Mol Breeding (2012) 30:831–843
DOI 10.1007/s11032-011-9667-6

Identification of marker-trait associations in the German winter barley breeding gene pool (*Hordeum vulgare* L.)

Jeannette Rode · Jutta Ahlemeyer · Wolfgang Friedt · Frank Ordon

Received: 20 May 2011 / Accepted: 29 October 2011 / Published online: 22 November 2011

Abstract A genome-wide association mapping approach for grain yield and traits of high agronomic relevance was carried out on basis of a set of 61 six-rowed and 48 two-rowed German winter barley (*Hordeum vulgare* L.) cultivars representing breeding progress in the period 1959–2003. Extensive phenotyping was conducted in field trials carried out at 12 locations in 3 years. Heritability was estimated at between 0.45 for grain yield and 0.94 for grains per spike. By using the Illumina Golden Gate Bead Array technology, 833 single nucleotide polymorphisms with an allele frequency higher than 5% were obtained. Linkage disequilibrium on the whole genome extends to 7.35 cM. Based on a mixed linear model approach taking into account the population structure estimated on the basis of 72 simple sequence repeat markers covering the whole barley genome, 91 significant marker-trait associations were detected, corresponding to 48 different genomic regions.

Keywords Barley · Association genetics studies · Linkage disequilibrium · Population structure · SNPs · Marker-assisted selection

Electronic supplementary material The online version of this article (doi:10.1007/s11032-011-9667-6) contains supplementary material, which is available to authorized users.

J. Rode · F. Ordon
Julius Kühn-Institute, Federal Research Centre for Cultivated Plants, Institute for Resistance Research and Stress Tolerance, Erwin-Baur-Str. 27, 06484 Quedlinburg, Germany

Present Address:
J. Rode (✉)
Martin-Luther-University Halle-Wittenberg, Institute of Agricultural and Nutritional Sciences, Betty-Heimann-Str. 3, 06120 Halle (Saale), Germany
e-mail: jeannette.rode@landw.uni-halle.de

J. Ahlemeyer · W. Friedt
Justus Liebig University, Department of Plant Breeding, Heinrich-Buff-Ring 26-32, 35392 Giessen, Germany

Present Address:
J. Ahlemeyer
Deutsche Saatveredelung AG (DSV), Weissenburger Str. 5, 59557 Lippstadt, Germany

Introduction

Barley (*Hordeum vulgare* L.) is the second most important crop species in Germany and ranks fourth in area cultivated worldwide. In barley, many monogenic traits as well as quantitative trait loci (QTL) have been mapped based on bi-parental mapping populations (for overview see Friedt and Ordon 2008), and the respective molecular markers are used for marker-assisted selection (MAS) in barley breeding. Since QTL detected in bi-parental populations are often not transferable to different genetic backgrounds, their general applicability for MAS is limited. In addition, in many cases QTL mapping is not performed in adapted breeding populations, so that breeders in

general question its relevance for applied barley breeding (Jannink et al. 2001). Marker-trait association analysis is an alternative to QTL mapping in biparental populations (Jun et al. 2008). While linkage studies can only use family data, association studies can also use data from unrelated individuals. Broader genetic variation in a more representative genetic background can be included in association studies (Neale and Savolainen 2004). Therefore, a higher proportion of molecular markers are polymorphic, thus association mapping may attain better genome coverage than any bi-parental map (Breseghello and Sorrells 2006).

One of the prerequisites of a marker-trait association analysis is knowledge of the degree of linkage disequilibrium (LD) in the population analysed. The level of LD varies according to the species and the loci investigated. For barley as a predominantly self-pollinating species, high levels of LD were estimated (Kraakman et al. 2004; Malysheva-Otto et al. 2006) compared to other species such as maize (Remington et al. 2001).

A major complication in LD mapping is the detection of false-positive marker-trait associations due to population structure. This is probably of particular importance in barley, as diversity studies have shown barley germplasm to be highly structured, predominantly due to spike morphology and vernalization requirement (Malysheva-Otto et al. 2006; Rostoks et al. 2006; Cockram et al. 2008; Comadran et al. 2009; Hamblin et al. 2009).

A genome-wide LD mapping approach requires an adequate marker density (Cardon and Bel 2001). However, the number of markers needed depends on the level of LD (Mohlke et al. 2010). Today, high-throughput marker technologies such as Illumina's Golden Gate™ Assay (Close et al. 2009) facilitate the estimation of the genome-wide genetic diversity and its use in barley breeding by linking the genetic to the phenotypic diversity. The Illumina Golden Gate oligonucleotide assays can detect single nucleotide polymorphisms (SNPs) by scoring the presence or absence of specific DNA alleles in a defined genomic representation through hybridization to microarrays (Oliphant et al. 2002).

Genome-wide association studies have already been conducted in many species, e.g. rice (Agrama et al. 2007), wheat (Breseghello and Sorrells 2006), perennial ryegrass (Skøt et al. 2005) and maize (Remington et al. 2001). In barley, Igartua et al. (1999) detected marker-trait associations for heading date based on restriction fragment length polymorphisms (RFLPs) in a population consisting of 32 barley cultivars. In addition, Ivandic et al. (2003) found associations between simple sequence repeat (SSR) markers and water-stress tolerance and powdery mildew resistance, respectively, in 52 wild barley lines. Using an LD mapping approach Kraakman et al. (2004, 2006) identified linkage between SSRs as well as amplified fragment length polymorphism (AFLP) markers with several agronomically relevant traits in 146 two-rowed spring barley cultivars. Pswarayi et al. (2008) studied changes in allele frequencies of Diversity Arrays Technology (DArT®) markers linked to yield QTL in 188 barley accessions (landraces, old and modern cultivars). Population structure and genome-wide LD were investigated in 192 *H. vulgare* accessions using nuclear microsatellite and DArT markers to identify associations with powdery mildew (Comadran et al. 2009). Inostroza et al. (2009) measured grain yield and plant height of 80 recombinant chromosome substitution lines of barley and marker-trait associations were identified using SSRs. DArT markers were used to identify associations for malting quality within a collection of 91 elite two-rowed malting barley lines (Beattie et al. 2010). Roy et al. (2010) genotyped 318 diverse wild barley accessions with DArT markers and SNP markers from the Illumina Golden Gate Assay for association mapping of spot blotch resistance. About 490 UK barley cultivars were genotyped using Golden Gate Bead Array technology to associate them with historical phenotypic data collected during varietal registrations (Cockram et al. 2010).

