Erstellung der entsprechenden Populationen in einem Züchtungsprogramm. Weiterhin spiegelt die erfasste genetische Variation nur die Variation in der jeweiligen Kruezungspopulation wider (maximal 2 Allele an einem Locus segregieren) und die Ergebnisse sind nicht immer direkt übertragbar in die Züchtung.

Assoziationskartierung setzt nicht die Erstellung spezifischer Kartierungspopulationen voraus. Sie wird wesentlich dazu genutzt, um beispielsweise ‚common diseases' in der Humangenetik zu untersuchen (Wellcome Trust Case Control Consortium, 2007, Salonen et al., 2007). Auch in der Tierzucht wurde diese Technik bereits etabliert (Barendse et al., 2007, Charlier et al., 2008). In der Pflanzenzucht waren zum Zeitpunkt der Antragstellung erst wenige Studien bekannt, die zusätzlich noch mit eng verwandtem Material, wenigen Markern und/oder wenigen Merkmalen durchgeführt worden waren (siehe auch Seite 4: Wissenschaftlicher und technischer Stand, an den angeknüpft wurde).

3 Planung und Ablauf des Vorhabens

Das Projekt gliederte sich in fünf verschiedene Arbeitspakete:

Arbeitspaket 1: Auswahl und Vermehrung des Materials

Die Sorten wurden von den Züchtern KWS-Lochow und Syngenta in Zusammenarbeit mit den anderen Partnern ausgewählt und von den Züchtungsfirmen vermehrt, um genug Material für die folgenden Feldversuche und die Genotypisierung vorzuhalten.

Arbeitspaket 2: Phänotypische Evaluierung

Feldversuche sollten an 2 Orten in Deutschland und in Frankreich in 2 hintereinander folgenden Jahren durchgeführt werden. Die züchterische Phänotypisierung wurde durch KWS-Lochow und Syngenta durchgeführt. Die Phänotypisierung bezüglich Krankheitsresistenzen (Fusarium, Septoria und Drechslera) wurden in einem Unterauftrag des IPKs durch das Julius-Kühn-Institut in 2 Jahren an je 2 Orten durchgeführt.

Arbeitspaket 3: Genotypisierung

Mehr als 800 Mikrosatelliten Marker wurden für die Genotypisierung ausgewählt und jeder Projektpartner genotypisierte etwa 200 Marker. Die Rohdaten wurden in einer Zusammenarbeit zwischen IPK und TG analysiert und bereinigt (Allelfrequenzen, monomorphe Marker, nicht-auswertbare Marker etc.) und allen Partnern zur Verfügung gestellt.

Arbeitspaket 4: Kandidatengene und SNPs

Aus der Literatur wurden Kandidatengene und –polymorphismen für die Genotypisierung der 372 Sorten ausgewählt. Diese wurden mit diversen Techniken am IPK und bei TraitGenetics genotypisiert. Ferner wurden bei TraitGenetics Methoden zur Identifikation von SNPs in Kandidatengenen und zur Hochdurchsatzgenotypiserung mit SNP-Markern entwickelt.

Arbeitspaket 5: Datenauswertung

Alle Daten wurden unter Federführung des IPKs gemeinsam mit den anderen Partnern hinsichtlich LD, Populationsstruktur und Marker-Merkmals-Assoziationen untersucht.

4 Wissenschaftlicher und technischer Stand, an den angeknüpft wurde

In Weizen waren zu Projektbeginn bereits einige Assoziationsstudien publiziert (Breseghello and Sorrells, 2006, Crossa et al., 2007, Tommasini et al., 2007). Diese waren jedoch mit geringerer Markerabdeckung durchgeführt worden oder konzentrierten sich auf wenige Merkmale und/oder einzelne Chromosomen. Als Markerrresource in hexaploidem Weizen waren bereits Mikrosatelliten entwickelt (Röder et al., 1998, Röder et al., 2002), während SNPs noch in der Entwicklungsphase waren (Ganal et al., 2009). Ferner waren die Möglichkeiten der Genotypisierung von SNP-Markern aufgrund der hexaploiden Natur von Weizen noch ungeklärt.

In der Gruppe Gen- und Genomkartierung am IPK und in Zusammenarbeit mit TraitGenetics sind vor Projektbeginn über 1000 Mikrosatelliten in Weizen entwickelt und kartiert worden. Weiterhin besitzen beide Partner große Expertise in der klassischen (biparentalen) Kartierung von QTL. Ferner besaßen die Züchtungsfirmen noch Zugang zu einer Reihe weiterer Mikrosatellitenmarker.

KWS-Lochow ist eine der größten Pflanzenzuchtfirmen in Europa mit Weizenzuchtprogrammen in Deutschland, Frankreich und Großbritannien. Erfahrung ist vorhanden in den Bereichen

Feldversuche, Sortenentwicklung, Biotechnologie (z. B. Entwicklung von Double Haploiden). Weiterhin konnten bereits von KWS-Lochow entwickelte Sorten in diesem Projekt Verwendung finden.

Syngenta hat sich in der Vergangenheit neben ihrem Schwerpunkt im Bereich Pflanzenschutz zu einem der größten Züchtungsunternehmen entwickelt. Syngenta hat, wie KWS-Lochow, eine Weizenzüchtung in Deutschland, Frankreich und Großbritannien. Weiterhin konnten auch hier bereits von Syngenta entwickelte Sorten verwendet werden.

5 Zusammenarbeit mit anderen Stellen

Die gesamten Arbeiten in dem Projekt wurden von allen Partnern zusammen in enger Interaktion generiert, da kein Partner alleine dieses Projekt hätte realisieren können. Neben regelmäßigen Projekttreffen (2 x jährlich) und weiteren ad-hoc Treffen von Untergruppen zu spezifischen Aufgaben, erfolgte auch eine enge Interaktion mit allen Partnern durch den regelmäßigen Austausch von Daten. Die Untersuchung auf Krankheitsresistenzen wurden im Rahmen eines Unterauftrages des IPKs an das Julius-Kühn-Institut (Zweigstelle Braunschweig, Herr Dr. Rodemann) durchgeführt.

Das IPK arbeitete in Arbeitspaket 3 mit den anderen Partnern zusammen um 372 Weizensorten mit SSR –Markern genomweit zu genotypisieren. Daneben führte das IPK die Kandidatengen-Analyse für alle Weizensorten durch (Arbeitspaket 4). Die Resistenztestungen für wichtige pilzliche Pathogene wurden als Unterauftrag für das IPK vom Julius Kühn-Institut in Braunschweig durchgeführt (Arbeitspaket 2). Eine wesentliche Rolle spielte das IPK in der statistischen Auswertung aller phänotypischen und genotypischen Daten, inklusive Analyse der Populationsstruktur, des ‚Linkage Disequilibrium' und der Berechnung der Marker-Merkmals-Assoziationen (Arbeitspaket 5).

Ansonsten erfolgte keine andere nennenswerte Zusammenarbeit mit andern Stellen außerhalb des Verbundes.

II. Eingehende Darstellung

1 Verwendung der Zuwendung und eingehende Darstellung der erzielten Ergebnisse

(As in the project outline the description of the results will be in English)

1A Genotyping with microsatellite markers

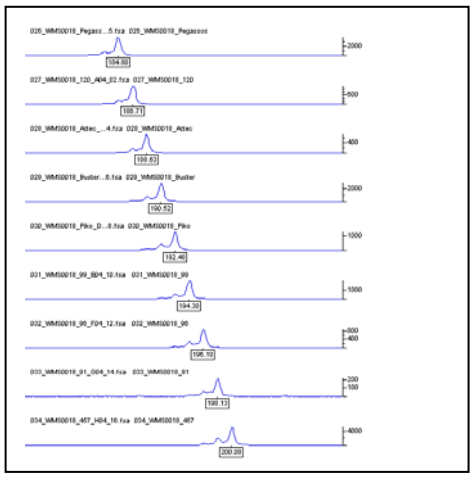
Initially more than 1000 different microsatellite markers were selected for genotyping. In a first step, approximately 200-300 microsatellite markers were selected by each of the partners (Syngenta = SYN, KWS-Lochow = KWS-L, Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung = IPK, TraitGenetics GmbH = TG) and they were tested for their quality on a limited set of lines through genotyping on different types of capillary sequencers. Concordance between labs was evaluated by TG through a ring test in which each of the labs genotyped the same 12 markers. After the elimination of some problematic markers, a slightly reduced set of microsatellite markers (approx. 850) was then used for genotying the entire set of lines (384 including controls). Ultimately, TraitGenetics provided full data sets for 239 markers, Syngenta for 211 markers, KWS-Lochow for 197 markers and IPK for 168 markers.

Genotyping of the microsatellite markers resulted in useful data for more than 1080 loci since some markers amplified more than one locus. Marker data were cleaned up by removing markers which amplified more than one locus and where the different loci could not clearly be separated or for monomorphic markers (164). After this clean-up of the marker set, 782 loci generated by 732 microsatellites remained in the dataset. With 4.8 % missing data points, 276844 genotype data points were available for the association analysis of which heterozygotes comprised 2.6 %. All together, 6198 different alleles could be detected corresponding to approximately 7.5 different alleles per marker.

This data set is to our knowledge the largest microsatellite marker data set for European wheat material. Microsatellites are multiallelic markers with each fragment representing a different allele. Compared to SNP markers which are biallelic markers that cannot reveal the full allelic diversity unless haplotype-specific markers are used, microsatellite markers do provide a deeper insight into the allelic diversity at a given locus. In Figure 1, an example is shown for an allelic ladder generated with a microsatellite marker.

Figure 1:

Display of the different alleles generated with a microsatellite marker and sorted according to the fragment size.



From this marker data set, all alleles with an allele frequency of less than 3% (or 11 lines) were subsequently eliminated from the marker/trait analysis since rare alleles can obscure the data due to the low significance of values obtained in such a small sample number. After this cleaning process of the rare alleles, 3204 alleles remained in the data set. The average number of alleles per locus changed then to 4.1 (ranging from 2 to 13 alleles). The final allele distribution is shown in Figure 2. The majority of markers showed between 2 and 4 alleles.

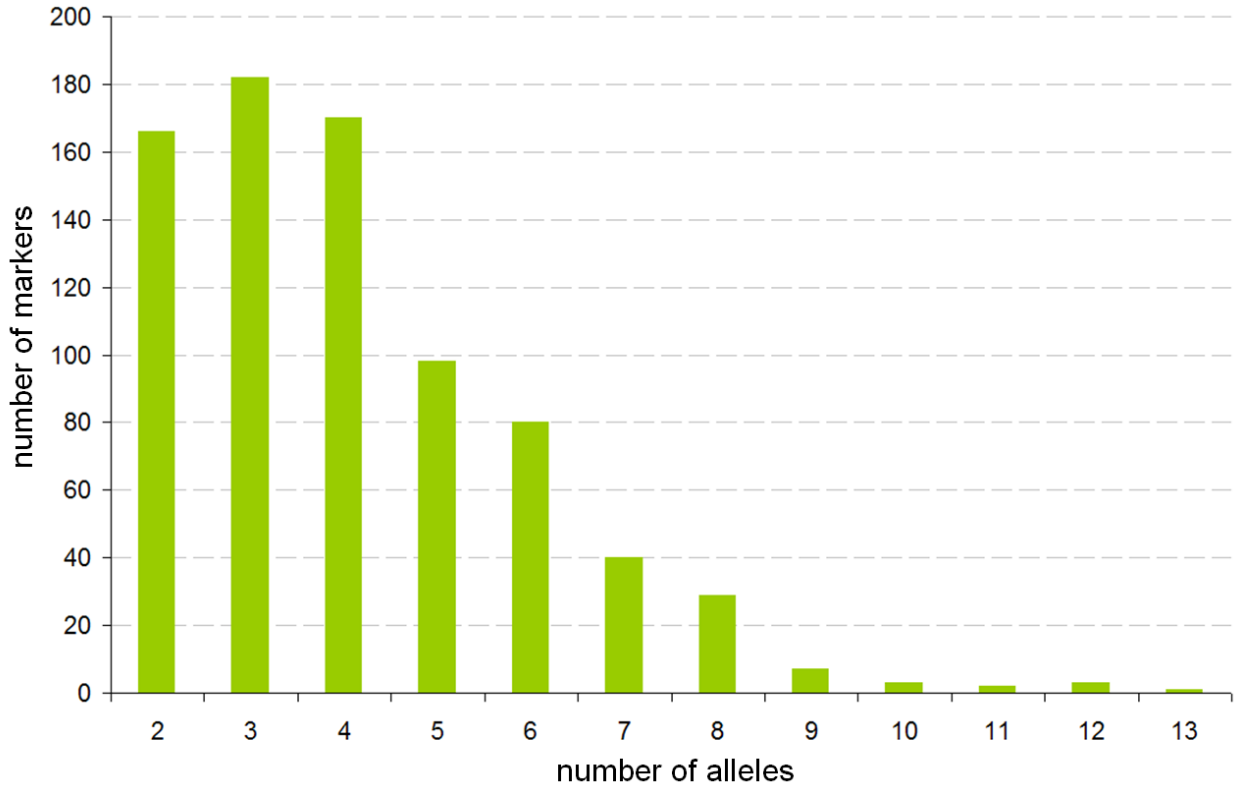


Figure 2: Distribution of allele numbers of 732 microsatellites genotyped in 372 wheat varieties.

The actual allele frequencies ranged from 5 % to 95 % (Figure 3). The majority of alleles displayed a frequency of approximately 10%. Since most of the markers have 2 to 4 alleles and assuming equal frequencies of alleles at a given locus, the majority of alleles should have a frequency of 0.25 to 0.5. The obtained marker data showed however frequently one major allele at a given locus.

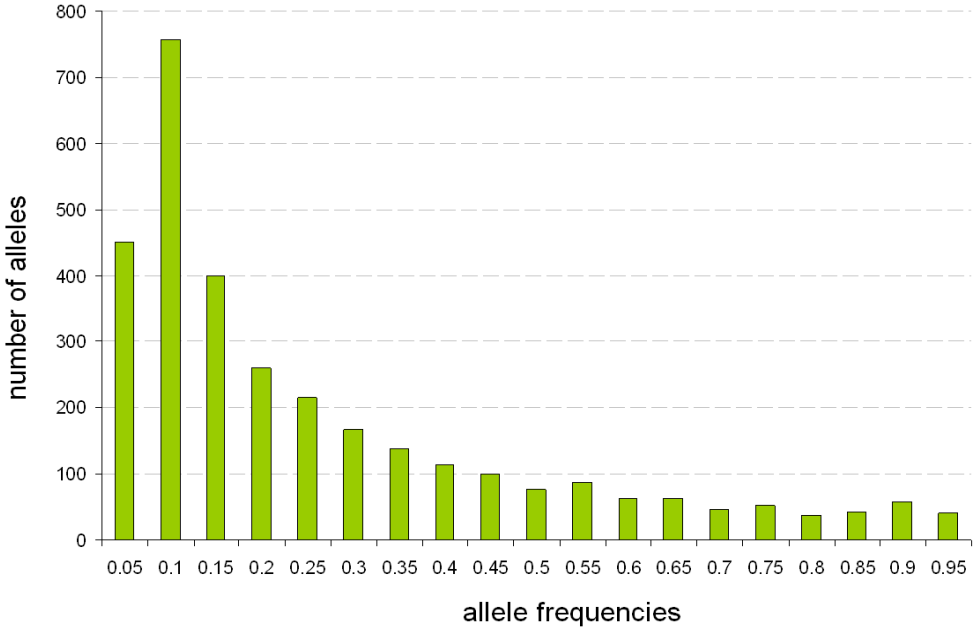


Figure 3: Distribution of frequencies of the 3204 alleles.

