Theor Appl Genet (2007) 115:697–708
DOI 10.1007/s00122-007-0601-6

ORIGINAL PAPER

Association mapping of *Stagonospora nodorum* blotch resistance in modern European winter wheat varieties

L. Tommasini · T. Schnurbusch · D. Fossati · F. Mascher · B. Keller

Received: 24 November 2006 / Accepted: 25 June 2007 / Published online: 19 July 2007
© Springer-Verlag 2007

Abstract Association mapping in populations relevant for wheat breeding has a large potential for validating and fine-mapping QTLs identified in F2- or DH (double haploid)-derived populations. In this study, associations between markers in the region of *QSng.sfr-3BS*, a major QTL for resistance to *Stagonospora nodorum* glume blotch (SNG), and SNG resistance were investigated by linkage and association analyses. After increasing marker density in 240 $F_{5:7}$ recombinant inbred lines (RILs), *QSng.sfr-3BS* explained 43% of the genetic variance and peaked 0.6 cM proximal from the marker SUN2-3B. Association between SNG resistance and markers mapped in the region of *QSng.sfr-3BS* was investigated in a population of 44 modern European winter wheat varieties. Two genetically distinct subpopulations were identified within these lines. In agreement with linkage analyses, association mapping by a least squares general linear model (GLM) at marker loci in the region of *QSng.sfr-3BS* revealed the highest association with SNG resistance for SUN2-3B ($p < 0.05$). Association mapping can provide an effective mean of relating genotypes to complex quantitative phenotypes in hexaploid wheat. Linkage disequilibrium (r^2) in chromosome 3B extended less than 0.5 cM in 44 varieties, while it extended about 30 cM in 240 RILs, based on 91 SSR and STS marker-pair comparisons. This indicated that the association mapping population had a marker resolution potential at least 390-fold higher compared to the RIL population.

Communicated by M. Sillanpää.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-007-0601-6) contains supplementary material, which is available to authorized users.

L. Tommasini · T. Schnurbusch · B. Keller (✉)
Department of Plant Molecular Biology, Institute of Plant Biology, University of Zürich, Zürich 8008, Switzerland
e-mail: bkeller@botinst.uzh.ch

Present Address:
L. Tommasini
Department of Botany and Plant Sciences,
University of California Riverside,
2150 Batchelor Hall, Riverside, CA 92521-0124, USA

Present Address:
T. Schnurbusch
University of Adelaide, Waite Campus,
Australian Centre for Plant Functional Genomics,
PMB1, Glen Osmond, SA 5064, Australia

D. Fossati · F. Mascher
Agroscope Changins, Station fédérale de recherches agronomiques, Nyon1 1260, Switzerland

Introduction

QTLs studies in wheat and many other crop species are mainly carried out by linkage analysis in F2- or DH (double haploid)-derived mapping populations. These populations are highly useful to determine marker alleles linked to the traits for which the populations were developed. However, markers linked to QTLs in these populations might not be able to detect QTL effects in a set of plant lines relevant for breeding due to variable levels of linkage disequilibrium (LD) in mapping and breeding populations. In addition, only QTLs that differ between two parental lines can be detected in each mapping population. Association mapping in complex-pedigree populations relevant for plant breeding could prove useful for validating QTLs previously identified by linkage analyses, but also for detecting a higher

number of QTLs than in F2- and DH-derived populations. Association mapping was first successfully used for the identification of alleles at loci contributing to susceptibility to human diseases (reviewed in Goldstein 2003). Recently this approach has been used in different plant species to identify markers and genes associated with a variety of phenotypes. In hexaploid wheat, Breseghello and Sorrels (2006) found significant SSR marker associations with kernel size. In barley, different marker types were identified to be associated with different traits such as yield, yield stability, heading date, flowering time, water-stress tolerance and salt tolerance (Kraakman et al. 2004; Igartua et al. 1999; Ivandic et al. 2002, 2003; Pakniyat et al. 1997). In rice, correlations between RAPD markers and different quantitative traits were found by association (Virk et al. 1996), while in maize, associations between *dwarf8* polymorphisms and flowering time (Thornsberry et al. 2001) as well as associations with other characters were detected. Association studies have also been successful in detecting markers in LD with resistance genes for *Verticillum dahliae* (Simko et al. 2004) and late blight (Gebhardt et al. 2004) in potato, and for powdery mildew resistance in barley (Ivandic et al. 2003).

Population structure, deriving from diversifying selection and local adaptation to the environment, and familial relatedness from recent coancestry are likely to be present in modern wheat varieties deriving from complex pedigree-based breeding programs. New computational methods have recently allowed more accurate detection of marker-trait associations in complex-pedigree populations by accounting for population structure and relatedness (Pritchard et al. 2000; Yu et al. 2006). Besides determining the ability of a marker for detecting a QTL, LD extent also determines marker resolution in a given population. When using recombinant inbred line (RIL) populations composed of about 200 lines, genetic resolution is typically about 10–20 cM (Tanksley 1993). Fine-mapping in segregating populations derived from a biparental cross such as RILs or near isogenic lines (NILs) recombining in the region of the QTL of interest typically requires populations of about 2,000 individuals to reach a potential resolution of less than 1 cM (Blair et al. 2003; Chunwongse et al. 1997). Developing such populations is time-consuming and expensive. In a mapping population composed of a diverse set of varieties, marker resolution is expected to be higher than in a RIL population of the same size, as diversity is higher and LD extent is shorter, because of the many recombination events that have accumulated. Earlier LD studies in different crop species suggested that association mapping might improve significantly the fine-mapping of quantitative trait loci in crop plants, but large variability in LD extent was detected in different crops, mapping populations and genomic regions. LD extent was less than 1 cM in chromosome 2D and approximately 5 cM in the centromeric region of chromosome 5A in 95 elite lines of hexaploid wheat from Eastern USA based on SSR markers (Breseghello and Sorrels 2006). In maize, faster rates of decline of LD with distance, ranging from 400 to 7,000 bp, were measured within coding sequences (Remington et al. 2001; Tenaillon et al. 2001; Ching et al. 2002; Palaisa et al. 2003). In a few studies considerably larger LD extent was measured. In rice, LD extended from 45 to 100 kb (Garris et al. 2003) for different genes and populations.

In hexaploid wheat (*Triticum aestivum* L.), association studies are limited to three traits, kernel size and milling quality, in a collection of winter wheat from eastern USA (Breseghello and Sorrels 2006), and the quantity of a high-molecular-weight glutenin (Ravel et al. 2006). Therefore, association studies on other traits important for breeding such as disease resistance are required for investigating the potential of the association mapping approach in wheat. Besides, studies on LD in wheat were only carried out in parts of chromosomes 2D and 5A (Breseghello and Sorrels 2006), and therefore LD studies are necessary for assessing marker resolution in the B genome. Furthermore, LD extent was only measured in a collection of American wheat lines, leading to the need for examining LD in other breeding material.

Stagonospora nodorum blotch is a devastating disease of wheat world-wide (Duczek et al. 1999; Halama 2002) caused by *Phaeosphaeria nodorum* (E. Müller) Hedjaroude, anamorph *S. nodorum* (Berk., Castellani and Germano). Breeding cultivars with high genetic resistance is an effective and environmentally sound method for controlling this disease (Shaner 1999). Resistance to *S. nodorum* blotch is generally controlled by several independently inherited loci, conferring leaf (*S. nodorum* leaf blotch) and spike (*S. nodorum* glume blotch, SNG) resistance (Fried and Meister 1987; Botswick et al. 1993; Wicki et al. 1999). These genes are subject to environmental and pleiotropic effects, for instance from plant height and heading time (van Ginkel and Rajaram 1999; Broennimann et al. 1973; Scott et al. 1982; Wicki et al. 1999). A major QTL for resistance to SNG resistance, *QSng.sfr-3BS*, explaining 31.2% of the phenotypic variance in a population of 240 $F_{5:7}$ RILs deriving from a cross between "Arina" and 'Forno', two Swiss winter wheat varieties, mapped in the terminal part of chromosome arm 3BS, peaking at SSR marker locus GWM389 (Schnurbusch et al. 2003).

In this study, associations between markers in the region of *QSng.sfr-3BS*, and SNG resistance were investigated by linkage and association analyses. After improving marker density in 240 RILs, *QSng.sfr-3BS* explained 43% of the genetic variance and peaked 0.6 cM proximal from SUN2-3B. Association between SNG resistance and markers mapped in the region of *QSng.sfr-3BS* was investigated in a

collection of 44 current European winter wheat varieties. Association mapping disclosed the highest association with SNG resistance for SUN2-3B, in agreement with linkage analyses. Here, we show that *QSng.sfr-3BS* was conserved in a set of modern European breeding varieties and that association mapping provides an effective mean of relating genotypes to complex quantitative phenotypes in hexaploid wheat.