Besides whole-genome association approaches, candidate gene-based studies are used to detect associations between SNPs located in or near candidate genes and traits of interest. Several candidate gene association studies have been conducted in barley; e.g. Caldwell et al. (2006) measured LD between and within gene loci of the chromosomal region harboring the hardness locus in three different gene pools of *H. vulgare* (64 cultivated, 23 landrace and 34 wild barley accessions). Stracke et al. (2007) analysed LD in the physical and genetic context of the barley gene *Hv-eIF4E*, which confers resistance to Barley yellow mosaic virus and Barley mild mosaic virus. Cockram et al. (2008) identified vernalization

genes by analysing 429 spring and winter barley cultivars. Associations between SNP markers detected for the α-amylase gene *amy1* and important malting quality parameters were discovered in a collection of 117 European spring and winter barley cultivars by Matthies et al. (2009a, b). Stracke et al. (2009) estimated associations between nucleotide diversity at three genes (*Ppd-H1*, *HvCO1*, and *HvFT1*) and variation in flowering time in a collection of 220 spring barley cultivars.

The aim of this study was to determine whether a whole-genome association mapping approach is suited to detecting marker-trait associations in a gene pool relevant for German winter barley breeding. Consequently, the German barley collection investigated represents the most important varieties from the period 1959–2003.

Materials and methods

Plant material

The association panel consisted of a collection of 109 German winter barley cultivars from 31 different breeding companies, grouped by ear type into 61 six-rowed and 48 two-rowed types. The cultivars have been released in Germany between 1959 and 2003 and were the most important cultivars in their decades of release. The complete list of cultivars used in this study, the year of release and the breeding companies are presented in Electronic Supplementary Material Table S1.

Field trials

Field experiments were conducted in cooperation with private breeding companies in the growing seasons 2003/2004, 2004/2005 and 2005/2006. The 109 winter barley cultivars were grown at 12 sites located throughout Germany. Each trial was arranged according to an alpha-lattice with two replications per site. The winter barley cultivars were grown according to local practice for sowing rate and other inputs.

For association studies, the following traits were considered: grain yield (t/ha), 1,000-kernel weight (g), grains per spike, spikes/m^2, hectolitre weight (kg/hl), plant height (cm), lodging at flowering (rating 1–9), lodging before harvest (rating 1–9), heading date (days after 1 May), brackling (rating 1–9), necking (rating 1–9), and sieve fraction (2.8 mm). The number of ears/m^2 was determined by counting the number of tillers in one meter of row. The number of grains per spike was determined by dividing the grain yield per plot by the 1,000-kernel weight and the number of spikes per plot. The visual ratings were based on a 1–9 rating scale, with 1 for minimum and 9 for maximum.

For the analysis of variance, Type I mean squares (MS) were calculated using the general linear models procedure in SAS (SAS software; SAS Institute 2004). Heritabilities (h^2) for all investigated traits were estimated as follows:

$$h^2 = (\mathrm{MS_G} - \mathrm{MS_{GE}} - \mathrm{MS_{GY}} + \mathrm{MS_E})/\mathrm{MS_G}$$

where $\mathrm{MS_G}$ = mean square of genotypes (cultivars), $\mathrm{MS_{GE}}$ = mean square of genotype × environment, $\mathrm{MS_{GY}}$ = mean square of genotype × year, $\mathrm{MS_E}$ = mean square of error.

DNA extraction

Genomic DNA was extracted from leaf tissue according to a modified protocol of Stein et al. (2001).

SSR analysis

The cultivars were genotyped by means of 72 SSR markers covering the whole barley genome (Ramsay et al. 2000). SSR amplification was performed according to Ramsay et al. (2000) in a total volume of 10 µl. PCR products were separated on 6% polyacrylamide gels on a LI-COR DNA Sequencer 4200 (MWG Biotech AG, Ebersberg, Germany). The bands on all gels were scored in a binary format, 1 for the presence and 0 for the absence of bands at polymorphic sites, respectively.

Genetic similarity estimation and principal coordinate analysis

Genetic similarity (GS) between the 109 winter barley cultivars was calculated by the Numerical Taxonomy and Multivariate Analysis System (NTSYS) ver. 1.70 computer software (Rohlf 2000) based on the 1–0 data from the SSR analysis. GS was calculated using the Dice (1945) coefficient. Graphical representations of the similarities between cultivars were obtained by

principal coordinate analysis (PCoA; Sneath and Sokal 1973).

Population structure

The SSR data were used to identify subpopulations within the barley collection by means of the software package STRUCTURE (ver. 2.2; Pritchard et al. 2000). Independent simulations were performed for a K setting between 2 and 10 using the admixture model and a burn-in of 500,000 followed by 1,000,000 iterations. The optimal number of subpopulations K was chosen on the basis of the second-order rate of change of the likelihood function with respect to K. Results were summarized in matrices of fractional subpopulation membership (Q matrices).

Genotyping with SNP markers

To investigate allelic diversity in the barley collection, whole-genome profiling using Illumina's Golden Gate Bead Array technology (Illumina Inc., San Diego, CA, USA; http://www.illumina.com) was carried out.

A set of 1,536 genome-wide biallelic SNPs from a customized oligonucleotide pool assay (IPK-OPA, unpublished) was tested on the DNA of the 109 cultivars using the Golden Gate BeadArray technology by Illumina Inc. The IPK-OPA has been mainly built on a selection of markers from two pilot assays (pOPA1, pOPA2) that are polymorphic between the two barley cultivars Barke and Morex. Raw data were transformed to genotype calls, initially using Illumina's GenCall software. The data from all 109 winter barley cultivars were visually inspected to manually set 1,536 archetypal clustering patterns. The cluster positioning was guided by knowledge that heterozygotes are nearly non-existent because all cultivars were regarded as pure lines. A small proportion of detected non-homozygous data points was treated as missing data. Genotype calls were exported as spreadsheets from GenCall to then parse to create input for statistics software.

