Genome-wide association studies for agronomical traits in a world wide spring barley collection

Raj K Pasam1, Rajiv Sharma1, Marcos Malosetti2, Fred A van Eeuwijk2, Grit Haseneyer1,3, Benjamin Kilian1 and Andreas Graner1*

Abstract

Background: Genome-wide association studies (GWAS) based on linkage disequilibrium (LD) provide a promising tool for the detection and fine mapping of quantitative trait loci (QTL) underlying complex agronomic traits. In this study we explored the genetic basis of variation for the traits heading date, plant height, thousand grain weight, starch content and crude protein content in a diverse collection of 224 spring barleys of worldwide origin. The whole panel was genotyped with a customized oligonucleotide pool assay containing 1536 SNPs using Illumina’s GoldenGate technology resulting in 957 successful SNPs covering all chromosomes. The morphological trait “row type” (two-rowed spike vs. six-rowed spike) was used to confirm the high level of selectivity and sensitivity of the approach. This study describes the detection of QTL for the above mentioned agronomic traits by GWAS.

Results: Population structure in the panel was investigated by various methods and six subgroups that are mainly based on their spike morphology and region of origin. We explored the patterns of linkage disequilibrium (LD) among the whole panel for all seven barley chromosomes. Average LD was observed to decay below a critical level (r2-value 0.2) within a map distance of 5-10 cM. Phenotypic variation within the panel was reasonably large for all the traits. The heritabilities calculated for each trait over multi-environment experiments ranged between 0.90-0.95. Different statistical models were tested to control spurious LD caused by population structure and to calculate the P-value of marker-trait associations. Using a mixed linear model with kinship for controlling spurious LD effects, we found a total of 171 significant marker trait associations, which delineate into 107 QTL regions. Across all traits these can be grouped into 57 novel QTL and 50 QTL that are congruent with previously mapped QTL positions.

Conclusions: Our results demonstrate that the described diverse barley panel can be efficiently used for GWAS of various quantitative traits, provided that population structure is appropriately taken into account. The observed significant marker trait associations provide a refined insight into the genetic architecture of important agronomic traits in barley. However, individual QTL account only for a small portion of phenotypic variation, which may be due to insufficient marker coverage and/or the elimination of rare alleles prior to analysis. The fact that the combined SNP effects fall short of explaining the complete phenotypic variance may support the hypothesis that the expression of a quantitative trait is caused by a large number of very small effects that escape detection. Notwithstanding these limitations, the integration of GWAS with biparental linkage mapping and an ever increasing body of genomic sequence information will facilitate the systematic isolation of agronomically important genes and subsequent analysis of their allelic diversity.

* Correspondence: graner@ipk-gatersleben.de
1Leibniz-Institute of Plant Genetics and Crop Plant Research (IPK), Corrensstr. 3, 06466 Gatersleben, Germany
Full list of author information is available at the end of the article

© 2011 Pasam et al; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
Background

Determining the genetic basis of agronomic traits has been one of the major scientific challenges in the process of crop improvement. Most of the agronomically important traits are quantitative, resulting in greater difficulty for discerning genetic differences underlying the phenotype of interest. Currently, linkage mapping (analysis) is the most common approach in plants to detect quantitative trait loci (QTL) corresponding to complex traits. In linkage mapping, linkage disequilibrium (LD) is generated by establishing a population from a cross between two parental lines. The co-segregation of alleles of mapped marker loci and phenotypic traits allows the identification of linked markers. Due to the restricted number of meiotic events that are captured in a biparental mapping population, the genetic resolution of QTL maps often remains confined, to a range of 10-30 cM [1,2]. Moreover, linkage analysis can only sample a small fraction of all possible alleles in a population from which the parents originated.

An alternative approach, association mapping (AM) known as LD mapping relies on existing natural populations or designed populations of plants to overcome the constraints inherent to linkage mapping. LD mapping exploits ancestral recombination events that occurred in genetic mapping studies in humans [3,4] and has been recently considered for plant research. By exploiting non-random associations of alleles at nearby loci (LD), it is possible to scoop out significantly associated genomic regions with a set of mapped markers. Success of mapping depends on the quality of phenotypic data, population size and the degree of LD present in a population [5,6]. In general, the power of association studies depends on the degree of LD between genotyped markers and the functional polymorphisms. The decay of LD varies greatly i) between species [7], ii) among different populations within one species and iii) also among different loci within a given genome [8,9].

LD mapping is based on two strategies: i) re-sequencing of selected candidate genes and ii) genome-wide association which exploits marker polymorphisms across all chromosomes [10]. Genome-wide association studies (GWAS) have become increasingly popular and powerful over the last few years in human and animal genetics. The emergence of more cost-effective, high-throughput genotyping platforms have rendered AM an increasingly attractive approach for QTL mapping in plants [11]. In the last few years, an increasing number of association studies based on the analysis of candidate genes have been published (reviewed in [7]). These include e.g. the Dwarf8 [12] and the phytoene synthase locus in maize [13], flowering time genes in barley [14], the PsyI-Al locus in wheat [15], the rhg-1 gene in soybean [16]; and a series of candidate genes in Arabidopsis [17,18].

Barley (Hordeum vulgare L.) was domesticated in the Fertile Crescent about 10,000 years ago [19-21]. Today barley is the fourth most important cereal crop after wheat, rice and maize. In addition to its agricultural importance, the barley genome is considered as a model for other crop species of the Triticeae tribe including wheat and rye [22,23]. In this regard an ever increasing repertoire of marker and sequence resources has been developed for barley which can be efficiently utilized [24-26]. Over the last few years candidate gene based AM studies were reported for barley [9,14,27]. GWAS with dense marker coverage are not yet conducted routinely for barley, albeit the potential of this approach has been demonstrated in some pilot studies [28-30].

Inbreeding crops such as barley are characterized by a high level of population structure caused by the impact of non random mating and subsequent selection. This is exemplified by two-rowed and six-rowed barley cultivars which form distinct subpopulations, because the corresponding breeding programs rely on different progenitors. The same applies to the subpopulations of spring and winter barley [31]. There are higher chances of occurrence of type I and type II errors in AM than in biparental QTL analysis due to the confounding effect of population structure in the panel [2,32-34]. Specific statistical approaches have been proposed to account for population structure in AM [35]. Yu et al. [36] described a mixed-linear model (MLM) approach which performs better than previous models [37]. Still these models have their individual shortcomings and care needs to be taken in controlling for population structure and balancing the rate of false positives and false negatives in the analysis.

In the present study, our main objective was to map genetic polymorphisms underlying complex agronomic traits such as heading date (HD), plant height (PHT), thousand grain weight (TGW), starch content (SC) and crude protein content (CPC) in spring barley using GWAS. We studied a diverse spring barley collection comprising 224 accessions from 52 countries previously described by Haseneyer et al. [38]. We provide a comprehensive overview on population structure and genetic diversity as well as their effects on GWAS. To study the dynamics of LD across the seven barley chromosomes we investigated the patterns of LD decay. Finally, we identify and locate a substantial number of known and novel QTL for the traits investigated.

Methods

Association mapping panel

The association mapping panel consists of 224 spring barley accessions selected from the Barley Core Collection (BCC) [39] and the barley Genebank collection.
Phenotypic evaluation
The accessions were planted in a 25 × 15 lattice design with three replications in the years 2004 and 2005 at the following locations: Stuttgart (Southwest Germany), Irlbach (Southeast Germany) and Wohlsde (Northern Germany). Heading date (HD) and plant height (PHT) were scored in field plots. Thousand grain weight (TGW) was estimated from sampled grains per plot. Starch content (SC) and crude protein content (CPC) were estimated using a near infrared reflectance spectrometer (NIRS) from ground seed samples from all environments. In order to convert the nitrogen content to crude protein values, we considered a factor of 6.25. We followed the methods described in Naumann and Bassler [40] to estimate the starch content and nitrogen content. Phenotypic data were analyzed using REML (Residual Maximum Likelihood) implemented in GenStat 9 software [41]. Variance components were calculated by fitting a mixed linear model (MLM) to multi-environment data. Heritabilities were estimated for all traits considering the percentages of genotypic variance, over the total phenotypic variance estimated for all traits considering the percentages of genotypic variance, over the total phenotypic variance.