Materials and methods

Plant material

Two-hundred and forty single seed descent (SSD)-derived lines ($F_{5:7}$) from a cross between two adapted Swiss winter wheat (*T. aestivum* L.) cultivars "Arina" and "Forno" (Paillard et al. 2003) were evaluated in field trials for SNG reaction. Disease phenotyping for the SSD $F_{5:7}$ population was described in detail in Schnurbusch et al. (2003). The association mapping population consisted of 44 modern European winter wheat cultivars (Table 1) from the Swiss national cereal gene bank at Agroscope Changins. The cultivars were released between 1980 and 2004 (with the exception of Zenith that was released in 1969), and represent a broad part of the European germplasm including genotypes with extreme resistance and susceptibility towards Stagonospora glume blotch. Field trials for Stagonospora glume blotch resistance were part of the Swiss winter wheat breeding programme. Inoculum for artificial infections consisted of a mix of 10–15 *S. nodorum* strains, isolated in different regions in Switzerland. The composition of the strain mix varied between years, combining old isolates and new isolates in order to maximize virulence. Spores were produced on sterilized wheat kernels as described in Fried and Meister (1987). Field trials consisted of at least three replicates per year over a period of 2–16 years in one environment under artificial infection. Disease severity notes were resumed into an area under disease progress curve (AUDPC) value according to the procedure described by Shaner and Finney (1997). In order to take into account the role of varietal earliness on disease development (Brönnimann, 1968), AUDPC values were corrected according to the earliness class (early, mid-season, late) of each variety (M. Winzeler and P.M. Fried, personal communication). In order to be able to compare disease resistance between years, corrected AUDPC were ranked using an index, by dividing the AUDPC value of each genotypes with the average of all genotypes tested in that year and multiplied by 100. In order to test for the systematic change in resistance over years, we examined the trend of resistance for four long term-grown varieties in our set, Arina, Galaxy, Titlis and Zenith, which were present in resistance nurseries for 5 years (Zenith), 10 years (Titlis), and 16 years (Arina and Galaxy). Based on regression analysis of resistance or susceptibility over years, no evidence for a systematic trend in resistance was detected at the 1% level. No evidence for *S. nodorum* genetic adaptation or co-evolution with the host in the field has been described in previous studies (Keller et al. 1997). The varieties used in this study were classified as resistant when indices were between 38.0 and 89.9 and as susceptible with indices ranging from 121.0 to 168.0. These indices were used in linkage and association analyses for SNG resistance.

Genotype analysis

DNA extraction, purification and quantification were performed using a CTAB method published in Stein et al. (2001). DNA for each variety was extracted from bulks of leaves from 10 plants at the first flag leaf stage for the association mapping population. SSR and STS markers located in the telomeric part of chromosome 3BS were identified from public databases (http://wheat.pw.usda.gov/GG2/inde.shtml). These are GWM1037, BARC68, BARC101, BARC102, GWM1034, BARC75, BARC133, SUN2-3B and STS3B1. Nullitetrasomic lines N3BT3A, N3AT3D and N3DT3B were used to confirm chromosomal localisation of the markers. Markers polymorphic between "Arina" and "Forno" were mapped in the "Arina"–"Forno" RIL population (Paillard et al. 2003). Nineteen SSR markers from 19 different chromosomes previously mapped in the "Arina"–"Forno" genetic map were chosen to determine population structure and the level of significant LD among unlinked markers likely due to population structure. In addition, 13 SSR and STS markers previously mapped in 3BS were amplified to evaluate the extent of LD. PCR reactions and fragment size analyses were performed as described by Paillard et al. (2003). SSR bands were scored with different letters, based on their relative fragment size, characterising each allele uniquely. Heterozygous loci were scored as missing data.

Genetic mapping and linkage analyses

The genetic linkage map and computational methods for linkage analysis used in this study were described by Paillard et al. (2003). Composite interval mapping (CIM) was carried out as described in Schnurbusch et al. (2003).

Allele diversity and population structure

Gene diversity and heterozygosity were calculated at 32 SSR and STS marker loci in 44 varieties using PowerMarker (http://www.powermarker.net). Gene diversity (H) at a locus is calculated from allele frequencies and is defined as

Table 1 Fourty-four European winter wheat varieties forming the association mapping population

	Variety name	Origin RAC	Subpopulation ownership[a]	Index SNG resistance	SNG resistance	Alleles at Sun2-3B[b]
1	IENA	France	1	38.0	Very good	A
2	MANHATTAN	Germany	2	42.2	Very good	A
3	ROMEO	Sweden	2	56.3	Very good	B
4	CHAMPTAL	France	2	58.5	Very good	B
5	ARINA	Switzerland	1	63.0	Very good	A
6	SOISSONS	France	1	64.7	Very good	A
7	ARBOLA	Switzerland	2	68.5	Very good	B
8	CONTRA	Germany	2	73.0	Very good	B
9	AROLLA	Switzerland	1	73.0	Very good	A
10	CAPO	Hungary	1	73.5	Very good	A
11	GARMIL	Switzerland	1	74.3	Very good	A
12	SEGOR	Switzerland	1	75.0	Very good	A
13	DRIFTER	Germany	2	76.8	Good	B
14	CHARGER	Great Britain	2	77.0	Good	A
15	EIGER	Switzerland	1	77.3	Good	A
16	BAROUDEUR	France	1	77.3	Good	B
17	BUSSARD	Germany	1	80.0	Good	B
18	FRANCO	Germany	1	81.2	Good	B
19	EPHOROS	Germany	1	82.6	Good	A
20	BISCAY	Germany	2	83.2	Good	B
21	VERGAS	Germany	2	84.8	Good	B
22	LUDWIG	Austria	2	85.7	Good	B
23	TRANSIT	Hungary	2	86.5	Good	B
24	OBELISK	Yugoslavia	2	86.5	Good	B
25	MUVERAN	Switzerland	1	86.5	Good	B
26	SORBAS	Germany	1	87.8	Good	B
27	RUNAL	Switzerland	1	89.9	Medium to good	B
28	TITLIS	Switzerland	1	89.9	Medium to good	A
29	RURIK	France	2	121.0	Susceptible	B
30	SKATER	Great Britain	2	122.0	Susceptible	B
31	GENIAL	France	1	122.3	Susceptible	B
32	ZENITH	Switzerland	2	124.6	Susceptible	B
33	BOSTON	France	1	126.5	Susceptible	B
34	BATIS	Germany	2	127.5	Susceptible	B
35	ORPIC	France	1	129.5	Susceptible	A
36	BOVAL	Switzerland	1	130.1	Susceptible	B
37	CAPNORD	France	1	130.6	Susceptible	B
38	TERZA	Switzerland	1	133.3	Susceptible	A
39	CERTO	Germany	1	133.6	Susceptible	B
40	FORNO	Switzerland	1	143.5	Very susceptible	B
41	SARDONA	Switzerland	1	152.0	Very susceptible	A
42	PITOMA	Croatia	1	154.0	Very susceptible	B
43	TAPIDOR	France	1	154.6	Very susceptible	B
44	AMELIO	France	1	168.0	Very susceptible	B

SNG Stagonospora nodorum glume blotch

[a] As defined by the STRUCTURE software

[b] sun2-3B was the most significantly associated marker ($p < 0.001$) with SNG resistance

$$H = 1 - \sum_{u} p^2 l$$

where p_{lu} is the frequency of the uth allele for the lth locus, squared and summed across all alleles in the locus. For the analysis of population structure, we performed a model-based (Bayesian) cluster analysis based on 19 markers from 19 different chromosomes (Table 2). This analysis was implemented in the software STRUCTURE (Pritchard et al. 2000), which identifies subgroups of accessions with distinct allele frequencies within the germplasm. STRUCTURE computes a Q matrix defined as an $n \times p$ population structure incidence matrix where n is the number of individuals assayed and p is the number of populations defined; Q is inferred from Pritchard's STRUCTURE (Pritchard et al. 2000) estimates with p populations (p is Pritchard's K). The model-based cluster analysis was used to test the hypotheses of one to four subpopulations ($K = 1$ to $K = 4$). No admixture and correlated allele frequencies in different subpopulations were assumed in our calculations. 100,000 iterations and a burn-in period of 100,000 were carried out for each run. Five independent STRUCTURE runs were made for $K = 1$ to $K = 4$ and an average likelihood value across five runs was calculated for each K. Significance of the difference between likelihood of the data for $K = 1$ to $K = 4$ was evaluated by t test. The highest likelihood of the data was observed for $K = 2$ clusters of plants ($p < 0.001$). Therefore, Q matrix as average of five runs for $K = 2$ was calculated. The Informativeness (I_n) of 19 markers for assigning individuals to the two subpopulations identified by STRUCTURE was estimated as described by Rosenberg et al. (2003). The optimal two-subpopulations model and correspondent averaged Q matrix was used in subsequent association analyses in TASSEL (http://www.maizegenetics.net/ bioinformatics/inde.htm). Correlation between the country of origin of the 44 varieties and their population structure as established with STRUCTURE was tested by one-way ANOVA using the JMP for Windows version 5.1 (SAS Institute, Cary, NC).

