



Genome wide association analyses for drought tolerance related traits in barley (*Hordeum vulgare* L.)

R.K. Varshney^{a,*}, M.J. Paulo^{b,c,1}, S. Grando^{d,1}, F.A. van Eeuwijk^{b,c}, L.C.P. Keizer^b, P. Guo^{d,2}, S. Ceccarelli^d, A. Kilian^e, M. Baum^d, A. Graner^a

^a Leibniz Institute of Plant Genetics & Crop Plant Research (IPK), Corrensstrasse 3, D-06466 Gatersleben, Germany

^b Wageningen University and Research Center, Applied Statistics, 6700 AC Wageningen, The Netherlands

^c Centre for Biosystems Genomics, P.O. Box 98, 6700 AB Wageningen, The Netherlands

^d International Centre for Agricultural Research in Dry Areas (ICARDA), P.O. Box 5466, Aleppo, Syria

^e Diversity Arrays Technology Pty Ltd, 1 Wilf Crane Crescent, Yarralumla, Canberra, ACT 2600, Australia

ARTICLE INFO

Article history:

Received 23 May 2011

Received in revised form 8 October 2011

Accepted 15 October 2011

Keywords:

Genome wide association analysis
Diversity Array Technology (DArT)
Single nucleotide polymorphisms (SNPs)
Simple sequence repeats (SSRs)
Barley

ABSTRACT

Genome wide association (GWA) analysis of yield, yield components, developmental, physiological and anatomical traits was conducted for a barley germplasm collection consisting of 185 cultivated (*Hordeum vulgare* L.) and 38 wild (*Hordeum spontaneum* L.) genotypes, originating from 30 countries of four continents. Phenotypic evaluations were performed at a dry (Breda) and wet (Tel Hadya) location in Syria. Genome wide association study was done with 816 markers comprised of 710 diversity array technology (DArT), 61 single nucleotide polymorphism (SNP) and 45 microsatellite or simple sequence repeat (SSR) markers. Diversity analysis revealed 5 groups of germplasm, related to origin (Middle East, North Africa), structural information (two-rows), and domestication (wild versus domesticated). Linkage disequilibrium (LD) decayed after 3 cM with a few exceptions at 10 and 15 cM. Although a few QTLs were identified that differed between the dry and wet site, these QTLs explained low phenotypic variation and could not unequivocally be related to drought tolerance when compared to earlier linkage mapping based QTL analysis studies. Therefore, GWA analysis seems to be not very effective for identification of major QTLs for complex traits like drought tolerance in highly structured germplasm collections.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Drought is an important abiotic stress limiting yield, and barley seems to be relatively well adapted to water deficit (Ceccarelli and Grando, 1996). Nevertheless, in times of climate changes some areas have been predicted to be more subjected to frequent drought. Therefore, efforts to improve drought tolerance need to be continued and strengthened. Barley proved a good model to understand drought and water-stress tolerance mechanisms. Although drought is a complex and polygenic trait with strong interactions between loci and strong genotype \times environment interactions (Ceccarelli, 2010), several biochemical, physiological and genetical studies were conducted in the past to identify the genes responsi-

ble for drought tolerance and its mechanism in barley (Zhu, 2002; Tuberosa and Salvi, 2004).

In the context of genetic studies molecular markers and mapping populations, generated after crossing of a drought tolerant genotype with drought sensitive genotype, facilitated identification of genomic regions associated with quantitative trait loci (QTLs) responsible for drought related traits (e.g. Teulat et al., 1998, 2001, 2003; Ellis et al., 2002; Baum et al., 2003; Forster et al., 2004; von Korff et al., 2008). Further, advanced backcross QTL (AB-QTL) studies have also been conducted to identify drought associated QTLs (e.g. Talame et al., 2004; von Korff et al., 2006; Baum et al., 2007). However, it is important to note that the mapping populations used for QTL identification are products of just a few cycles of recombination and this limits the resolution of genetic maps and identified QTLs. Furthermore, the segregation of major developmental genes in each cross, which either accounted for much of the phenotypic variation makes it difficult to determine whether or not the identified loci are robust and transient (Forster et al., 2004). In addition, in the majority of linkage mapping based QTL studies, QTLs identified for physiological traits involved in drought did not co-locate with the yield and agronomic QTLs (Comadran et al., 2008).

* Corresponding author. Present address: International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru 502 324, Greater Hyderabad, A.P., India. Tel.: +91 40 3071 3305; fax: +91 40 3074 3075.

E-mail address: r.k.varshney@cgiar.org (R.K. Varshney).

¹ Authors should be considered as joint first authors.

² Present address: College of Life Science, Guangzhou University, Guangzhou, China.

In recent years, QTL identification in populations with no regular genetic structure has become possible with the development of analytical approaches exploiting linkage disequilibrium (LD) between markers and closely linked, functionally polymorphic QTLs present in large numbers of accessions (Yu et al., 2006; Malosetti et al., 2007; Ersoz et al., 2007). Under this approach, marker-trait association is expected for markers close to QTL, because typically the accumulated recombination events occurring during the development of the lines will have broken down LD between markers and QTL beyond a certain threshold distance. This approach based on LD in arbitrary structured populations has been and remains an important genetic tool in human and animal genetics. Human disease genetics, indeed, was the first area for which the association mapping approach was developed and where significant successes were achieved (e.g. Lander and Schork, 1994; Jorde, 2000; Carlson et al., 2004). In case of animal systems, the major concern was about the LD patterns in breeding populations to determine to what extent LD holds and the marker density required to fine map trait/genes: e.g. in cattle (Farnir et al., 2000; Koning et al., 2007; Charlier et al., 2008), in pig (Nsengimana et al., 2004; Amaral et al., 2008), sheep (McRae et al., 2002; Meadows and Kijas, 2009) and chicken (Andreescu et al., 2007). In plant genetics, though some association mapping studies have been conducted, the majority of them are on model plant systems such as Arabidopsis (e.g. Zhao et al., 2007), maize (e.g. Palasia et al., 2003; Ersoz et al., 2009) and rice (e.g. Garris et al., 2003; Agrama et al., 2007). Association mapping is gaining importance also in other crop species, due to the development of high-throughput marker-genotyping systems, next generation sequencing technologies and availability of genome sequences. It has been possible to investigate LD patterns and conduct association studies in some crop species as well, e.g. in wheat (Bresghehlo and Sorrells, 2006; Neumann et al., 2010), liliium (Sköt, 2005), grapevine (Barnaud et al., 2006), soybean (Hyten et al., 2007), potato (D'hoop et al., 2008), and tomato (Berloo et al., 2008).

In case of barley, a few studies have been published recently. For example, Kraakman et al. (2004) identified QTLs for yield and yield stability in a collection of modern spring barley cultivars using AFLP (amplified fragment length polymorphism) markers. Rostoks et al. (2006) used high-throughput SNP genotyping assays and demonstrated that the LD present in elite germplasm, after accounting for population structure, can be effectively exploited to map traits by using whole-genome association scans with several hundreds to thousands of markers. By using DArT (Diversity Array Technology) markers on a collection of 192 barley genotypes that represented landraces, old- and contemporary cultivars sampling key regions around the Mediterranean basin and Europe, Comadran et al. (2009) investigated patterns of genetic diversity and LD and found that the collection was a suitable resource for association mapping. By combining the DArT marker data with the yield data in mixed model analyses, Comadran et al. (2008, 2011) identified several QTLs for yield under drought conditions in Mediterranean basin. Another association study by using DArT and SNP markers identified several QTLs for spot blotch resistance in barley (Roy et al., 2010).

Similarly in another study using 1536 SNP arrays on 500 UK barley cultivars having a strong population structure, interesting marker-trait associations were assessed for some phenotypes by careful application of mixed model technology (Cockram et al., 2010). In another partial follow up study using a slightly extended set of genotypes (615 cultivars) and phenotypes (32 morphological and 10 agronomic traits), Wang et al. (2011) concentrated on comparing various statistical approaches for association mapping in barley. They showed the superiority of mixed model methodology for GWA analysis to assess marker-trait association for even complex traits in barley.

The present study was undertaken to find marker-trait associations for yield and agronomic traits under drought conditions by harnessing the potential of allelic diversity present in the germplasm collection of barley held in the ICARDA genebank through an LD-based GWA mapping approach. A set of 223 accessions sampled from ICARDA genebank was grown in two contrasting environments in Syria, namely in a favourable site (Tel Hadya) and a dry site (Breda). The genotyping of the present germplasm collection with SSR and SNP markers has been described in an earlier study (Varshney et al., 2010). The present paper describes the genotyping of the germplasm set with additional 710 DArT markers, and the statistical analysis of population structure, LD decay and marker-trait associations based on 816 markers including 710 DArTs, 61 SNPs and 45 SSRs by a mixed model that corrects for population structure.

2. Materials and methods

2.1. Plant material

A set of 223 barley (*Hordeum vulgare* L.) accessions representing landraces (122), cultivars (66) and wild genotypes (35) from 30 countries from four continents were used (Table S1). DNA was isolated from these genotypes as mentioned in Thiel et al. (2003).