1B Linkage disequilibrium

LD has been calculated between all marker pairs with the programme TASSEL v. 2.1 (Bradbury et al., 2007) and is represented by the coefficient of determination r^2. LD was evaluated chromosome-wise for all possible marker pairs and for neighbouring marker pairs only, as well as for unlinked markers defined as being located on different chromosomes. LD decay was examined by graphs of pair-wise distances (cM) versus r^2. The LD decay was estimated at the point where a second degree Loess curve, which was calculated with the function loess.smooth in R, intercepted with the threshold of the critical LD. Critical LD was evaluated following Breseghello and Sorrells (2006). Briefly, r^2 values for all available marker pairs on different chromosomes were square root transformed and the 95th percentile was chosen as the threshold, above which LD is likely attributable to physical linkage.

Details for 8146 marker pairs which returned r^2 values are shown in Table 1. Of these 62.9 % show statistically significant LD ($p < 0.05$) as determined by 10000 permutations. The average r^2 for loci on different chromosomes (unlinked loci) is 0.031. LD is in general low with being higher between neighbouring marker pairs (0.03 on chromosome 6A to 0.14 on chromosome 4B) than between all marker pairs on the same chromosome (0.02 on chromosomes 1B, 2D, 3A, 3B, 3D, 5D, 6A, 6D, 7A and 7D to 0.08 on chromosome 4D). LD reached its lowest value for marker distances above 50 cM with 0.01 between all marker pairs on the same chromosome. For distances below 10 cM, it was slightly higher (0.09 for all marker pairs on the same chromosome and 0.11 for neighbouring marker pairs) than for distances between 10 to 20 cM (0.05 for all marker pairs on the same chromosome and 0.03 between neighbouring marker pairs). The critical value for LD is 0.176. Figure 4 shows the LD in r^2 versus physical distance for all marker pairs. Since the Loess curve is well below the threshold for LD due to physical linkage, the distance at which LD decay occurs was frequently below 2.5 cM which is slightly below the resolution which we achieved with our marker set (average distance between

markers approximately 7 cM). This indicated that in the utilized material, a higher marker density would have been beneficial.

Table 1: Summary of LD (r^2) for 8146 marker pairs

	all marker pairs				neighbouring marker pairs		
genetic distance	number of marker pairs	average r^2	number of significant marker pairs	% of significant marker pairs	number of marker pairs	average r^2	number of significant marker pairs
< 10 cM	854	0.089	701	82.1	357	0.105	294
10 cM – 20 cM	721	0.053	581	80.6	91	0.031	68
20 cM – 50 cM	1706	0.028	1241	72.7	27	0.033	18
> 50 cM	4865	0.011	2598	53.4	1	0.014	1

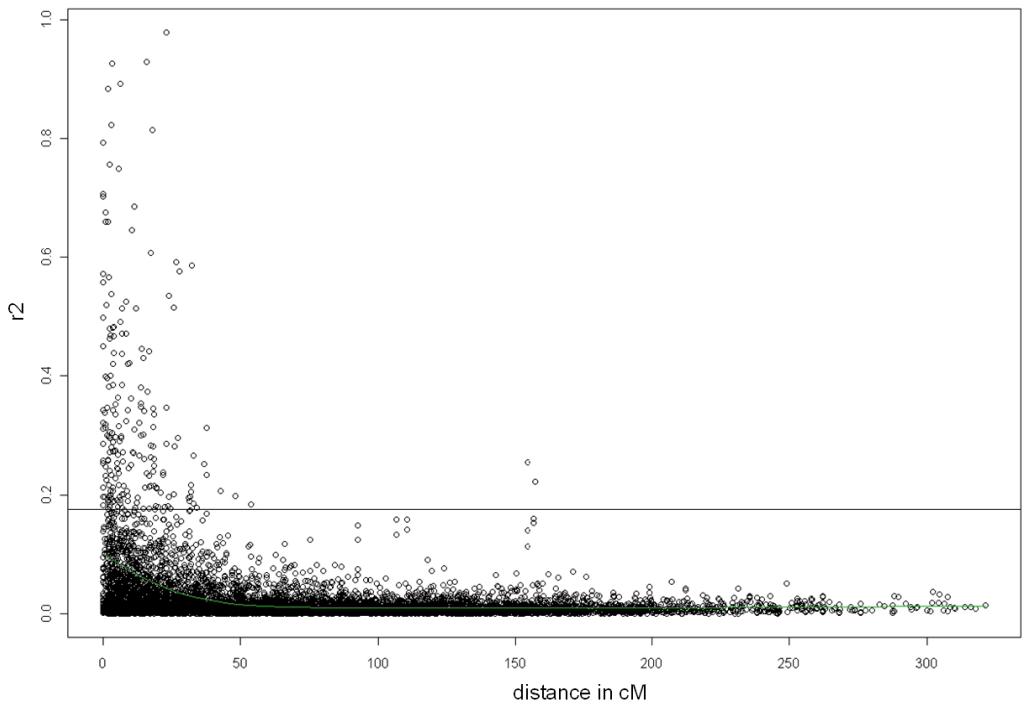
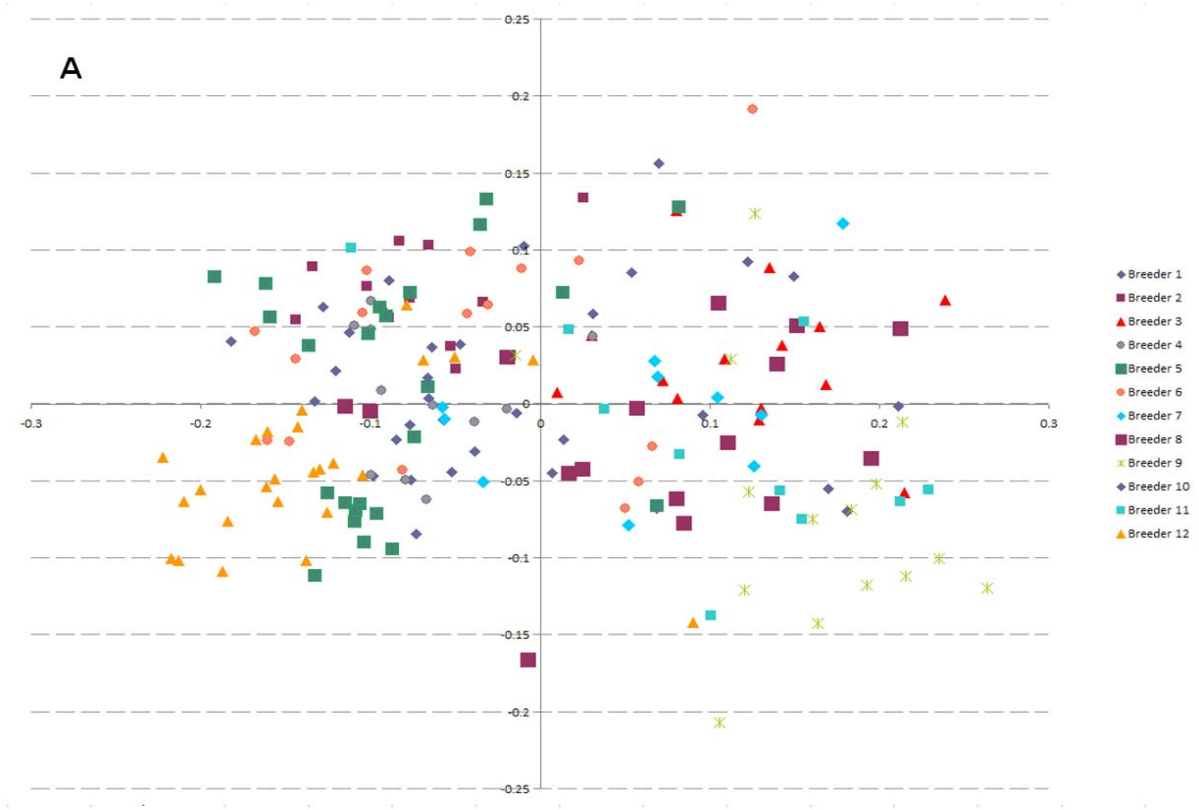


Figure 4: LD as r^2 versus genetic distance in cM. The horizontal line marks the threshold above which LD is likely due to physical linkage. The green curve represents the second degree Loess curve.

1C Population structure and kinship matrix

Population structure was assessed using different numbers of markers and different methods. The program STRUCTURE (Pritchard et al., 2000), which is based on Bayesian methods, was employed with 56 markers, distributed across 20 of the 21 chromosomes of wheat. No population structure was detected although the number of assumed subpopulations ranged from 1 up to 25 (data not shown). A principal coordinate analysis was also performed with the software NTSYS (Rohlf, 2008) using all markers. No apparent population structure was also detected there. Finally 155 markers were chosen regarding their distribution across the genome and reliability in terms of number of missing genotypes and heterozygotes. With these markers, Rogers' modified distance (Wright, 1987) was calculated for all pairs of varieties and a principal coordinate analysis performed. In Figure 4A, the varieties have been marked for those breeders, who had more than 15 varieties in the project. There were no distinct clusters visible. Varieties of different breeders were closely together in the plot, with only a few breeders showing any overlap (for example breeder 9 and breeder 12). In Figure 5B, the spring wheat varieties, which were thought of to constitute an outgroup, are marked. The 14 spring wheat varieties were not clustered closely together or away from the 358 winter wheat varieties. The fact that there is no apparent population structure in the material makes this population suited for association analysis (Rafalski, 2010) since population substructures could give rise to false positive results.



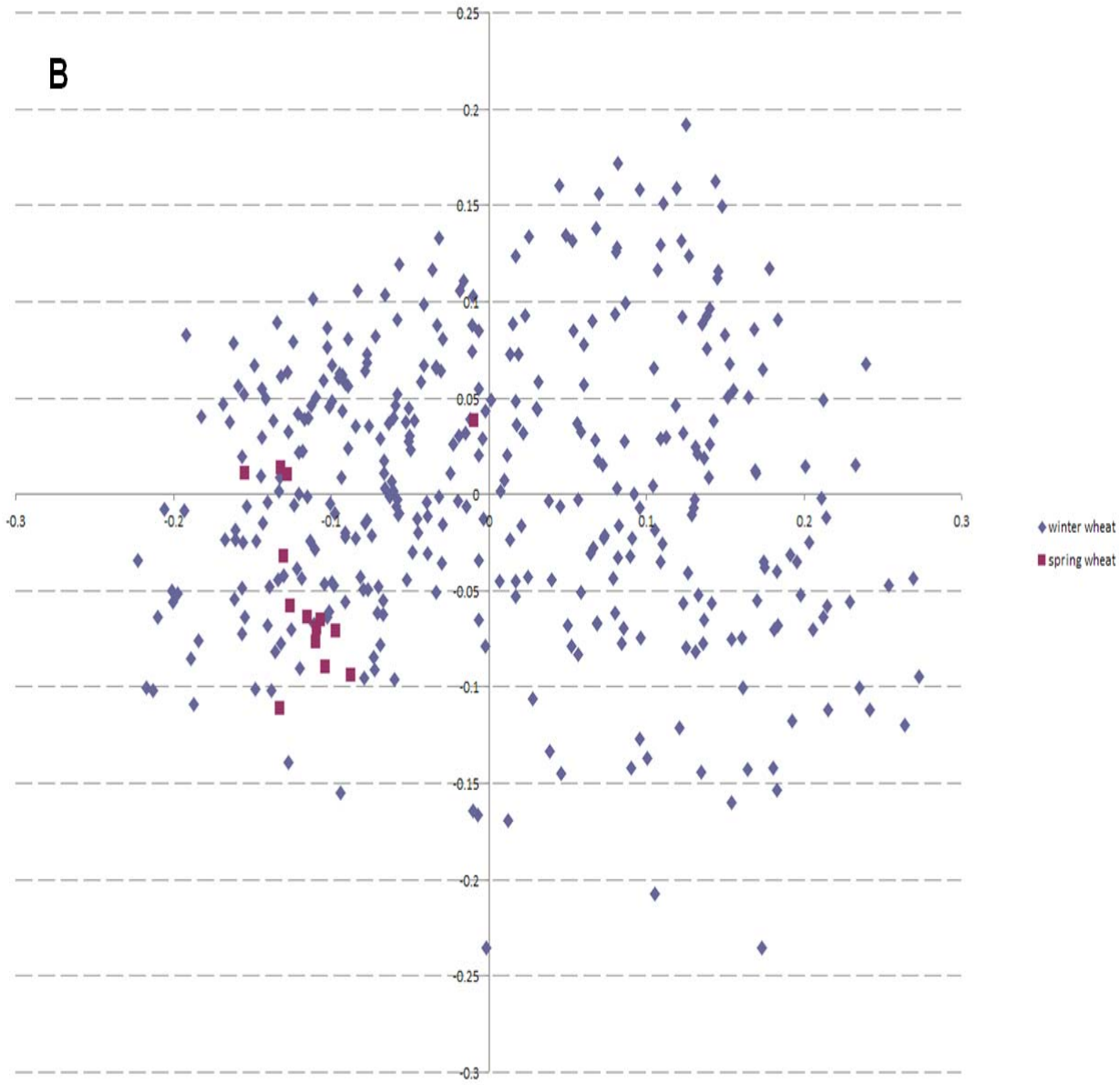


Figure 5: Principal coordinate analysis with 155 markers

A) distinguishes the results by breeder, showing 12 breeders which had the most varieties in the trial. B) shows the distinction between 14 spring wheat and 358 winter wheat varieties.

To account for familial relatedness a kinship matrix was calculated with the programme SPAGeDi 1.3 (Hardy and Vekemans, 2002) according to Loiselle et al. (1995) and utilized for the association mapping. From Figure 6, it becomes clear that the majority of the material (37252 variety pairs from 69006 possible pairs, amounting to 54 %) showed very low (<0.1) kinship coefficients, hinting at no relatedness. Only about 5 % show kinship coefficients of 0.3 or greater indicating familial relatedness. Nevertheless, the kingship coefficients were important for the analysis of marker/trait associations since correcting for kinship reduced the number of identified marker/trait associations in the subsequent analyses significantly.