Table 2 SSR and STS markers used for improving the map of the Arina × Forno RIL population and for association mapping

Marker	Chromosome position	Allele number	Gene diversity	Informativeness (I_n)
gwm389	3BS	7	0.79	
gwm1034a	3BS	11	0.81	
barc75	3BS	3	0.51	
sun2-3B	3BS	2	0.45	
barc133	3BS	5	0.61	
cfd79c	3BS	2	0.48	
gwm1037	3BS	3	0.62	
gwm77	3BS	5	0.48	
cfd4a	3BS	13	0.90	
cfa2134b	3BS	9	0.78	
gwm131b	3BS	7	0.83	
gwm383b	3BS	2	0.24	
gwm326	3BS	6	0.64	
gwm357	1A	3	0.57	0.00
gwm11	1B	10	0.75	0.16
wmc216a	1D	5	0.32	0.06
gwm382	2A	9	0.86	0.14
gwm257	2B	4	0.54	0.06
gwm261	2D	4	0.62	0.03
wmc264	3A	5	0.72	0.03
gwm340	3B	7	0.72	0.14
gwm645	3D	8	0.78	0.19
gwm148	4A	5	0.62	0.09
gwm538	4B	3	0.45	0.03
cfa2173	4D	6	0.74	0.09
cfa2141	5A	7	0.79	0.10
gwm234	5B	4	0.68	0.02
gwm174	5D	9	0.85	0.25
gwm169	6A	10	0.84	0.12
gwm219	6B	8	0.76	0.19
gwm469	6D	6	0.76	0.05
gwm44	7D	7	0.78	0.14
Mean		6.09	0.62	0.10

Marker–traits associations

To calculate marker association with SNG resistance, we tested our data set for the appropriateness of different association analysis models, as described in Supplementary Material (S1–S4). The least square solution to the fixed effects general linear model (GLM-Q, Searle 1987) showed the best fit for our data set and was used for association analyses, as implemented in the TASSEL software, taking into consideration the Q matrix produced by STRUCTURE for two subpopulations. To select appropriate significance thresholds for association analyses, probability values of association between single alleles from nineteen unlinked neutral SSR marker loci and SNB resistance based on the GLM-Q association model were permuted 1,000 times (Churchill and Doerge 1994). A polymorphic site was deemed to have a significant association if the p-value was below the 5% empirically derived value. Only alleles present at least in nine samples were considered in the analyses. Single alleles were tested for association.

Candidate loci located in the region of *QSng.sfr-3BS* (SSRs BARC75 and BARC133, and STS SUN2-3B) were

evaluated for association with SNB resistance using the same method described above.

Linkage disequilibrium

TASSEL was used to measure the extent of LD as squared allele frequency correlations estimates (r^2, Weir 1996) and to measure significance of r^2 for each pair of loci. For multiple alleles, a weighted average of r^2 between each locus pair was calculated (Farnir 2000). Only alleles with frequencies equal or greater than 0.05 were considered for LD calculations (Thornsberry et al. 2001). Significance of LD for SSR pairs was determined by 100,000 permutations for each pair (Weir 1996). The number of marker pairs with LD probability values less than threshold values of 0.01 and 0.001 were counted. LD extent was estimated separately for unlinked loci and for loci on the same chromosome. Thirteen SSR and one STS loci were used for calculating LD extent in chromosome 3BS. The genetic distance (cM) between marker pairs used for calculating LD decay was based on the "Arina"–"Forno" RIL mapping population. Decay of LD with distance between markers was evaluated by nonlinear regression (PROC NLIN) using JMP. To measure the extent of LD across unlinked markers, data from 19 SSR unlinked polymorphic sites were used, and LD calculations were carried out similarly as described above. LD calculations were carried out within all 44 lines, within each subpopulation identified using STRUCTURE, and in the 240 RILs used for linkage analyses.

Results

Linkage analyses for Stagonospora resistance on chromosome 3BS

To improve the genetic map around *QSng.sfr-3BS*, seven markers polymorphic between "Arina" and "Forno" and located in chromosome 3BS (GWM1037, GWM1034, BARC75, SUN2-3B, BARC133 and BARC68) were mapped using the RIL population derived from "Arina" and "Forno". The improved map of chromosome 3B is composed of 24 markers, covering a total genetic distance of 183.1 cM, with an average density of one marker per 7.6 cM, as shown in Fig. 1. Composite interval mapping for Stagonospora glume blotch index was performed with a LOD threshold of 4.5, as described in Schnurbusch et al. (2003). One of the newly mapped markers, SUN2-3B, was located 0.6 cM from the *QSng.sfr-3BS* peak, increasing the percentage of explained phenotypic variance of *QSng.sfr-3BS* from 31.2 to 43%. *QSng.sfr-3BS* was defined by SSR GWM1034 and BARC133 located 2.1 and 3.2 cM from the *QSng.sfr-3BS* peak, respectively (Fig. 1).

A winter wheat population for association mapping

A population of 44 modern winter wheat varieties representing a large part of the current European germplasm was assembled for association analyses, with the goal to compare the linkage mapping approach to association mapping. Association mapping requires population structure to be taken into account for avoiding false positive associations. Therefore, the 44 winter wheat varieties were genotyped at 19 SSR loci (Table 2), mapped on 19 different chromosomes for assessing genetic diversity and population structure. In total, 210 alleles were detected in the association mapping population at 32 loci. In average, 6.0 alleles per locus were detected, ranging from 2 (*cfd79c* and *sun2-3B*) to 13 (*cfd4a*). The diversity index ranged from 0.24 (*gwm383b*) to 0.90 (*cfd4a*), with a mean of 0.65 (Table 2). Overall, 6.6% of the alleles were null or missing, and 2.2% of the loci were heterozygous and were considered as nulls in association and structure analyses. Population structure was assessed using a model-based clustering method to test the hypotheses of one to four subpopulations ($K = 1$ to $K = 4$). The average logarithm of the probability of data likelihood (Ln P(D)) for $K = 1$, $K = 2$, $K = 3$ and $K = 4$ subpopulations within 44 lines was Ln P(D) = –1291.6, Ln P(D) = –1248.0, Ln P(D) = –1286.1 and Ln P(D) = –1377.4, respectively. Therefore, the highest likelihood of the data was observed when the number of subpopulations was set to two (Ln $P(D)_{K=2}$ < Ln $P(D)_{K=1}$ at the 0.001 p level). For $K = 2$ the association mapping population was divided in two subgroups composed of 28 and 16 lines, considering 50% ownership of a variety to a subpopulation as threshold for population assignment (Table 1). Even considering a more conservative threshold of 60% ownership of a variety to a subpopulation as for population assignment, most of the individuals (38) could be confidently assigned to a subpopulation. Figure 2 shows the percentage of ownership of each variety to the first and second subpopulation, for $K = 2$ based on the average Q matrix across five runs obtained from STRUCTURE. A sufficient number of markers (at least 9) showed sufficiently high I_n values ($I_n > 0.1$) for confident assignment of 44 individuals to the two subpopulations identified by STRUCTURE (Table 2).