2.2. Phenotypic data

The germplasm collection was planted along with four checks in the cropping season 2004–2005 at ICARDA's research stations located near Tel Hadya (36°01'N; 37°20'E, elevation 300 m asl, long term average annual rainfall = 338 mm) and near Breda (35°56'N; 37°10'E, elevation 354 m asl, long term average annual rainfall = 269 mm) in Syria. The experimental design was un-replicated augmented by four repeated checks randomly included; entries were fully randomized with different randomizations in the two sites. The checks (2 six-rowed and 2 two-rowed) were used to adjust for spatial variation. Plots were arranged in a row and column grid (10 rows, 28 columns). Trials were sown on plots of 3 m², 1.2 m wide (4 rows 30 cm apart), 2.5 m long and grown under rainfed conditions. Total rainfall was 303 mm in Tel Hadya and 268 mm in Breda. Although the total rainfall difference in the cropping season 2004–2005 was only 35 mm (303 and 268 mm) between Tel Hadya and Breda, both sites have shown good differentiation for drought adaptation over the years. Not only the total amount of rainfall but rainfall distribution and soil structure at both sides are important. While Tel Hadya has heavy clay soils with good water holding capacity, Breda has largely sandy soil with little water holding capacity. For this reason average rainfed barley yield in years with average rainfall are around 4.4 t/ha in Tel Hadya whereas it is 1.4 t/ha in Breda.

The following agronomic traits were recorded: growth vigour in Tel Hadya (gv.th) as a visual score at the 5–6 leaf stage, using a scale from 1 = good vigor to 5 = poor vigor; growth habit in Tel Hadya (gh.th) as a visual score at the 5–6 leaf stage, using a scale from 1 = erect to 5 = prostrate; days to heading in Tel Hadya (dh.th) as the number of days from emergence to awns appearance in 50% of the plants in a plot; plant height in Tel Hadya (ph.th) and Breda (ph.br) measured in cm from ground level to the base of the spike, at maturity; peduncle length in Tel Hadya (pl.th) and in Breda (pl.br) in cm from the last node to the bottom of the spike, at maturity; peduncle extrusion in Tel Hadya (pe.th) and Breda (pe.br) in cm from the ligule of the flag leaf to the bottom of the spike, at maturity (this measure can be negative in the case all or part of spike remains in the booth); spike length in Tel Hadya (sl.th) and Breda (sl.br) in cm excluding the awns; leaf colour in Tel Hadya (spad.th) and

Table 1

Description of the traits in the present study, in terms of their minimum, maximum, mean, standard deviation and coefficient of variation (standard deviation divided by mean) values. We also show the number of QTLs found for each trait. TH = Tel Hadya (wet), BR = Breda (dry).

Trait	Description	Unit/scale	Minimum	Maximum	Mean	St.Dev.	Coef.Var.	No. of QTLs alpha = 0.05)
gv.th	Growth vigor (TH)	1 = good; 5 = bad	0.9	5	2.3	1	0.43	2
gh.th	Growth habit (TH)	1 = erect; 5 = prostrate	1.2	5	2.5	1	0.39	3
dh.th	Days to heading (TH)		72.2	104	84.8	5.2	0.06	0
ph.th	Plant height at the base of the spike (TH)	cm	72.6	119	92.6	8.5	0.09	1
ph.br	Plant height at the base of the spike (BR)	cm	35.7	88	55.7	8.3	0.15	0
pl.th	Peduncle length (TH)	cm	14.5	46	27.9	6	0.21	1
pl.br	Peduncle length (BR)	cm	7.6	39	20.1	6.2	0.31	6
pe.th	Peduncle extrusion (TH)	cm	-8.6	25	6.2	5.4	0.88	3
pe.br	Peduncle extrusion (BR)	cm	-12.2	15	-1.7	5.5	3.27	6
sl.th	Spike length (TH)	cm	4.1	13	8	1.6	0.2	7
sl.br	Spike length (BR)	cm	5.1	12	8.2	1.3	0.16	1
spad.th	Leaf colour (TH)		34.3	53	42.9	3.5	0.08	2
spad.br	Leaf colour (BR)		38	58	46.5	3.8	0.08	2
by.th	Total above ground biomass yield (TH)	kg/ha	3402.6	14209	10355.2	2240.8	0.22	3
by.br	Total above ground biomass yield (BR)	kg/ha	1360.4	6433	4247.8	867.8	0.2	3
gy.th	Grain yield (TH)	kg/ha	363.5	7781	4419.4	1735.9	0.39	1
gy.br	Grain yield (BR)	kg/ha	342.1	2440	1436.4	511.1	0.36	3
tkw.th	Thousand kernel weight (TH)	g	26.4	62	42.9	6.2	0.15	7
tkw.br	Thousand kernel weight (BR)	g	23.7	48	36.5	5	0.14	8

Breda (spad.br) determined using a chlorophyll meter (SPAD-502, Minolta, Japan) in the middle of two random leaves (the average was used for the analysis) at full heading stage before any symptoms of senescence were visible, between 09:00 and 12:00 h (solar time); total above ground biomass in Tel Hadya (by.th) and Breda (by.br) in kg/ha, measured by hand harvesting each plot at maturity, before shattering of the brittle rachis entries; grain yield in Tel Hadya (gy.th) and Breda (gy.br) in kg/ha, measured after threshing the harvested sample; thousand kernel weight in Tel Hadya (tkw.th) and Breda (tkw.br) in g, measured as the average of three samples of 100 kernels per plot; the scored and measured traits are listed in Table 1. Individual trials were analyzed with a GENSTAT (Payne and Arnold, 2002) program for spatial analysis of unreplicated trials in which the response of the checks provides the basis for modelling the spatial variability in the field and to adjust the genotypes' performance. Best Linear Unbiased Predictors (BLUPs) for each one of the traits in each environment were estimated. The BLUPs for the two locations, Tel Hadya and Breda, formed the basis for marker-trait association analyses.

2.3. Genotyping and marker analysis

Three sets of markers were used for genotyping the germplasm collection. Genotyping data for 61 SNP (Kota et al., 2008) and 45 SSR (Varshney et al., 2007) markers have been reported in an earlier study (Varshney et al., 2010). In addition, the germplasm collection was genotyped with Diversity Array Technology (DArT) markers and marker data were generated for 710 DArT loci, which are analysed in the present study. Positions for SNP and SSR markers were obtained from an integrated map by Stein et al. (2007), while positions for the DArT markers came from Wenzl et al. (2006). The set of 710 DArT markers provided a good coverage of the seven linkage groups (Fig. S1). The DArT markers are dominant markers scored either '1' if a fragment is present or '0' if a fragment is absent. The present set of DArT markers contained each between 0 and 20% missing observations. In a pre-selection all unmapped markers were excluded. As a result, we ended up with 61 SNP, 45 SSR and 710 DArT markers with a known map position, in a total of 816 markers, which are analysed in the present study. The 61 SNP markers were co-dominantly scored with two alleles, and the 45 SSR markers were co-dominantly scored and had several alleles.

2.4. Diversity analysis and population structure

The genetic diversity structure of the entire collection was investigated in various ways, both by distance-based methods and by a model-based approach, using SNP, SSR and DArT markers. Prior to the analysis of diversity and of population structure, a selection of the complete set of 816 mapped markers was made to ensure that the genome was evenly covered and not overrepresented at some regions. Therefore, a subset of 212 markers was taken, corresponding to loci that were at least 5 cM apart and that covered the entire genome. The distance of 5 cM was chosen as to represent the distance beyond which linkage disequilibrium between loci could be considered to be unimportant. The resulting sample of markers was composed of 80% DArT, 10% SNP and 10% SSR markers. (The influence of the sampling procedure for the marker loci was investigated by repeating the procedure a few times and comparing the results of subsequent analyses. These analyses showed a high degree of agreement.)

The program *Dissimilarity Analysis and Representation for Windows*, also known as DARwin (Perrier et al., 2003) was run for the analysis of principal coordinates (PCO) and for clustering analysis by neighbor-joining (NJ) method. The subset of 212 markers described above was imported into DARwin and a distance matrix was calculated for the present collection of 223 accessions. The distances between accessions were scaled to lie between 0 and 1 and were calculated using the complements of simple matching coefficients. Markers were given equal weight.

The same marker set was used to investigate population structure by Bayesian model based clustering as implemented in the *Structure* software version 2.1 (Pritchard et al., 2000). The *Structure* algorithm constructs groups characterized by Hardy-Weinberg equilibrium and linkage equilibrium for variation at the loci within the groups, whereas LD is present between the groups (Pritchard et al., 2000; Falush et al., 2003). The program was run assuming a population admixture model with independent allele frequencies between subpopulations. The number of assumed subpopulations (K) was set to vary between 1 and 11, and for each value of K an independent MCMC (Markov Chain Monte Carlo) of 100,000 iterations was run, of which the first 10,000 were discarded as burn-in. The likelihood of the data for a given number of assumed subpopulations K is provided by the software, and the value of K with the highest likelihood can be interpreted to correspond to an estimate

for the underlying number of clusters. Within the admixture model we can obtain the membership probabilities of each genotype to each cluster.