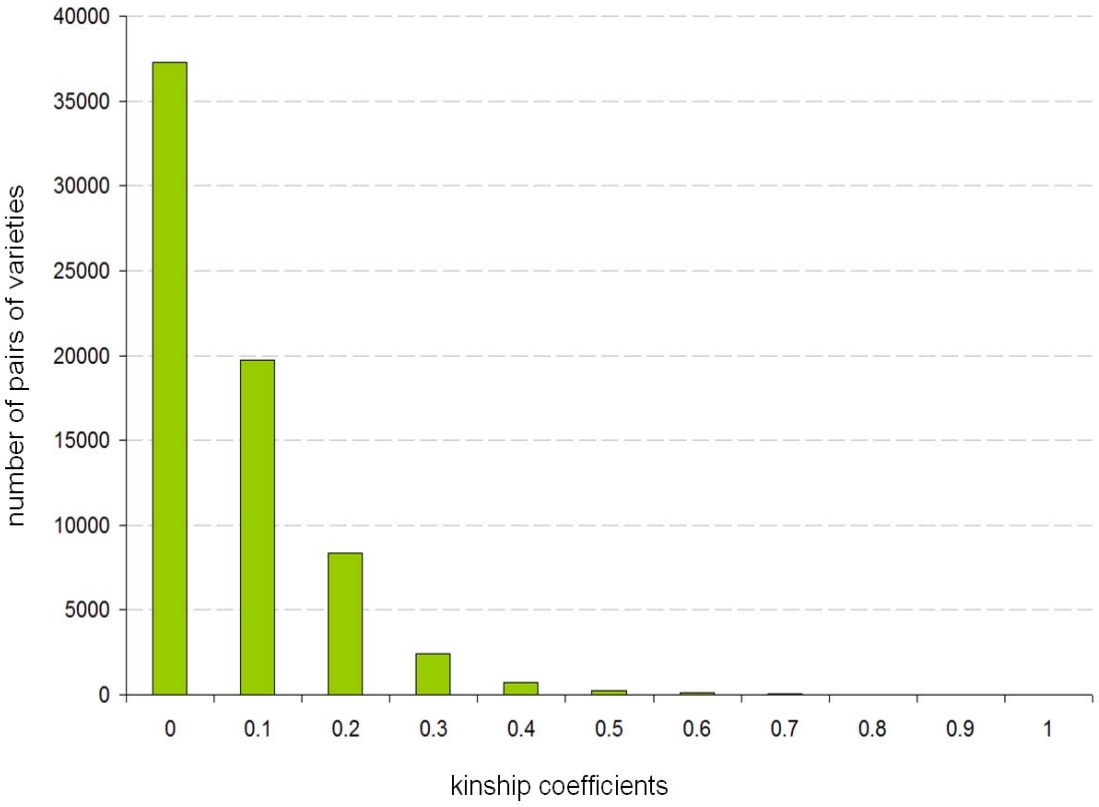


Figure 6: Distribution of pairwise kinship coefficients between 372 varieties calculated with 155 markers

1D Genetic map

Most of the markers could be mapped on the ITMI mapping population (W7984 x Opata 85) with the program MAPMAKER 3.0 (Lander et al., 1987) using the Kosambi mapping function (Kosambi, 1994) with a LOD score of three as the threshold for linkage. A total of 620 markers could be placed on the map with 19 markers mapping to more than one position in the genome, resulting in a map of 4470 cM length (Table 2). This resulted in an average marker distance of 7.1 cM (ranging from 5.1 cM on chromosome 4B to 11.4 cM on chromosome 6A). The genome was evenly covered, with the D-genome being slightly overrepresented (37.1 % of the markers) and homoeologous group 6 being slightly underrepresented (9.8 % of markers). The average number of alleles per marker was 4.1. This ranged from 3.1 on chromosome 4D to 5.4 on chromosome 6A. The three genomes exhibited very similar average numbers of alleles per marker with 4.3, 4.0 and 4.1 for the A-, B- and D-genome, respectively. Also the homoeologous groups are very similar (minimum of 4.1 to maximum of 4.7 for groups 5 and 1, respectively) with the exception of group 4 which showed an average number of alleles per marker of 3.4. A number of gaps larger than 10 cM were observed on every chromosome. This number ranged from 3 on chromosome 4B to 13 on chromosome 5D. Possibly some of the 112 markers which could not be mapped (mostly due to no segregation in the ITMI mapping population) will fill these gaps.

Table 2: Characteristics of the genetic map and mapped markers

chromosome	number of markers	length of map	average marker distance	stdev	min*	max	average number of alleles per marker	number of gaps > 10 cM
1A	22	177.6	8.1	7.7	1.5	33	4.8	6
1B	33	182.5	5.5	4.8	0.4	13.5	5.0	6
1D	24	209.6	8.7	8.8	1.3	35.5	4.3	7
2A	35	238.1	6.8	6.5	0.1	26	4.1	9
2B	35	226.0	6.5	7.7	0.6	40.1	3.8	8
2D	37	216.1	5.8	5.7	0.8	23.2	4.6	6
3A	33	213.2	6.5	6.9	1.5	32.6	4.2	4
3B	39	287.9	7.4	7.2	0.6	32.6	4.4	9
3D	40	214.8	5.4	5.4	0.1	22.4	3.9	7
4A	32	224.0	7.0	5.9	0.6	21.1	3.2	8
4B	21	106.1	5.1	4.8	0.5	17.8	3.8	3
4D	19	140.2	7.4	8.6	0.5	24.2	3.1	6
5A	30	234.3	7.8	8.1	1.0	37.1	3.8	10
5B	36	219.8	6.1	5.9	0.9	26.1	3.9	7
5D	46	321.4	7.0	5.4	0.7	21.1	4.5	13
6A	16	183.1	11.4	10.0	1.2	35.2	5.4	7
6B	18	126.1	7.0	6.3	0.9	19.7	3.7	6
6D	29	242.7	8.4	9.0	10.0	36.4	3.8	9
7A	31	250.2	8.1	9.3	0.1	43.3	4.9	10
7B	24	177.4	7.4	9.8	0.9	37.9	3.5	5
7D	44	278.9	6.3	5.5	0.7	21.7	4.3	11

* smallest value above 0

1E Candidate genes/polymorphisms and SNP analysis

As a proof-of-principle for the association genetics approach and to establish SNP genotyping procedures, candidate genes and polymorphisms were chosen from literature. Emphasis was placed on markers for which associations have been reported for the phenotypic and disease resistance traits that were scored in the project. The selected genes are listed in Table 2. In some cases haplotypes (e.g. from 3 reduced height genes Rht-B1b, Rht-D1b, Rht8) were constructed and tested for association.

Table 2: Candidate genes and polymorphisms for which the 372 varieties were genotyped

Gene/polymorphism name	Abbreviation	Candidate for	Reference
Matrix attachment region of the Bx7 gene	Bx_MARB	dough strength	Butow et al., 2004, Theor Appl Genet
	Glu-B1-1d	flour quality	Schwarz et al., 2004, Theor Appl Genet
Eyespot resistance gene	Pch1	resistance to eyespot	Groenewald et al., 2003, Plant Breeding
Puroindoline-a	Pina	grain hardness	Gautier et al., 1994, Plant Mol Biol
Puroindoline-b	Pinb	grain hardness	Huang and Röder, 2005, J Agric Food Chem
Photoperiod response locus	Ppd-D1a	sensitivity to day length	Beales et al., 2007, Theor Appl Genet
Dwarfing gene	Rht8	plant height	Chebotar et al., 2001, Genetika
Dwarfing gene	Rht-B1b	plant height	Ellis et al., 2002, Theor Appl Genet
Dwarfing gene	Rht-D1b	plant height	Ellis et al., 2002, Theor Appl Genet
	SBCMV	resistance to the soil borne cereal mosaic virus	Perovic et al., 2009, Mol Breeding
	Stb6	resistance to Septoria tritici blotch	Chartrain et al., 2005, Plant Pathol
	TAGW2	grain width and thousand grain weight	Su et al., 2011, Theor Appl Genet
Subunit Ax2* of Glu-A1	UMN19	flour quality	Liu et al., 2008, Theor Appl Genet
Dx2 and Dx5 subunits of Glu-A1	UMN25	flour quality	Liu et al., 2008, Theor Appl Genet
Dy10 and Dy12 subunits of Glu-A1	UMN26	flour quality	Liu et al., 2008, Theor Appl Genet
Viviparous-1B gene	Vp1B	pre-harvest sprouting	Xia et al., 2008, Euphytica
Vernalization gene	Vrn-B1	growth habit	Zhang et al., 2008, Crop Sci
Vernalization gene	Vrn-D1	growth habit	Zhang et al., 2008, Crop Sci

Genotyping of the candidate genes by the IPK was mostly performed with the Pyrosequencing technology. However, this technology was very costly and not amenable to very high throughput. Furthermore, it can be expected that in the future, there will be a switch from microsatellite markers to SNP markers in wheat. While in other crop plants, such as maize, SNP markers have already been used at the time of the start of the project, large scale SNP analysis in wheat was not possible due to two main points:

The first point was the lack of available SNP markers. SNP identification requires the comparative sequencing of lines or varieties and the identification of SNPs between these individuals. In contrast to diploid organisms, the identification of useful SNPs is hampered by the fact that in an allohexaploid species, SNPs detected between the three genomes have to be discriminated from SNPs observed between individual lines. SNPs between genomes, also termed intergenomic SNPs, are not useful but more prevalent.

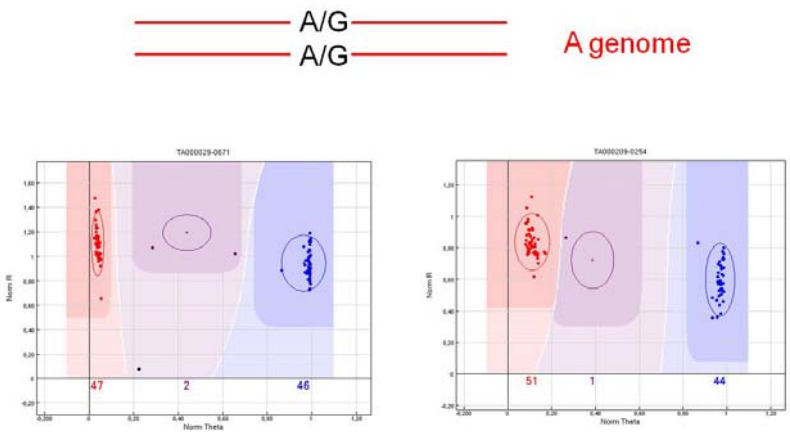
The second point was that most large scale SNP analysis procedures have been tailored towards the analysis of diploid species such as human or maize. In a hexaploid species such as wheat, it was not clear at the beginning of the project whether some SNP analysis technologies were suitable for large scale SNP analysis in wheat.

In order to evaluate the future use of large scale SNP analysis in wheat for association studies, TraitGenetics and the IPK have collaborated on an exploratory study regarding the analysis of the wheat genome with the Golden Gate SNP analysis technology. For this, SNPs were selected from a small scale SNP identification project that was performed in the framework of this project based on amplicon resequencing. Furthermore, additional SNPs were selected from the candidate gene set described in Table 2. Two sets of 192 SNPs were selected that fulfilled the Golden Gate assay design criteria from Illumina and these SNPs were subsequently analyzed on the 372 wheat lines and some additional controls as well as the ITMI population using the Veracode platform.

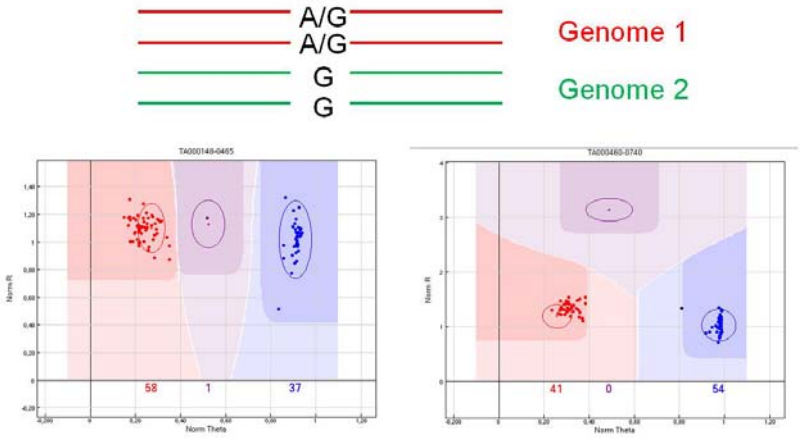
Results from these experiments clearly indicated that SNP analysis was possible in wheat using high throughput SNP analysis technologies such as the Golden Gate technology. However due to the hexaploid nature of wheat, the analysis of such SNPs was more difficult. In hexaploid wheat, an SNP assay can, depending on the chosen assay primers and the respective genomic sequence in the three genomes, detect one, two or all three genomes. This resulted in a number of different pattern types. If only one of the three genomes was detected, the cluster pattern was similar to that of a diploid organism (e.g. AA, AG and GG) as shown in Figure 7A. If two of the three genomes were detected by the SNP assay and one of the genomes was monomorphic in the analyzed material, then there was one monomorphic genome in the background which resulted in a shift of the clusters to one side. Thus the actual detected situation was for the three possible alleles GGGG, GGGA and GGAA as shown in Figure 7B. If all three genomes were detected by the SNP assay and two of the genomes were monomorphic in the analyzed material, then there were two monomorphic genomes in the background which resulted in a very strong shift of the clusters to one side. Thus the actual detected situation was for the three possible alleles GGGGGG, GGGGGA and GGGGAA as shown in Figure 7C. If an intergenomic SNP was analyzed, all material appeared to be heterozygous.

Based on these experiments, it could clearly be stated that SNP analysis in hexaploid wheat using high throughput technologies will be possible but the resulting SNP calling will need more time and efforts. Similar data have also been published elsewhere (Chao et al., 2010).

A



B



C

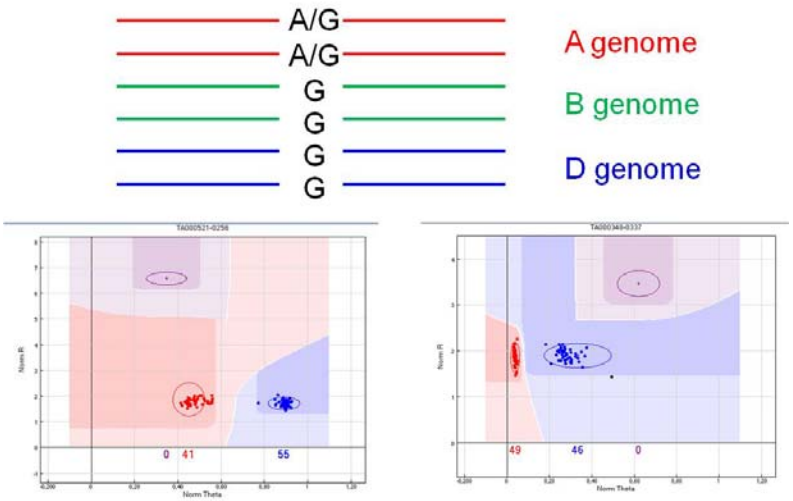


Figure 7: Cluster pattern of different types of SNPs in wheat lines

A) Cluster pattern for SNP assays detecting only one genome. B) Cluster pattern for SNP assays detecting two genomes. C) Cluster pattern for SNP assays detecting all three genomes.