The optimal two-subpopulation model and Q matrix was used in subsequent association analyses. It was possible to identify origin-related causes for the subpopulations observed. In fact, ANOVA analyses indicated that the presence of the two subpopulations identified using STRUCTURE could partially be explained by the country of origin of the 44 varieties screened ($F = 3.32$, $p < 0.01$). Varieties from five countries were included within the

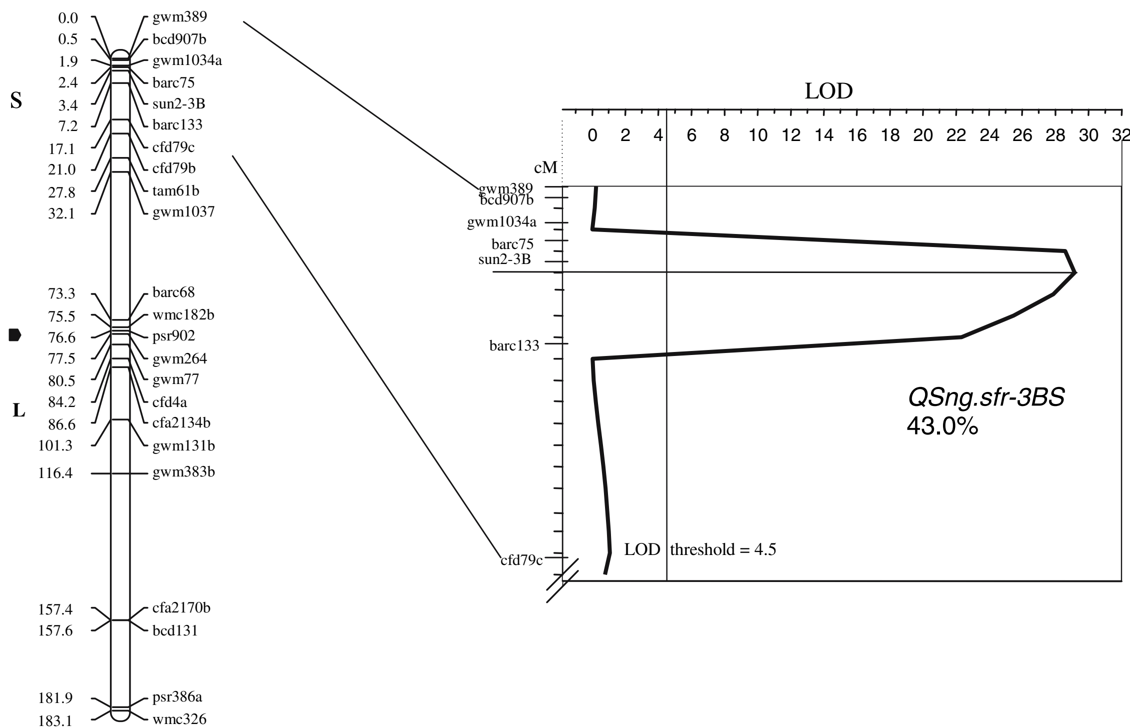


Fig. 1 Chromosome 3B genetic map of the Arina–Forno RIL population (*left*). Genetic distances are indicated in cM on the *left* of the map and the corresponding marker names are indicated on the *right*. On the *right* graph, composite interval analysis of *QSng.sfr-3BS* for the AUDPC (*Stagonospora nodorum*) with improved resolution. Each *dash* on the cM scale represents 1 cM

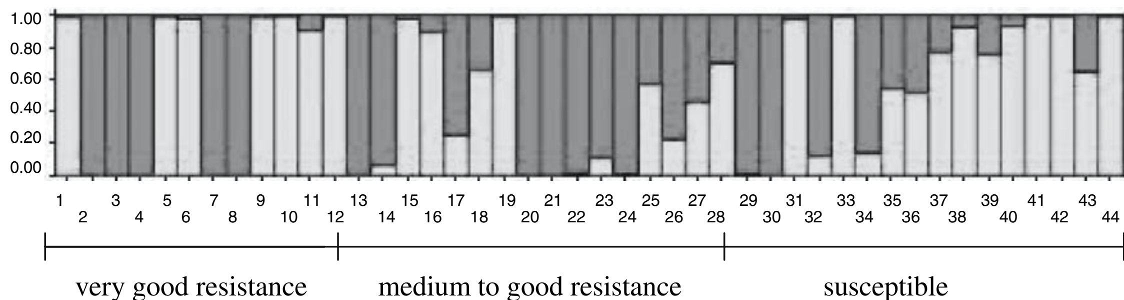


Fig. 2 Proportion of ownership to subpopulation 1 (*light grey*) and subpopulation 2 (*dark grey*) for each of the 44 varieties examined (lanes *1–44*) based on 19 SSR markers used to build the Q matrix from model-based clustering. Lanes *1–44* 44 varieties ordered based on decreasing SNG resistance level

subpopulation composed of 28 lines, most Swiss (12) and French (9) varieties were included in this subgroup. Varieties from eight different countries were included within the subpopulation composed of 16 varieties, mainly represented by varieties from Germany (Table 1).

Marker-trait associations

Association of candidate loci located in the region of *QSng.sfr-3BS* (SSRs BARC75 and BARC133, and STS SUN2-3B) with SNB resistance showed that only one candidate marker, SUN2-3B, was significantly associated at the 5% probability level ($p = 0.0007$). SUN2-3B was also the closest marker to *QSng.sfr-3BS* (0.6 cM distal from *QSng.sfr-3BS*) and therefore the most predictive marker for *QSng.sfr-3BS* both in the "Arina"–"Forno" RIL population, as well as in the association mapping population. SUN2-3B has only two alleles. Twelve resistant varieties (index values ranging from 38.0 to 89.9) carried "Arina"s *sun2-3B* allele (Table 1). These varieties had a mean resistance index of 69.2. Sixteen resistant varieties did not carry "Arina"s *sun2-3B* resistance allele. These varieties had a mean resistance index of 78.9, ranging from 56.3 to 89.9. Arina's *sun2-3B* allele was detected also in three susceptible varieties that had a mean resistance index of 138.2 ranging from 129.5 to 152.0.

LD extent

To measure marker resolution of association mapping and RIL populations, 13 SSR and one STS loci providing 91 marker-pair comparisons, were used for calculating LD extent in chromosome 3B. The markers used in 3B covered the whole length of the chromosome. The shortest distance at which LD was measured between any two markers was 0.5 cM. In chromosome 3B, LD (r^2) ranged from 0.001 to 0.139, with a mean of 0.034 within 44 lines in the association mapping population, while in the RIL population it ranged from 0.001 to 0.680, with a mean of 0.146 (Figs. 3, 4). Lack of LD ($r^2 < 0.1$) was detected at distances as short as 0.5 cM in 44 varieties and in 96% of marker pair comparisons. r^2 was greater than 0.1 for four marker pairs located at distances ranging from 2.4 to 83.2 cM, possibly due to convergent selection pressures between distant markers (Fig. 3). In contrast, considerably larger LD extent was observed in the "Arina"–"Forno" RIL population, where LD only decayed ($r^2 < 0.1$) at distances greater than 30 cM (Fig 4). LD extent within chromosome 3B was strongly correlated to genetic distance ($F = 86.7$, $p < 0.001$) within the "Arina"–"Forno" RIL population, while no correlation could be observed within the whole association mapping population ($F = 3.57$, $p = 0.062$), due to the decay of LD at distances shorter than distances between the marker pairs examined and to the influence of breeding and evolutionary processes on the association mapping population. In 44 varieties, LD (r^2) was significant in 11.5 % of marker pair comparisons at the 1% level, while in the RIL population LD was significant in 49% of the marker pair comparisons (Figs. 3, 4). A tendency to higher LD in the region of *QSng.sfr-3BS*, where markers were more densely spaced, could be observed on the LD matrix in the association mapping population (Fig. 4). At the 0.01% level, only three marker pairs were significantly associated in 44 varieties compared to 38 marker pairs in 240 RILs (Fig. 4).

To compare the level of LD of linked and unlinked markers and estimate the number of positive associations not due to close physical distance, LD across 19 unlinked SSR polymorphic sites corresponding to 171 marker pair comparisons, was measured in RIL and association mapping populations (Table 3). Within 44 lines, LD was significant for 2.9% of the unlinked marker pair comparisons at the 0.01 p level, whilst within the subpopulations of 16 and 26 lines LD was significant for 2.9 and 2.3% of the marker pairs, respectively. This was very similar to the RIL population, where LD was significant for 3.5% of the marker pair comparisons. Thus, the percentage of unlinked marker pairs in significant LD was lower than in 3B.