LD analysis

LD was measured using the squared correlation coefficient r^2 between pairs of SNP markers (Pritchard and Przeworski 2001). As the majority of SNPs of the IPK-OPA have been included in a barley consensus map (Close et al. 2009), it was possible to compare LD at known genetic map distances. Values of r^2 were calculated using the TASSEL software 2.1 (http://www.maizegenetics.net) between all pairs of loci and plotted against the genetic distance in centimorgans (cM). A smooth line was drawn using a 1,000-period moving linear regression. Background LD was assessed as the 95% percentile of physically unlinked markers. To determine the distance at which LD decays below background LD, functions for r^2 measurements of genetic distances were calculated using a sliding-window approach with a window size of 500 adjacent pairwise SNPs. All calculations were done for the whole dataset and for six- and two-rowed cultivars, separately.

Association analysis

For the association mapping approach, only polymorphic markers with allele frequencies larger than 5% were taken into account. The K matrix was generated based on 72 SSRs using the kinship matrix function in TASSEL software. Based on the least square means for the cultivar performance and 833 Illumina SNPs, marker-trait associations were calculated in a mixed linear model (MLM) approach with population affiliations as covariates and ear morphology as a factor using TASSEL. The threshold for significance was based on a false discovery rate of 0.1% ($P \leq 0.001$). For all associated markers, the average difference of the phenotypic values of both marker classes was estimated using the UNIANOVA procedure of SPSS (SPSS 19.0, SPSS Inc., Chicago, IL, USA). Association profiles were created by plotting $-\log_{10}(P)$ values against chromosomal positions. The percentage of variance explained by each SNP (R^2) was obtained using the TASSEL software.

Results

Phenotypic variation

Summary statistics for each trait are presented in Table 1. The 109 winter barley cultivars phenotyped in this study in 3 years' field trials showed a broad variation for all traits measured, e.g. grain yield of the

Table 1 Agronomical traits of 109 winter barley cultivars estimated in 3 years' field trials

Traits	Number of environments tested in			61 Six-rowed winter barley cultivars			48 Two-rowed winter barley cultivars			h^2
	2003/2004	2004/2005	2005/2006	Min	Max	Mean	Min	Max	Mean	
Grain yield (t/ha)	12	10	11	6.28	8.98	8.06	6.71	8.15	7.65	0.46
1,000-kernel weight (g)	10	10	10	31.74	50.28	43.44	44.13	61.22	52.21	0.81
Grains per spike	7	6	6	28.68	44.62	36.47	17.15	22.51	19.48	0.94
Spikes/m^2	8	6	7	480.22	703.47	548.99	662.57	918.88	789.31	0.84
Hectolitre weight (kg/hl)	10	10	10	59.78	69.14	64.90	67.59	73.46	69.37	0.70
Plant height (cm)	10	9	9	95.04	123.09	109.59	89.77	111.80	101.42	0.63
Lodging at flowering (1–9)	9	6	5	1.04	6.69	2.46	1.02	4.14	2.07	0.73
Lodging before harvest (1–9)	6	5	2	2.03	7.16	4.10	1.77	5.74	3.96	0.63
Heading date (days after 1 May)	10	9	7	16.23	20.51	18.53	17.62	22.56	19.75	0.82
Brackling (1–9)	1	4	7	1.83	5.10	3.38	1.39	4.50	2.75	0.47
Necking (1–9)	3	5	7	1.89	4.82	3.16	1.59	4.74	2.80	0.60
Sieve fraction 2.8 mm	10	10	10	9.07	66.87	43.01	27.55	84.09	65.21	0.79

six-rowed cultivars ranged between 6.28 and 8.98 t/ha. A broad variation was also detected between six- and two-rowed cultivars, implying that the complete sample represents the phenotypic diversity present within the German winter barley breeding pool.

Variance components were calculated by SAS. The results confirmed that the genotypic variance was significant for all traits ($P < 0.0001$). Detailed results of the analysis of variance are presented in Electronic Supplementary Material Table S2. Heritabilities (h^2) were high and ranged from 0.46 for grain yield to 0.94 for grains per spike (Table 1).

Principal coordinate analysis

GS of 109 winter barley cultivars was assessed using 72 SSRs. A total of 404 loci were detected by this approach. The average number of alleles per locus was 5.4, ranging from 2.0 to 13.0. Dice's GS within the two groups ranged from 0.290 (six-rowed) versus 0.333 (two-rowed) to 0.956 (six-rowed) versus 0.910 (two-rowed), with a mean of 0.590 (six-rowed) versus 0.575 (two-rowed).

Principal coordinate analysis was conducted to visualize the GS of the subpopulations. The samples formed two clusters separating two- and six-rowed cultivars based on GS (Electronic Supplementary Material Fig. S1).

Population structure

To avoid spurious associations, the genetic structure among all 109 barley cultivars was investigated on the basis of the 72 SSR markers randomly distributed over the genome. In a first set of runs, two groups separating the two- and six-rowed cultivars were detected. Because principal coordinate analysis also pointed to the existence of these two groups, the population structure of the six-rowed and the two-rowed cultivars was calculated separately and, as a result, four subgroups could be distinguished within each of the two groups (Fig. 1).