1F Phenotypic evaluation of the material in the field

A total number of 358 winter wheat varieties and 14 spring wheat varieties were chosen by the breeders (SYN and KWS-L). Attention had been paid to the material being recent (of importance in the present breeding process) and of European origin. The 372 wheat varieties accessions were grown in two seasons. Season 1 was sown in 2008 and harvested 2009 (2008/2009). Season 2 was sown 2009 and harvested 2010 (2009/2010). In Germany, the varieties have been grown in two locations, Wohlde (WOH) and Seligenstadt (SEL), in both seasons. In France, the varieties have been evaluated in the season 2008/2009 in one location, Andelu (AND) since the second location could not be analyzed due to slug damage. To compensate that in season 2009/2010, three locations in France, Andelu (AND), Saultain (SAU), Janville (JAN) were used for the field evaluation.

Phenotyping was done by the breeders for a total of 13 traits. The availability of traits per season and location is shown in Figure 8.

Trait	Season	Andelu (AND)	Janville (JAN)	Saultain (SAU)	Seligenstadt (SEL)	Wohlde (WOH)
EW	2008/2009					
	2009/2010					
GH	2008/2009					
	2009/2010					
GPE	2008/2009					
	2009/2010					
HAG	2008/2009					
	2009/2010					
HD	2008/2009					
	2009/2010					
PC	2008/2009					
	2009/2010					
PH	2008/2009					
	2009/2010					
PY	2008/2009					
	2009/2010					
SDS	2008/2009					
	2009/2010					
STC	2008/2009					
	2009/2010					
SW	2008/2009					
	2009/2010					
TKW	2008/2009					
	2009/2010					
YIE	2008/2009					
	2009/2010					
ZEL	2008/2009					
	2009/2010					
	Disease inoculation trials					

trait	season	Ahlum (AHL)	Lafferde (LAF)	Cecilienkoog (CEC)	Halle-Bodenwerder (BOD)	
SEP	2008/2009					
	2009/2010					
DTR	2008/2009					
	2009/2010					
FHB	2008/2009					
	2009/2010					
	Available					
	Not available					

Figure 8: Availability of phenotyped traits per location and season

Phenotypes were available for all 8 season x location-combinations (or environments) for the traits heading date (HD), protein content (PC), plant height (PH), plot yield (PY) and from this yield (YIE) has been calculated, specific weight (SW), thousand kernel weight (TKW) and Zeleny volume (ZEL). The traits grain hardness (GH) and sedimentation index (SDS) were only determined at the French locations. The traits ear weight (EW), number of grains per ear (GPE), Hagberg falling number (HAG) and starch content (STC) were available for only a few of the environments.

Disease resistance has been evaluated in separate inoculation trials (subcontract to the Julius Kühn-Institut) at different locations in the same two seasons in Germany. Resistance to Fusarium head blight (FHB) has been evaluated in Ahlum (AHL) in both seasons, in Cecilienkoog (CEC) in season 2008/2009 and in Halle-Bodenwerder (BOD) in season 2009/2010. Resistance to Septoria tritici blotch (SEP) has been evaluated in CEC in both seasons and in Lafferde (LAF) in the season 2008/2009. Resistance to Drechslera tritici repentis (DTR) has been evaluated in AHL and in LAF in both seasons (see also Figure 8).

1G Association of microsatellite marker alleles with traits

Adjusted means were calculated for each environment employing the following formula:

$Y_{ilm} = \mu + g_i + r_l + b_{lm} + e_{ilm}$,

where y_{ilm} is the vector of phenotypic observations for variety i in the lth replicate in block m, μ is the intercept term, g_i is the genetic effect of the ith variety, r_l is the lth replication, b_{lm} is the mth block in replication l and e_{ilm} is the residual. The terms variety and replication are considered fixed and block is a random term.

BLUP values across all locations for a given trait were calculated with the following formula:

$Y_{ijklm} = \mu + g_i + y_j + l_k + r_{jkl} + b_{jklm} + e_{ijklm}$,

Where y_{ijklm} is the vector of phenotypic observations for variety i in year j at location k in the lth replicate in the mth block, g_i is the genetic effect of the ith variety, y_j is the jth year, l_k is location k, r_{jkl} is the lth replication at location k in year j, b_{jklm} is the mth block in the lth replication at location k in year j and e_{ijklm} is the residual. Year, location and replicate were considered fixed effects and genotype and block were considered random effects.

For the calculation of genotype-phenotype associations, microsatellite data were converted to SNP data. Each allele was considered as a single marker and genotypes were coded as 'A' when the variety possessed the allele and 'G' if it contained another allele. We applied a two-step association mapping approach using the adjusted means and BLUP values across all environments as trait values. Associations were calculated separately for all environments and the BLUP values. Calculations were performed with the software packages GenStat 13th edition (VSN International 2010) and the software package TASSEL version 2.0.1.

In GenStat a script provided by M. Malosetti (personal communication) was utilized. The statistical model employed was as follows:

$$Y_{il} = \mu + g_i + m_l + e_{il},$$

where y_{il} is the vector of adjusted means or BLUP values, μ is the intercept term, g_i is the ith variety, m_l denotes the genotype of the lth marker (that is allele) and e_{il} is the residual term. Marker genotypes were regarded as fixed effects and genotypes as random effects with the kinship matrix as the covariance structure of the random genotypes.

In TASSEL, the mixed linear model (MLM) with the same kinship matrix was used (Yu et al., 2006). The EMMA option (Kang et al., 2008) was chosen and the other parameters left at the default setting.

Here only marker-trait associations were considered which showed at least a $-\log_{10}(p)$ of 3, corresponding to a p-value of below 0.001 (Table 3). The number of markers associated for each trait per chromosome can be seen in Table 4. Figure 9 A – P show the Manhattan plots for the single traits). Associations with a $-\log_{10}(p)$ above 4.82 were still statistically significant after Bonferroni correction. On average 87 % of the associations were 'lost' through correction via Bonferroni. But Bonferroni correction has been considered to render significance thresholds which are too conservative (Bland and Altman, 1995). So the further presentation of results will consider all found associations.

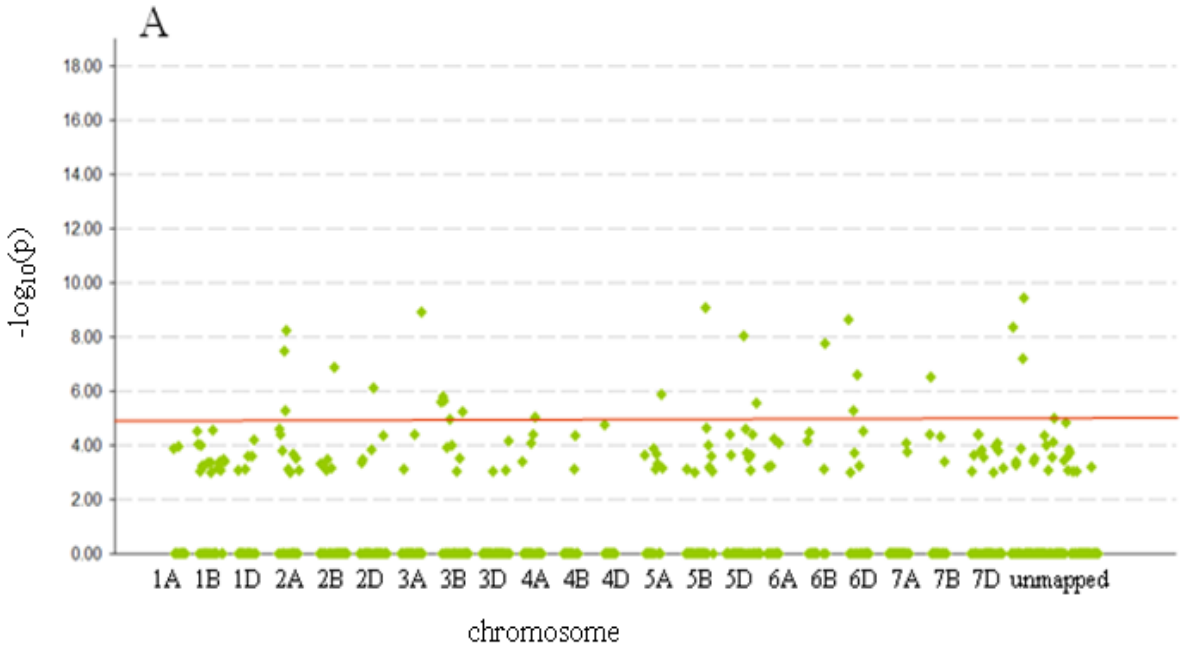
Table 3: Association summary for all microsatellite markers for 16 traits

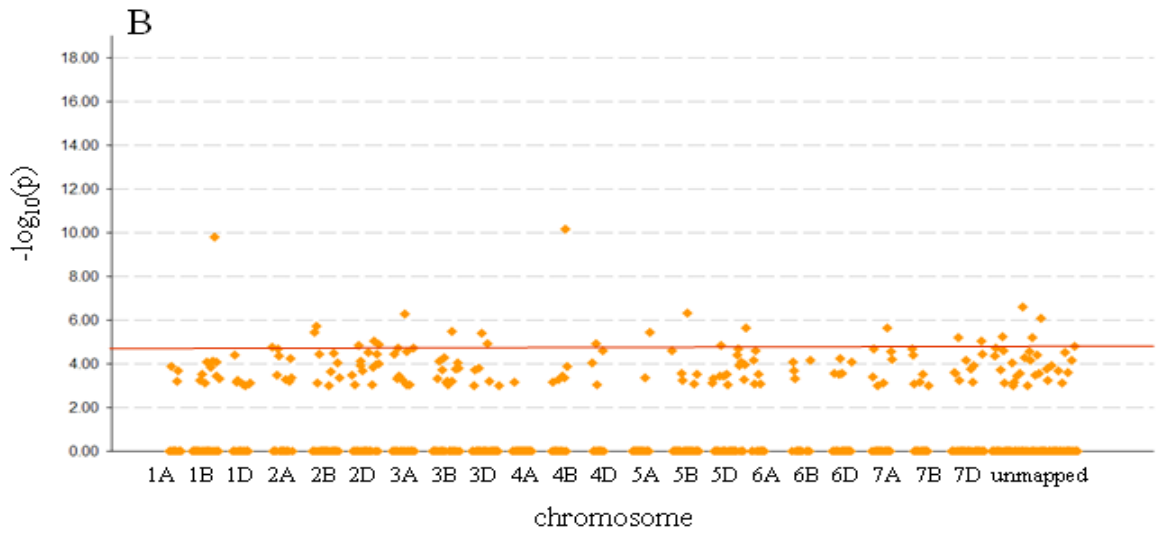
					per environment			per marker		
	Number of associated markers	Number of markers with $-\log_{10}(p) > 4.82$	Number of associations with BLUP values	Number of markers detected with BLUP values only	Average number of markers	min	max	Average number of environments	min	max
DTR	91	11	30	27	22.3	9	32	1.1	1	2
EW	30	1	14	5	13.0	3	23	1.0	1	2
FHB	79	9	26	25	18.0	15	21	1.4	1	4
GH	62	8	24	0	22.3	18	28	1.5	1	4
GPE	92	9	46	1	33.3	29	40	1.1	1	2
HAG	122	15	31	3	31.0	21	55	1.1	1	2
HD	152	26	68	0	62.8	40	84	3.4	1	8
PC	222	45	85	11	55.2	29	94	2.1	1	8
PH	185	24	75	0	65.9	39	91	2.9	1	8
SDS	118	21	51	4	42.5	34	51	1.6	1	4
SEP	87	10	33	3	30.7	28	35	1.2	1	2
STC	143	28	67	2	66.3	56	85	1.5	1	3
SW	196	14	40	1	46.1	25	67	1.9	1	7
TKW	136	7	30	1	27.1	20	37	1.7	1	7
YIE	202	28	56	17	33.0	17	49	1.6	1	4
ZEL	151	27	46	2	43.4	30	52	2.4	1	8

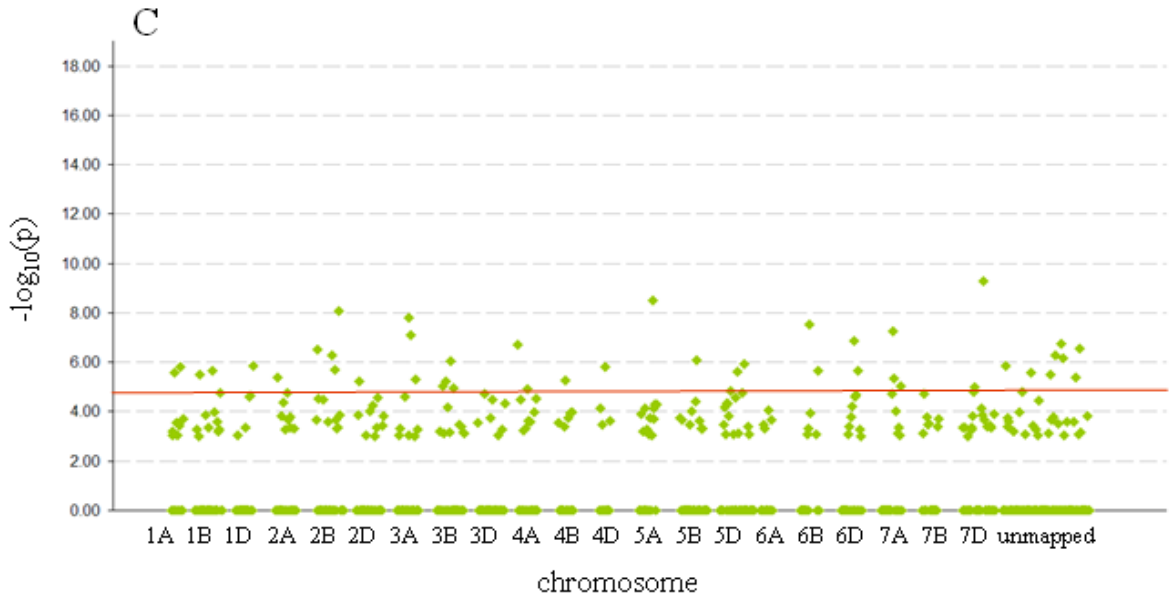
Table 4: Number of associated markers per trait and chromosome