Discussion

Genetic diversity and structure

Even though our sample was restricted to 44 elite European winter wheat lines, the level of detected diversity was relatively high, with an average of 6.0 alleles per locus. The SSR allele diversity found in our population for association mapping represents approximately half the total diversity present in most of the European wheat elite germplasm (10.5 SSR alleles per locus detected in 502 European hexaploid wheat varieties, Röder et al. 2002), or a third of the diversity found in the wheat germplasm repository (18.1 alleles per locus in 998 accessions of hexaploid wheat from the IPK gene bank in Germany, Huang et al. 2002). A similar number of alleles per locus as in our study was found in 60 hexaploid wheat cultivars from Eastern Europe (4.8 alleles, Stachel et al. 2000), in a collection of 95 cultivars of soft winter wheat from Eastern USA (4.8 alleles, Breseghello and Sorrels 2006), and in 134 durum wheat cultivars of diverse geographical origins, representing the main gene pools (5.5 alleles, Maccaferri et al. 2005).

Using the model-based clustering method, for $K = 2$ the highest data likelihood could be reached. This provided evidence for the presence of two subpopulations. The country of origin of the 44 varieties examined was in agreement with this subdivision, with almost all Swiss and French varieties belonging to the larger subpopulation,

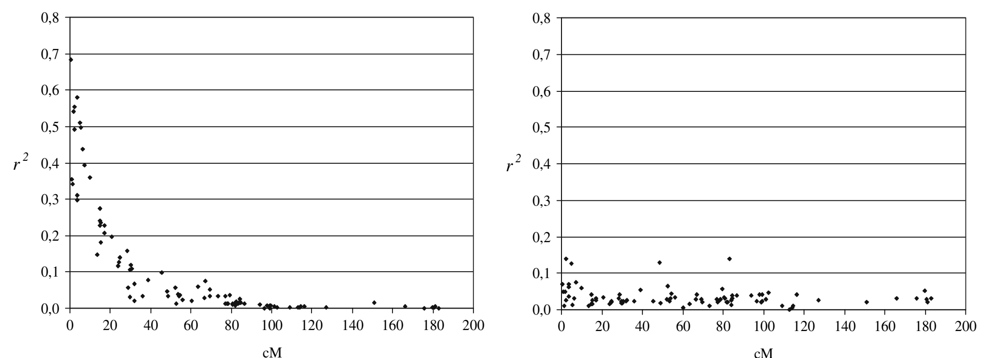


Fig. 3 Linkage disequilibrium (r^2) as a function of genetic distance between SSR and STS marker-pairs in chromosome 3B in 240 recombinant inbred lines (*left*) and 44 varieties (*right*). The genetic distance between markers was determined using the Arina–Forno RIL population

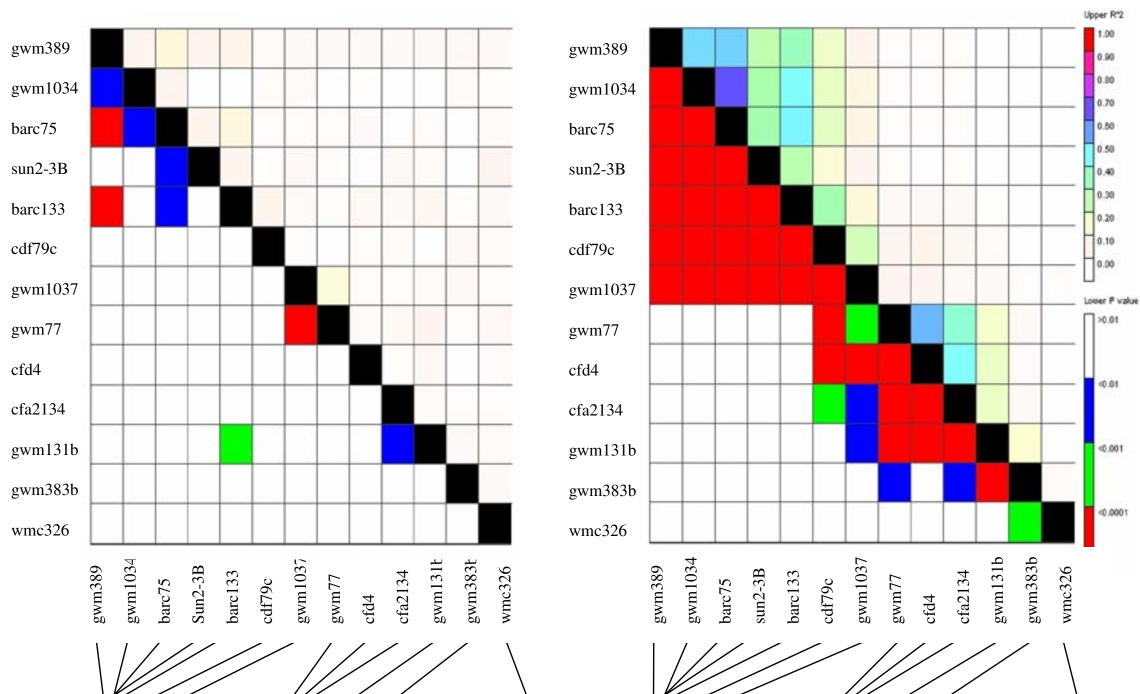


Fig. 4 Linkage disequilibrium measure (r^2, *above diagonal line*) and probability value (*P, below diagonal line*) for 14 markers in chromosome 3BS in 44 varieties (*left*) and 240 recombinant inbred lines (*right*). The picture represents all pair-wise comparisons of polymorphic sites. The genetic map location of STS and SSR markers are shown at the *bottom* of the graph

while German varieties were equally split between the two subgroups. Similarly, separation between French and German varieties based on RFLP markers was observed in a collection of European winter wheat lines (Siedler et al. 1994). A set as small as 20 microsatellites was previously shown to be sufficient for estimating genetic diversity in 55 elite wheat genotypes (Prasad et al. 2000).

Marker–trait associations

Association analyses of candidate markers detected the highest marker association with SNG resistance for *sun2-3B* ($p < 0.05$), the marker which was closest to *QSng.sfr-3BS*. These results indicated that marker-trait associations

Table 3 Percentage SSR loci pairs in LD ($p < 0.01$) in 44 varieties

Population	SSR pairs	Number of lines	Percent of marker pairs in LD
44 varieties	Chromosome 3BS	44	13
Subpopulation 1		28	8.8
Subpopulation 2		16	2.2
44 varieties	Unlinked	44	2.9
Subpopulation 1		28	2.3
Subpopulation 2		16	2.9

at *QSng.sfr-3BS* found in the “Arina”–“Forno” RIL mapping populations are maintained in the current European winter wheat breeding germplasm pool. Thus, both linkage and LD mapping approaches were successful in detecting correct associations. The mean SNG resistance index over years was 83.6 for lines carrying the A allele (deriving from the resistant parent “Arina”) and 104.2 for lines carrying the B allele (deriving from the susceptible parent ‘Forno’). As a consequence, *sun2-3B* is a useful marker for identifying SNG resistance alleles in wheat breeding material, assisting in the selection of more resistant progeny and parents for new crosses, and represents the first diagnostic markers for resistance to SNG in a large germplasm pool.

Our study identified a number of varieties carrying “Arina”s *sun2-3B* allele and showing good SNG resistance level. However, some resistant varieties did not carry the *sun2-3B* resistance allele of Arina. These might have (a) resistance gene(s) different than *QSng.sfr-3BS*, which was/were not in linkage disequilibrium to SUN2-3B and *QSng.sfr-3BS.* Alternatively, these varieties carried *QSng.sfr-3BS* but not the *sun2-3B* “Arina” resistance allele due to recombination events between or at SUN2-3B and *QSng.sfr-3BS* sites. “Arina”s *sun2-3B* allele was also detected in three susceptible varieties. In these varieties, *QSng.sfr-3BS* was probably present but might not be able to confer a sufficient level of resistance (relatively to the

other varieties screened) due to allelic interactions in these genetic backgrounds. Alternatively, *QSng.sfr-3BS* was present but not detected by SUN2-3B due to recombination or mutation events between or at SUN2-3B and *QSng.sfr-3BS*.

However, knowledge about the presence/absence of *QSng.sfr-3BS* in this set of cultivars will assist in the selection of parental lines in order to combine different SNG resistance sources (e.g. "Iena", "Arina" and "Romeo" and "Champtal") enhancing the overall level of SNG resistance within the European winter wheat germplasm.