Genotypic data analysis and population structure
Polymorphic information content (PIC) values were calculated for each SNP using Powermarker 3.25 [44]. Major allele frequency, minor allele frequency (MAF), gene diversity and Nei’s genetic distance (d) [45] were calculated and a NJ (Neighbor-Joining) dendrogram (data not shown) based on d was computed. From the 957 SNPs, a final set comprising 918 SNPs with MAF larger than 0.05 was used for analysis of population structure, LD and marker trait associations.

To estimate the number of subgroups in the panel, different methodologies and different software packages were employed and compared in order to determine the appropriate population structure in collection. For the quantitative assessment of the number of groups in the panel, a Bayesian clustering analysis was performed using a model based approach implemented in the software package STRUCTURE v2.2 [46,47]. This approach uses multi-locus genotypic data to assign individuals to clusters or groups (K) without prior knowledge of their population affinities and assumes loci in Hardy-Weinberg equilibrium. The program was run with 918 SNP markers for k-values 1 to 15 (hypothetical number of subgroups), with 100000
burnin iterations followed by 50000 MCMC (Markov Chain Monte Carlo) iterations for accurate parameter estimates. To verify the consistency of the results we performed 5 independent runs for each k. An admixture model with correlated allele frequencies was used. The most probable number of groups was determined by plotting the estimated likelihood values \([\ln P(D)]\) obtained from STRUCTURE runs against \(k\). \(\ln P(D)\) is the log likelihood of the observed genotype distribution in \(k\) clusters and is an output by STRUCTURE simulation. The \(k\) value best describes the population structure based on the criteria of maximizing the log probability of data or in other words the value at which \(\ln P(D)\) reaches a plateau [46]. STRUCTURE results with the SNP marker dataset were confirmed with the results from STRUCTURE runs using a set of Diversity Array Technology (DArT) markers (Pasam et al. unpubl, Additional file 3: Figure S1). In a second approach principal coordinate analysis (PCoA) based on the dissimilarity matrix was performed using DARwin (Diversity Analysis and Representation for windows) [48]. In a third approach a NJ dendrogram based on Nei’s genetic distance matrix was constructed. The substructure in the collection using different methodologies was compared and the final \(k\) value using STRUCTURE was ascertained. For this \(k\) value, the Q-matrix (population membership estimates) was extracted from STRUCTURE runs. This matrix provides the estimated membership coefficients for each accession in each of the subgroups.

### Linkage disequilibrium analysis

The extent of LD effects both the number of markers required for GWAS and the resolution of mapping the trait. LD is in many cases influenced by population structure resulting from the demographic and breeding history of the accessions. Genome-wide LD analysis was performed among the panel and subgroups by pair wise comparisons among the SNP markers using HAPLOVIEW [49]. LD was estimated by using squared allele frequency correlations \((r^2)\) between the pairs of loci [50]. The loci were considered to be in significant LD when \(P < 0.001\), the rest of \(r^2\) values were not considered as informative. The pattern and distribution of intra-chromosomal LD was visualized and studied from LD plots generated for each chromosome by HAPLOVIEW. To investigate the average LD decay in the whole genome among the panel, significant intra-chromosomal \(r^2\) values were plotted against the genetic distance (cM) between markers. The smoothing second degree LOESS curve was fitted using GENSTAT [41]. A critical value for \(r^2\) was estimated by square root transforming of unlinked \(r^2\) values to obtain a normally distributed random variable, and the parametric 95th percentile of that distribution was taken as a critical \(r^2\) value [32]. Unlinked \(r^2\) refers to the \(r^2\) between the marker loci with a genetic distance greater than 50 cM or on independent linkage groups.

### Association analysis

Different statistical models were used to calculate \(P\)-values for associating each marker with the trait of interest, along with accounting for population structure to avoid spurious associations by TASSEL v.2.1 (http://www.maizegenetics.net). We followed the formula \(y = X\beta + M + Zu + e\), where \(y\) is a response vector for phenotypic values, \(\beta\) is a vector of fixed effects regarding population structure, \(\alpha\) is the vector of fixed effect for marker effects, \(u\) is the vector of random effects for co-ancestry and \(e\) is the vector of residuals. \(X\) can be either the Q-matrix or the PCs from Principal Component Analysis (PCA), \(M\) denotes the genotypes at the marker and \(Z\) is an identity matrix. Six models comprising both general linear models (GLM) and mixed linear models (MLM) were selected to test the marker-trait-associations (MTA). Results were compared to determine the best model for our analysis. PCA was conducted with TASSEL. The first ten significant PCs explained 43% of the cumulative variance of all markers. A kinship matrix (K-matrix), the pair-wise relationship matrix which is further used for population correction in the association models was calculated with 918 SNP markers using TASSEL [51]. The following models were tested: i) Naive model: GLM without any correction for population structure; ii) Q-model: GLM with Q-matrix as correction for population structure; iii) P-model: GLM with PCs as correction for population structure; iv) QK-model: MLM with Q-matrix and K-matrix as correction for population structure; v) PK-model: MLM with PCs and K-matrix as correction for population structure and vi) K-model: MLM with K-matrix as correction for population structure [35,36,52,53]. All SNP markers were re-mapped by association mapping to determine the mapping resolution of the panel as suggested by [24]. The critical \(P\)-values for assessing the significance of MTAs were calculated based on a false discovery rate (FDR) separately for each trait [54], which was found to be highly stringent. Considering the stringency of the model used for accounting for population structure, most of the false positives were inherently controlled. Thus, we considered a more liberal approach as proposed by Chan et al. [55] for determining the threshold level for significant MTAs. It was suggested that the bottom 0.1 percentile distribution of the \(P\)-values can be considered as significant, which in our analysis resulted in threshold levels of 0.05 to 0.09 for individual traits. Alternatively, as a compromise between the two approaches an arbitrary threshold \(P\)-value of 0.03 was used for all traits and all models. This rather rough estimate was obtained by arranging-\log_{10} P\)-values in a descending order, and the value at which the curve starts to
flatten is determined as the threshold value. All association models with all traits were re-analyzed using GENSTAT [41] to check for any discrepancy.

**Results**

**Phenotypic data**

Large phenotypic variation was observed for all traits. Outliers in the data were identified based on the residuals derived from the data of all environments and were removed from further analysis. For the trait heading date, data from year 2004 was excluded from the analysis due to differences in scoring this trait between the individual locations. Variance components were calculated by REML. The results confirmed that genotypic variance was significant for all traits ($P < 0.001$). GxE interactions were also significant ($P < 0.001$) but represented only a small fraction of the total variance. Heritabilities ranged between 0.90-0.95 indicating the robustness of the data and the low error rate. Year-wise means, ranges and heritabilities over all environments for the traits HD, PHT, TGW, SC and CPC are presented in Table 1 and their frequency distributions are illustrated in Additional file 4: Figure S2. The correlation exhibited by the agronomic traits between each other is outlined in Table 2. The traits SC and CPC are highly correlated (-0.7) and other traits showed moderate to weak correlation among each other. PHT was shown to be weakly correlated with HD and also with SC and CPC. TGW is found to be positively correlated with SC and negatively correlated with CPC. Substantial phenotypic differences were reported between two-rowed and six-rowed genotypes. The means for all traits were significantly different between the two groups (Additional file 5: Table S3). The variation observed was larger for all traits in six-rowed barleys than in two-rowed barleys. The greatest influence of spike morphology (two-rowed vs. six-rowed) on phenotypic variation was seen for TGW, whereas the greatest influence of population structure was observed for PHT (Additional file 6: Table S4).

Best Linear Unbiased Estimates (BLUEs) of genotypic means were calculated from the fixed genotypic effects to avail unbiased mean estimates. Using Best Linear Unbiased Predictors (BLUPs) is less suitable as it would cause double shrinking [56]. Henceforth we used BLUEs in our further analysis. However, comparison of both BLUPs and BLUEs revealed very high concordance between both estimates, which is a direct consequence of the high heritabilities (Additional file 7: Figure S3).