2.5. Investigation of population structure by principal components analysis/Eigen analysis

Population structure can be investigated in various ways. In addition to PCO and *Structure*, we investigated population structure by the methodology as proposed by Patterson et al. (2006) that is based on an principal components analysis/eigen value decomposition of the normalized genotypic data matrix containing the marker genotype scores. Principal components are tested for significance and significant principal components (PCs) can be included as covariates in models for association analysis or tests for LD between markers. The principal components analysis/eigen value decomposition of the genotypic data matrix and the testing for significance of the principal components is implemented in the procedure EIGENANALYSIS, available in GenStat's QTL-library (GenStat 13th edition). For the present matrix of DArT markers the first 23 principal components (PC) were significant.

2.6. Analysis of LD between markers

LD between pairs of markers was studied for pairs of DArT markers. The approach used a logistic regression model (Hosmer and Lemeshow, 2000) in which, on a logistic scale, variation in a first marker, marker (1), was explained by the variation in a second marker, marker (2), after correction for population structure. The response, marker (1), is a binary vector of length 223, corresponding to marker scores: 1 if a fragment is present or 0 if a fragment is absent. The correction for population structure follows from the inclusion of the significant principal components of the genotypic data matrix (see above) as covariates. The predictor, marker (2), is a binary vector of marker scores, having the value 1 if the fragment is present and 0 otherwise. The association between a marker (1) and (2) can be assessed by inspection of the deviance difference (log-likelihood ratio) between a logistic regression model including population structure + predictor marker versus a similar model without the predictor marker. For all pairs of markers, the resulting *P*-values were saved and transformed to a $-\log_{10}(P)$ and then plotted against the corresponding genetic distances between the markers (Fig. 1, Fig. S2). This way of investigating LD is available in procedure QLDDECAY from GENSTAT's QTL-library, available in GENSTAT 13th edition (Payne and Arnold, 2002).

2.7. Marker-trait association analysis

Association analysis was performed on 816 DArT, SNP and SSR markers, after recoding each SSR marker as a binary vector for the presence or absence of its most frequent allele. The analysis was based on a mixed model that corrected for population structure, and was carried out in GENSTAT using restricted maximum likelihood (REML). The following mixed model was used for association analysis:

Phenotypic response = structure + marker + error.

The phenotypic response is a 223-length vector containing the BLUPs obtained either for the location Tel Hadya (wet site) or for Breda (dry site). This model corrects for population structure by including the significant principal components of the genotypic data matrix containing the DArT markers as covariates. The effects for those covariates were assumed to follow a normal distribution with a common variance. The markers to be tested for association with the phenotypic response were included as follows. For DArT

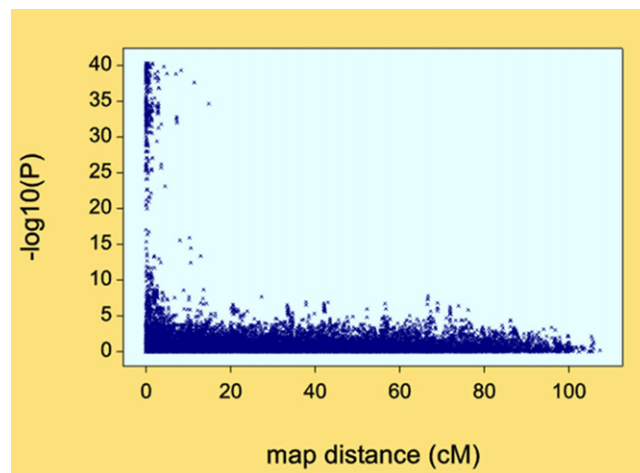


Fig. 1. Linkage disequilibrium (LD) between pairs of DArT markers, after correction for population structure, across all chromosomes. LD is expressed in terms of the $-\log_{10}(P)$ value of a deviance test comparing a logistic regression model explaining the variation in one marker by the variation in another marker after correction for population structure. The plateau in $-\log_{10}(P)$ values beyond 4–5 cM corresponds to an r^2 of about 0.1.

markers and most frequent SSR allele it was a vector having the value 1 if the band, or allele, was present and 0 otherwise. For SNP markers it was a vector coded 0 for (0, 0), 1 for (0, 1) and (1, 0), and 2 for (1, 1). The fixed marker terms were tested by a Wald statistic (Verbeke and Molenberghs, 2000; Payne et al., 2006), and corresponding effect, standard error of the effect, and *P*-value were saved. Proportions of explained variance for individual markers were estimated by the increments in R^2 statistic after fitting fixed individual markers in a model with fixed covariates for population structure.

2.8. Correction for multiple testing

A threshold for the *P*-values resulting from the association analysis was estimated by the method proposed by Li and Ji (2005) and Cheverud (2001). For this analysis we restricted the DArT markers to those containing less than 5% missing observations, which resulted in a subset of 216 DArT markers. With this subset we calculated the effective number of independent loci, and subsequently applied the Bonferroni correction based on the effective number of independent tests. For the subset of 216 DArT markers the effective number of independent tests (or loci) was estimated as 138, thus for a genome-wide error type I alpha = 0.05 we set for each test a threshold of $P = 0.05/138 = 0.00036$ or $-\log_{10}(P) = 3.44$. To arrive at this threshold, we used procedure QTHRESHOLD from GENSTAT's QTL-library, available at GENSTAT's 13th edition.

3. Results and discussion

3.1. Marker analysis

The germplasm collection was genotyped with 45 SSR and 61 SNP markers as mentioned in an earlier study (Varshney et al., 2010). With an objective to enhance marker data on this collection, DArT markers were used and genotyping data were obtained for 710 DArT marker loci in this study. The 710 DArT markers provided average genome coverage of 101 marker loci per linkage group (LG) with a minimum of 51 markers for LG 4H and a maximum of 129 for LG 2H, as shown in Table 2. Such more sparse coverage of LG 4H was observed in the DArT consensus map (Wenzl et al., 2006). As a result, like in the study of Comadran et al. (2009), several monomorphic blocks of DArT markers from LG 2H were observed.

Table 2
Details of the molecular markers (DArT, SNP and SSR) used in the study.

Chromosome	DArT			SNP			SSR		
	Number of markers	Average inter-marker distance (standard deviation)	Average PIC value (standard deviation)	Number of markers	Average inter-marker distance (standard deviation)	Average PIC value (standard deviation)	Number of markers	Average inter-marker distance (standard deviation)	Average PIC value (standard deviation)
1H	108	1.6 (2.7)	0.32 (0.05)	6	25.7 (23.8)	0.34 (0.08)	7	17.7 (13.9)	0.35 (0.03)
2H	129	1.4 (2.6)	0.32 (0.07)	8	16.0 (15.0)	0.28 (0.10)	5	34.4 (10.5)	0.37 (0.01)
3H	114	1.9 (3.9)	0.33 (0.05)	9	16.1 (28.2)	0.40 (0.17)	6	16.5 (15.7)	0.36 (0.03)
4H	51	3.4 (7.1)	0.31 (0.07)	9	10.2 (10.3)	0.41 (0.08)	7	21.5 (8.1)	0.36 (0.03)
5H	93	2.2 (3.6)	0.32 (0.06)	15	14.0 (14.4)	0.40 (0.15)	6	32.6 (17.0)	0.35 (0.04)
6H	92	1.9 (3.6)	0.33 (0.05)	8	17.9 (18.6)	0.33 (0.08)	7	15.5 (11.4)	0.34 (0.07)
7H	123	1.7 (3.3)	0.34 (0.04)	6	26.1 (15.5)	0.36 (0.12)	7	20.0 (12.3)	0.34 (0.04)

Originally a total of 1105 marker loci were assembled, including DArTs, SSRs and SNPs on the germplasm collection. Prior to our analysis we excluded the markers with unknown map positions, so that the mapping positions of all markers taken into the present analysis were known through published genetic maps for SNP and SSR (Stein et al., 2007) or DArT (Wenzl et al., 2006) markers. In terms of linkage groups, a higher number of marker loci belonged to LG 2H (142 marker loci) and LG 7H (136 marker loci), for which marker genotyping data were assembled, while LG 4H contributed to a small number (67) of marker loci, that were irregularly spaced as well (see standard deviation in Table 2).

Polymorphism information content (PIC) value indicates informativeness of a marker locus or marker system, with the minimum being 0 and the maximum being 0.50 for bi-allelic markers and 1 for multi-allelic markers. The PIC values of DArT markers varied from 0.13 to 0.38 with an average PIC value 0.33. For SNP markers, PIC values varied between 0.08 and 0.50 and were on average 0.35. In case of the most frequent SSR allele, PIC values ranged from 0.19 to 0.38 with an average of 0.35. It was interesting to note that DArT and SSR markers showed similar trends for PIC values across the different linkage groups (Table 2). However, SNP marker loci from LG 2H showed lower PIC values while the most frequent SSR allele at loci from LG 2H had high PIC values.