Our association analyses were based on single-gene models. Simulation data showed that using multiple-gene models for analyzing complex traits influenced by a set of multiple loci with effects of different magnitudes can be beneficial (Kilpikari and Sillanpää 2003). A main advantage of multilocus association consists in the avoidance of multiple testing problems. However SNG resistance has been shown to be oligogenic (Schnurbusch et al. 2003) and therefore our analysis would benefit less from the multiple-gene model approach. Besides, it is not clear how robust are the multipoint analyses based on genetically mapped SSR markers in complex genomes as wheat's, as marker map distances vary from cross to cross. Also, in our map the *QSng.sfr-3BS* region is well covered and therefore little extra detection power will be obtained from "interval" mapping, especially considering that the population size is small. Certainly an interval analysis would provide a peak indicating the most likely location, but the confidence interval of the peak would still be large with a small population and thus no better than what we likely have obtained with single-marker analysis.

The "Arina"–"Forno" RIL population was highly useful for the detection of the major SNG resistance QTL, *QSng.sfr-3BS*. However, the analysis of a RIL population can not reveal if a QTL is conserved in current breeding material, a considerable limitation when testing markers for marker-assisted selection. Our data suggest that these limitations can be overcome by using linkage disequilibrium approaches to validate QTLs first identified by biparental linkage analysis. Increasing the size of the association mapping population and marker density in the region of interest could be useful for detecting association of rare alleles or of alleles associated with small phenotypic effects. Exploiting the complementary strength and weaknesses of both approaches will allow efficient development of markers for breeding.

LD extent and pattern

The limited LD between unlinked marker pairs within homogenous subpopulations and within 44 varieties (2.3–2.9% of the marker pair comparisons), provided evidence for the suitability of the assembled population for association mapping. The similarity of LD values within subpopulations and within the whole population of 44 varieties was in agreement with the limited amount of structure detected in 44 varieties. In 44 varieties, LD (r^2) declined below 0.1 within 0.5 cM on chromosome 3B. These results showed that properly selected populations of diverse lines can give high resolution for association mapping in wheat. Similar to our findings, LD extent in chromosome 2D was less than 1 cM in 95 elite hexaploid wheat lines from Eastern USA, based on SSR loci pairwise comparisons (Breseghello and Sorrels 2006). The slightly higher LD extent detected by Breseghello and Sorrels (2006) compared to our study, could be due to a higher number of alleles per locus in our study (6.0 vs 4.8) and to the presence of higher level of structure within the sample examined. Similar LD extent as in our study was also found in *Arabidopsis* (Nordborg et al. 2002), where LD significantly decayed at about 1 cM (250 kb) in a global set of 76 accessions and at 30 kb when the sample was increased to 196 accessions (Hagenbald et al. 2004). Faster LD decay varying from 45–100 kb within different subpopulations and genomic regions was observed in candidate gene regions in rice (Garris et al. 2003). Considerably shorter LD extent was found in several maize coding regions, where LD significantly declined starting from 100–2,000 bp in different studies (Tenaillon et al. 2001; Remington et al 2001; Ching et al. 2002; Palaisa et al. 2003; Wilson et al. 2004). To estimate if LD decays at distances shorter than 0.5 cM and examine LD at the gene size scale in wheat, more densely spaced markers should be generated from several hundred kb of sequence data in the region of *QSng.sfr-3BS* in the population studied. This would allow estimating if LD mapping in wheat could provide sufficient resolution to identify functional nucleotide polymorphisms like in maize (Thornsberry et al. 2001).

While in the 44 varieties LD significantly declined within 0.5 cM, in 240 RIL lines LD declined at about 30 cM. Considering that (1) the size of the association mapping population is 6.5 times smaller than the size of the RIL, (2) the expected value of r^2 is a function of the parameter $\rho = 4N_e c$, where c is the recombination rate between the two markers and N_e is the effective population size, and (3) for large ρ, $E(r^2) \approx 1/\rho$ (reviewed by Hudson 2001), the resolution power in the association mapping population was at least 390 times higher than in the RIL population. This suggests that our association mapping population could contribute to fine-mapping *QSng.sfr-3BS* and other genes located in 3B within genomic areas previously identified by linkage in RIL lines, but not for genome-wide marker scans which would require screening an unmanageable number of markers.

Conclusion

This study allowed to improve marker resolution of a major QTL for SNG resistance and showed that association mapping using SSR or STS markers and genetically diverse germplasm provides an effective mean of relating genotypes to complex quantitative phenotypes in wheat. By building on previous linkage mapping analyses, the association approach confirmed that *QSng.sfr-3BS* is conserved in a set of modern European breeding varieties and allowed validating this QTL. These markers will allow the assessment of genetic potential of specific genotypes prior to phenotypic evaluation and allow breeders to incorporate desirable alleles efficiently into wheat germplasm.

Acknowledgments Special thanks to Dr. Flavio Breseghello for helpful comments on the manuscript, and to Prof. Clay Sneller, Dr. Peter Morrell, Dr. Gael Pressoir, Dr. Edward Buckler, Dr. Jianming Yu for suggestions on statistical analyses. This work was supported by the Swiss National Science Foundation for Scientific Research (Grant No. 105620) and by the University of Zürich.

References

Blair MA, Garris A, Iyer A, Chapman B, Kresovich S et al (2003) High resolution genetic mapping and candidate gene identification at the *xa5* locus for bacterial blight resistance in rice (*Oryza sativa* L.). Theor Appl Genet 107:62–73

Botswick ED, Ohm HM, Shaner G (1993) Inheritance of Septoria glume blotch resistance in wheat. Crop Sci 33:439–443

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

Brönnimann A (1968) Zur Kenntnis von Septoria nodorum Berk., dem Erreger der Spelzenbräune und einer Blattdürre des Weizens. J Phytopathol 61:101–146

Brönnimann A, Fossati A, Hani F (1973) Spreading of *Septoria nodorum* (Berk) and damage to artificially induced mutants of winter wheat cultivar Zenith (*Triticum aestivum*). Plant Breed 70:230–245

Ching A, Caldwell KS, Jung M, Dolan M, Smith OS et al (2002) SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genet 3:19

Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. Genetics 138:963–971

Chunwongse J, Doganlar S, Crossman C, Jiang J, Tanksley SD (1997) High-resolution genetic map of the *Lv* resistance locus in tomato. Theor Appl Genet 95:220–223

Duczek LJ, Sutherland KA, Reed SL, Bailey KL, Lanford GP (1999) Survival of leaf spot pathogens on crop residues of wheat and barley in Saskatchewan. Can J Plant Pathol 21:165–173

Farnir F, Coppieters W, Arranz J-J, Berzi P, Cambisano N et al (2000) Extensive genome-wide linkage disequilibrium in cattle. Genome Res 10:220–227

Fried PM, Meister E (1987) Inheritance of leaf and head resistance of winter wheat to *Septoria nodorum* in diallel cross. Phytopathology 77:1371–1375

Garris AJ, McCouch SR, Kresovich S (2003) Population structure and its effects on haplotype diversity and linkage disequilibrium surrounding the *a5* locus of rice *Oryza sativa* L. Genetics 165:759–769

Gebhardt C, Ballvora A, Walkemeier B, Oberhagemann P, Shueler K (2004) Assessing genetic potential in germplasm collections of crop plants by marker-trait association: a case study for potatoes with quantitative variation of resistance to late blight and maturity type. Mol Breed 13:93–102

Goldstein DB, Tate SK, Sisodiya SM (2003) Pharmacogenetics goes genomic. Nat Rev Genet 4:937–947

Hagenblad J, Tang C, Molitor J, Werner J, Zhao K et al (2004) Haplotype structure and phenotypic associations in the chromosomal regions surrounding two *Arabidopsis thaliana* flowering time loci. Genetics 168:1627–1638

Halama P (2002) Mating relationships between isolates of *Phaeosphaeria nodorum* (anamorph *Stagonospora nodorum*) from geographical locations. Eur J Plant Pathol 108:593–596

Huang Q, Borner A, Roder MS, Ganal MW (2002) Assessing genetic diversity of wheat (*Triticum aestivum* L.) germplasm using microsatellite markers. Theor Appl Genet 105:699–707

Hudson RR (2001) Linkage disequilibrium and recombination. In: Balding D, Bishop M, Cannings C (eds) Handbook of statistical genetics. Wiley, New York, pp 309–324

Igartua E, Casas AM, Ciudad F, Montoya JL, Romagosa I (1999) RFLP markers associated with major genes controlling heading date evaluated in a barley germ plasm pool. Heredity 83:551–559

Ivandic V, Hackett CA, Nevo E, Keith R, Thomas WTB et al (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