**Population structure and genetic diversity**

From the high quality 985 SNPs, 957 markers had been genetically mapped and therefore were considered for this study. Of these, 39 SNPs (4%) were excluded because of a MAF below 0.05. Of the remaining SNPs, the majority revealed a MAF between 0.1 to 0.5 (Figure 1). These SNP markers were distributed over all seven chromosomes with an average spacing of 1.18 cM. The distribution of SNP markers is not exactly uniform and varies within and among chromosomes with a minimum of 105 markers on chromosome 4H and a maximum of 164 markers on 5H (Table 3). Diversity statistics computed for each SNP are summarized in Additional file 8: Table S5. PIC values for SNPs ranged from 0.09 to 0.5 with an average of 0.30. Most of the markers (726) displayed PIC values exceeding 0.25, demonstrating the informativeness of these markers in our panel. The average PIC values of the markers on each chromosome ranged between 0.29 (5H) to 0.33 (6H). The mean gene diversity value for the whole panel was 0.39 and spread within a range of 0.09 to 0.5. It was reported in several studies that the stratification of barley cultivars is concordant with spike morphology, mainly as a result of breeding history [57,58]. Therefore, similar molecular diversity statistics were generated separately for two-rowed and six-rowed barley groups within our panel and for the six subgroups. Observed mean PIC values are higher for the six-rowed group (0.31) than for two-rowed barleys (0.27). Similarly,

### Table 1 Estimation of mean, minimum (Min), maximum (Max) and heritabilities ($h^2$) of all traits

| Trait                  | 2004 Min | Max | Mean | 2005 Min | Max | Mean | $h^2$ (%) | GxE |
|------------------------|----------|-----|------|----------|-----|------|----------|-----|
| Plant height (PHT)     | 20       | 120 | 77.04| 30       | 120 | 73.69| 92.82    |     |
| Heading date (HD)      | *        | *   | **   | *        | *   | 56   | 92.1     |     |
| Thousand grain weight (TGW) | 17.77   | 67.23 | 44.92| 20.1     | 62.6 | 42.43| 92.9     |     |
| Starch content (SC)    | 408      | 64.58| 56.85| 44.01    | 65.31| 56.91| 96.3     |     |
| Protein content (CPC)  | 9.74     | 25.74| 14.88| 10.35    | 25.18| 14.90| 92.1     |     |

Heritabilities were calculated on entry mean basis.

### Table 2 Correlation coefficients among different traits estimated across all environmental data

|          | HD  | PHT | CPC | SC  |
|----------|-----|-----|-----|-----|
| HD       | 0.29**|
| PHT      | *   | -0.25**|
| CPC      | -0.43**| 0.17*|
| SC       | 0.43**| -0.76**|
| TGW      | 0.04| -0.30**| 0.33**|

*highly significant at $P < 0.001$, * significant at $P < 0.01$, rest are not significant
average gene diversity estimated was higher in six-rowed (0.38) than in two-rowed accessions (0.33).

The population structure in the panel of 212 barley accessions was analyzed using 918 SNP markers and a model based approach in STRUCTURE. The LnP(D) appeared to be an increasing function of $k$ for all the values observed. But the most significant increase of LnP(D) was observed when $k$ was increased from 1 to 2 (Figure 2). At $k = 2$ the panel is clearly categorized into two-rowed and six-rowed barleys with few exceptions. The two main groups were further divided yielding six subgroups in total as LnP(D) values nearly reached a plateau at $k = 6$. Hence, we chose a value of $k = 6$ for our analysis as minimum number of groups present in the panel. Different values of $k$ are still possible but will not qualitatively affect the results. An accession was assigned to a subgroup if at least 50% of the genome information was estimated to belong to one group. The accessions clustered into groups mostly according to their spike morphology and their geographical origins, as was demonstrated already by Haseneyer et al. [38]. The six groups are defined as: Group 1 (G1): 24 six-rowed barleys mostly from AM and WANA; G2: 31 accessions mostly six-rowed barley from EA; G3: 31 accessions mostly six-rowed barleys from EU; G4: 24 accessions mostly two-rowed from EU; G5: 79 accessions mostly two-rowed barleys from EU; G6: 23 accessions mostly two-rowed from WANA and AM (Figure 3). The dominant stratification of the population according to spike morphology is confirmed by PCoA (Additional file 9: Figure S4) and NJ dendrogram (not shown). In the PCoA, it is obvious that the primary axis separates the accessions based on row type and further grouping is related to the region of origin. Overall, the clustering of accessions was consistent among various methods and we further explored the genetic diversity within these groups. The summary statistics for each group with 918 SNP markers is reported in Table 4. Observed gene diversity values ranged from 0.27 in G5 to 0.35 in G1; PIC values ranged from 0.22 in G5 to 0.29 in G1. Pairwise genetic distances ranged from 0.006 to 0.628, with an overall mean of 0.39. The average overall genetic distance between groups has been calculated, and the largest genetic distance of 0.36 was observed between the groups G2 (six-rowed, EA) and G5 (Two-rowed, EU). Similarly G4 (six-rowed, EU) and G5 (six-rowed, EU) are found to be closely related groups with an average genetic distance of 0.17 (Table 5).

**Table 3** SNP coverage and distribution across all chromosomes

| Chromosome | cM   | Markers | Marker coverage | PIC  |
|------------|------|---------|-----------------|------|
| 1H         | 139.79 | 117     | 1.19            | 0.31 |
| 2H         | 156.72 | 146     | 1.07            | 0.29 |
| 3H         | 173.17 | 151     | 1.15            | 0.32 |
| 4H         | 123.29 | 105     | 1.17            | 0.32 |
| 5H         | 195.42 | 119     | 1.19            | 0.29 |
| 6H         | 129.38 | 119     | 1.09            | 0.33 |
| 7H         | 166.56 | 116     | 1.44            | 0.30 |
| Total      | 1084.33| 918     | 1.18            | 0.31 |

Average PIC values for all SNPs on each chromosome are represented.

**Linkage disequilibrium**

LD analysis was performed using 918 SNPs for i) entire panel, ii) separately for two-rowed and six-rowed barleys, and iii) each of the six subgroups. Pairwise LD was estimated using the squared-allele frequency correlations ($r^2$) and was found to decay rapidly with the genetic distance. We studied different aspects of LD in our panel and observed that LD varies along the chromosomes with regions of high LD interspersed with regions of...
Figure 3 Population sub-structuring in the panel. Bayesian clustering of the 212 barley accessions into six defined groups (G1, G2, G3, G4, G5, G6) based on 918 SNP markers. The number of accessions per group and their respective geographical origin and row type is presented.

| Group | Average major allele frequency | No. genotypes | Gene diversity | PIC   |
|-------|--------------------------------|---------------|---------------|-------|
| Whole panel | 0.6978                                | 212           | 0.3904        | 0.3079 |
| 2-rowed group | 0.7551                                | 122           | 0.3325        | 0.2714 |
| 6-rowed group | 0.7064                                | 90            | 0.3852        | 0.3108 |
| G1    | 0.7359                                | 24            | 0.3551        | 0.2903 |
| G2    | 0.7933                                | 31            | 0.3060        | 0.2338 |
| G3    | 0.7773                                | 31            | 0.3106        | 0.2536 |
| G4    | 0.7746                                | 24            | 0.2971        | 0.2297 |
| G5    | 0.7976                                | 79            | 0.2791        | 0.2681 |
| G6    | 0.7547                                | 23            | 0.2865        | 0.2681 |

PIC values are given as the mean values of the corresponding marker partitions.
A critical value of $r^2$, or basal LD, was calculated from inter-chromosomal LD analysis and is estimated to be 0.2 beyond which LD is assumed to be caused by genetic linkage. The point at which the LOESS curve intercepts the critical $r^2$ is determined as the average LD decay of the population. Based on these criteria the intra-chromosomal LD decayed between 5-10 cM for individual chromosomes and average LD decay of the whole genome was observed to be at 7 cM (Figure 4). Extensive variability in the magnitude of $r^2$ at a given genetic distance was detected reflecting the wide local variation in the extent of LD across the chromosomes. The correlation between $r^2$ and marker distance was found to be significantly negative ($r = -0.40$) for markers below a distance of 10 cM, whereas marker pairs with larger distance showed no significant correlation with $r^2$. Significant intra-chromosomal $r^2$ values ($P < 0.001$) ranged from 0.02 to 1 with an average of 0.12 for the whole panel. Among all significant loci in LD, 13.7% of the loci are above the critical $r^2$ value of 0.2 in the whole panel.