Linkage disequilibrium

The squared allele-frequency correlations (r^2) representing LD were assessed for all pairs of Illumina SNP markers. Electronic Supplementary Material Fig. S2 shows pairwise r^2 values in relation to the genetic distance (cM). A critical value of r^2, beyond which LD is likely to be caused by genetic linkage in the absence of population structure, was estimated at 0.13 for all winter barley cultivars (Table 2). Similar estimates of background LD were calculated for the six-rowed (0.11) and two-rowed (0.14) cultivars. LD differed markedly for the different chromosomes, e.g. within the six-rowed cultivars LD ranged from 3.03 (2H) to 17.78 cM (3H). A rapid LD decay below the critical r^2

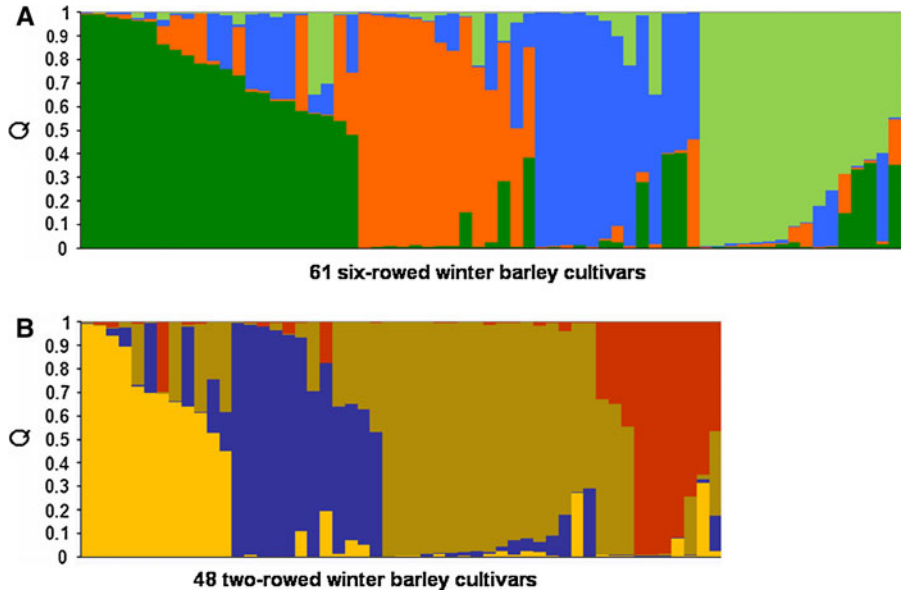

Fig. 1 Calculated population structure based on 72 SSR loci. The population was partitioned into *color-coded* subpopulations. Each *bar* represents a single cultivar, and the *colored segments* within each *bar* reflect the proportional contribution of the subpopulations to that cultivar. Cultivars are not ordered (**A** six-rowed, **B** two-rowed winter barley cultivars). (Color figure online)

Table 2 Estimates of background LD and distances at which LD decays below background LD for six- and two-rowed cultivars and the complete set of genotypes

Background r^2	Six-rowed 0.11	Two-rowed 0.14	Complete set 0.13
Distance at which r^2 decays below background (cM) for chromosome			
1H	6.06	7.93	2.83
2H	3.03	7.69	7.58
3H	17.78	9.16	9.72
4H	7.07	7.61	4.01
5H	10.30	10.69	3.63
6H	12.21	4.86	4.95
7H	6.89	5.58	5.21
Whole genome	10.75	8.24	7.35

was estimated for distances >7.35 cM within the complete set of winter barley cultivars. A slower LD decay was observed in the subgroups than in the whole set of genotypes.

In a few cases high r^2 values between theoretically unlinked markers (>50 cM) were observed (Electronic Supplementary Material Fig. S2).

Association mapping

The main objective of this study was the detection of marker-trait associations in the German winter barley breeding gene pool. Table 3 summarizes the results, and details are shown in Electronic Supplementary Material Table S3.

In summary, 91 significant associations were detected using the MLM approach with population structure as a covariate. The 91 marker-trait associations correspond to 66 different SNPs with a maximum of four significant associations for one SNP (marker 649) with different traits.

Most of the marker-trait associations were detected for plant height (25), lodging at flowering (12), lodging before harvest (11), and grain yield (9). Only one SNP could be detected for grains per spike, which showed a significant association within the 109 winter barley cultivars.

For all associated SNPs the map position was known, e.g. SNP 1462 significantly associated with hectolitre weight ($P = 0.0006$) is located on chromosome 1H at 101.45 cM. Most of the associated markers are mapped on chromosome 5H (43 SNPs) and marked 15 genomic regions. No marker-trait associations were detected on chromosome 7H.

The associated SNPs explained between 0.2% (e.g. SNP marker 2184 associated with spikes/m^2) and 15.4% (e.g. SNP marker 4876 associated with plant height) of the phenotypic variance for individual traits. A high number of the significant SNP markers proved to be closely linked and map at same positions, e.g. four markers associated with plant height are located

Table 3 Summary of all marker-trait associations found in a collection of 109 barley cultivars

Traits	Number of genomic regions associated with one trait on the seven chromosomes*							Total number of genomic regions
	1H	2H	3H	4H	5H	6H	7H	
Grain yield			42 cM (1 SNP)		51 cM (6 SNPs)			4 (9 SNPs)
			169 cM (1 SNP)		61 cM (1 SNP)			
1,000-kernel weight		59 cM (1 SNP)			124 cM (1 SNP)	6 cM (1 SNP)		5 (5 SNPs)
		83 cM (1 SNP)				129 cM (1 SNP)		
Grains per spike	41 cM (1 SNP)							1 (1 SNP)
Spikes/m^2	52 cM (2 SNPs)				133 cM (4 SNPs)	2 cM (1 SNP)		3 (7 SNPs)
Hectolitre weight	101 cM (2 SNPs)							2 (3 SNPs)
	121 cM (1 SNP)							
Plant height	41 cM (1 SNP)	113 cM (2 SNPs)	39 cM (6 NPs)	84 cM (1 SNP)	85 cM (1 SNP)			10 (25 SNPs)
	91 cM (2 SNPs)		56 cM (1 SNP)		108 cM (1 SNP)			
					133 cM (7 SNPs)			
					142 cM (3 SNPs)			
Lodging at flowering			42 cM (1 SNP)	97 cM (1 SNP)	36 cM (1 SNP)	71 cM (1 SNP)		6 (12 SNPs)
					50 cM (2 SNPs)			
					61 cM (6 SNPs)			
Loading before harvest		138 cM (1 SNP)		97 cM (2 SNPs)	36 cM (1 SNP)	71 cM (1 SNP)		5 (11 SNPs)
					61 cM (6 SNPs)			
Heading date	41 cM (1 SNP)		60 cM (1 SNP)	49 cM (3 SNPs)				3 (5 SNPs)
Brackling	121 cM (1 SNP)			97 cM (1 SNP)	62 cM (1 SNP)	65 cM (2 SNPs)		4 (5 SNPs)
Necking		59 cM (3 SNPs)						1 (3 SNPs)
Sieve fraction		89 cM (1 SNP)			124 cM (2 SNPs)	73 cM (1 SNP)		4 (5 SNPs)
						129 cM (1 SNP)		
Total number of genomic regions	8 (11 SNPs)	6 (9 SNPs)	6 (11 SNPs)	5 (8 SNPs)	15 (43 SNPs)	8 (9 SNPs)		48 (91 SNPs)

* One genomic region up to c. 7.35 cM. The chromosomal position in cM was taken from the respective SNP with the highest R^2

on chromosome 5H at 132.63 cM. In total, the 66 associated SNPs represented 48 genomic regions (Table 3).