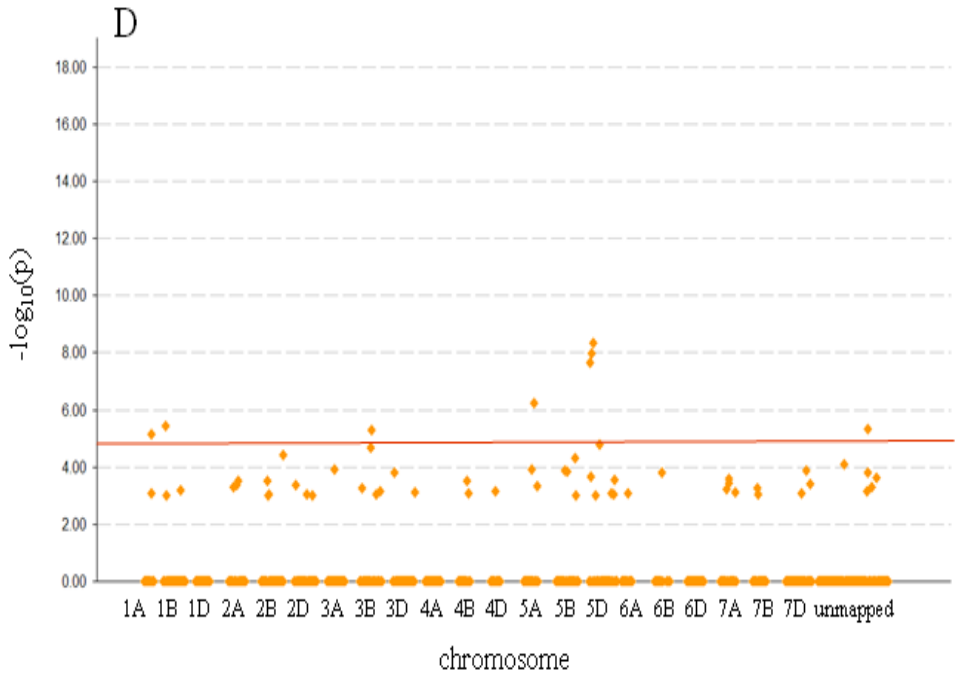
Chromo-some	# loci mapped	DTR	EW	FHB	GH	GPE	HAG	HD	PC	PH	SDS	SEP	STC	SW	TKW	YIE	ZEL
1A	22	4	0	2	2	3	3	2	8	3	5	1	2	4	2	2	5
1B	33	1	0	2	3	6	12	17	11	10	9	7	9	8	4	12	11
1D	24	4	1	1	0	1	3	5	5	7	6	2	3	4	2	7	6
2A	35	3	6	5	3	7	3	11	9	8	4	4	2	10	5	11	10
2B	35	3	1	7	3	4	4	6	14	9	7	3	11	8	7	10	7
2D	37	6	0	6	3	1	3	5	10	14	5	3	8	9	10	12	9
3A	33	6	0	2	1	2	3	3	8	10	3	5	4	9	4	7	3
3B	39	4	2	3	5	5	8	9	12	11	3	6	5	12	7	4	4
3D	40	4	0	2	2	1	3	3	7	6	5	4	4	11	7	8	9
4A	32	4	0	0	0	4	3	4	10	1	2	3	7	7	7	7	1
4B	21	1	1	2	2	2	3	2	6	6	4	6	8	6	2	6	3
4D	19	9	0	1	1	6	1	1	4	4	1	0	3	4	3	6	1
5A	30	4	2	7	4	5	5	9	12	3	5	4	6	7	6	8	11
5B	36	5	0	2	4	2	9	8	9	6	6	4	10	6	7	6	6
5D	46	6	4	3	9	7	8	10	15	14	7	6	8	13	10	12	14
6A	16	2	1	2	1	3	3	5	4	5	1	3	0	3	5	5	4
6B	18	0	0	0	1	1	1	2	4	3	2	0	2	3	4	4	4
6D	29	3	1	4	0	5	7	7	10	5	3	3	4	9	4	7	5
7A	31	3	2	3	4	2	10	2	11	7	9	0	5	8	3	6	7
7B	24	1	5	5	2	5	3	4	8	6	2	1	7	6	4	10	2
7D	44	2	0	5	2	4	7	12	16	10	6	6	11	11	5	10	8
unm*	138	16	4	15	10	16	20	25	29	37	23	16	24	38	28	42	21

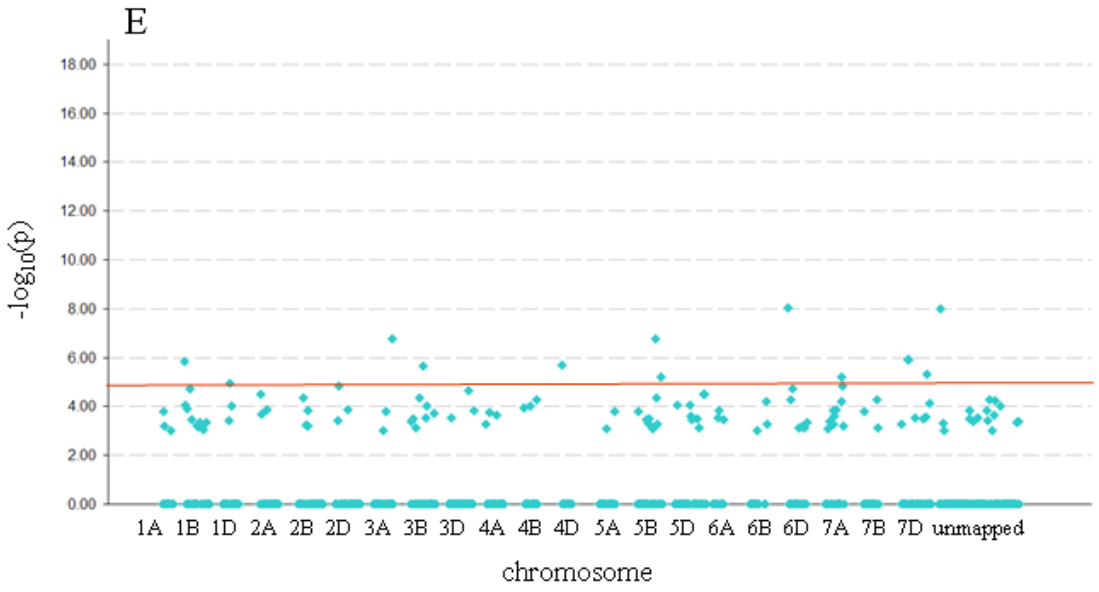
* unmapped markers

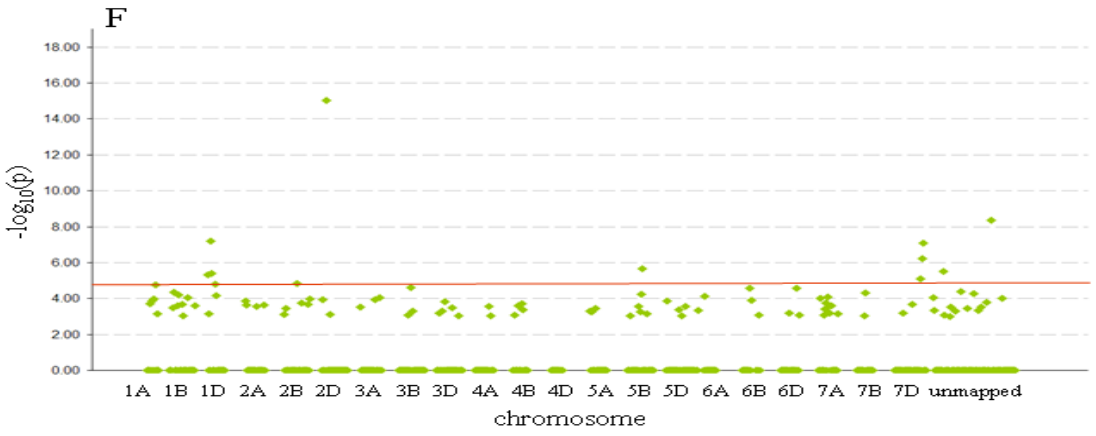

A
18.00
16.00
14.00
12.00
10.00
8.00
6.00
4.00
2.00
0.00
-log10(p)
1A 1B 1D 2A 2B 2D 3A 3B 3D 4A 4B 4D 5A 5B 5D 6A 6B 6D 7A 7B 7D unmapped
chromosome

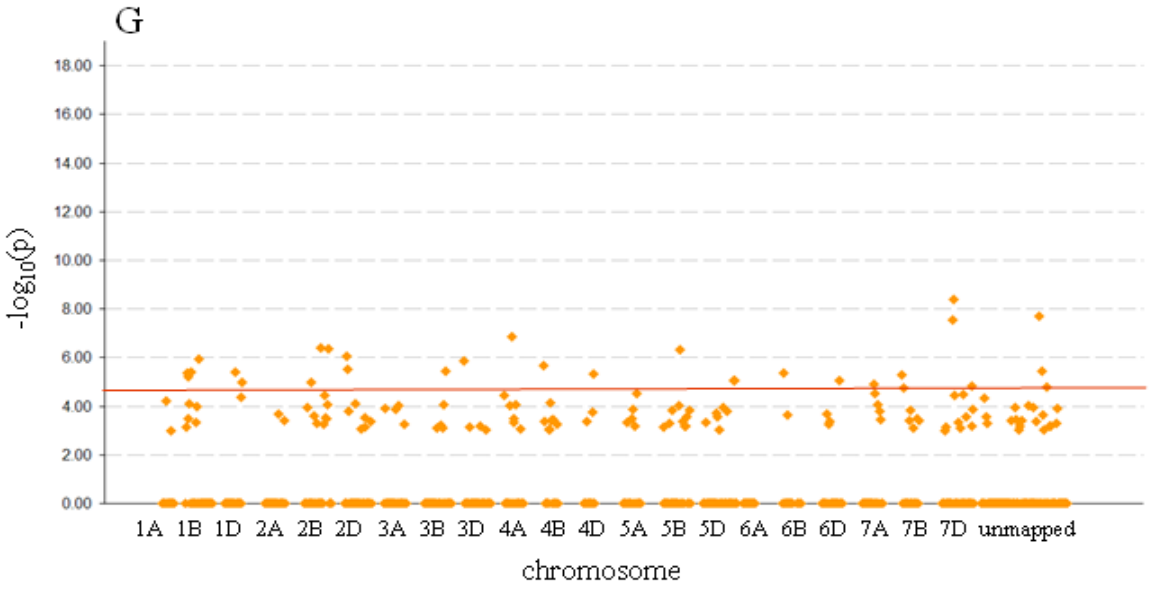

B
18.00
16.00
14.00
12.00
10.00
8.00
6.00
4.00
2.00
0.00
-log10(p)
1A 1B 1D 2A 2B 2D 3A 3B 3D 4A 4B 4D 5A 5B 5D 6A 6B 6D 7A 7B 7D unmapped
chromosome

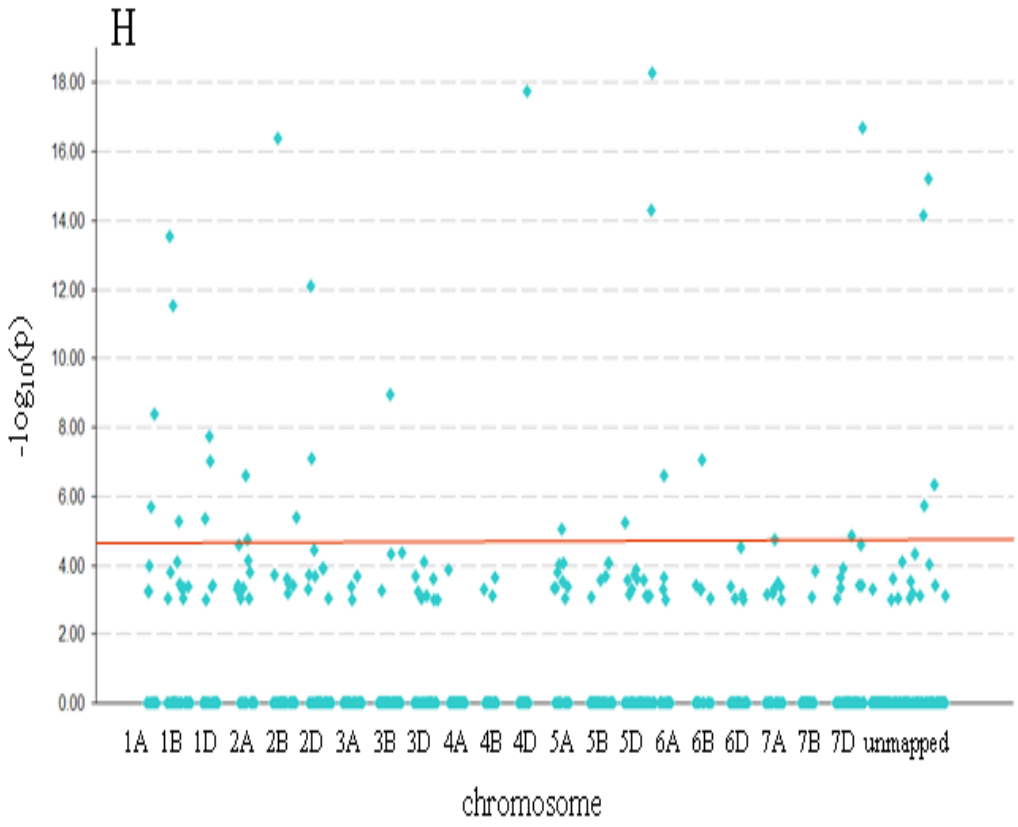

C
18.00
16.00
14.00
12.00
10.00
8.00
6.00
4.00
2.00
0.00
-log10(p)
1A 1B 1D 2A 2B 2D 3A 3B 3D 4A 4B 4D 5A 5B 5D 6A 6B 6D 7A 7B 7D unmapped
chromosome

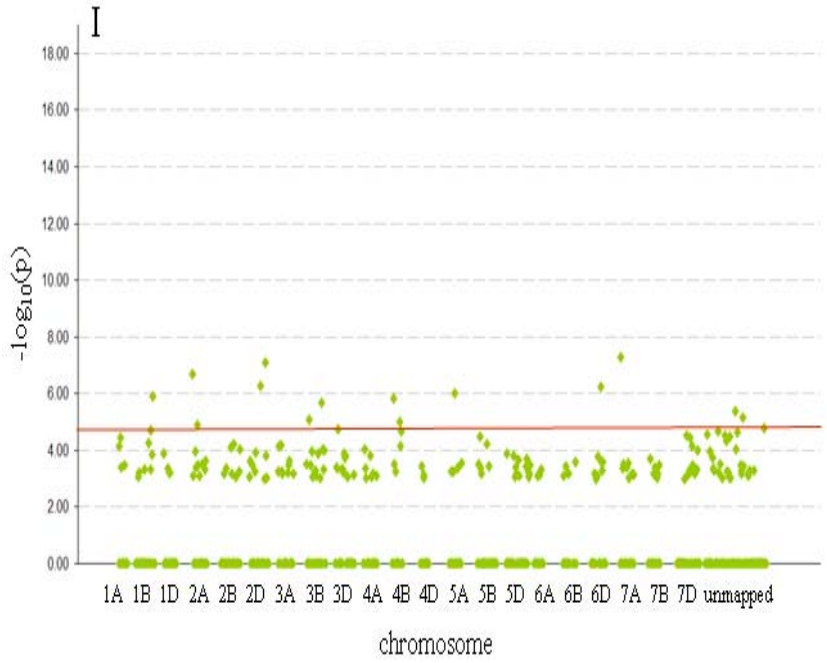

D
18.00
16.00
14.00
12.00
10.00
8.00
6.00
4.00
2.00
0.00
-log10(p)
1A 1B 1D 2A 2B 2D 3A 3B 3D 4A 4B 4D 5A 5B 5D 6A 6B 6D 7A 7B 7D unmapped
chromosome