Patterns of linkage disequilibrium within subgroups

At the intra-chromosomal level, mean $r^2$ values for two-rowed and six-rowed barley groups ranged between 0.18 and 0.17, which is slightly more than the mean $r^2$ of whole panel. The percentages of significant $r^2$ values were higher in the two-rowed than in the six-rowed subgroup for all classes of marker pairs except for the independent markers. This pattern is also similar to LD values above the basal level of 0.2, and a slightly slower LD decay was observed for two-rowed barley compared to the group of six-rowed types and to the whole panel. Similarly, the mean $r^2$ values were estimated for individual subgroups where they ranged from 0.3 (G5) to 0.49 (G4).

The LD decay in the subgroups was much slower than in the whole panel. In Figure 5, binned $r^2$ values are mapped against the recombination distance (cM) across the genome. In the whole panel the average LD decays below a basal level (0.2) within 5 cM, while in the two-rowed and six-rowed groups the basal level is reached between 10-15 cM and with LD in six-rowed barley decaying faster than in two-rowed barley. Within G5 LD decays to the basal level within 20-25 cM, while it does not reach the basal level in the remaining subgroups (G1,2,3,4,6).

Average LD decay graphs for each group showed different patterns. Specifically, in the subgroups G4 and G5 at distances 45 and 74 cM we observed larger LD peaks. Scrutinizing these peaks revealed that high LD in these regions was caused by markers with low allele frequencies. The consequence of the reduced population size of the individual subgroups is that the presence of a solitary allele in single accession already might show a MAF above the critical threshold. Varying patterns of LD decay in different sub-populations are likely reflecting their breeding
Table 6 LD overview for the whole panel and the subgroups of two-rowed and six-rowed barley

|                  | Total pairs | % significant | Significant pairs | Mean $r^2$ | Pairs in complete LD | % of pairs in LD > 0.2 | Mean of $r^2$ > 0.2 |
|------------------|-------------|---------------|-------------------|------------|----------------------|------------------------|----------------------|
| Whole panel      |             |               |                   |            |                      |                        |                      |
| total            | 62222       | 39.4          | 24567             | 0.12       | 59                   | 13.72                  | 0.36                 |
| 0-10 cM          | 10602       | 62.2          | 6554              | 0.20       | 59                   | 33.70                  | 0.42                 |
| 11-20 cM         | 8028        | 45.1          | 3626              | 0.10       | 0                    | 10.21                  | 0.29                 |
| 21-50 cM         | 19066       | 38.3          | 7310              | 0.09       | 0                    | 8.40                   | 0.27                 |
| > 50 cM          | 24526       | 28.5          | 7077              | 0.08       | 0                    | 4.00                   | 0.25                 |
| 2-rowed          |             |               |                   |            |                      |                        |                      |
| total            | 48803       | 21.6          | 10544             | 0.18       | 94                   | 23.74                  | 0.43                 |
| 0-10 cM          | 8183        | 50.0          | 4098              | 0.29       | 94                   | 48.00                  | 0.48                 |
| 11-20 cM         | 6244        | 29.9          | 1869              | 0.13       | 0                    | 13.31                  | 0.30                 |
| 21-50 cM         | 15066       | 17.2          | 2601              | 0.12       | 0                    | 9.59                   | 0.28                 |
| > 50 cM          | 19310       | 10.2          | 1976              | 0.10       | 0                    | 3.81                   | 0.26                 |
| 6-rowed          |             |               |                   |            |                      |                        |                      |
| total            | 58356       | 20.2          | 11801             | 0.17       | 95                   | 22.40                  | 0.36                 |
| 0-10 cM          | 9947        | 36.8          | 3661              | 0.24       | 95                   | 40.37                  | 0.43                 |
| 11-20 cM         | 7439        | 21.1          | 1569              | 0.14       | 0                    | 18.26                  | 0.27                 |
| 21-50 cM         | 17768       | 16.9          | 3016              | 0.14       | 0                    | 15.04                  | 0.27                 |
| > 50 cM          | 23202       | 15.3          | 3555              | 0.13       | 0                    | 12.14                  | 0.25                 |

LD statistics are given for the total number of locus pairs and for different marker linkage classes (for details see text).
histories [1] and may impinge on the QTL mapping resolution of the panel.

Evaluation of the association panel
All 918 SNPs were re-mapped using an LD approach. A model with kinship accounting for population structure was used for validating the genetic map position of the markers. We used each marker information as an individual trait and ran the analysis with the remaining SNPs to find the most significantly associated markers. The map distance between the target marker in question and the most highly associated marker was used to evaluate the resolution of the panel. More than 85% of the SNP markers had their genetic map position within 0-10 cM distance of their original map position and the majority of them re-mapped at the same position (Figure 6). This re-mapping of markers shows that the resolution of QTL captured by AM approach in our panel will be within a range of 5-10 cM.

Association analysis
Comparison of models
We tested several models to detect associations between SNP markers and agronomic traits. Owing to the complexity and the considerable amount of population structure present in our panel, we observed numerous spurious associations when using the naive (simple) model for AM. Hence, we assessed the usefulness of various linear models to account for population structure by comparing their ability to reduce the inflation of false positive associations. To this end ranked P-values from GWAS were plotted in a cumulative way for each model by using spike morphology as phenotypic trait (Figure 7). As demonstrated by Kang et al. [53] the distribution of P-values ideally should follow a uniform distribution with less deviation from the expected P-values. The models QK, PK and K showed a good fit for P-values, while the other models were characterized by the excess of small P-values which is tantamount to an abundance of spurious associations. This is particularly obvious in the case of the “naive” model, where nearly half of the P-values are smaller than 0.01. On the other hand, the K-model performed similar to the PK and QK model in displaying a highly uniform distribution
of P-values and at the same time requiring less computational time. Irrespective of the model, major marker trait associations were constantly detected. However, the more stringent the model was the less spurious background associations were detected. All models considered for GWAS are presented for the trait spike morphology (Additional file 11: Figure S6). For all other traits only results from the K-model will be presented and discussed.

Barley spike morphology (row type)
Apart from comparing different AM models, we aimed to examine the spike morphology trait as a proof of concept for GWAS and to evaluate the resolution of the association panel. According to its spike morphology barley is classified as two-rowed and six-rowed types and the genes for this trait have been well documented [29,30,60]. We scored the row type character in the panel and considered 918 markers for AM using all models. A marker trait association was considered when the marker main effect was significant at 0.03 [-log10 (0.03) = 1.5]. This results in a total of 34 markers that are significantly associated with the trait row type by using the K-model. (Additional file 11: Figure S6). The results are congruent with previous row type studies (see Figure 8).

Heading date
Thirty-four markers were found to be significantly associated with heading date (HD). These were grouped into 19 QTL located on all chromosomes. Significant marker trait associations within a genetic distance of 5-10 cM are delineated into a single QTL. Chromosome 2H harbors the maximum number of markers associated with the trait (Figure 9a). Some of these association results with the SNP markers effectively correspond to genomic regions of previously mapped flowering time QTL. These include genomic regions of various prominent flowering pathway genes like Ppd-H1, HvFT1, HvCO1 and HvCO3 (see Table 7).

Plant height
Thirty-two markers displayed significant associations with plant height (PHT). These markers detected 19 QTL (Table 8). Except for chromosome 1H, significantly associated markers were found on all chromosomes with the majority located on 2H and 3H (Figure 9b).