The identification of associations was supported by association profiles, i.e. from plots of P values [$-\log_{10}(P)$] against chromosomal positions. The significance of the association between markers and phenotypic traits was tested with a value of $-\log_{10}(P) = 3$, which indicates a probability value of 0.001. As an example, Fig. 2 shows the associations of Illumina SNPs with grain yield, 1,000-kernel weight, and spikes/m^2, respectively. Most of the associated markers for grain yield were mapped on chromosome 5H in a region between 50 and 60 cM. Only five SNPs were associated with 1,000-kernel weight mapped on chromosomes 2, 5 and 6H. Seven Illumina SNPs located on three different chromosomes were associated with spikes/m^2.

Discussion

In the first step of our study, phenotyping was performed in 3 years' field trials at 12 locations in Germany. The 109 winter barley cultivars used in this study showed a broad phenotypic variation for all traits measured, i.e. yield, yield components and lodging. High values of heritability indicate that these traits can be improved by breeding and that QTL can be detected using such an experimental design.

Microsatellite (SSR) markers evenly distributed over the seven barley linkage groups were effectively used to identify population structure among the 109 winter barley cultivars. SSRs are very well suited for the estimation of genetic diversity due to their multiallelic nature. In addition, SSRs have a higher polymorphism information content (PIC) than other markers, which indicates the higher efficiency of these

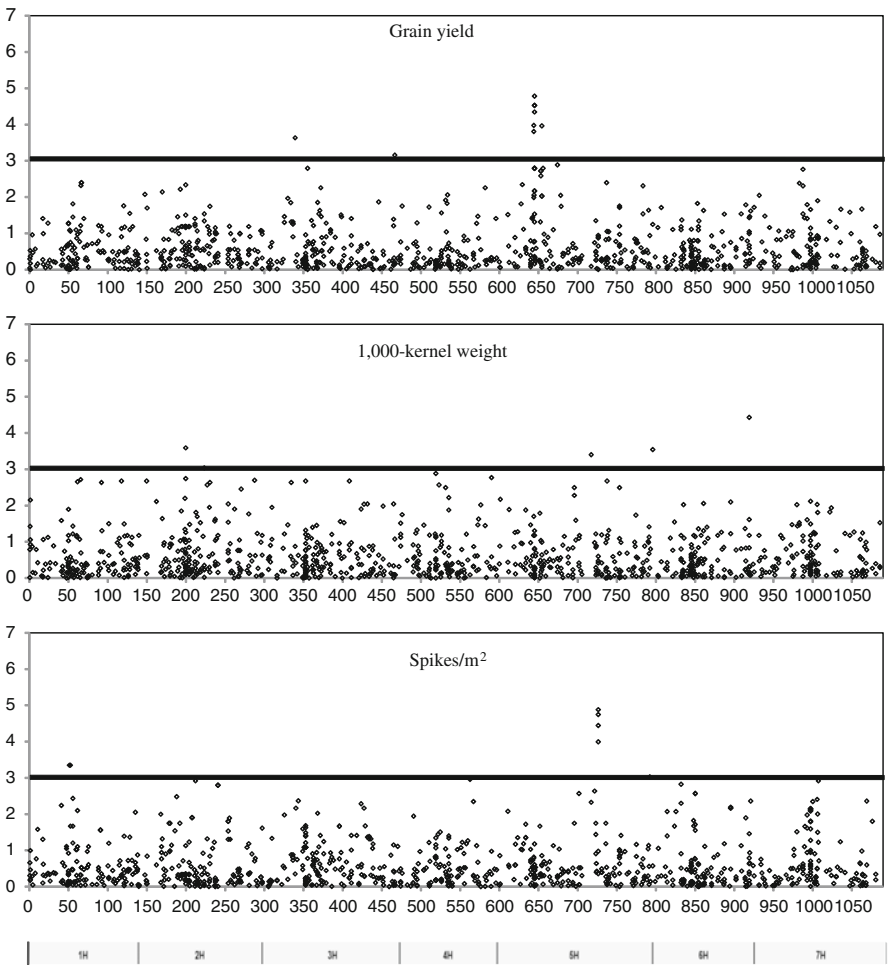

Fig. 2 Association profiles showing $-\log_{10}(P)$ of correlation between Illumina SNPs and grain yield, 1,000-kernel weight, and spikes/m^2, respectively, against the position of the markers on all seven barley chromosomes. SNP markers were arranged on the x axis by their position in cM

markers in showing the existence of diversity among the barley genotypes (Chaabane et al. 2009). Our results show that estimation of population structure using 72 SSR markers yielded clusters/subpopulations according to the ear type of the barley cultivars. In Ordon et al. (2005), the 109 winter barley cultivars were genotyped with at first 30 of the 72 SSRs and PCoA carried out on these data already separated two- and six-rowed cultivars. Grouping by ear type has been noted in other studies for spring (Brantestam et al. 2007; Hayes and Szucs 2006; Cockram et al. 2008; Hamblin et al. 2009; Stracke et al. 2009; Zhang et al. 2009) and winter barley cultivars (Stracke et al. 2007; Cockram et al. 2008; Hamblin et al. 2009). The ear morphology is encoded by the major gene *v* (*Vrs1*) which is located on chromosome 2H (Komatsuda et al. 2007). The strong separation between two- and six-rowed barley cultivars could be explained by the fact that breeders tend to work preferentially within each of these two groups and exploit crosses only to a lower extent between different row-types (Zhang et al. 2009).