3.2. Population structure in the germplasm collection

The selection of the number of groups in the population was based on the likelihood of the data given the number of groups, K . K was varied between 1 and 11. Independent runs of *Structure* for each K value showed strong improvement up to $K=5$, and smaller improvements for larger K . Therefore, we chose $K=5$ as the number of clusters in the population. The composition of each cluster in terms of country of origin is shown in Table 3, and the trait means and standard deviations for each cluster are shown in Table S2. Cluster 1 (MEast) represents Middle Eastern and Asian landraces and contains mainly early genotypes (taking approximately 64 days to flower, see Table S2). For Cluster 1 (MEast), six-row types (39) were more abundant than two-row types (21). Cluster 5 (NAfr) is mainly composed of North African landraces and breeding lines that are predominantly of the six-row types (44) as compared to two-row types (9). This group also includes early genotypes (65 days to flower) that produce the heaviest grain ($tkw.th=46.1$ g, $tkw.br=38.7$ g) of the population. Cluster 3 (2Row) mostly includes two-rowed barley (36 two-rows versus 11 six-rows) and breeding lines, mainly from Syria. The genotypes in this cluster take an average of 68 days to flower. Cluster 4 (Wild) is exclusively composed of wild barley genotypes from Middle East and Asia, that, naturally, are all two-row types. These accessions are characterized by a much longer time to flowering (75 days in average), lower yield ($by.th=6.5$ t/ha, $by.br=3.2$ t/ha, $gy.th=1.0$ t/ha

and $gy.br=0.6$ t/ha), taller plants ($ph.th=97.6$ cm, $ph.br=65.6$ cm), and longer peduncles ($pl.th=35.8$ cm and $pl.br=29.4$ cm). Finally Cluster 2 (Rest) is a small heterogeneous group comprising landraces and breeding lines from Middle East, North Africa, Asia and Europe, with 14 two-row types and 12 six-row types. This group of early genotypes (65 days to flower) has the highest yield ($by.th=11.8$ t/ha, $by.br=4.6$ t/ha, $gy.th=5.5$ t/ha and $gy.br=1.5$ t/ha) and shorter plants ($ph.th=89.9$ cm, $ph.br=51.8$ cm) and peduncles ($pl.th=24.5$ cm and $pl.br=16.4$ cm). The most contrasting clusters in terms of trait values are the Wild group and the Rest group.

An analysis of principal coordinates (PCO) and of clustering analysis by neighbor-joining was performed using program DARwin. The two first dimensions shown in Fig. 2 explained 17% of the

Table 3
Composition of five sub populations as obtained by model based Bayesian clustering (*Structure*), in terms of entries belonging to particular countries of origin.

Country ^a	MEast (1)	NAfr (5)	2Row (3)	Wild (4)	Rest (2)
AFG	7			2	
AUS			1		2
AZE	2			1	1
CHN	3				
CYP		1			
DZA		8			1
EGY	1	8	3	1	1
ERI	2				
ETH	2		1		
FRA					1
GRC		1			
ICARDA Genebank	2		10		4
IND	2				
IRN	15	1	2	2	
IRQ	2	1		1	
JOR	1	1		12	1
LBN			1		
LBY		11	1		1
MAR		5	1		
OMN	3	3	1		
PAK	8			1	
PAL		1	1	3	
RUS					1
SAU	3				1
SYR	1	5	20	6	6
TJK		2			
TKM	3			6	1
TUN		4			2
TUR			3	1	2
UZB	1				1
YEM	2	1			

^a Standard code for country of origin, e.g. AFG=Afghanistan, AUS=Australia, AZE=Azerbaijan, CHN=China, CYP=Cyprus, DZA=Algeria, EGY=Egypt, ETH=Ethiopia, ERI=Eretria, GRC=Greece, FRA=France, IND=India, IRN=Iran, IRQ=Iraq, JOR=Jordan, LBN=Lebanon, LBY=Libya, MAR=Morocco, OMN=Oman, PAK=Pakistan, PAL=Palestine, SAU=Saudi Arabia, SYR=Syria, TJK=Tajikistan, TKM=Turkmenistan, TUR=Turkey, TUN=Tunisia, UZB=Uzbekistan, YEM=Yemen.

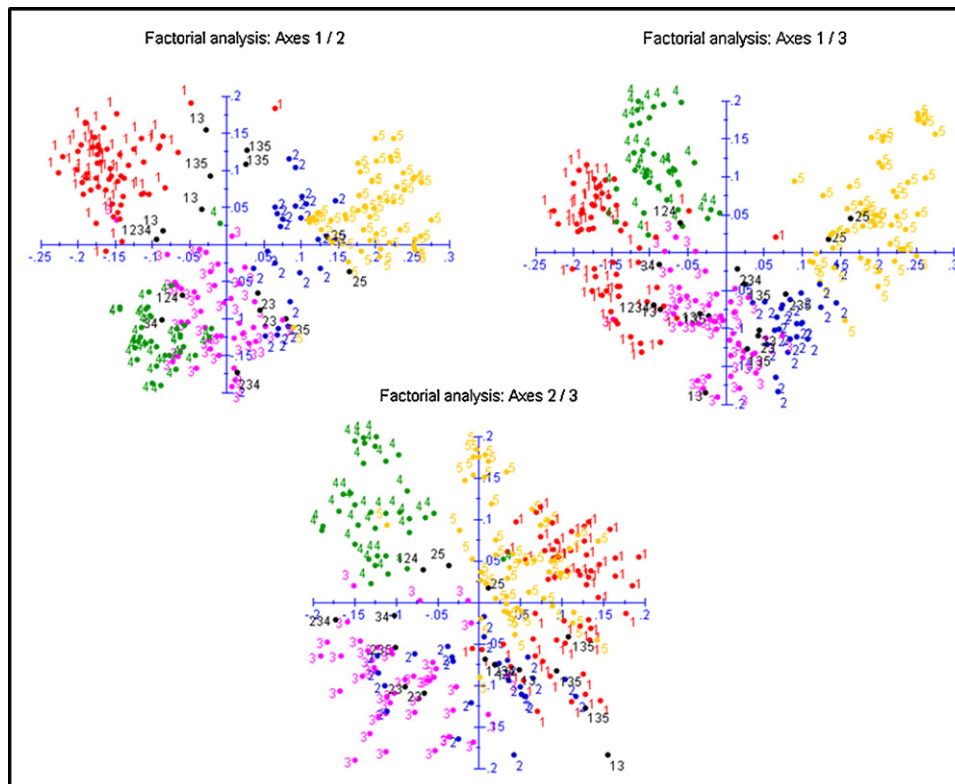


Fig. 2. Analysis of population structure of the germplasm collection used in the present study by principal coordinate (PCO) analysis. Each genotype was labeled and coloured according to its *Structure* group. Five sub groups were identified in the present population, namely Middle East (1), North Africa (5), Two Rows (3), Wild (4) and Rest (2). The first three coordinates, which represent 21% of the genotypic variation, have been shown.

variation present in the genetic distance calculated between genotypes. The five clusters classified by *Structure* can be identified in the PCO plot. Genotypes that are very admixed were labeled as a string containing the cluster numbers, for example a genotype classified as a mixture of clusters Rest (2) and NAfr (5) was labeled “25”. The first two PCO axes show three main groups, namely the MEast accessions, the NAfr accessions, and finally 2Row and Wild accessions (which are also all of the two-row type). In this two-dimensional display cluster Rest partly overlaps with 2Row and NAfr. There is in general good agreement between the *Structure* classification and the PCO visualization (Fig. 2). The population structure in our data can be ascribed to a combination of origin, domestication or breeding history, and structure, two-rows versus six-rows. These driving forces behind population structure in barley were also identified in the recent studies (Cockram et al., 2010; Hamblin et al., 2010).

3.3. Analysis of LD between markers

The analysis of LD decay, after correction for population structure groups (Fig. 1, Fig. S2) shows the genome-wide LD, with $-\log_{10}(P)$ values (y -axis) plotted against genetic distance in cM (x -axis). Fig. 1 shows that the strongest, and significant, LD is observed at very short distances, till 3 cM. This is close to the LD extent that has recently been reported (2.5–3.5 cM) by Comadran et al. (2009) and Zhang et al. (2009). Beyond 3 cM, LD becomes more or less constant at a value around a squared correlation of about 0.1, with some incidental jumps to 0.2. Around 10 cM and 15 cM, we observe a few exceptional cases of significant LD. LD varied slightly between chromosomes. In Fig. 1 we pooled the LD values across the whole genome. LD at 10 cM and beyond was observed on chromosomes

2H, 6H and 7H (data not shown). After 15 cM there was hardly any LD between loci.

3.4. Association analysis of quantitative traits

Grain yield had an average of 1.4 t/ha in the dry site (Breda) and an average of 4.4 t/ha in the favourable or wet site (Tel Hadya). These values are similar to those observed in Breda and in Tel Hadya for a different germplasm collection (Pswarayi et al., 2008; Comadran et al., 2008). Also the average biomass yield varied between 4.2 t/ha (Breda) and 10.4 t/ha (Tel Hadya), and the average thousand kernel weight (*tkw*) was 36.5 g in Breda versus 42.9 g in Tel Hadya.