Thousand grain weight
Thirty-six markers yielding 21 QTL were significantly associated with Thousand Grain Weight (TGW, Figure 9c). Markers significantly associated with the trait were present on all chromosomes. As expected some of the TGW related QTL overlapped with QTL for spike morphology. The markers SNP56, SNP215, SNP385 and SNP458 are co-localized to the same region as Vrs3, VrsI, Vrs4 and Int-c genomic regions (Table 9).

Starch content
We found thirty-five markers to be significantly associated with the trait Starch Content (SC). These markers formed a total of 25 QTL (Figure 9d). Significantly associated markers for starch content were present on all chromosomes. Similar to TGW markers corresponding to the Vrs3 region (SNP56 & SNP66) are significantly associated with starch content. Several significant markers, co-localized with previously mapped genes and QTL for SC (Table 10).

Protein content
We found thirty-four markers to be significantly associated with crude protein content (CPC). These markers detected a total of 23 QTL (Figure 9e) and were distributed over all chromosomes. Some of the QTL for protein content overlapped with the QTL regions identified for CPC (Table 11).

Discussion
In the present study we describe the application of whole genome association mapping in a panel of diverse spring barley genotypes for agronomic traits. For each of the analyzed traits we identified 19 to 25 QTL. A substantial portion of the derived QTL locations are congruent with previously identified QTL in various biparental mapping populations (Tables 7, 8, 9, 10, 11). GWAS are strongly influenced by the quality of the phenotypic data [79]. In the present study heritabilities for all traits exceeded 0.9 and phenotypic means reflected a broad variation in the panel. The observed differences for two-rowed and six-rowed groups were expected due to their different breeding histories and the pleiotropic effects of spike morphology (Additional file 5: Table S3). Phenotypic variation observed for all traits is higher in the six-rowed group than in the two-rowed group, which is in accordance with the higher genetic diversity
of this subgroup (Table 4). A more detailed analysis of population structure revealed six subgroups, which were mostly defined by spike morphology and geographical origin, both of which are known to impinge on the expression of agronomic traits.

Genetic diversity and population structure
Arguably an association mapping panel should suffice both phenotypic and molecular diversity for the outcome of reliable association results. Owing to the availability of a large number of mapped SNP markers that can be interrogated in a multiparallel manner [26], we were able to achieve a high marker coverage amounting to 1 marker per 1.18 cM. The average PIC (0.30) and Gene diversity (0.33) values observed in this panel of accessions are comparable with the results in previous studies using bi-allelic markers. PIC values differed among chromosomes and among different germplasm subgroups (Tables 3 & 4). Among all chromosomes, the highest average PIC value (0.33) was detected for chromosome 6H—which corresponds to the observations made by Rostoks et al. [24] in a set of European barley cultivars. We determined the population structure in our panel by implementing various approaches (STRUCTURE, PCoA and NJ-dendrogram) and found similar results. Several previous studies e.g. Maliysheva-Otto et al. [57], Rostoks et al. [24], Zhang et al. [58] and Hamblin et al. [80] have shown that growth habit, spike morphology and geographical origin are the major factors that mirror population structure in barley. Since the present study has been restricted to spring barley, spike morphology and geographical origin were the fundamental determinants for population substructuring (G1 to G6) (Figure 3). The 55 landrace accessions included in this panel were distributed among all groups. The subgroups G1, G2 and G3 are mainly six-rowed barleys and the subgroups G4, G5 and G6 include mainly two-rowed barleys. Two-rowed barleys in the panel are more closely related to each other and less diverse than the six-rowed accessions, which is in contrast to the findings of Zhang et al. [58] for Canadian germplasm. While in our panel two-rowed barleys even outnumbered the six-rowed accessions, the reason for their limited diversity might be that the majority originated from Europe. The geographical distribution of the accessions has a major influence on the diversity of alleles sampled in the population [57]. In Europe, two-rowed barley is mainly grown as raw material for malt
production. Malting quality is a quantitative trait. The use of a limited number of principal progenitors in the corresponding breeding programs has resulted in the reduction of genetic diversity and in the concomitant formation of a distinct subpopulation as it is seen in our present panel [81].

LD configuration and consequences

The resolution of LD mapping depends on the extent of LD across the genome and the rate of LD decay with genetic distance [82,83]. Genome-wide LD studies for barley have been previously reported in various populations using different molecular markers such as AFLP, SSR and DArT [57,58,84], with few studies, however, relying on more than 1000 markers. In our panel of spring barley accessions of worldwide origin, intra-chromosomal whole genome LD decays below the critical $r^2$ value (0.2) within a genetic distance of 5 cM. It needs to be kept in mind that this is an average value, which summarizes substantial intra-chromosomal LD variation. The extent of intra-chromosomal LD for different chromosomes ranges from 5 - 10 cM with varying patterns along each chromosome (Additional file 10: Figure S5). Previous studies found various levels of LD decay in different barley populations [9,29,83] and among different chromosomes [24]. The LD decay was more

| SNP     | Chr | Position | MAF | P-value | -log10(P) | R² (%) | Effect | QTL   | Reference QTL | Literature |
|---------|-----|----------|-----|---------|-----------|--------|--------|-------|---------------|------------|
| SNP111  | 1H  | 128.14   | 0.19 | 0.0032  | 2.49      | 0.63   | -2.51 | QTL1_HD | HvFT3         | [61]       |
| SNP119  | 1H  | 138.92   | 0.23 | 0.0198  | 1.70      | 0.33   |
| SNP129  | 2H  | 27.29    | 0.37 | 0.0099  | 2.00      | 0.38   |
| SNP130  | 2H  | 28.44    | 0.36 | 0.0080  | 2.10      | 0.39   | -1.29 | QTL2_HD | PpdH1        | [62]       |
| SNP133  | 2H  | 31.02    | 0.38 | 0.0280  | 1.55      | 0.39   |
| SNP135  | 2H  | 33.73    | 0.10 | 0.0262  | 1.58      | 0.51   |
| SNP142  | 2H  | 41.66    | 0.27 | 0.0096  | 2.02      | 0.40   |
| SNP148  | 2H  | 53.53    | 0.34 | 0.0043  | 2.37      | 0.47   | -1.37 | QTL3_HD         |
| SNP170  | 2H  | 63.53    | 0.32 | 0.0077  | 3.10      | 0.88   |
| SNP174  | 2H  | 63.53    | 0.41 | 0.0011  | 2.96      | 0.68   |
| SNP177  | 2H  | 63.53    | 0.32 | 0.0013  | 2.89      | 0.55   |
| SNP173  | 2H  | 63.53    | 0.33 | 0.0033  | 2.48      | 0.53   |
| SNP183  | 2H  | 66.83    | 0.44 | 0.0265  | 1.58      | 0.39   | -2.32 |
| SNP191  | 2H  | 71.12    | 0.44 | 0.0043  | 1.0061    | 0.0012 |
| SNP196  | 2H  | 73.04    | 0.10 | 0.0061  | 2.21      | 0.65   |
| SNP199  | 2H  | 73.75    | 0.14 | 0.0012  | 2.92      | 0.49   | 2.53  |
| SNP242  | 2H  | 115.78   | 0.39 | 0.0207  | 1.68      | 0.54   | -1.98 | QTL6_HD         |
| SNP284  | 3H  | 8.23     | 0.24 | 0.0111  | 1.95      | 0.42   | -1.45 | QTL7_HD         |
| SNP340  | 3H  | 59.89    | 0.35 | 0.0047  | 2.33      | 0.38   | 1.80  | QTL8_HD | HvGI         | [61]       |
| SNP520  | 4H  | 82.42    | 0.32 | 0.0198  | 1.70      | 0.47   | -1.15 | QTL9_HD         |
| SNP543  | 4H  | 123.29   | 0.26 | 0.0024  | 2.62      | 0.67   | -1.59 | QTL10_HD        |
| SNP559  | 5H  | 39.97    | 0.46 | 0.0146  | 1.84      | 0.32   | -1.20 | QTL11_HD | HvCO3        | [65]       |
| SNP630  | 5H  | 100.28   | 0.20 | 0.0203  | 1.69      | 0.58   |
| SNP635  | 5H  | 102.06   | 0.13 | 0.0278  | 1.56      | 0.34   | -2.02 | QTL12_HD        |
| SNP636  | 5H  | 103.92   | 0.28 | 0.0278  | 1.56      | 0.35   |
| SNP639  | 5H  | 108.18   | 0.38 | 0.0236  | 1.63      | 0.53   |
| SNP728  | 6H  | 28.39    | 0.43 | 0.0132  | 1.88      | 0.28   | -1.43 | QTL13_HD        |
| SNP778  | 6H  | 60.23    | 0.49 | 0.0306  | 1.51      | 0.37   | -2.15 | QTL14_HD        |
| SNP829  | 6H  | 124.85   | 0.33 | 0.0281  | 1.55      | 0.37   | 1.52  | QTL15_HD        |
| SNP854  | 7H  | 37.55    | 0.36 | 0.0060  | 2.22      | 0.57   | 2.50  | QTL16_HD | HvFT1        | [61,63]   |
| SNP855  | 7H  | 38.32    | 0.35 | 0.0009  | 3.01      | 0.54   |
| SNP875  | 7H  | 68.46    | 0.06 | 0.0180  | 1.74      | 0.48   | -2.33 | QTL17_HD        |
| SNP908  | 7H  | 84.92    | 0.35 | 0.0266  | 1.58      | 0.37   | 1.66  | QTL18_HD | HvCO1        | [61,65]   |
| SNP921  | 7H  | 104.78   | 0.35 | 0.0131  | 1.88      | 0.59   | -1.79 | QTL19_HD        |