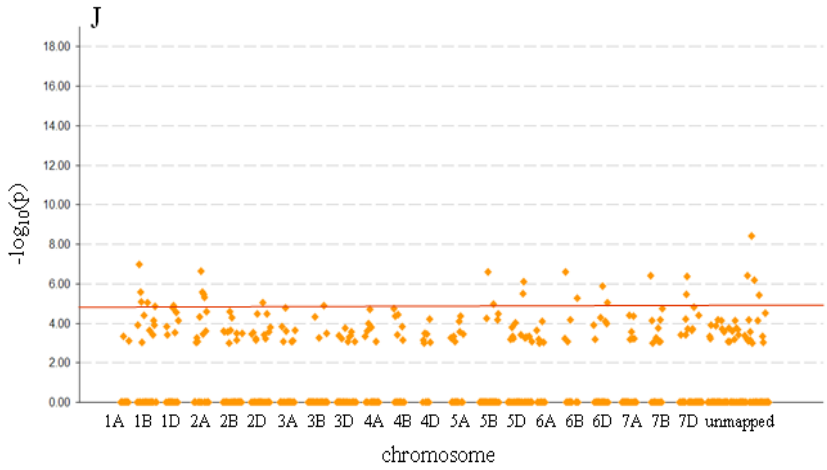

E
18.00
16.00
14.00
12.00
10.00
8.00
6.00
4.00
2.00
0.00
-log10(p)
1A 1B 1D 2A 2B 2D 3A 3B 3D 4A 4B 4D 5A 5B 5D 6A 6B 6D 7A 7B 7D unmapped
chromosome

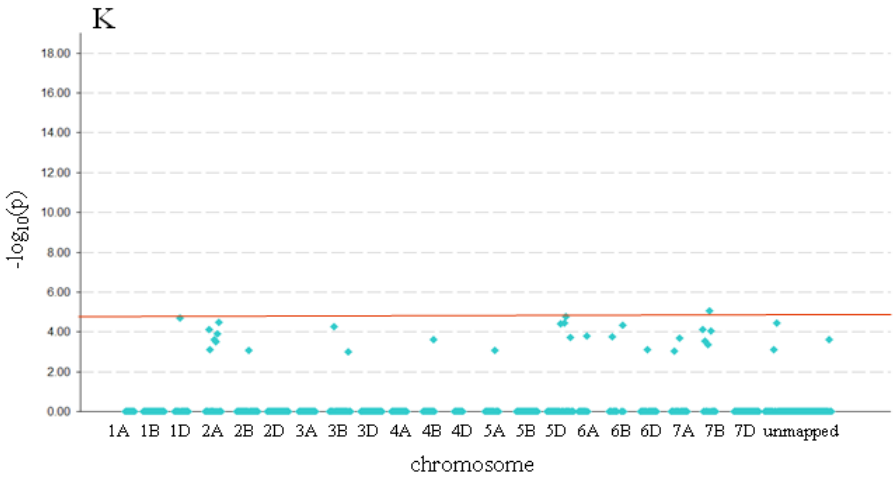

F
18.00
16.00
14.00
12.00
10.00
8.00
6.00
4.00
2.00
0.00
-log10(p)
1A 1B 1D 2A 2B 2D 3A 3B 3D 4A 4B 4D 5A 5B 5D 6A 6B 6D 7A 7B 7D unmapped
chromosome

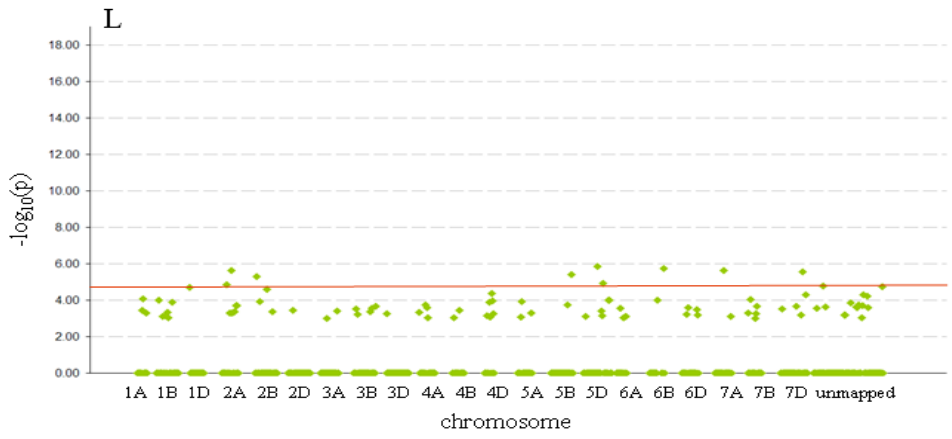

G
18.00
16.00
14.00
12.00
10.00
8.00
6.00
4.00
2.00
0.00
-log10(p)
1A 1B 1D 2A 2B 2D 3A 3B 3D 4A 4B 4D 5A 5B 5D 6A 6B 6D 7A 7B 7D unmapped
chromosome

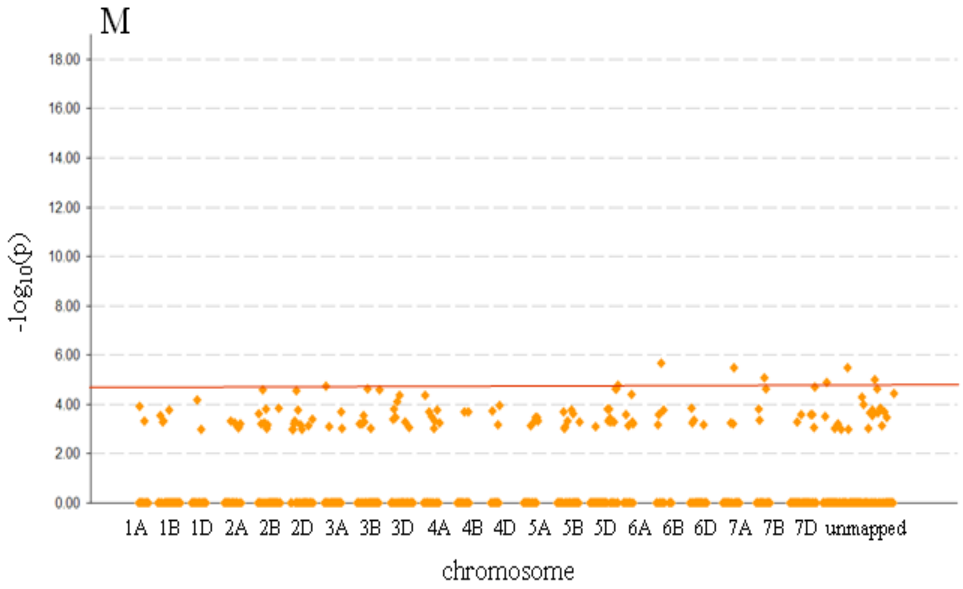

H
18.00
16.00
14.00
12.00
10.00
8.00
6.00
4.00
2.00
0.00
-log10(p)
1A 1B 1D 2A 2B 2D 3A 3B 3D 4A 4B 4D 5A 5B 5D 6A 6B 6D 7A 7B 7D unmapped
chromosome

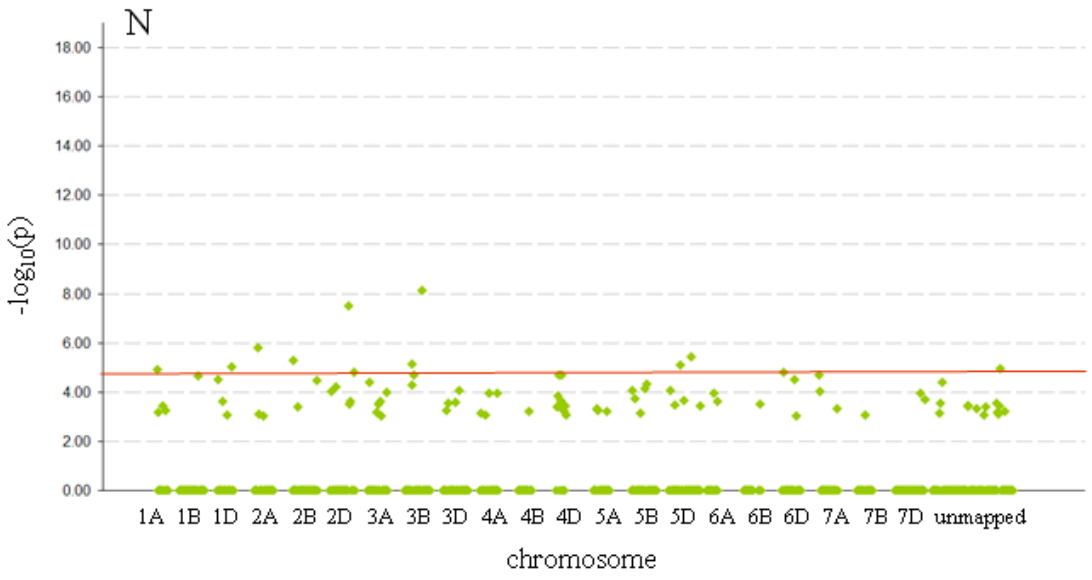

I
18.00
16.00
14.00
12.00
10.00
8.00
6.00
4.00
2.00
0.00
-log10(p)
1A 1B 1D 2A 2B 2D 3A 3B 3D 4A 4B 4D 5A 5B 5D 6A 6B 6D 7A 7B 7D unmapped
chromosome


J
18.00
16.00
14.00
12.00
10.00
8.00
6.00
4.00
2.00
0.00
-log10(p)
1A 1B 1D 2A 2B 2D 3A 3B 3D 4A 4B 4D 5A 5B 5D 6A 6B 6D 7A 7B 7D unmapped
chromosome


K
18.00
16.00
14.00
12.00
10.00
8.00
6.00
4.00
2.00
0.00
-log10(p)
1A 1B 1D 2A 2B 2D 3A 3B 3D 4A 4B 4D 5A 5B 5D 6A 6B 6D 7A 7B 7D unmapped
chromosome


L
18.00
16.00
14.00
12.00
10.00
8.00
6.00
4.00
2.00
0.00
-log10(p)
1A 1B 1D 2A 2B 2D 3A 3B 3D 4A 4B 4D 5A 5B 5D 6A 6B 6D 7A 7B 7D unmapped
chromosome


M
18.00
16.00
14.00
12.00
10.00
8.00
6.00
4.00
2.00
0.00
-log10(p)
1A 1B 1D 2A 2B 2D 3A 3B 3D 4A 4B 4D 5A 5B 5D 6A 6B 6D 7A 7B 7D unmapped
chromosome


N
18.00
16.00
14.00
12.00
10.00
8.00
6.00
4.00
2.00
0.00
-log10(p)
1A 1B 1D 2A 2B 2D 3A 3B 3D 4A 4B 4D 5A 5B 5D 6A 6B 6D 7A 7B 7D unmapped
chromosome

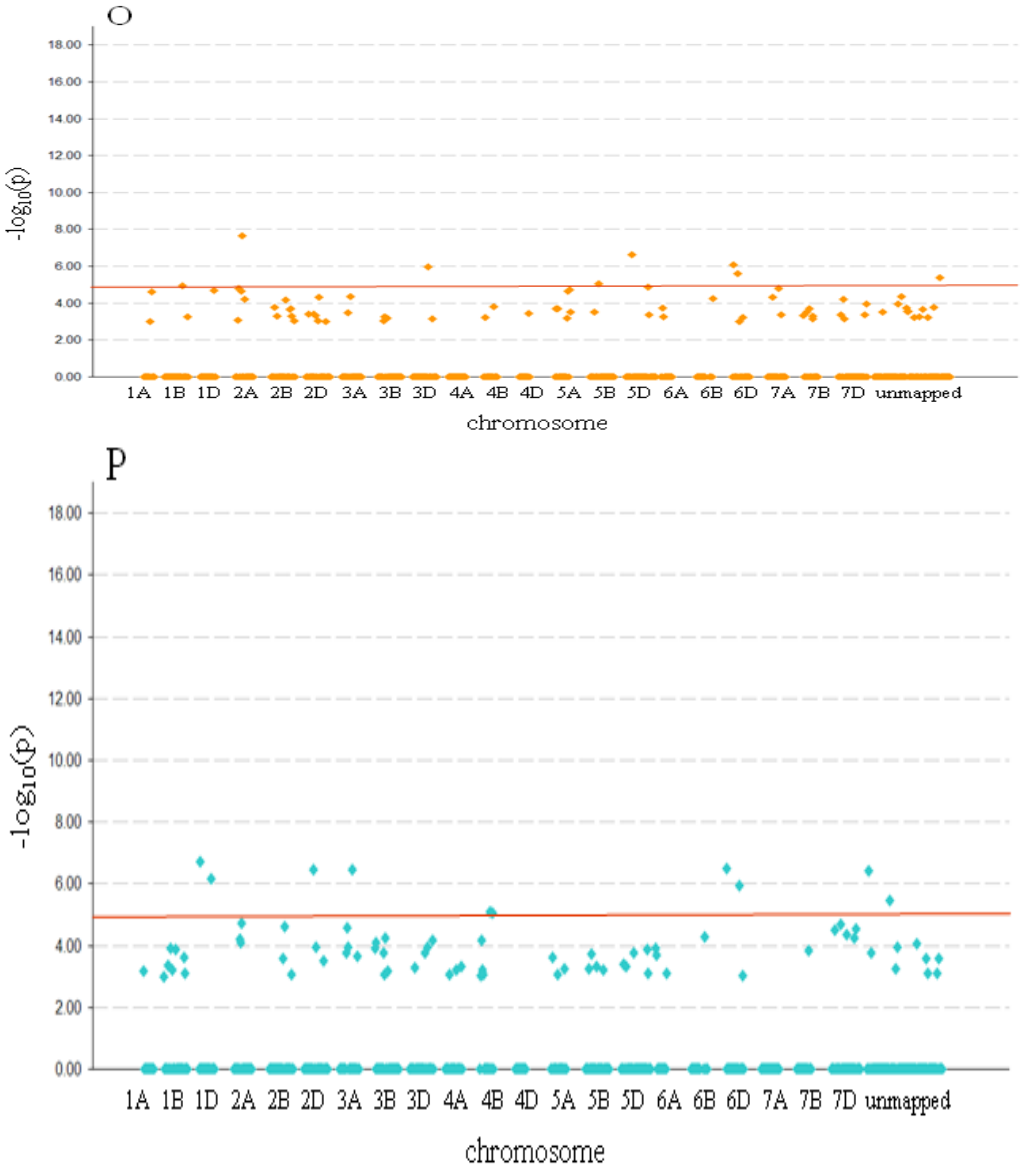


Figure 9: Manhattan plots of associations, the orange line indicates the threshold of $-\log_{10}(p)$ for multiple correction, all markers which are not associated or associated with a $-\log_{10}(p)$ below 3 are set to 0, A: heading date, B: plant height, C: protein content, D: grain hardness, E: Hagberg falling number, F: sedimentation index, G: starch content, H: Zeleny volume, I: specific weight, J: yield, K: ear weight, L: number of grains per ear, M: thousand kernel weight, N: resistance to Drechslera tritici repentis, O: resistance to Fusarium head blight, P: resistance to Septoria tritici.