The total number of markers significantly associated with each trait was between 0 and 8, as shown in Table 1. These numbers are very comparable to the study of Cockram et al. (2010), who found between one and three significant QTLs for their set of 15 highly heritable traits (mean $h^2 = 0.58$), whereas for 17 lowly heritable traits (mean $h^2 = 0.17$) no QTLs were found. In Table 4 we show the most highly associated DArT marker with each phenotype, its location (chromosome and position in cM), the effect of a double allele substitution: band present versus band absent in the case of DArT markers, genotype (1, 1) versus genotype (0, 0) in the case of SNP markers, and most frequent allele present versus absent in the case of SSR markers, standard error of the effect, ratio of the effect by its standard error, and percentage of the trait's variance explained by the marker, the latter assessed by a fixed model as described in Section 2. A marker explained at most 6.7% of the variation. The location of all significant QTLs by trait is shown in Fig. 3, and detailed information on all significant QTLs is provided in Table S3.

Table 4Details on the most significant markers associated with the targeted traits using a significance level corresponding to $\alpha=0.05$.

Trait	Marker name	Linkage group	Position (cM)	Frequency of allele 1	% of Missing value	$-\log_{10}(P)$	Effect of double allele or genotype (allele = 1)	Standard error of double allele effect	Ratio of double allele effect and standard error	Double allele substitution effect (% trait range)	% Phenotypic variance explained
<i>gv.th</i>	bPb.8907	3	205	47	9	3.9	-0.5	0.1	4.0	5.9	3.0
<i>gh.th</i>	GBM1363	5	144	58	0	7.2	-0.7	0.1	5.4	9.2	3.6
<i>ph.th</i>	bPb.2379	7	127	37	8	3.9	-5.1	1.3	3.8	5.5	1.1
<i>pl.th</i>	bPb.5355	3	222	28	6	4.4	-3.7	0.9	4.1	5.8	3.6
<i>pl.br</i>	bPb.9746	3	97	43	7	6.0	-4.9	1.0	4.9	7.8	3.7
<i>pe.th</i>	bPb.2580	5	151	87	1	3.9	-4.2	1.1	3.8	6.3	3.2
<i>pe.br</i>	bPb.0994	2	116	25	5	6.3	-4.5	0.9	5.1	8.2	6.7
<i>sl.th</i>	bPb.2406	3	144	65	6	5.9	-1.3	0.3	4.8	7.2	2.8
<i>sl.br</i>	GBS0143	2	88	17	2	4.4	-2.0	0.3	8.2	14.9	4.1
<i>spad.th</i>	bPb.6477	6	180	37	4	3.7	-2.2	0.6	3.7	5.9	5.0
<i>spad.br</i>	bPb.2325	5	139	21	6	4.1	-2.6	0.7	3.9	6.5	3.5
<i>by.th</i>	bPb.4616	3	153	67	3	4.3	1382	343	4.0	6.4	2.3
<i>by.br</i>	bPb.4616	3	153	67	3	3.8	565	150	3.8	5.6	2.9
<i>gy.th</i>	bPb.4616	3	153	67	3	4.8	929	216	4.3	6.3	1.8
<i>gy.br</i>	GBS0469	1	133	65	3	3.7	-143	39	3.7	3.4	3.8
<i>tkw.th</i>	bPb.4269	6	134	74	5	4.3	-4.4	1.1	4.1	6.2	0.8
<i>tkw.br</i>	bPb.4875	2	46	40	8	8.5	-3.9	0.7	6.0	8.1	0.6

Three significant associations were found for biomass yield in Tel Hadya (*by.th*) and also three in Breda (*by.br*). The first QTL is located in chromosome 3H, at 153 cM, and explains 2.3% and 2.9% of the variation observed in *by.th* and *by.br* respectively. Another four QTLs were found on LG 5H, at 95 cM and at 156 cM, explaining respectively 1.6% and 2.4% of the variation for *by.th*, and at 144 cM and 147 cM for *by.br* explaining 1.4% and 1.2% of the variation. The latter very likely represent the same locus.

A single significant association was found for grain yield in Tel Hadya (*gy.th*), located on LG 3H at 153 cM. This QTL explained 1.8% of the variation in *gy.th*, and it was the same QTL found for *by.th* and *by.br*. QTLs for yield on the long arm of LG 3H have been attributed to the presence of the *denso/sdw1* locus (short straw) commonly available in germplasm from Near East and North Africa. The role of barley *sdw1/denso* gene to control plant height as well as yield and quality is already well documented (Jia et al., 2011). Three significant associations were found for grain yield in Breda (*gy.br*). A locus explaining around 3.8% of the phenotypic variation was found on LG 1H at 133 cM, and two other associations were found on LG 5H, at 122 cM and 144 cM, explaining 1.9% and 1.8% of the trait's variation. QTLs for yield have been previously reported on LG 1H and 5H (Baum et al., 2003; Long et al., 2003; Pillen et al., 2003; Talame et al., 2004; Li et al., 2005, 2006; von Korff et al., 2006). Some QTLs reported in our study are also in the close proximity of *Vrn-H1* and this region has been shown to be associated with QTLs

for grain yield in earlier studies (Pillen et al., 2003, 2004; von Korff et al., 2006). Also, in a recent study, one of these two QTLs on LG 5H coincided with a locus reported in another work (Comadran et al., 2008).

For thousand kernel weight (*tkw*) a total of 15 associations were found significant, corresponding to just 7 distinct loci. For *tkw.th*, we found 7 QTLs representing 4 distinct loci, namely on LG 2H at 46 cM (<1% explained variance), on LG 6H at 33 cM (3% explained variance) and at 134 cM (<1% explained variance) and on LG 7H at 153 cM (<1% explained variance). In Breda, we had 8 significant results for *tkw.br*, representing 5 loci, namely a locus on LG 2H at 46 cM (<1% explained variance), another on LG 3H at 46 cM (2.3% explained variance), on LG 5H at 43 cM (2.8% explained variance), on LG 6H at 33 cM (1% explained variance, shared with *tkw.th*), and on LG 7H at 131 cM (<1% explained variance). Some of these QTLs were found in the same region on LGs 6H and 7H that harboured QTLs for yield, relative water content, heading date, etc. in earlier studies (Pillen et al., 2003; Talame et al., 2004; Li et al., 2006; von Korff et al., 2006). It is also interesting to note that most of the loci identified for the yield traits were also found significant for other traits, such as growth habit (*gh*) and growth vigour (*gv*), as shown in Fig. 3.

It is important to note that in this study the explained variation, R^2 , for all of the identified QTLs was rather low, ranging from 0.1% (*days to flower*) to 6.7% (*pe.br*, peduncle extrusion). Also,

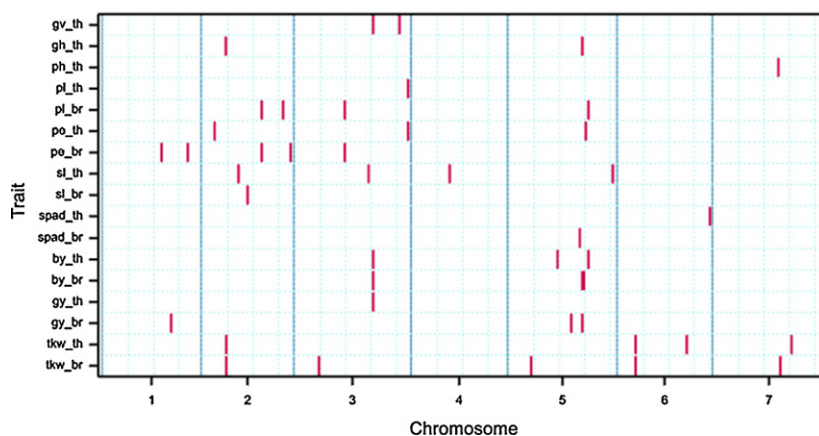


Fig. 3. Location of QTLs for individual trait, identified by association mapping approach, on the seven chromosomes of barley. Details of these QTLs have been given in Table 4.

expressing the effects of QTL alleles in alternative ways, by the effect of a double allele substitution as a percentage of the range, emphasized the impression that QTL effects were small to modest. However, the numbers for the explained variation by QTLs as observed in the current study are not uncommon in association studies in cereals. For example, Maccaferri et al. (2011), studying a diverse durum wheat panel across 15 environments with varying water availability, found QTLs explaining between 5 and 10% for highly heritable traits and only 2.5–4.2% for yield. In the two recent studies on GWA analysis in barley (Cockram et al., 2010; Wang et al., 2011), marker-trait association contributing larger phenotypic variation was shown for the traits that are highly heritable, and almost qualitative traits. In contrast, QTLs explaining >10% variances seem hard to be identified for the complex traits like drought tolerance in GWA analyses in barley. QTLs identified in the present study were not only small, but also very little consistent across environments. The suggestion is made by Maccaferri et al. (2011) that similar yield levels under stress conditions can be obtained via very different genetic pathways, thereby making it hard to detect consistent QTLs with clear effects. While inspecting the pattern of trait correlations for Tel Hadya and Breda (Table S4 and Figs. S3 and S4), we observe that in both locations grain yield was strongly correlated with biomass and weakly correlated with thousand kernel weight. Also, in both locations peduncle height, peduncle length and peduncle extrusion were strongly correlated with each other. In Tel Hadya the yield-biomass-thousand kernel weight group of traits had a strong negative correlation with the peduncle traits, whereas in Breda these groups of traits were unrelated. Therefore some differentiation seems to occur between the two sites, but it is unclear whether this differentiation is responsible for the QTLs being different between the locations or whether the latter is just a consequence of the fact that the probability for small QTLs to be consistently picked up is low. In this respect, our results are very comparable to earlier studies (Comadran et al., 2008, 2011) who evaluated a widely diverse barley association panel for 2 years at seven locations, at each location a 'dry' and a 'wet' trial. QTL effects for yield were somewhere between 5 and 10% and no QTLs could be identified that were specific for either dry or wet conditions.