Significant markers associated for heading date with K-model, corresponding MAF, P value of association, variance explained by marker ($R^2$), effect of the most significant marker within the QTL interval, name of the QTL, and the reference QTL or gene from literature.
rapid in the study of Comadran et al. [85] probably due to the inclusion of landraces in the collection. Caldwell et al. [9] also showed that LD decays more rapidly in barley landraces compared to elite barley cultivars. Less extensive LD beyond 10 cM has been found in our panel, as the majority of significant LD values above the basal level (33.7%) are due to tightly linked markers. Significant inter-locus LD values of unlinked markers (4%) may be the result of population structure (Table 6). We found some closely linked markers that are in complete Linkage Equilibrium (LE), while some distantly linked markers exhibited high LD values. This reflects the dynamic variation of LD patterns along the chromosomes as it has been shown in this panel at the sequence level for several transcription factors [27]. As to the individual subgroups, the portion of significant $r^2$-values above the basal level (0.2) is higher within six-rowed than in two-rowed groups indicating high LD in these groups. Interestingly, LD in all subgroups extended beyond 30 cM except for G5 where LD extended to about 20-25 cM (Figure 5). This is most likely because of the larger population size of G5 compared to the other subgroups. The extensive LD observed in the subgroups is probably due to their decreased population size and a concomitant increase in relatedness.

**Genome-wide association mapping**

Despite the advantages of GWAS to pinpoint genetic polymorphisms underlying agronomic traits, this approach may suffer from an inflation of false positives

| Marker | Chr | Position | MAF | P-value | -log10(P) | $R^2$ (%) | Effect | QTL | Reference QTL | Literature |
|--------|-----|----------|-----|---------|-----------|-----------|--------|-----|---------------|------------|
| SNP122 | 2H  | 8.57     | 0.13| 0.0016  | 2.80      | 0.94      | -5.64  | QTL12_PHT    |             |
| SNP136 | 2H  | 33.74    | 0.35| 0.0138  | 1.86      | 0.54      |        |     |               |            |
| SNP137 | 2H  | 38.03    | 0.35| 0.0044  | 2.36      | 0.84      | -5.23  | QTL13_PHT    | Ph2         |
| SNP168 | 2H  | 59.21    | 0.16| 0.0229  | 1.64      | 0.46      |        |     |               |            |
| SNP171 | 2H  | 63.53    | 0.10| 0.0155  | 1.81      | 0.51      | 6.73   | QTL14_PHT    | HT (marker:B15c) |
| SNP175 | 2H  | 63.53    | 0.10| 0.0155  | 1.81      | 0.51      |        |     |               |            |
| SNP199 | 2H  | 73.75    | 0.14| 0.0224  | 1.65      | 0.49      | 4.72   | QTL15_PHT    | sdw3        |
| SNP200 | 2H  | 74.37    | 0.25| 0.0162  | 1.79      | 0.54      |        |     |               |            |
| SNP254 | 2H  | 130.01   | 0.20| 0.0117  | 1.93      | 0.56      | -4.57  | QTL16_PHT    | HT (StMo-2H2) |
| SNP256 | 2H  | 131.77   | 0.28| 0.0175  | 1.76      | 0.5       |        |     |               |            |
| SNP295 | 3H  | 36.49    | 0.13| 0.0124  | 1.91      | 0.55      |        |     |               |            |
| SNP303 | 3H  | 43.23    | 0.23| 0.0083  | 2.08      | 0.61      | -5.57  | QTL17_PHT    | HT:HalMo-3H |
| SNP304 | 3H  | 46.31    | 0.35| 0.0141  | 1.85      | 0.54      |        |     |               |            |
| SNP312 | 3H  | 52.50    | 0.25| 0.0002  | 3.55      | 1.15      |        |     |               |            |
| SNP313 | 3H  | 52.50    | 0.42| 0.0220  | 1.66      | 0.48      | -5.80  | QTL18_PHT    | uzr         |
| SNP404 | 3H  | 126.27   | 0.39| 0.0129  | 1.89      | 0.55      | 4.11   | QTL19_PHT    | sdw1/denso  |
| SNP406 | 3H  | 127.10   | 0.37| 0.0061  | 2.21      | 0.65      |        |     |               |            |
| SNP427 | 3H  | 155.09   | 0.06| 0.0120  | 1.92      | 0.56      | -2.68  | QTL20_PHT    |             |
| SNP430 | 3H  | 162.06   | 0.05| 0.0053  | 2.28      | 0.75      |        |     |               |            |
| SNP519 | 4H  | 80.79    | 0.18| 0.0306  | 1.51      | 0.4       | 3.70   | QTL21_PHT    | QHei.pil-4H.S |
| SNP575 | 5H  | 50.27    | 0.31| 0.0028  | 2.55      | 0.8       | 5.30   | QTL22_PHT    |             |
| SNP588 | 5H  | 51.30    | 0.41| 0.0159  | 1.80      | 0.5       |        |     |               |            |
| SNP623 | 5H  | 85.93    | 0.16| 0.0164  | 1.79      | 0.51      | 4.95   | QTL23_PHT    |             |
| SNP643 | 5H  | 110.26   | 0.10| 0.0133  | 1.88      | 0.41      | -5.58  | QTL24_PHT    | HT           |
| SNP654 | 5H  | 132.63   | 0.44| 0.0235  | 1.63      | 0.46      | -4.15  | QTL25_PHT    | QHei.pil-Sh.1 |
| SNP722 | 6H  | 12.54    | 0.42| 0.0229  | 1.64      | 0.41      | -5.11  | QTL26_PHT    |             |
| SNP724 | 6H  | 16.97    | 0.30| 0.0033  | 2.48      | 0.8       |        |     |               |            |
| SNP757 | 6H  | 55.36    | 0.09| 0.0060  | 2.24      | 0.64      | -8.54  | QTL27_PHT    |             |
| SNP766 | 6H  | 55.36    | 0.32| 0.0092  | 2.04      | 0.6       |        |     |               |            |
| SNP831 | 6H  | 124.85   | 0.28| 0.0038  | 2.42      | 0.74      | -4.89  | QTL28_PHT    |             |
| SNP882 | 7H  | 73.75    | 0.45| 0.0210  | 1.68      | 0.46      | -4.23  | QTL29_PHT    | HT           |
| SNP947 | 7H  | 144.45   | 0.41| 0.0301  | 1.52      | 0.43      | -3.77  | QTL30_PHT    |             |
due to population structure [4,52,86]. Several statistical models to correct for the effect of population structure have been proposed and tested in previous studies [37,52,87]. Since we detected a considerable amount of structure in the present panel we used linear models to control for population structure and to reduce the false positive associations. Similar to the previous studies of comparing GWAS models in allogamous and autogamous species [37,52], our results suggest that K-model, QK model and PK model performed better than others (Figure 7). Moreover, for the K-model computational time is faster and no additional steps like identifying appropriate population structure (Q-matrix) in the panel are required. Since in an exploratory analysis mostly consistent results were obtained for all three approaches, the K-model was employed in the complete analysis of all traits to avoid redundancy of data. Still it should be kept in mind that correcting for population structure not only reduces the frequency of false positives but also may entail false negatives in situations where a