The maximum number of markers associated with a trait was 222 for protein content (Table 3). After Bonferroni correction this trait still shows the highest number of associated markers with 45. The minimum number of markers associated with ear weight was 30, leaving one marker statistically significant after correction. This trait was measured in only two locations.

For each trait there were some markers associated in all of the investigated environments. This reached from an average of 13 markers being associated in both possible environments for EW (for EW only two environments were evaluated), with a minimum of 3 markers and a maximum of 23 markers being associated in a given environment and with 30 markers being associated altogether.

This left 4 markers (30-23-3) which were associated in both environments. This number of markers being associated per environment reached up to 66.3 for starch content.

Traits had been measured in 2, 3, 4 and 8 environments (Figure 8). The number of possible environments a marker was associated in could be thought of as a measure of repeatability. For heading date markers were associated on average in 3.4 environments from 8 possible ones (Table 3). For plant height, Zeleny volume and protein content markers were on average associated in 2.9, 2.4 and 2.1 environments, respectively. For all traits which were measured in 2, 3 and 4 environments this average was below 2.

Some markers showed only associations with BLUP values which had been calculated across all environments. For GH, HD and PH all of the markers associated with BLUP values were in addition associated with adjusted means in at least one environment (Table 3). For DTR from the 91 associated markers, 30 were associated with BLUP values and 27 of those were associated only with BLUP values.

Table 4 shows the number of markers associated for each trait on each chromosome plus the unmapped markers. This does not include candidate genes/polymorphisms. Markers which were associated and were in close vicinity to candidate genes/polymorphisms may reflect the effect of the candidate gene. This could be the case for chromosomes 1B and 1D for SEP where this trait showed the highest number of associated markers. For 2D on which Ppd-D1a is located most of the traits showed an association, and for 5D where 5 of the traits (GH, PH, SW, TKW, YIE) showed their highest number of associated markers for a chromosome and which includes the grain-hardness locus puroindoline b (Table 5). The highest number of markers for which a trait showed associations on a single chromosome is 14 for PH on chromosomes 2D and 5D and for ZEL on chromosome 5D. In all cases this could be due to the presence of candidate genes for which both traits showed associations on the mentioned chromosomes.

1F Associations of candidate genes/polymorphisms

From 19 candidate genes/polymorphisms (Table 5) four are not associated with any of the traits. For the trait ear weight none of the candidate genes/polymorphisms was associated. This trait was only available for 2 environments. The two puroindoline genes a and b were candidates for grain hardness. Puroindoline a was not associated with grain hardness, but only with yield, whereas puroindoline b was associated with grain hardness as expected and in addition with the other quality traits PC, SDS, STC and ZEL. It was also associated with PH, SW and YIE. Four other polymorphisms which were candidates for flour quality were genotyped. One of these was associated with the two quality traits SDS and ZEL. Another polymorphism was only associated with GPE and not with any of the quality traits. The remaining two polymorphisms showed an identical pattern of association. They were associated with the quality traits GH, SDS, SW and ZEL. The candidate gene for pre-harvest sprouting (Vp1b) was associated with HAG which in turn was dependent on protein composition which is influenced by the physiological processes during sprouting. Three dwarfing genes were genotyped and a haplotype matrix was constructed from these genotypes. One of those (Rht-B1b = Rht1) was only associated with TKW but not with PH. Rht8 was associated with PH as expected and with the yield traits SW, TKW and YIE as well as with SEP and ZEL. Rht-D1b (=Rht2) was associated with PH, all the disease traits (FHB, SEP, DTR), all the yield traits (GPE, SW, TKW, YIE), the quality

traits PC, SDS and ZEL. From 14 different haplotypes constructed from the three dwarfing genes, three showed a good frequency (22.9 %, 32.6 %, 14.4 %) these were associated with YIE only, with YIE, FHB, GPE, HD, PH and SW and with FHB, HD, PH, SW and YIE, respectively. Another 2 haplotypes had an acceptable frequency of 10.5 % and 6.6 %, respectively. These are associated with FHB, PC, STC, SW, YIE, ZEL and PC, PH, SW, YIE, ZEL, respectively. TaGW2 is in rice a candidate for thousand grain weight and grain width (Su et al., 2011). In our study it was associated with TKW and YIE. Three candidate genes for different disease resistances had genotyped. The gene for eyespot resistance was not associated with any trait, but eyespot had not been evaluated in the study. The polymorphism for resistance to the soil-borne-mosaic-virus was associated with SEP. Stb6 as a candidate for resistance to Septoria tritici blotch was genotyped and was associated with SEP as expected. It was also associated with FHB, PC, PH, STC, TKW and YIE. Most of these associations were due to rare and super-rare alleles. Two major alleles constituted most of the genotypes. With frequencies of 63.4 % and 14.4 % these were associated with TKW and YIE, respectively.

1G Correlation between traits

Correlations between traits have been calculated as the Pearsons' correlation coefficient. In Table 6 values above the diagonal show the number of markers for which both traits are jointly associated and values below the diagonal show the correlation coefficient for significant correlations only. The minimum number of markers associated for the same two traits was 2 and was between SEP and EW. The maximum number of markers associated in common between two traits was 101 and this accounted for the highly negatively correlated traits YIE and PC. On average, any two traits shared 28.1 associated markers. Traits which were highly correlated (correlation coefficient above 0.3) shared on average more markers with 38.3 than traits which were not highly correlated (correlation coefficient below 0.3) with 21.1 markers. The maximum correlation of 0.89 was found between SDS and ZEL. These two traits shared 65 markers. Expected negative correlations between FHB and HD, FHB and PH, GPE and TKW, PC and YIE had been confirmed. Expected positive correlations between DTR and FHB, DTR and SEP, EW and GPE, EW and YIE, FHB and SEP, GPE and YIE were also confirmed. Not confirmed were the expected positive correlations between SW and YIE, which turned out to be negative, and the expected negative correlations between DTR and YIE and FHB and YIE which turned out to be positive.

Table 5: Summary of associations of 16 traits with candidate genes/polymorphisms

Shown are the highest $-\log_{10}(p)$-values, orange are those associations with a $-\log_{10}(p)$-value above 4.82 which remain significant after Bonferroni correction.

gene	candidate for	DTR	EW	FHB	GH	GPE	HAG	HD	PC	PH	SDS	SEP	STC	SW	TKW	YIE	ZEL
Bx_MARB	dough strength																
Glu-B1-1d	flour quality										6.85						5.53
hapRHts	plant height			4.21		3.69		7.80	4.32	11.54		3.25	3.28	6.36	3.45	55.61	4.68
Pch1	resistance to eyespot																
Pina	grain hardness															4.98	
Pinb	grain hardness				40.90				3.89	3.24	9.80		4.75	3.78		3.19	20.41
Ppd-D1a	sensitivity to day length	3.12		3.50		6.45	12.45	43.93	3.88	10.23		11.50		7.65	14.60	4.20	
Rht8	plant height									3.18		3.25		3.14	4.31	3.33	4.31
Rht-B1b	plant height														3.44		
Rht-D1b	plant height	10.01		13.77		4.65	3.84		3.74	24.08	4.92	3.23		8.72		3.70	
SBCMV	resistance to soil-borne-mosaic-virus											3.82					
Stb6	resistance to Septoria tritici blotch			5.92					7.43	4.23		6.51	5.04		4.00	5.87	
TaGW2	grain width and thousand grain weight														3.52	56.43	
UMN19	flour quality					3.37										3.82	
UMN25	flour quality				3.63						4.86			3.20			7.70
UMN26	flour quality				3.30						4.87			3.68			7.47
Vp1B	pre-harvest sprouting						3.83										
Vrn-B1	growth habit																
Vrn-D1	growth habit																

Table 6: Trait correlations

Values above the diagonal show the number of markers with which both traits are associated, values below the diagonal are Pearsons' correlation coefficients shown only for significant ($p < 0.05$) correlations, green are positive correlations, orange are negative correlations, yellow are correlations which range from negative to positive.

	DTR	EW	FHB	GH	GPE	HAG	HD	PC	PH	SDS	SEP	STC	SW	TKW	YIE	ZEL
DTR		4	22	9	21	13	16	36	34	14	16	20	27	25	37	15
EW	0.13		3	3	13	6	15	13	12	6	2	4	15	8	16	6
FHB	0.12/0.33*	017/0.26		9	18	13	22	30	37	8	16	23	20	22	28	16
GH	0.12/0.13	ns#	0.14/0.29		18	20	17	30	24	21	9	17	23	16	25	30
GPE	0.11/0.12	0.21/0.69	0.11/0.37	0.11/0.23		27	36	43	26	26	11	27	25	38	44	28
HAG	0.11/0.18	0.12/0.13	0.1/0.31	0.1/0.25	0.11/0.13		48	38	31	28	18	31	35	28	30	39
HD	0.13/0.21	0.12/0.2	0.1/0.42	0.25	0.12/0.45	0.22/0.45		59	32	22	32	39	41	35	54	38
PC	0.1/0.18	0.11/0.23	0.13/0.46	0.11/0.4	0.11/0.43	0.11/0.28	0.52/0.18		68	31	32	80	66	47	101	67
PH	0.15/0.34	ns	0.38/0.68	0.16/0.31	0.1/0.27	0.1/0.28	0.13/0.33	0.15/0.47		34	25	40	83	44	70	43
SDS	0.11/0.13	0.12/0.13	0.11/0.29	0.3/0.62	0.14/0.3	0.1/0.35	0.14/0.32	0.18/0.53	0.17/0.43		16	32	45	28	38	65
SEP	0.15/0.2	ns	0.12/0.48	0.12	ns	0.12/0.32	0.11/0.47	0.24/0.13	0.2/0.44	0.1		16	32	19	32	14
STC	0.11/0.12	ns	0.19/0.25	0.23/0.34	0.13/0.21	0.11/0.23	0.22/0.37	0.32/0.65	0.1/0.23	0.14/0.29	0.2/0.1		37	38	56	40
SW	0.1/0.31	0.11/0.17	0.17/0.56	0.23/0.45	0.14/0.4	0.12/0.31	0.1/0.55	0.16/0.59	0.35/0.71	0.33/0.53	0.11/0.34	0.11/0.3		46	62	54
TKW	0.14/0.11	0.1/0.26	0.12/0.14	0.1/0.22	0.26/0.53	0.21/0.13	0.12/0.58	0.11/0.36	0.11/0.32	0.1/0.28	0.14/0.28	0.25/0.14	0.1/0.4		48	32
YIE	0.11/0.17	0.11/0.22	0.12/0.41	0.1/0.35	0.1/0.3	0.26/0.14	0.35/0.5	0.22/0.71	0.1/0.4	0.11/0.45	0.35/0.13	0.23/0.57	0.11/0.46	0.29/0.35		52
ZEL	0.11/0.17	0.11/0.12	0.12/0.31	0.41/0.62	0.1/0.35	0.11/0.28	0.1/0.41	0.27/0.68	0.27/0.41	0.52/0.86	0.11	0.2/0.48	0.39/0.69	0.1/0.39	0.12/0.53	

* minimum/maximum

not significant

2 Wichtige Positionen des zahlenmäßigen Nachweises

0812 Wissenschaftler:	151.488,10 €
0817 Techn. Assistent/in:	119.988,76 €
0835 Vergabe von Aufträgen:	68.502,00 €
0843 Sachmittel:	109.449,31 €
0846 Dienstreisen:	7.182,40 €
Summe:	456.610,57 €

3 Notwendigkeit und Angemessenheit der geleisteten Arbeit

Die benötigte Zeit für die Genotypisierung war notwendig und angemessen, da es sich um die Etablierung einer routinemäßigen Anwendung eines neuen Gerätes handelte. Arbeitsabläufe mussten etabliert und standardisiert werden, sowie die Auswertung der Ergebnisse erarbeitet werden. Ferner musste Erfahrung mit der Auswertung der Daten gesammelt werden. Dies hat zu den beschriebenen Ergebnissen geführt.

Über das gesamte Verbundprojekt kann gesagt werden, dass alle Ziele und Meilensteine aus dem ursprünglichen Antrag voll erreicht werden konnten. Die erhaltenen Ergebnisse stellen aktuell vermutlich den umfangreichsten Datensatz für europäischen Winterweizen dar.

4 Voraussichtlicher Nutzen und Verwertbarkeit der Ergebnisse

Die gewonnenen Erkenntnisse über Marker-Merkmals-Asosziationen müssen validiert werden. Dies geschieht unter anderem durch das Folgeprojekt VALID, das vom BMBF unter der Nummer 0315947 mit denselben Verbundpartnern gefördert wird. Diese Ergebnisse können von den Züchtungsfirmen im Rahmen ihrer Zuchtprogramme mittels der markergestützen Selektion eingesetzt werden. Marker-Merkmals-Assoziationen, die nicht von den Züchtern genutzt werden können, oder die Bestätigungen bereits bestehender Ergebnisse (z. B. aus QTL-Studien) darstellen, können vom IPK als Forschungseinrichtung in Zusammenarbeit mit den anderen Verbundpartnern publiziert werden. Für eine weitere zukünftige Nutzung ergeben sich für die einzelnen Verbundpartner folgende Möglichkeiten:

IPK

Das IPK erhält gut charakterisiertes Weizenmaterial, welches im Rahmen weiterführender Studien von Nutzen sein kann. Ein Teil des Materials ist bereits Gegenstand weiterer Anträge.

TraitGenetics

TG kann die Daten, die mit proprietären Markern verknüpft sind, effektiver im Rahmen von Dienstleistungen für Kunden nutzen. Weiterhin ist die gewonnene Erfahrung im Bereich der

Marker/Merkmalsanalyse wichtig für Projekte mit Partnern in anderen Pflanzenarten. Ferner hat TraitGenetics Erfahrungen im Bereich der Nutzung von SNP-Markern in polyploiden Pflanzen gesammelt.