In addition to the extreme environmental conditions triggering different genetic pathways, another explanation for the low explained variation by QTLs in wide cereal association panels is that variation for yield and yield components is strongly conditioned by earliness. If the genotype does not fit in the growing window, it will produce little or no yield. Population structure typically coincides strongly with adaptation, origin, breeding program, growth habit (winter versus spring types), besides the two-row versus six-row contrast (Cockram et al., 2010; Hamblin et al., 2010). Also in our case, population structure was driven by a combination of origin, coinciding with adaptation and earliness, and two-rows versus six-rows. When inspecting the amount of variation between and within groups (Table S2), it can be expected that correction for structure and kinship will remove a large part of the signal that may lead to the discovery of QTLs. Hamblin et al. (2010) encountered strong population structure in a large panel of cultivated barleys together with inconsistency of correlation of marker alleles across structure groups. Therefore some concerns were raised about the feasibility of association mapping in a strongly heterogeneous set of structure groups and they proposed to focus on one or a few large groups that each by themselves would be large enough to perform association mapping with sufficient power. Similar recommendations were made by Waugh et al. (2011). On the other hand, two other studies (Cockram et al., 2010; Wang et al., 2011) showed optimism and claimed that careful use of mixed model methodology will allow to perform association mapping for complex traits even in the presence of strong population structure. However QTLs identified in their studies were for high heritability traits.

QTL mapping studies in biparental populations show larger QTL effects and explained variation. This may look somewhat surprising, but it is not unexpected. The QTL mapping studies in biparental populations involve crosses with relatively small population sizes that are likely to overestimate the effects of QTLs (Beavis, 1998; Melchinger et al., 1998). Secondly, as many QTLs were identified in this study, the presence of tolerance alleles at more than one QTL in any one germplasm accession could affect the estimate of the QTL effects. As per Roy et al. (2010), if most of the QTLs identified had a relatively large effect and most accessions carried the tolerance allele at more than one of these loci, the R^2 for any one QTL would be lower.

In summary, this study together with some other studies undertaken in barley (Kraakman et al., 2004; Rostoks et al., 2006; Comadran et al., 2008, 2009, 2011; Roy et al., 2010; Hamblin et al., 2010) raises the concerns on the usefulness of GWA approach to identify the major QTLs for deployment in molecular breeding strategies, although a few studies still show optimism (Cockram et al., 2010; Wang et al., 2011). Surely, at the positive side, genome wide genotyping scans have become possible due to availability of large scale DaT (Wenzl et al., 2006) and SNP (Close et al., 2009) marker systems. This study certainly provides genetic information which merits further attention and subsequent confirmation studies. For instance, with some effort a set of drought tolerant accessions may be identified that carries tolerance alleles at different QTLs. Such accessions can be used as donor parents in marker-assisted backcrossing (MABC) programme to transfer the tolerant alleles in elite barley lines. The near-isogenic lines generated that way can be characterized further to confirm the QTLs, provide more precise estimates of allele effects and assess whether individual QTL are equally effective to drought stress (Roy et al., 2010). It is also important to mention here that the identified QTLs may not be suitable for direct application in molecular breeding for drought tolerance. Some follow up studies including complex crosses like multi-parent advanced generation inter cross (MAGIC) population with well chosen parents on the basis of results obtained in this study are required to assess the allele effects precisely and identify suitable markers to efficiently pyramid multiple QTLs into elite cultivars through MABC or marker-assisted recurrent selection (MARS) programme for developing more stable drought tolerant barley varieties.

Acknowledgements

Thanks are due to Ms. U. Beier for her excellent technical assistance. This research work was financially supported by a GTZ project (no. 2002.7860.6-001.00 and contract no. 1060503) sponsored by BMZ (Bundesministerium für Wirtschaftliche Zusammenarbeit und Entwicklung), Germany. Work by M.J. Paulo and F.A. van Eeuwijk was (co)financed by the Centre for BioSystems Genomics (CBG) which is part of the Netherlands Genomics Initiative/Netherlands Organisation for Scientific Research, and by the Generation Challenge Program/IBP project 2.2.1.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.fcr.2011.10.008.

References

- Agrama, H.A., Eizenga, G.C., Yan, W., 2007. Association mapping of yield and its components in rice cultivars. *Mol. Breed.* 19, 341–356.
- Amaral, A.J., Megens, H.J., Crooijmans, R.P.M.A., Heuven, H.C.M., Groenen, M.A.M., 2008. Linkage equilibrium decay and haplotype block structure in the pig. *Genetics* 179, 569–579.