Population structure may have a major impact on the outcome of association studies, resulting in false-positive associations (Pritchard et al. 2000). Many species have highly structured populations, e.g. maize (Thornsberry et al. 2001), wheat (Breseghello and Sorrells 2006), rice (Agrama et al. 2007) and barley (Comadran et al. 2009). Population structure is obviously the consequence of multiple events, e.g. modalities of domestication as shown in maize (Matsuoka et al. 2002) or geographical regions of origins as demonstrated in rice (Garris et al. 2005).

The set of 109 winter barley cultivars described in this publication was considered to be composed of two groups (based on ear type), each of them consisting of four subgroups.

Based on the information of 30 of the 72 SSR markers also used in this study, three populations were identified by Ordon et al. (2005), comprising the two-rowed cultivars and the six-rowed cultivars separately. Consequently, we could show that the implementation of more markers resulted in a better characterization of population structure. The observed population structure was taken into account during the identification of marker-trait associations.

A genome-wide LD mapping requires many markers, the number depending on the level of LD. LD decays faster with increasing genetic map distances. Therefore, a very large number of markers are required for effective whole-genome association genetic scans if the LD is low. LD is influenced by the recombination rate, allele frequency, population structure and selection (Flint-Garcia et al. 2003). Out-breeding species generally exhibit low LD, whereas self-fertilizing inbreeding species theoretically exhibit high LD. Therefore, the extent of genome coverage in different germplasm can vary from short distances of a few hundred base pairs up to genetic regions of tens of cM.

Our set of 109 German winter barley cultivars was explored using the Illumina Golden Gate assay, such that 918 polymorphic SNP markers could be used for determining LD. Mean LD values higher than $r^2 = 0.13$ extended up to 7.35 cM within the whole set of cultivars, which suggests that there is sufficient marker coverage for attempting a whole-genome association approach in barley. The LD decay turned out to be slower in the six-rowed and two-rowed cultivars than in the whole set of genotypes. Varying patterns of LD decay in two- and six-rowed barley cultivars probably reflect their different breeding histories (Flint-Garcia et al. 2003). A separate analysis of marker-trait associations within the six-rowed and two-rowed barley cultivars may therefore partly lead to different results, but was not conducted in this study due to the small sample size.

However, in comparison with other species the LD is high, e.g. in the out-breeding species maize LD diminished within a few hundred to 2,000 bp, depending on whether landrace or cultivars are analysed (Remington et al. 2001; Tenaillon et al. 2001; Palaisa et al. 2003). In sugar beet, LD extended up to 3 cM (Kraft et al. 2000). In contrast, in *Arabidopsis* as an inbreeding species, a higher LD of about 50 cM was found by Nordborg et al. (2002), but this research was done in populations with only a few genotypes and after extreme inbreeding. The most probable reason for the high level of LD observed in barley is the fact that it is an inbreeding species (Morrell et al. 2005). This observation is confirmed by Kraakman et al. (2004), who estimated LD over large distances >50 cM in some regions in related barley germplasm. Malysheva-Otto et al. (2006) reported on genome-wide LD, which extends up to 10 cM in a set of 953 cultivated barley accessions. Highly significant LD extended over >60 cM in a set of European spring and winter barley cultivars analysed by Rostoks et al.

(2006). Comadran et al. (2009) used DArT markers to analyse 192 *H. vulgare* accessions and a lower degree of LD was estimated (extending up to 3.12 cM) due to a higher average density of markers per chromosome than in other association studies. Rostoks et al. (2006) estimated that approximately 1,000 evenly distributed markers are necessary to ensure good coverage of the barley genome in order to detect genetic linkage to QTL. For map-based cloning, low LD and a higher number of markers may be required, depending on factors such as genome location, which can influence the relationship between physical and genetic distance (Zhang et al. 2009). New high-density marker technologies such as Diversity Array Technology and Illumina Golden Gate BeadArray technology facilitate this. Zhang et al. (2009) genotyped 170 Canadian barley cultivars using 942 DArT loci to estimate LD up to 2.6 cM. In our study we used nearly the same number of SNP loci (918 Illumina), but the extent of LD was higher. Cockram et al. (2010) measured marker-trait associations between 1,111 Illumina SNPs and 500 barley varieties from the UK. They too detected high levels of long-range LD. One reason may be that there is a lower genetic diversity in the respective sets of barley varieties. It is known that LD would be expected to extend over much shorter distances when analysing a very diverse set of genotypes (ancestral populations).

In some cases in wild barley (*Hordeum spontaneum*) as well as in landraces, LD has low values in genes not subjected to the strong directional selection forces of plant breeding. Morrell et al. (2005) determined that LD of wild barley decays at rates similar to maize (Remington et al. 2001). On the other hand, Caldwell et al. (2006) identified LD over a region of about 200 kb encompassing the *Ha* locus controlling grain hardness in barley. Consequently, genes that have been under breeding selection pressure will be in higher LD.

In this study, we explored the possibility of mapping QTL in a set of barley cultivars instead of using a segregating population derived from a bi-parental cross. In summary, 66 significant SNP markers were selected using the software package TASSEL.

Many of the associated markers were found in a region where a candidate gene or QTL or a marker-trait association for one of the traits investigated has been reported previously, e.g. two SNP markers (147, 840) associated with plant height on chromosome 1H are mapped in a region where the photoperiod gene *Ppd-H2* (Laurie et al. 1995) is located. The vernalization gene *VrnH1* is located on chromosome 5H (Yan et al. 2005) where several SNPs associated with plant height (4,098, 2,184, 4,106, 4,874, 5,322, 70, 1,718, 2,180) were detected. Schmalenbach et al. (2009) and Qi et al. (1998) also reported the detection of plant height QTL within a region in which QTL/genes for flowering are located. They detected a high correlation between flowering time and plant height within a set of wild barley introgression lines and recombinant inbred lines, respectively. In our study, we estimated a significant correlation ($r = 0.51$) between these traits, too. This correlation may be due to pleiotropic effects of single genes or due to close linkage of several genes controlling these traits.