Ivandic V, Thomas WTB, Nevo E, Zhang Z, Forster BP (2003) Associations of simple sequence repeats with quantitative trait variation including biotic and abiotic stress tolerance in *Hordeum spontaneum*. Plant Breed 122:300–304

Keller SM, McDermott JM, Pettway RE, Keller SM, McDermott JM, Pettway RE et al (1997) Gene flow and sexual reproduction in the wheat glume blotch pathogen *Phaeosphaeria nodorum* (Anamorph *Stagonospora nodorum*). Phytopathology 87:353–358

Kilpikari R, Sillanpää MJ (2003) Bayesian analysis of multilocus association in quantitative and qualitative traits. Genet Epidemiol 25:122–135

Kraakman ATW, Rients EN, Petra MM, Van den Berg M, Stam P et al (2004) Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. Genetics 168:435–446

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

Nordborg M, Borevitz JO, Bergelson J, Berry CC, Chory J et al (2002) The extent of linkage disequilibrium in *Arabidopsis thaliana*. Nat Genet 30:190–193

Paillard S, Schnurbusch T, Winzeler M, Messmer M, Sourdille P et al (2003) An integrative genetic linkage map of winter wheat (*Triticum aestivum* L.). Theor Appl Genet 107:1235–1242

Pakniyat H, Powell W, Baird E, Handley LL, Robinson D et al (1997) AFLP variation in wild barley (*Hordeum spontaneum* C. Koch) with reference to salt tolerance and associated ecogeography. Genome 40:332–341

Palaisa KA, Morgante M, Williams M, Rafalski A (2003) Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci. Plant Cell 15:1795–1806

Prasad M, Varshney RK, Roy JK, Balyan HS, Gupta PK (2000) The use of microsatellites for detecting DNA polymorphism, genotype identification and genetic diversity in wheat. Theor Appl Genet 100:584–592

Pritchard JK, Stephens M, Rosenberg NA, Donnelly P (2000) Association mapping in structured populations. Am J Hum Genet 67:170–181

Ravel C, Praud S, Murigneux A, Linossier L, Dardevet M, Balfourier F, Dufour P, Brunel D, Charmet G (2006) Identification of Glu-B1-1 as a candidate gene for the quantity of high-molecular-weight glutenin in bread wheat (*Triticum aestivum* L.) by means of an association study. Theor Appl Genet 112:738–743

Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR et al (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc Natl Acad Sci USA 98:11479–11484

Röder MS, Wendehake K, Korzun V, Bredemeijer G, Laborie D et al (2002) Construction and analysis of a microsatellite-based database of European wheat varieties. Theor Appl Genet 106:67–73

Rosenberg NA, Li LM, Ward R, Pritchard JK (2003) Informativeness of genetic markers for inference of ancestry. Am J Hum Genet 73:1402–1422

Schnurbusch T, Paillard S, Fossati D, Messmer M, Schachermayr G et al (2003) Detection of QTLs for Stagonospora glume blotch resistance in Swiss winter wheat. Theor Appl Genet 107:1226–1234

Scott PR, Benedikz PW, Cox CJ (1982) A genetic-study of the relationship between height, time of ear emergence and resistance to *Septoria nodorum* in wheat. Plant Pathol 31:45–60

Searle SR (1987) Linear models for unbalanced data. Wiley, New York

Shaner G, Finney RE (1997) The effect of nitrogen fertilisation on the expression of slow-mildewing resistance in Knox wheat. Phytopathology 67:1051–1056

Shaner G (1999) Breeding for resistance to septoria and stagonospora blotches in winter wheat in the United States. In: van Ginkel M, McNab A, Krupinsky J (eds) *Septoria* and *Stagonospora* diseases of cereals: a compilation of global research. Proceedings of the fifth international septoria workshop, CYMMIT, Mexico

Siedler H, Messmer MM, Schachermayr GM, Winzeler H, Winzeler M (1994) Genetic diversity in European wheat and spelt breeding material based on RFLP data. Theor Appl Genet 88:994–1003

Simko I, Costanzo S, Haynes KG, Christ BJ, Jones RW (2004) Linkage disequilibrium mapping of a *Verticillium dahliae* resistance quantitative trait locus in tetraploid potato (*Solanum tuberosum*) through a candidate gene approach. Theor Appl Genet 108:217–224

Stachel M, Lelley T, Grausgruber H, Vollmann J (2000) Application of microsatellites in wheat (*Triticum aestivum* L.) for studying genetic differentiation caused by selection for adaptation and use. Theor Appl Genet 100:242–248

Stein N, Herren G, Keller B (2001) A new DNA extraction method for high-throughput marker analysis in a large-genome species such as *Triticum aestivum*. Plant Breed 120:354–356

Tanksley SD (1993) Mapping polygenes. Annu Rev Genet 27:205–233

Tenaillon MI, Sawkins MC, Long AD, Gaut RL, Doebley JF et al (2001) Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). Proc Natl Acad Sci USA 98:9161–9166

Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D et al (2001) *Dwarf8* polymorphisms associate with variation in flowering time. Nat Genet 28:286–289

van Ginkel M, Rajaram S (1999) Breeding for resistance to the Septoria/stagonospora blights of wheat. In: van Ginkel M, McNab A, Krupinsky JM (eds) Septoria and stagonospora diseases of cereals: a compilation of global research. CYMMIT, Mexico, pp 117–126

Virk PS, Ford-Lloyd BV, Jackson MT, Pooni HS, Clemeno TP (1996) Predicting quantitative variation within rice germplasm using molecular markers. Heredity 76:296–304

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

Wicki W, Winzeler M, Schmid JE, Stamp P, Messmer MM (1999) Inheritance of resistance to leaf and glume blotch caused by *Septoria nodorum* Berk. In winter wheat. Theor Appl Genet 99:1265–1272

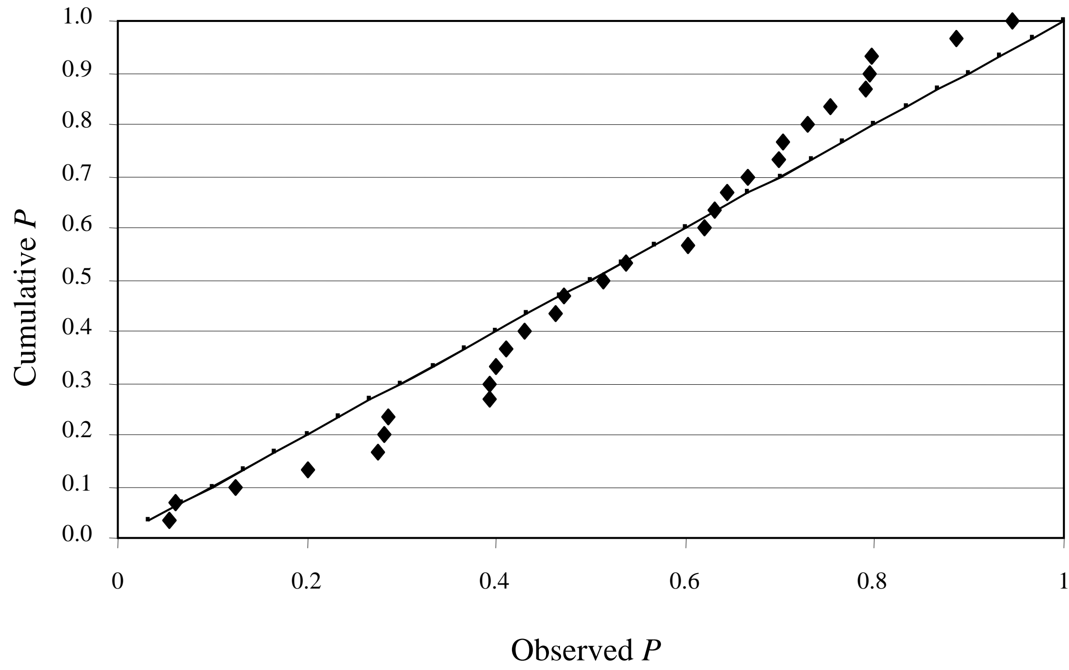
Wilson LM, Whitt SR, Ibanez AM, Rocheford TR, Goodman MM et al (2004) Dissection of maize kernel composition and starch production by candidate gene association. Plant Cell 16:2719–2733