- Andreescu, C., Avendano, S., Brown, S.R., Hassen, A., Lamont, S.J., Dekkers, J.C.M., 2007. Linkage disequilibrium in related breeding lines of chickens. *Genetics* 177, 2161–2169.
- Barnaud, A., Lacombe, T., Doligez, A., 2006. Linkage disequilibrium in cultivated grapevine (*Vitis vinifera* L.). *Theor. Appl. Genet.* 112, 708–716.
- Baum, M., Grando, S., Backes, G., Jahoor, A., Sabbagh, A., Ceccarelli, S., 2003. QTLs for agronomic traits in the Mediterranean environment identified in recombinant inbred lines of the cross 'Arta' × *H. spontaneum* 41-1. *Theor. Appl. Genet.* 107, 1215–1225.
- Baum, M., von Korff, M., Guo, P., Lakew, B., Hamwieh, A., Lababidi, S., Udupa, S.M., Sayed, H., Choumane, W., Grando, S., Ceccarelli, S., 2007. Molecular approaches and breeding strategies for drought tolerance in barley. In: Varshney, R.K., Tuberosa, R. (Eds.), *Genomics-Assisted Crop Improvement Genomics Applications in Crops*, 2. Published by Springer, The Netherlands, pp. 51–80.
- Beavis, W.B., 1998. QTL analyses: power, precision and accuracy. In: Patterson, A.H. (Ed.), *Molecular Dissection of Complex Traits*. CRC Press, Boca Raton.
- Berloo, R., Zhu, A., Ursem, R., Verbakel, H., Gort, G., Eeuwijk, F.A., 2008. Diversity and linkage disequilibrium analysis within a selected set of cultivated tomatoes. *Theor. Appl. Genet.* 117, 89–101.
- Bresheghello, F., Sorrells, M.E., 2006. Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. *Genetics* 172, 1165–1177.
- Carlson, C.S., Eberle, M.A., Kruglyak, L., Nickerson, D.A., 2004. Mapping complex disease loci in whole-genome association studies. *Nature* 429, 446–452.
- Ceccarelli, S., Grando, S., 1996. Drought as a challenge for the plant breeder. *Plant Growth Regul.* 20, 149–155.
- Ceccarelli, S., 2010. Drought and Drought Resistance. *Encyclopedia of Biotechnology in Agriculture and Food*, vol. 1, pp. 205–207.
- Charlier, C., Coppieters, W., Rollin, F., Desmecht, D., Agerholm, J.S., Cambisano, N., Carta, E., Dardano, S., Dive, M., Fasquelle, C., Frennet, J.C., Hanset, R., Hubin, X., Jorgensen, C., Karim, L., Kent, M., Harvey, K., Pearce, B.R., Simon, P., Tama, N., Nie, H., Vandeputte, S., Lien, S., Longeri, M., Fredholm, M., Harvey, R.J., Georges, M., 2008. Highly effective SNP-based association mapping and management of recessive defects in livestock. *Nat. Genet.* 40, 449–454.
- Cheverud, J.M., 2001. A simple correction for multiple comparisons in interval mapping genome scans. *Heredity* 87, 52–58.
- Close, T.J., Bhat, P.R., Lnard, S., Wu, Y., Rostoks, N., Ramsay, L., Druka, A., Stein, N., Svensson, J.T., Wanamaker, S., Bozdog, S., Roose, M.L., Moscou, M.J., Chao, S., Varshney, R.K., Szucs, P., Sato, K., Hays, P.M., Mathews, D.E., Kleinhafs, A., Muehlbauer, G.T., DeYong, J., Marshal, D.F., Modishetty, K., Fenton, R.D., Condamine, P., Graner, A., Waugh, R., 2009. Development and implementation of high-throughput SNP genotyping in barley. *BMC Genomics* 10, 582.
- Comadran, J., Russell, J.R., van Eeuwijk, F.A., Ceccarelli, S., Grando, S., Baum, M., Stanca, A.M., Pecchioni, N., Mastrangelo, A.M., Akar, T., Al-Yassin, A., Benbelkacem, A., Choumane, W., Ouabbou, H., Dahan, R., Bort, J., Araus, J.L., Pswarayi, A., Romagosa, I., Hackett, C.A., Thomas, W.T., 2008. Mapping adaptation of barley to droughted environments. *Euphytica* 161, 35–45.
- Comadran, J., Thomas, W.T., van Eeuwijk, F.A., Ceccarelli, S., Grando, S., Stanca, A.M., Pecchioni, N., Akar, T., Al-Yassin, A., Benbelkacem, A., Ouabbou, H., Bort, J., Romagosa, I., Hackett, C.A., Russell, J.R., 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.
- Comadran, J., Russell, J.R., Booth, A., Pswarayi, A., Ceccarelli, S., Grando, S., Stanca, A.M., Pecchioni, N., Akar, T., Al-Yassin, A., Benbelkacem, A., Ouabbou, H., Bort, J., van Eeuwijk, F.A., Thomas, W.T., Romagosa, I., 2011. Mixed model association scans of multi-environmental trial data reveal major loci controlling yield and yield related traits in *Hordeum vulgare* in Mediterranean environments. *Theor. Appl. Genet.* 122, 1363–1373.
- Cockram, J., White, J., Zuluaga, D.L., Smith, D., Comadran, J., Macaulay, M., Luo, Z., Kearsey, M.J., Werner, P., Harrap, D., Tapsell, C., Liu, H., Hedley, P.E., Stein, N., Schulte, D., Steuernagel, B., Marshall, D.F., Thomas, W.T.B., Ramsay, L., Mackay, I., Balding, D.J., The AGOUEB, Consortium, Waugh, R., O'Sullivan, D.M., 2010. Genome-wide association mapping to candidate polymorphism resolution in the unsequenced barley genome. *Proc. Natl. Acad. Sci. U.S.A.* 107, 21611–21616.
- D'hoop, B.B., Paulo, M.J., Mank, R.A., van Eck, H.J., van Eeuwijk, F.A., 2008. Association mapping of quality traits in potato (*Solanum tuberosum* L.). *Euphytica* 161, 47–60.
- Ellis, R.P., Forster, B.P., Gordon, D.C., Handley, L.L., Keith, R.P., Lawrence, P., Meyer, R., Powell, W., Robinson, D., Scrimgeour, C.M., Young, G., Thomas, W.T.B., 2002. Phenotype/genotype associations for yield and salt tolerance in a barley mapping population segregating for two dwarfing genes. *J. Exp. Bot.* 53, 1163–1176.
- Ersoz, E.S., Yu, J., Buckler, E.S., 2007. Applications of linkage disequilibrium and association mapping in crop plants. In: Varshney, R.K., Tuberosa, R. (Eds.), *Genomics Assisted Crop Improvement*, vol. 1. *Genomics Approaches and Platforms*. Springer, Netherlands, pp. 97–119.
- Ersoz, E.S., Yu, J., Buckler, E.S., 2009. Applications of linkage disequilibrium and association mapping in maize. In: Kriz, A.L., Larkins, B.A. (Eds.), *Molecular Genetic Approaches to Maize Improvement*, vol. 3. *Biotechnology in Agriculture and Forestry*. Springer, Berlin/Heidelberg, pp. 173–195.
- Falush, D., Stephens, M., Pritchard, J.K., 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164, 1567–1587.
- Farnir, F., Coppieters, W., Arranz, J.J., Berzi, P., Cambisano, N., Grisart, B., Karim, L., Marcq, F., Moreau, L., Mni, M., Nezer, C., Simon, P., Vanmanshoven, P., Wagenaar, D., Georges, M., 2000. Extensive genome-wide linkage disequilibrium in cattle. *Genome Res.* 10, 220–227.
- Forster, B., Ellis, R., Moir, J., Talame, V., Sanguineti, M., Tuberosa, R., This, D., Teulat-Merah, B., Ahmed, I., Mariy, S., Bahri, H., El-Ouahabi, M., Zoumarou-Wallis, N., El-Fellah, M., Salem, M., 2004. Genotype and phenotype associations with drought tolerance in barley tested in North Africa. *Ann. Appl. Biol.* 144, 157–168.
- Garris, A.J., McCouch, S.R., Kresovich, S., 2003. Population structure and its effect on haplotype diversity and linkage disequilibrium surrounding the xa5 locus of rice (*Oryza sativa* L.). *Genetics* 165, 759–769.
- Hambli, M.T., Close, T.J., Bhat, P.R., Chao, S., Kling, J.G., Abraham, K.J., Blake, T., Brooks, W.S., Cooper, B., Griffey, C.A., Hayes, P.M., Hole, D.J., Horsley, R.D., Obert, D.E., Smith, K.P., Ullrich, S.E., Muehlbauer, G.J., Jannink, J.L., 2010. Population structure and linkage disequilibrium in U.S. barley germplasm: implications for association mapping. *Crop Sci.* 50, 556–566.
- Hosmer, D.W., Lemeshow, S., 2000. *Applied Logistic Regression*, 2nd ed. John Wiley & Sons, New York.
- Hyten, D.L., Choi, I.Y., Song, Q., Shoemaker, R.C., Nelson, R.L., Costa, J.M., Specht, J.E., Cregan, P.B., 2007. Highly variable patterns of linkage disequilibrium in multiple soybean populations. *Genetics* 175, 1937–1944.
- Jia, Q., Zhang, X.Q., Westcott, S., Broughton, S., Cakir, M., Yang, J., Lance, R., Li, C., 2011. Expression level of a gibberellin 20-oxidase gene is associated with multiple agronomic and quality traits in barley. *Theor. Appl. Genet.* 122, 1451–1460.
- Jorde, J.B., 2000. Linkage disequilibrium and the search for complex disease gene. *Genome Res.* 10, 1435–1444.
- Koning, D.J., Cabrera, C.P., Haley, C.S., 2007. Genetical genomics: combining gene expression with marker genotypes in poultry. *Poult. Sci.* 86, 1501–1509.
- Kota, R., Varshney, R.K., Prasad, M., Zhang, H., Stein, N., Graner, A., 2008. An EST-derived single nucleotide polymorphism markers for assembling genetic and physical maps of the barley genome. *Funct. Int. Genomics* 8, 223–233.
- Kraakman, A.T.W., Niks, R.E., van den Berg, P.M.M.M., Stam, P., van Eeuwijk, F.A., 2004. Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. *Genetics* 168, 435–446.
- Lander, E.S., Schork, N.J., 1994. Genetic dissection of complex traits. *Science* 265, 2037–2048.
- Li, J., Ji, L., 2005. Adjusting multiple testing in multi locus analyses using the eigen values of a correlation matrix. *Heredity* 95, 221–227.
- Li, J., Huang, X.Q., Heinrichs, F., Ganai, M.W., Roeder, M.S., 2005. Analysis of QTLs for yield, yield components and malting quality in a BC₃-DH population of spring barley. *Theor. Appl. Genet.* 110, 356–363.
- Li, J., Huang, X.Q., Heinrichs, F., Ganai, M.W., Roeder, M.S., 2006. Analysis of QTLs for yield components, agronomic traits and disease resistance in an advanced backcross population of spring barley. *Genome* 49, 454–466.
- Long, N.R., Jeffries, S.P., Warner, P., Karakousis, A., Kretschmer, J.M., Hunt, C., Lim, P., Eckermann, P.J., Barr, A., 2003. Mapping and QTLs analysis of barley population Mundah × Keel. *Aust. J. Agric. Res.* 54, 1163–1171.
- Maccafferri, M., Sanguineti, M.C., Demontis, A., El-Ahmed, A., Garcia del Moral, L., Maalouf, F., Nachit, M., Nserallah, N., Ouabbou, H., Rhouma, S., Royo, C., Villegas, D., Tuberosa, R., 2011. Association mapping in durum wheat grown across a broad range of water regimes. *J. Exp. Bot.* 62, 409–438.
- Malosetti, M., Linden, C.G., Vosman, B., Eeuwijk, F.A., 2007. A mixed model approach to association mapping using pedigree information with an illustration to resistance for *Phytophthora infestans* in potato. *Genetics* 175, 879–889.
- McRae, F., McEwan, J.C., Dodds, K.G., Wilson, T., Crawford, A.M., Slate, J., 2002. Linkage disequilibrium in domestic sheep. *Genetics* 160, 1113–1122.
- Meadows, J.R.S., Kijas, J.W., 2009. Re-sequencing regions of the ovine Y chromosome in domestic and wild sheep reveals novel paternal haplotypes. *Anim. Genet.* 40, 119–123.
- Melchinger, A.E., Utz, H.F., Schoen, C.C., 1998. Quantitative traits locus (QTL) mapping using different testers and independent population samples in maize reveals low power of QTL detection and large bias in estimates of QTL effects. *Genetics* 149, 383–403.
- Neumann, K., Kobyljski, B., Dencic, S., Varshney, R.K., Boerner, A., 2010. Genome-wide association mapping: a case study in bread wheat (*T. aestivum* L.). *Mol. Breed.* 27, 37–58.
- Nsengimana, J., Baretta, P., Haley, C.S., Visscher, P.M., 2004. Linkage disequilibrium in the domesticated pig. *Genetics* 166, 1395–1404.
- Palasia, K.A., Morgante, M., Williams, M., Rafalskia, A., 2003. Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci. *Plant Cell* 15, 1795–1806.
- Patterson, N., Price, A.L., Reich, D., 2006. Population structure and EIGENANALYSIS. *PLoS Genet.* 2, 2074–2093.
- Payne, R.W., Arnold, G.M., 2002. *GenStat Release 6.1. Reference Manual – Part 3: Procedure Library PL14. VSN International, Oxford, UK.*
- Payne, R.W., Harding, S.A., Murray, D.A., Soutar, D.M., Baird, D.B., Welham, S.J., Kane, A.F., Gilmour, A.R., Thompson, R., Webster, R., Tunnicliffe Wilson, G., 2006. *The Guide to GenStat Release 9, Part 2: Statistics*. VSN International, Hemel Hempstead.
- Perrier, X., Flori, A., Bonnot, F., 2003. Data analysis methods. In: Hamon, P., Seguin, M., Perrier, X., Glaszmann J.C. (Eds.), *Genetic Diversity of Cultivated Tropical Plants*. Enfield, Science Publishers, Montpellier, pp. 43–76.
- Pillen, K., Zacharias, A., Léon, J., 2003. Advanced backcross QTL analysis in barley (*H. vulgare* L.). *Theor. Appl. Genet.* 107, 340–352.
- Pillen, K., Zacharias, A., Léon, J., 2004. Comparative AB-QTL analysis in barley using a single exotic donor of *Hordeum vulgare* ssp. *spontaneum*. *Theor. Appl. Genet.* 108, 1591–1601.
- Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959.