However, a couple of new loci were found for all traits measured. Coincidence of the detected marker-trait associations with known QTL positions might indicate a candidate gene locus, but the QTL were detected in different mapping populations using different marker types. Consequently, it is not certain that the associations identified in this study and known QTL indicate the same genes.

The work presented indicates that association mapping approaches can be a useful alternative to classical QTL approaches based on bi-parental crosses. The analysis was performed with barley cultivars which represent breeding progress in Germany during the last five decades, so the results are of great relevance for barley breeding.

Identification of marker-trait associations could be followed by the development of segregating populations, polymorphic for the involved loci, in which the associations are confirmed. Verified trait-associated markers can be used in MAS in combination with traditional field selection to enhance barley breeding. In addition, the isolation of the genes underlying the QTL by map-based cloning will be essential in unravelling the genetic basis of phenotypic variation among barley genotypes.

Acknowledgments The authors wish to thank the co-operating breeding companies Deutsche Saatveredelung AG, Dieckmann GmbH & Co. KG, Limagrain GmbH, KWS Lochow GmbH, Nordsaat Saatzucht GmbH, Pflanzenzucht Oberlimpurg, Saatzucht Bauer GmbH & Co. KG, Lantmännen SW Seed GmbH, Saatzucht Josef Breun GmbH & Co. KG, Saatzucht Streng GmbH & Co. KG, Secobra Saatzucht GmbH and W. von

Borries-Eckendorf GmbH & Co. KG for their support. This work is financed by the German Research Foundation (DFG) as part of the Collaborative Research Centre (SFB) 299 at the University of Giessen and by the Federal Ministry of Education and Research (BMBF) within the framework of GABI-GENOBAR (FKZ 03115066 A–E).

References

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

Beattie AD, Edney MJ, Scoles GJ, Rossnagel BG (2010) Association mapping of malting quality data from western Canadian two-row barley cooperative trails. Crop Sci 50:1649–1663

Brantestam AK, von Bothmer R, Dayteg C, Rashal I, Tuvesson S, Weibull J (2007) Genetic diversity changes and relationships in spring barley (*Hordeum vulgare* L.) germplasm of nordic and baltic areas as shown by SSR markers. Genet Resour Crop Evol 54:749–758

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

Caldwell KS, Russell J, Langridge P, Powell W (2006) Extreme population-dependent linkage disequilibrium detected in an inbreeding plant species, *Hordeum vulgare*. Genetics 172:557–567

Cardon LR, Bel JI (2001) Association study designs for complex diseases. Nat Rev Genet 2:91–99

Chaabane R, El Felah M, Salah HB, Naceur MB, Abdelly C, Ramla D, Nada A, Saker M (2009) Molecular characterization of tunisian barley (*Hordeum vulgare* L.) genotypes using microsatellites (SSRs) markers. Eur J Sci Res 36:6–15

Close TJ, Bhat PR, Lonardi S, Wu Y, Rostoks N, Ramsay L, Druka A, Stein N, Svensson JT, Wanamaker S, Bozdag S, Roose ML, Moscou MJ, Chao S, Varshney RK, Szűcs P, Sato K, Hayes PM, Matthews DE, Kleinhofs A, Muehlbauer GJ, De Young J, Marshall DF, Madishetty K, Fenton RD, Condamine P, Graner A, Waugh R (2009) Development and implementation of high-throughput SNP genotyping in barley. BMC Genomics 10:582

Cockram J, White J, Leigh FJ, Lea VJ, Chiapparino E, Laurie DA, Mackay IJ, Powell W, O'Sullivan DM (2008) Association mapping of partitioning loci in barley. BMC Genet 9:16

Cockram J, White J, Zuluaga DL, Smith D, Comadran J, Macaulay M, Luo Z, Kearsey MJ, Werner P, Harrap D, Tapsell C, Liu H, Hedley PE, Stein N, Schulte D, Steuernagel B, Marshall DF, Thomas WTB, Ramsay L, Mackay I, Balding DJ, The AGOUEB Consortium, Waugh R, O'Sullivan DM (2010) Genome-wide association mapping to candidate polymorphism resolution in the unsequenced barley genome. Proc Natl Acad Sci USA 107:21611–21616

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

Dice LR (1945) Measures of the amount of ecological association between species. Ecology 26:297–302

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

Friedt W, Ordon F (2008) Molecular markers for gene pyramiding and resistance breeding in barley. In: Varshney R, Tuberosa R (eds) Genomics-assisted crop improvement, vol 2. Genomics applications in crops. Springer, Berlin, pp 81–101

Garris AJ, Tai TH, Coburn J, Kresovich S, McCouch S (2005) Genetic structure and diversity in *Oryza sativa* L. Genetics 169:1631–1638

Hamblin MT, Close TJ, Bhat PR, Chao S, Kling JG, Abraham KJ, Blake T, Brooks WS, Cooper B, Griffey CA, Hayes PM, Hole DJ, Horsley RD, Obert DE, Smith KP, Ullrich SE, Muehlbauer GJ, Jannink J-L (2009) Population structure and linkage disequilibrium in US barley germplasm: implications for association mapping. Crop Sci 50:556–566

Hayes P, Szucs P (2006) Disequilibrium and association in barley: thinking outside the glass. Proc Natl Acad Sci USA 103:18385–18386

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

Inostroza L, del Pozo A, Matus I, Castillo D, Hayes P, Machado S, Corey A (2009) Association mapping of plant height, yield, and yield stability in recombinant chromosome substitution lines (RCSLs) using *Hordeum vulgare* subsp. *spontaneum* as a source of donor alleles in a *Hordeum vulgare* subsp. *vulgare* background. Mol Breed 23:365–376

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

Jannink J-L, Bink MCAM, Jansen RC (2001) Using complex plant pedigrees to map valuable genes. Trends Plant Sci 6:337–342

Jun T-H, Van K, Kim MY, Lee S-H, Walker DR (2008) Association analysis using SSR markers to find QTL for seed protein content in soybean. Euphytica 162:179–191