Table 9 GWAS results for trait thousand grain weight

| Marker | Chr | Position | MAF   | P-value | -log10(P) | R^2 (%) | Effect | QTL | Reference QTL | Literature |
|--------|-----|----------|-------|---------|-----------|---------|--------|-----|----------------|-----------|
| SNP48  | 1H  | 55.49    | 0.47  | 0.0288  | 1.56      | 0.34    | -2.19  | QTL1_TGW |               |           |
| SNP56  | 1H  | 61.35    | 0.26  | 0.0128  | 1.92      | 0.39    |        |      |                |           |
| SNP62  | 1H  | 66.70    | 0.25  | 0.0019  | 2.77      | 0.70    | 2.59   | QTL2_TGW |               |           |
| SNP68  | 1H  | 72.43    | 0.09  | 0.263   | 1.60      | 0.37    |        |      |                |           |
| SNP76  | 1H  | 87.62    | 0.21  | 0.0225  | 1.67      | 0.38    |        |      |                |           |
| SNP78  | 1H  | 88.23    | 0.26  | 0.0019  | 2.79      | 0.69    | 2.29   | QTL3_TGW |               |           |
| SNP81  | 1H  | 92.04    | 0.28  | 0.0208  | 1.70      | 0.38    |        |      |                |           |
| SNP137 | 2H  | 38.03    | 0.35  | 0.0259  | 1.61      | 0.40    | -1.20  | QTL4_TGW |               |           |
| SNP171 | 2H  | 63.53    | 0.10  | 0.006   | 2.26      | 0.50    |        |      |                |           |
| SNP174 | 2H  | 63.53    | 0.41  | 0.029   | 1.55      | 0.35    | 2.27   | QTL5_TGW | QTgw.pil-2H.2 | [67,72] |
| SNP175 | 2H  | 63.53    | 0.10  | 0.006   | 2.26      | 0.50    |        |      |                |           |
| SNP210 | 2H  | 82.75    | 0.36  | 0.0081  | 2.13      | 0.51    |        |      |                |           |
| SNP215 | 2H  | 86.63    | 0.32  | 0.0267  | 1.59      | 0.38    | 1.33   | QTL6_TGW |               |           |
| SNP245 | 2H  | 117.94   | 0.42  | 0.0084  | 2.11      | 0.51    | -1.66  | QTL7_TGW |               |           |
| SNP262 | 2H  | 139.65   | 0.31  | 0.0091  | 2.07      | 0.49    | -1.62  | QTL8_TGW |               |           |
| SNP305 | 3H  | 47.09    | 0.16  | 0.0232  | 1.67      | 0.39    | 3.01   | QTL9_TGW |               |           |
| SNP385 | 3H  | 98.49    | 0.37  | 0.0131  | 1.91      | 0.45    | 1.94   | QTL10_TGW |               |           |
| SNP395 | 3H  | 111.42   | 0.37  | 0.0302  | 2.61      | 0.40    | 1.75   | QTL12_TGW | int-c         | [29,60] |
| SNP458 | 4H  | 26.19    | 0.34  | 0.0224  | 1.67      | 0.40    |        |      |                |           |
| SNP460 | 4H  | 26.66    | 0.26  | 0.0034  | 2.52      | 0.63    |        |      |                |           |
| SNP467 | 4H  | 40.36    | 0.33  | 0.0007  | 3.21      | 0.74    |        |      |                |           |
| SNP469 | 4H  | 40.36    | 0.17  | 0.0006  | 3.28      | 0.82    | 2.52   | QTL13_TGW | QTgw.pil-4H.3 | [72]    |
| SNP643 | 5H  | 110.26   | 0.10  | 0.0312  | 1.52      | 0.28    | -3.00  | QTL14_TGW | QTgw.pil-5H.2 | [72]    |
| SNP663 | 5H  | 142.2    | 0.16  | 0.0004  | 3.45      | 0.87    |        |      |                |           |
| SNP664 | 5H  | 142.2    | 0.16  | 0.0002  | 3.79      | 1.00    | 4.47   | QTL15_TGW | QGwe.TaER-5H.2 | [74]    |
| SNP666 | 5H  | 142.2    | 0.15  | 0.0012  | 3.00      | 0.74    |        |      | QTgw.pil-5H.4 | [72]    |
| SNP709 | 5H  | 187.38   | 0.28  | 0.0082  | 2.12      | 0.50    | 2.02   | QTL16_TGW | QTgw.pil-5H.4 | [72]    |
| SNP739 | 6H  | 43.83    | 0.08  | 0.016   | 1.82      | 0.43    |        |      |                |           |
| SNP740 | 6H  | 44.77    | 0.42  | 0.0041  | 2.43      | 0.6     |        |      |                |           |
| SNP741 | 6H  | 44.77    | 0.41  | 0.0064  | 2.23      | 0.55    | -1.91  | QTL17_TGW |               |           |
| SNP770 | 6H  | 55.94    | 0.31  | 0.003   | 2.58      | 0.54    |        |      |                |           |
| SNP851 | 7H  | 34.82    | 0.43  | 0.0056  | 2.29      | 0.52    |        |      | QGwe.HaTR-7H.1 | [75]    |
| SNP854 | 7H  | 37.55    | 0.36  | 0.0277  | 1.58      | 0.32    | -1.88  | QTL18_TGW |               |           |
| SNP919 | 7H  | 88.65    | 0.13  | 0.0164  | 1.81      | 0.43    | 3.01   | QTL19_TGW |               |           |
| SNP934 | 7H  | 129.91   | 0.24  | 0.0048  | 2.36      | 0.59    | 1.84   | QTL20_TGW |               |           |
| SNP944 | 7H  | 143.68   | 0.12  | 0.0315  | 1.52      | 0.27    | -1.37  | QTL21_TGW | QTgw.HaTR-7H.1 | [72]    |

Significant markers associated for TGW, their MAF, P-value of association, variance explained by marker (R^2), effect of most significant marker within the QTL interval, name of the QTL, and reference QTL or gene from literature.
character state is strongly correlated with population structure [28].

In order to confirm the efficiency and resolution of the panel for association mapping using the range of available markers, we remapped all 918 SNPs using the K-model. From 918 SNPs, 783 were remapped within 10 cM of their original positions. Only 14% of the markers mapped beyond 10 cM. Among the successfully remapped markers more than 95% markers are within 5 cM distance from the original map position indicating the mapping resolution of our panel (see Figure 6).

Rostoks et al. [24] has used the same approach to evaluate their barley collection for GWAS with a subset of markers and successfully mapped 80% of the markers.