- Pswarayi, A., van Eeuwijk, F.A., Ceccarelli, S., Grando, S., Comadran, J., Russell, J.R., Pecchioni, N., Tondelli, A., Akar, T., Al-Yassin, A., Benbelkacem, A., Ouabbou, H., Thomas, W.T., 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.
- Rostoks, N., Ramsay, L., MacKenzie, K., Cardle, L., Bhat, P.R., Roose, M.L., Svensson, J.T., Stein, N., Varshney, R.K., Marshall, D.F., Graner, A., Close, T.J., Waugh, R., 2006. Recent history of artificial outcrossing facilitates whole-genome association mapping in elite inbred crop varieties. *Proc. Natl. Acad. Sci. U.S.A.* 103, 18656–18661.
- Roy, J.K., Smith, K.P., Muehlbauer, G.J., Chao, S., Close, T.J., Steffenson, B.J., 2010. Association mapping of spot blotch resistance in wild barley. *Mol. Breed.* 26, 243–256.
- Sköt, L.L., 2005. An association mapping approach to identify flowering time genes in natural populations of *Lolium perenne* (L.). *Mol. Breed.* 15, 233–245.
- Stein, N., Prasad, M., Scholz, U., Thiel, T., Zhang, H., Wolf, M., Kota, R., Varshney, R.K., Perovic, D., Grosse, I., Graner, A., 2007. A 1,000-loci transcript map of the barley genome: new anchoring points for integrative grass genomics. *Theor. Appl. Genet.* 114, 823–839.
- Talame, V., Sanguineti, M.C., Chiapparino, E., Bahri, H., Ben Salem, M., Forster, B.P., Ellis, R.P., Rhouma, S., Zoumarou, W., Waugh, R., Tuberosa, R., 2004. Identification of *Hordeum spontaneum* QTL alleles improving field performance of barley grown under rainfed conditions. *Ann. Appl. Biol.* 144, 309–319.
- Teulat, B., Merah, O., Souyris, I., This, D., 2001. QTLs for agronomic traits from a Mediterranean barley progeny grown in several environments. *Theor. Appl. Genet.* 103, 774–787.
- Teulat, B., This, D., Khairallah, M., Borries, C., Ragot, C., Sourdille, P., Leroy, P., Monneveux, P., Charrier, A., 1998. Several QTLs involved in osmotic adjustment trait variation in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.* 96, 688–698.
- Teulat, B., Zoumarou-Wallis, N., Rotter, B., Ben Salem, M., Bahri, H., This, D., 2003. QTL for relative water content in field-grown barley and their stability across Mediterranean environments. *Theor. Appl. Genet.* 108, 181–188.
- Thiel, T., Michalek, W., Varshney, R.K., Graner, A., 2003. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.* 106, 411–422.
- Tuberosa, R., Salvi, S., 2004. QTLs and genes for tolerance to abiotic stress in cereals. In: Varshney, R.K., Gupta, P.K. (Eds.), *Cereal Genomic*. The Netherlands Kluwer Academic Publishers, Dordrecht, pp. 253–315.
- von Korff, M., Wang, H., Léon, J., Pillen, K., 2006. AB-QTL analysis in spring barley: II Detection of favourable exotic alleles for agronomic traits introgressed from wild barley (*H. vulgare* ssp. *spontaneum*). *Theor. Appl. Genet.* 112, 1221–1231.
- von Korff, M., Grando, S., Del Greco, A., This, D., Baum, M., Ceccarelli, S., 2008. Quantitative trait loci (QTL) associated with adaptation to Mediterranean dryland conditions in barley. *Theor. Appl. Genet.* 117, 653–669.
- Varshney, R.K., Marcel, T.C., Ramsay, L., Russell, J., Röder, M., Stein, N., Waugh, R., Langridge, P., Niks, R.E., Graner, A., 2007. A high density barley microsatellite map with 775 SSR loci. *Theor. Appl. Genet.* 114, 1091–1103.
- Varshney, R.K., Baum, M., Guo, P., Grando, S., Ceccarelli, S., Graner, A., 2010. Features of SNP and SSR diversity in a set of ICARDA barley germplasm collection. *Mol. Breed.* 26, 229–242.
- Verbeke, G., Molenberghs, 2000. *Linear Mixed Models for Longitudinal Data*. Springer Verlag, New York.
- Wang, M., Jiang, N., Jia, T., Leach, L., Cockram, J., Waugh, R., Ramsay, L., Thomas, B., Luo, Z., 2011. Genome-wide association mapping of agronomic and morphologic traits in highly structured populations of barley cultivars. *Theor. Appl. Genet.*, doi:10.1007/s00122-011-1697-2.
- Waugh, R., Marshall, D., Thomas, B., Comadran, J., Russell, J., Close, T., Stein, N., Hayes, P., Muehlbauer, G., Cockram, J., O'Sullivan, D., Mackay, I., Flavell, A., BarleyCAP, A.G.O.U.E.B., Ramsay, L., 2011. Whole-genome association mapping in elite inbred crop varieties. *Genome* 53, 967–972.
- Wenzl, P., Li, H., Carling, J., Zhou, M., Raman, H., Paul, E., Hearnden, P., Maier, C., Xia, L., Caig, V., Ovesná, J., Cakir, M., Poulsen, D., Wang, J., Raman, R., Smith, K.P., Muehlbauer, G.J., Chalmers, K.J., Kleinhofs, A., Huttner, E., Kilian, A., 2006. A high-density consensus map of barley linking DArT markers to SSR RFLP and STS loci and phenotypic traits. *BMC Genomics* 7, 206.
- Yu, J., Pressoir, G., Briggs, W.H., Bi, I.V., Yamasaki, M., Doebley, J.F., McMullen, M.D., Gaut, B.S., Holland, J.B., Kresovich, S., Buckler, E.S., 2006. A unified mixed-model method for association mapping accounting for multiple levels of relatedness. *Nat. Genet.* 38, 203–208.
- Zhang, L.Y., Marchand, S., Tinker, N.A., Belzile, F., 2009. Population structure and linkage disequilibrium in barley assessed by DArT markers. *Theor. Appl. Genet.* 119, 43–52.
- Zhao, K., Aranzana, M.J., Kim, S., Lister, C., Shindo, C., Tang, C., Toomajian, C., Zheng, H., Dean, C., Marjoram, P., Nordborg, M., 2007. An *Arabidopsis* example of association mapping in structured samples. *PLoS Genet.* 3, e4.
- Zhu, J.K., 2002. Salt and drought stress signal transduction in plants. *Annu. Rev. Plant Biol.* 53, 247–273.