Züchtungsfirmen

Die Züchtungsfirmen profitieren von der Entwicklung neuer diagnostischer Marker, sowie der Entdeckung neuer Marker/Merkmalsassoziationen. Eventuell können solche Marker direkt für die markergestützte Selektion verwendet werden. Die in diesem Projekt gewonnene Erfahrung kann auch auf andere Fruchtarten (z. B. Roggen) übertragen werden.

Die mit Mikrosatelliten gewonnen Erkenntnisse können später mit SNP Sets komplementiert werden. In diesem Projekt wurde gezeigt, dass es möglich ist, mit der bereits durch Mikrosatelliten gewährleisteten Markerabdeckung, nutzbare Ergebnisse zu erzielen.

5 Bekannt gewordene Fortschritte auf dem Gebiet des Vorhabens bei anderen Stellen

Während der Laufzeit des Projektes wurden mehrere Artikel über Populationsstruktur und Assoziationskartierung in Weizen publiziert. Die meisten dieser Publikationen hatten allerdings eine deutlich geringere Mikrosatellitenmarkerdichte als in der vorliegenden Arbeit oder verwendeten DArT-Marker. Diese Marker liefern weniger Informationen als die verwendeten Mikrosatellitenmarker, da es sich dabei, wie bei SNPs, nur um biallele Marker handelt. Ferner sind diese Marker zu einem gewissen Maß im Bereich der Zentromere geclustert. Zusätzlich wurden in nahezu allen dieser Arbeiten deutlich weniger Linien oder Sorten untersucht. Ein wichtiger Aspekt für die Zukunft ist auch der Vergleich der in dem Projekt erhaltenen Daten mit den von anderen veröffentlichten Daten. Wie im Bereich der humanen Assoziationsanalyse wird es interessant sein zu sehen in wie weit die Ergebnisse in unterschiedlichen Materialstichproben übereinstimmen.

Bordes J, Ravel C, Le Gouis J, Lapierre A, Charmet G, Balfourier F: Use of a global wheat core collection for association analysis of flour and dough quality traits. Journal of Cereal Science 54:137-147 (2011).

Chao SM, Dubcovsky J, Dvorak J, Luo MC et al.: Population- and genome-specific patterns of linkage disequilibrium and SNP variation in spring and winter wheat (*Triticum aestivum* L.). BMC Genomics 11: 727 (2010).

Emebiri LC, Oliver JR, Mrva K, Mares D: Association mapping of late maturity alpha-amylase (LMA) activity in a collection of synthetic hexaploid wheat. Molecular Breeding 26: 39-49 (2010).

Le Couviour F, Faure S, Poupard B, Flodrops Y, DubreuilP, Praud S: Analysis of genetic structure in a panel of elite wheat varieties and relevance for association mapping. Theoretical and Applied Genetics 123:715-727 (2011).

Le Gouis J, Bordes J, Ravel C, Heumez E, Faure S, Praud S, Galic N, Remoué C, Balfourier F, Allard V, Rousset M: Genome-wide association analysis to identify chromosomal regions determining components of earliness in wheat. Theoretical and Applied Genetics (available online).

Miedaner T, Würschum T, Maurer HP, Korzun V, Ebmeyer E, Reif JC: Association mapping for Fusarium head blight resistance in European soft winter wheat. Molecular Breeding 28:647-655 (2011).

Neumann K, Kobiljski B, Dencic S, Varshney RK, Börner A: Genome-wide association mapping: a case study in bread wheat (Triticum aestivum L.). Molecular Breeding 27:37-58 (2011).

Reif JC, Gowda M, Maurer HP, Longin CFH, Korzun V, Ebmeyer E, Bothe R, Pietsch C, Würschum T: Association mapping for quality traits in soft winter wheat. Theoretical and Applied Genetics 122:961-970 (2011).

Reif JC, Maurer HP, Korzun V, Ebmeyer E, Miedaner T, Würschum T: Mapping QTLs with main and epistatic effects underlying grain yield and heading time in soft winter wheat. Theoretical and Applied Genetics 123:283-292 (2011).

Yu L-X, Lorenz A, Rutkoski J, Singh RP, Bhavani S, Huerto-Espino J, Sorrels ME: Association mapping and gene-gene interaction for stem rsut resistance in CIMMYT spring wheat germplasm. Theoretical and Applied Genetics 123:1257-1268 (2011).

6 Erfolgte und geplante Veröffentlichungen der Ergebnisse

Die bisherige Veröffentlichung der Ergebnisse erfolgte in Form von Postern und Vorträgen, die im Folgenden aufgelistet werden:

Poster:

Kollers S, Rodemann B, Ganal MW, Plieske J, Argillier O, Sawkins M, Korzun V, Ebmeyer E, Röder MS: Whole genome association mapping of Fusarium head blight resistance in European winter wheat (Triticum aestivum L.), Institutstag IPK-Gatersleben, 2011.

Kollers S, Röder M, Rodemann B, Korzun V, Argillier O, Pietsch C, Ebmeyer E, Plieske J, Joaquim, Ganal MW: Associations with Fusarium head blight in a set of elite Western-European germplasm of hexaploid wheat (Triticum aestivum L.), 11th GABI Status Seminar, Potsdam, 2011.

Kollers S, Luerßen H, Czihal R, Ganal MW, Röder MS: SNP genotyping in hexaploid wheat (Triticum aestivum L.) with Illumina BeadXpress, Institutstag IPK-Gatersleben, 2010 and 10th Gatersleben Research Conference, Gatersleben, 2010.

Kollers S, Röder MS, Korzun V, Ebmeyer E, Argillier O, Joaquim P, Plieske J, Rodemann B, Ganal MW: Research into population structure in a collection of elite Western-European germplasm of hexaploid wheat (Triticum aestivum L.), 10th GABI Status Seminar, Potsdam 2010 and 10. GPZ Haupttagung, Freising-Weihenstephan, 2010.

Kollers S, Röder MS, Korzun V, Ebmeyer E, Argillier O, Joaquim P, Kulosa D, Rodemann B, Ganal MW: Gabi WHEAT – A whole-genome association study in hexaploid wheat (Triticum aestivum L.), 19th International Triticeae Mapping Initiative – 3rd COST Tritigen joint Metting, Clermont-Ferrand, Frankreich, 2009.

Kollers S, Röder MS, Ebmeyer E, Korzun V, Argillier O, Joaquim P, Kulosa D, Rodemann B, Ganal MW: Association mapping in hexaploid wheat – A genome wide and candidate gene approach, 9th GABI Status Seminar, Potsdam, 2009.

Vorträge auf wissenschaftlichen Tagungen (Vortragender fett gedruckt)

Röder M, Ebmeyer E, Argillier O, Rodemann B, Kollers S, Plieske J, Kulosa D, Sawkins M, Hinze M, Joaquim P, Pietsch C, Korzun V, **Ganal M**: Establishment of a platform for genome-wide association mapping in wheat. 11th GABI Status Seminar, Potsdam 2011

Kollers S, Röder MS, Korzun V, Ebmeyer E, Argillier O, Joaquim P, Plieske J, Rodemann B, Ganal MW: Research into population structure in a collection of elite Western-European germplasm of hexaploid wheat (Triticum aestivum L.). International Conference on Green Plant Breeding Technologies, Wien, 2010.

In Vorbereitung ist die Publikation der Ergebnisse für Fusarium head blight. Desweiteren sollen in einer weiteren Publikation die Ergebnisse von Septoria tritici blotch und Drechslera tritici repentis publiziert werden.

III. Zitierte Literatur

Barendse W, Reverter A, Bunch RJ, Harrison BE, Barris W, Thomas MB: A validated whole-genome association study of efficient food conversion in cattle. Genetics 2007, 176(3):1893-1905.

Beales J, Turner A, Griffiths S, Snape J, Laurie DA: A Pseudo-Response Regulator is misexpressed in the photoperiod insensitive Ppd-D1a mutant of wheat (Triticum aestivum L.). Theoretical and Applied Genetics 2007, 115:721-733.

Bland JM, Altman DG: Multiple significance tests: the Bonferroni method. British Medical Journal 1995, 310:170.

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

Breseghello F, Sorrells ME: Association mapping of kernel size and milling quality in wheat (Triticum aestivum L.) cultivars. Genetics 2006, 172(2):1165-1177.

Butow BJ, Gale KR, Ikea J, Juhasz A, Bedö Z, Tamas L, Gianibelli MC: Dissemination of the highly expressed Bx7 glutenin subunit (Glu-B1 al allele) in wheat as revealed by novel PCR markers and RP-HPLC. Theoretical and Applied Genetics 2004, 109(7):1525-1535.

Chao SM, Dubcovsky J, Dvorak J, Luo MC et al.: Population- and genome-specific patterns of linkage disequilibrium and SNP variation in spring and winter wheat (*Triticum aestivum* L.). BMC Genomics 11: 727 (2010).

Charlier C, Coppieters W, Rollin F, Desmecht D, Agerholm JS and et al.: Highly effective SNP-based association mapping and management of recessive defects in livestock. Nature Genetics 2008, 40(4):449-454.

Chartrain L, Brading PA, Brown JKM: Presence of the Stb6 gene for resistance to septoria tritici blotch (Mycosphaerella graminicola) in cultivars used in wheat-breeding programmes worldwide. Plant Pathology 2005, 54:134-143.

Chebotar SV, Korzun VN, Sivolap YM: Allele distribution at locus WMS261 marking the dwarfing gene Rht8 in common wheat cultivars of Southern Ukraine. Russian Journal of Genetics 2001, 37:894-898.

Crossa J, Burgueno J, Dreisigacker S, Vargas M, Herrera-Foessel SA, Dillemo M, Singh RP, Trethowan R, Warburton M, Franco J, Reynolds M, Crouch JH, Ortiz R: Association analysis of historical bread wheat germplasm using additive genetic covariance of relatives and population structure. Genetics 2007, 177(3):1889-1913.

Ellis MH, Spielmeyer W, Gale KR, Rebetzke GJ, Richards RA: “Perfect” markers for the Rht-B1b and Rht-D1b dwarfing genes in wheat. Theoretical and Applied Genetics 2002, 105:1038-1042.

Ganal MW, Altmann T, Röder MS: SNP identification in crop plants. Current Opinion in Plant Biology 2009, 12(2):211-217.

Gautier M-F, Aleman M-E, Guirao A, Marion D, Joudrier P: Triticum aestivum puroindolines, two basic cystine-rich seed proteins: cDNA sequence and developmental gene expression. Plant Molecular Biology 1994, 25:43-57.

Groenewald JZ, Marais AS, Marais GF: Amplified fragment length polymorphism-derived microsatellite sequence linked to the Pch1 and Ep-D1 loci in common wheat. Plant Breeding 2003, 122:83-85.

Hardy OJ, Vekemans X: SPAGeDi: a versatilecomputer program to analyse spatial genetic structure at the individual or population levels. Molecular Ecology Notes, 2:618-620.

Huang X-Q, Röder MS: Development of SNP assays for genotyping the puroindoline b gene for grain hardness in wheat using pyrosequencing. Journal of Agricultural and Food Chemistry 2005, 53:2070-2075.

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

Kosambi DD: The estimation of map distances from recombination values. Annual Eugenics 1994, 12:172-175.

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L: MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1987, 1:174-181.

Liu S, Chao S, Anderson JA: New DNA markers for high molecular weight glutenin subunits in wheat. Theoretical and Applied Genetics 2008, 118:177-183.

Loiselle BA, Sork VL, Nason J, Graham C: Spatial genetic structure of a tropical understory shrub, Psychotria officinalis (Rubiaceae). American Journal of Botany 1995, 82(11):1420-1425.

Perovic D, Förster J, Devaux P, Hariri D, Guilleroux M, Kanyuka K, Lyons R, Weyen J, Feuerhelm D, Kastirr U, Sourdille P, Röder M, Ordon F: Mapping and diagnostic marker edevelopment for Soil-borne cereal mosaic virus resistance in bread wheat. Molecular Breeding 2009, 23:641-653.

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

Rafalski JA: Association genetics in crop improvement. Current Opinion in Plant Biology 2010, 13(2):174-180.

Röder MS, Korzun V, Wendehake K, Plaschke J, Tixier M-H, Leroy P, Ganal MW: A microsatellite map of wheat. Genetics 1998, 149(4):2007-2023.

Röder MS, Wendehake K, Korzun V, Bredemeijer G, Laborie D, Bertrand L, Isaac P, Rendell S, Jackson J, Cooke RJ, Vosman B, Ganal MW: Construction and analysis of a microsatellite-based database of European wheat varieties. Theoretical and Applied Genetics 2002, 106(1):67-73.

Rohlf FJ 2008: NTSYSpc: Numerical Taxonomy System, ver 2.20 Exeter Publishing Ltd.: Setauket, NY.

Salonen KT, Uimari P, Aalto J-M, Pirskanen M, Kaikkonen J and et al.: Type 2 diabetes whole-genome association study in four populations: The DiaGen Consortium. The American Journal of Human Genetics 2007, 81:338-345.

Schwarz G, Felsenstein FG, Wenzel G: Development and validation of a PCR-based marker assay for negative selection of the HMW glutenin allele Glu-B1-1d (Bx-6) in wheat. Theoretical and Applied Genetics 2004, 109:1064-1069.

Su Z, Hao C, Wang L, Dong Y, Zhang X: Identification and development of a functional marker of TaGW2 associated with grain weight in bread wheat (Triticum aestivum L.). Theoretical and Applied Genetics 2011, 122:211-223.

Tommasini L, Schnurbusch T, Fossati D, Mascher F, Keller B: Association mapping of Stagnospora nodorum blotch resistance in modern European winter wheat varieties. Theoretical and Applied Genetics 2007, 115(5):697-708.

VSN International (2010). GenStat for Windows 13th Edition. VSN International, Hemel Hempstead, UK. Web page: GenStat.co.uk.

Wellcome Trust Case Control Consortium: Association scan of 14,500 nsSNPs in four common diseases identifies variants involved in autoimmunity. Nature Genetics 2007, 39(11):1329-1337.

Wright S: Evolution and genetics of populations. Vol. IV, 1987. The Univ. of Chicago Press.

Xia LQ, Ganal MW, Shewry PR, He ZH, Yang Y, Röder MS: Exploiting the diversity of Viviparous-1 gene associated with pre-harvest sprouting tolerance in European wheat varieties. Euphytica 2008, 159:411-417.

Yu J, Pressoir G, Briggs WH, Vroh Bi L, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES: A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature Genetics 2006, 38(2):203-208.

Zhang XK, Xiao YG, Zhang Y, Xia XC, Dubcovsky J, He ZH: Allelic variation at the vernalization genes Vrn-A1. Vrn-B1, Vrn-D1, and Vrn-B3 in Chinese wheat cultivars and their association with growth habit. Crop Science 2008, 48:458-470.