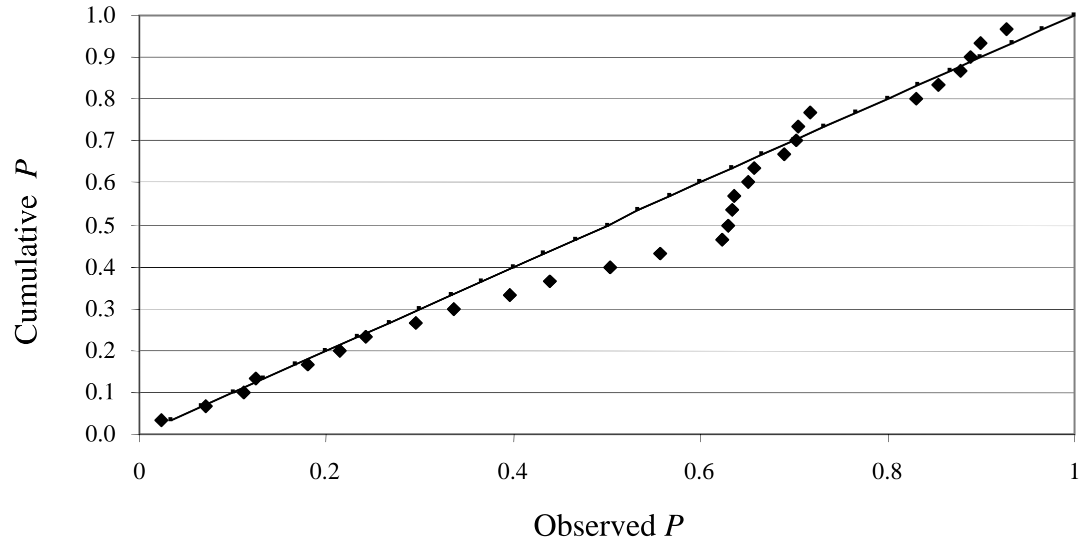
Yu J, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen D M, Holland JB, Kresovich S, Buckler ES (2005) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203–208

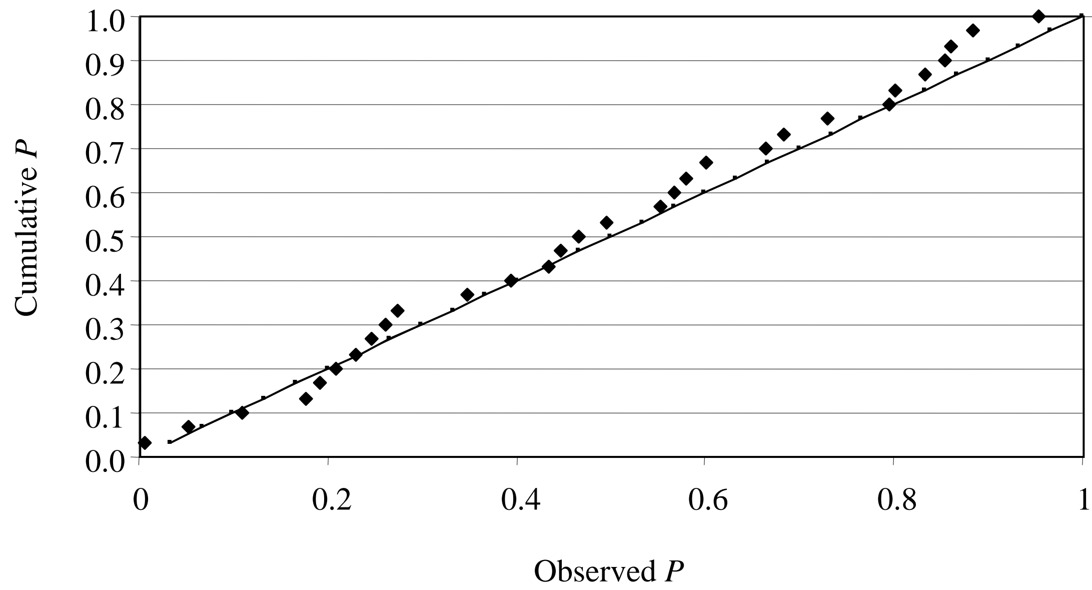
The appropriateness of three different association analysis models, the least square solution to the fixed effects General Linear Model (GLM-Q, Searle 1987), the Mixed Linear Model accounting for structure and kinship (MLM-QK, Yu et al. 2006), and the Mixed Linear Model accounting only for kinship (MLM-K, Yu et al. 2006), was tested for our data set. To assess the presence of relatedness in our population, the relative kinship matrix (K) was calculated using the software package SPAGeDi (Hardy and Vekemans, 2002), as described in Yu et al. (2006), based on the same 19 SSRs used for estimating population structure. We found that the relatedness component was significant in the population screened (in the KQ model without markers Pr Z of K<0.001; in the QK model with markers Pr Z of K<0.05). To evaluate which of these three models for association analysis is the most appropriate for our mapping population, we did a Kolmogorov-Smirnov goodness-of-fit test for uniform distribution for each model, using molecular data from the same 19 unlinked loci used for evaluating structure and kinship. The best fit for uniform distribution was found for the GLM-Q model (accounting for structure only) for which p=0.08. The worst fit was observed for the MLM-K model (accounting for relatedness only, MLM-K p=0.18). The MLM-KQ model (accounting for relatedness and structure) showed an improved fit compared to the MLM-K model (MLM-QK p=0.16). The simple association model that does not account neither for structure nor for kinship, was included here only to show the effect of not accounting for population structure and family relationships, and showed the worst fit (GLM-noQK p=0.183). To graphically evaluate the goodness-of-fit of the model tested, Observed *P* values from each association model tested were plotted over Cumulative *P* values (Suppl. Fig. S1-S4). From these plots it is evident that the GLM-Q model shows an almost perfectly uniform distribution and alignment along the plot diagonal line, indicating that it is the best model for our data set.

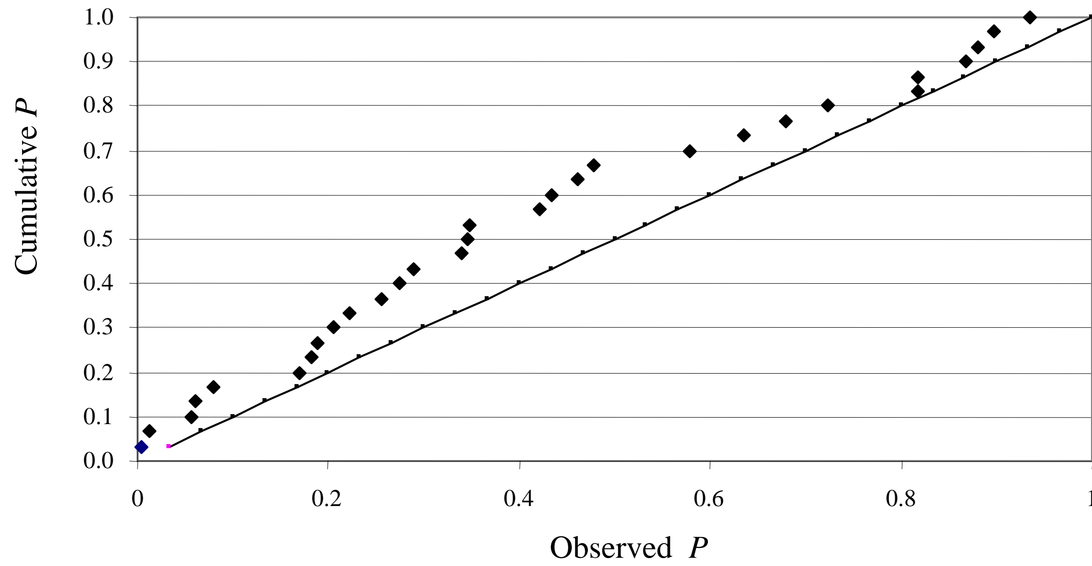
Figure S1-S4

Evaluation type I error rates for the MLM-QK (S.1), MLM-K (S.2), GLM-Q (S.3) and GLM-noQK (S.4) association models, based on19 unlinked SSR loci. The cumulative distributions of Observed *P* values are presented for these models. Approaches that control for type I errors should have a uniform distribution of P values (a diagonal line in these cumulative plots).


S1. MLM-QK model
Cumulative P
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
0.2
0.4
0.6
0.8
1
Observed P


S2. MLM-K model
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Cumulative P
0
0.2
0.4
0.6
0.8
1
Observed P


S3. GLM-Q model
Cumulative P
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
0.2
0.4
0.6
0.8
1
Observed P



S4. GLM-noQK model
Cumulative P
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
0.2
0.4
0.6
0.8
1
Observed P