Komatsuda T, Pourkheirandish M, He C, Azhaguvel P, Kanamori H, Perovic D, Stein N, Graner A, Wicker T, Tagiri A, Lundqvist U, Fujimura T, Matsuoka M, Matsumoto T, Yano M (2007) Six-rowed barley originated from a mutation in a homeodomain-leucine zipper I-class homeobox gene. Proc Natl Acad Sci USA 104:1424–1429

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

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

Kraft T, Hansen M, Nilsson N-O (2000) Linkage disequilibrium and fingerprinting in sugar beet. Theor Appl Genet 101:323–326

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

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

Matsuoka Y, Vigouroux Y, Goodman M, Sanchez JG, Buckler E, Doebley J (2002) A single domestication for maize shown by multilocus microsatellite genotyping. Proc Natl Acad Sci USA 99:6080–6084

Matthies IE, Weise S, Roeder M (2009a) Association of haplotype diversity in the *a*-amylase gene amy1 with malting quality parameters in barley. Mol Breed 23:139–152

Matthies IE, Weise S, Foerster J, Roeder M (2009b) Association mapping and marker development of the candidate genes (1 → 3), (1 → 4)-*β*-D-glucan-4-glucanohydrolase and (1 → → → 4)-*β*-xylan-endohydrolase 1 for malting quality in barley. Euphytica 170:109–122

Mohlke KL, Lange EM, Valle TT, Ghosh S, Magnuson VL, Silander K, Watanabe RM, Chines PS, Bergman RN, Tuomilehto J, Collins FS, Boehnke M (2010) Linkage disequilibrium between microsatellite markers extends beyond 1 cM on chromosome 20 in finns. Genome Res 11:1221–1226

Morrell PL, Toleno DM, Lundy KE, Clegg MT (2005) Low levels of linkage disequilibrium in wild barley (*Hordeum vulgare* ssp. *spontaneum*) despite high rates of self-fertilization. Proc Natl Acad Sci USA 102:2442–2447

Neale DB, Savolainen O (2004) Association genetics of complex traits in conifers. Trends Plant Sci 9:1360–1385

Nordborg M, Borevitz JO, Bergelson J, Berry CC, Chory J, Hagenblad J, Kreitman M, Maloof JN, Noyes T, Oefner PJ, Stahl EA, Weigel D (2002) The extent of linkage disequilibrium in *Arabidopsis thaliana*. Nat Genet 30:190–193

Oliphant A, Barker DL, Stuelpnagel JR, Chee MS (2002) BeadArray technology: enabling an accurate, cost-effective approach to high-throughput genotyping. Biotechniques 32:56–61

Ordon F, Ahlemeyer J, Werner K, Koehler W, Friedt W (2005) Molecular assessment of genetic diversity in winter barley and its use in breeding. Euphytica 146:21–28

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

Pritchard JK, Przeworski M (2001) Linkage disequilibrium in humans: models and data. Am J Hum Genet 69:1–14

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

Pswarayi A, van Eeuwijk FA, Ceccarelli S, Grando S, Comadran J, Russell JR, Pecchioni N, Tondelli A, Akar T, Al-Yassin A, Benbelkacem A, Ouabbou H, Thomas WTB, Romagosa I (2008) Changes in allele frequencies in landraces, old and modern barley cultivars of marker loci close to QTL for grain yield under high and low input conditions. Euphytica 163:435–447

Qi X, Niks RE, Stam P, Lindhout P (1998) Identification of QTLs for partial resistance to leaf rust (*Puccinia hordei*) in barley. Theor Appl Genet 96:1205–1215

Ramsay L, Macaulay M, Degli Ivanissevich S, MacLean K, Cardle L, Fuller J, Edwards KJ, Tuvesson S, Morgante M, Massari A, Maestri E, Marmiroli N, Sjakste T, Ganal M, Powell W, Waugh R (2000) A simple sequence repeat-based linkage map of barley. Genetics 156:1997–2005

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

Rohlf FJ (2000) NTSYS-pc: numerical taxonomy and multivariate analysis system, version 2.1. Exeter Software, Setauket

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

Roy JK, Smith KP, Muehlbauer GJ, Chao S, Close TJ, Steffenson BJ (2010) Association mapping of spot blotch resistance in wild barley. Mol Breed 26:243–256

SAS Institute (2004) SAS® 9.1.2 qualification tools user's guide. SAS Institute, Cary

Schmalenbach I, Léon J, Pillen K (2009) Identification and verification of QTLs for agronomic traits using wild barley introgression lines. Theor Appl Genet 118:483–497

Skøt L, Humphreys MO, Armstead I, Heywood S, Skøt KP, Sanderson R, Thomas ID, Chorlton KH, Hamilton NRS (2005) An association mapping approach to identify flowering time genes in natural populations of *Lolium perenne* (L.). Mol Breed 15:233–245

Sneath PHA, Sokal RR (1973) Numerical taxonomy: the principles and practice of numerical classification. W.H. Freeman, San Francisco

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

Stracke S, Prester T, Stein N, Perovic D, Ordon F, Graner A (2007) Effects of introgression and recombination on haplotype structure and linkage disequilibrium surrounding a locus encoding *bymovirus* resistance in barley. Genetics 175:805–817

Stracke S, Haseneyer G, Veyrieras J-B, Geiger HH, Sauer S, Graner A, Piepho HP (2009) Association mapping reveals gene action and interactions in the determination of flowering time in barley. Theor Appl Genet 118:259–273

Tenaillon MI, Sawkins MC, Long AD, Gaut RL, Doebley JF, Gaut BS (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, Buckler ES (2001) Dwarf8 polymorphisms associate with variation in flowering time. Nat Genet 28:286–289

Yan L, von Zitzewitz J, Skinner JS, Hayes PM, Dubcovsky J (2005) Molecular characterization of the duplicated meristem identity genes HvAP1a and HvAP1b in barley. Genome 48:905–912

Zhang LY, Marchand S, Tinker NA, Belzile F (2009) Population structure and linkage disequilibrium in barley assessed by DArT markers. Theor Appl Genet 119:43–52