To demonstrate the suitability of the panel and the model for GWAS, we first analyzed spike morphology (row type) (Figure 8). This trait can be easily scored and is important from the agronomic and the domestication point of view. The genetic basis of row type is already well known and several QTL have been mapped and genes have been cloned [29,60,88]. We identified 34 marker-trait associations for this trait (Figure 8).
identified marker-trait associations for row type are current with all previously identified major loci—vrs1 [88], vrs2, vrs3, vrs4 and int-c [29,30]. Additional, less significant associations detected for row type could not be associated to any known major loci, and need to be further explored. These results for row type act as a proof of concept for GWAS in our spring barley panel and reflect the efficiency of GWAS for high resolution QTL mapping in inbreeding species. Some of the row type QTL overlapped with associated regions for other traits, especially with the traits TGW, SC and CPC (Additional file 12: Figure S7). As expected, two-rowed barley has higher TGW than the six-rowed types, as the number of sink organs (kernels) in two-rowed spikes is smaller than in six-rowed spikes. While the effect of spike architecture on TGW is clearly pleiotropic, its influence on SC and CPC is the result of breeding history and end use quality. In case of malting barley, varieties are generally bred for high starch and low protein content [22]. As a result, the two-rowed types in our panel have higher starch content and lower protein content than

| SNP  | Chr | Position | MAF  | P-Value | -log10 (P) | R^2 (%) | Effect | QTL | Reference QTL | Literature |
|------|-----|----------|------|---------|------------|---------|--------|-----|---------------|------------|
| SNP47 | 1H  | 55.49    | 0.50 | 0.0044  | 2.36       | 0.78    | 0.74  | QTL1_CPC     |             |
| SNP97 | 1H  | 114.84   | 0.25 | 0.0139  | 1.86       | 0.56    | -0.85 | QTL2_CPC     |             |
| SNP136 | 2H  | 33.74    | 0.35 | 0.0310  | 1.51       | 0.45    | 0.39  | QTL3_CPC     | QPc.nab-2H.1,Qcp2a [67,76] |
| SNP170 | 2H  | 63.53    | 0.32 | 0.0190  | 1.72       | 0.54    |        |     |               |            |
| SNP173 | 2H  | 63.53    | 0.33 | 0.0115  | 1.94       | 0.62    | 0.72  | QTL4_CPC     | QGpc.StMo-2H.2 [67,75] |
| SNP174 | 2H  | 63.53    | 0.41 | 0.0055  | 2.26       | 0.74    |        |     |               |            |
| SNP177 | 2H  | 63.53    | 0.32 | 0.0116  | 1.94       | 0.61    |        |     |               |            |
| SNP200 | 2H  | 74.37    | 0.25 | 0.0071  | 2.15       | 0.72    | -0.56 | QTL5_CPC     |             |
| SNP205 | 2H  | 78.03    | 0.40 | 0.0296  | 1.53       | 0.47    |        |     |               |            |
| SNP226 | 2H  | 96.82    | 0.23 | 0.0056  | 2.25       | 0.66    | -0.90 | QTL6_CPC     | QPc.nab-2H.1 [67] |
| SNP227 | 2H  | 96.82    | 0.19 | 0.0160  | 1.80       | 0.55    |        |     |               |            |
| SNP244 | 2H  | 116.49   | 0.24 | 0.0242  | 1.62       | 0.48    | -0.47 | QTL7_CPC     | QGpc.HaMo-2H.2 [75] |
| SNP272 | 2H  | 147.94   | 0.26 | 0.0103  | 1.99       | 0.62    | -0.52 | QTL8_CPC     |             |
| SNP305 | 3H  | 47.09    | 0.16 | 0.0020  | 2.70       | 0.86    | -1.466 | QTL9_CPC     |             |
| SNP322 | 3H  | 55.57    | 0.18 | 0.0093  | 2.03       | 0.64    |        |     |               |            |
| SNP357 | 3H  | 72.26    | 0.32 | 0.0159  | 1.80       | 0.51    | -0.47 | QTL10_CPC    |             |
| SNP401 | 3H  | 122.14   | 0.24 | 0.0062  | 2.21       | 0.68    | 0.60  | QTL11_CPC    | Qcp3a [76] |
| SNP409 | 3H  | 130.82   | 0.41 | 0.0146  | 1.84       | 0.54    |        |     |               |            |
| SNP518 | 4H  | 79.58    | 0.45 | 0.0006  | 3.22       | 1.09    | -0.75 | QTL12_CPC    | QGpc.HaTR-4H.2 [77] |
| SNP531 | 4H  | 97.06    | 0.11 | 0.0281  | 1.55       | 0.45    | 0.79  |        |               |            |
| SNP534 | 4H  | 101.62   | 0.16 | 0.0025  | 2.60       | 0.87    |        |     | QGpc.HaMo-4H [69] |
| SNP537 | 4H  | 108.70   | 0.21 | 0.0160  | 2.80       | 0.92    |        |     |               |            |
| SNP616 | 5H  | 74.78    | 0.51 | 0.0107  | 1.97       | 0.54    | -0.814 | QTL14_CP     | QGpc.HaMo-5H [67,75] |
| SNP623 | 5H  | 85.93    | 0.16 | 0.0082  | 2.09       | 0.61    |        |     |               |            |
| SNP643 | 5H  | 110.26   | 0.10 | 0.0055  | 2.26       | 0.73    | 1.52  | QTL15_CPC    | QGpc.DiMo-5H.2 [78] |
| SNP699 | 5H  | 171.66   | 0.11 | 0.0219  | 1.66       | 0.53    |        |     | QTL16_CPC     |             |
| SNP807 | 6H  | 83.89    | 0.25 | 0.0020  | 2.70       | 0.91    | 0.67  | QTL17_CPC    |             |
| SNP844 | 7H  | 12.42    | 0.10 | 0.0214  | 1.67       | 0.51    | 0.79  | QTL18_CPC    |             |
| SNP855 | 7H  | 38.32    | 0.35 | 0.0019  | 2.72       | 0.91    | -0.86 | QTL19_CPC    |             |
| SNP860 | 7H  | 46.19    | 0.26 | 0.0314  | 1.50       | 0.45    |        |     |               |            |
| SNP871 | 7H  | 61.32    | 0.24 | 0.0285  | 1.55       | 0.46    | -0.50 | QTL20_CPC    | QPc.nab-7H [67] |
| SNP904 | 7H  | 80.94    | 0.22 | 0.0026  | 2.44       | 0.85    | -0.68 | QTL21_CPC    | QGpc.HaTR-7H [77] |
| SNP925 | 7H  | 112.46   | 0.37 | 0.0210  | 1.68       | 0.49    | 0.41  | QTL22_CPC    |             |
| SNP930 | 7H  | 121.09   | 0.07 | 0.0001  | 4.73       | 1.57    | 1.54  | QTL23_CPC    |             |

Significant markers associated for CPC with K-model, their MAF, P-value of association, variance explained by marker (R^2), effect of the most significant marker within the QTL interval, name of the QTL, and the reference QTL or gene from literature.
six-rowed types (Additional file 5: Table S3). As expected, the landraces included in the panel did not show this stratification as they did not underly this selection pressure.

Heading date (HD) reflects the adaptation of a plant to its environment and is a complex trait effected by numerous QTL both in outbreeding [89] and in inbreeding species [61]. Many SNP markers were found to be associated with the trait HD (Figure 9a) and we report a total of 34 significant SNPs defining 19 QTL. Some of these QTL hit genomic regions that were previously reported to harbor major genes including HvFT3, PpdH1, HvFT4, eps2, HvGl1, HvCO3, HvFT1 and HvCO1 (Table 7). In a previous study using the same panel, fragments from three flowering time candidate genes were re-sequenced and SNPs within the gene PpdH1 revealed the largest effects on HD [14]. In the present GWAS, SNPs located in the vicinity (ca. 2 cM) of PpdH1 showed significant associations with HD (Table 7). By further including all PpdH1 SNPs from Stracke et al. [14] into our GWAS, these SNPs revealed the highest association of all markers used (Figure 10). These findings lend strength to the hypothesis that a further increase in marker coverage will either lead to the detection of additional associations or improve the significance of existing QTLs.

For the trait PHT we found 19 putative QTL regions located on chromosomes 2H, 3H, 4H, 5H, 6H and 7H comprising 32 marker trait associations. Semi-dwarf and dwarf cultivars have been developed worldwide to reduce lodging and to improve the harvest index. Different genes/alleles have been deployed in different geographic regions: the GA sensitive sdw1 dwarfing gene has been deployed in America and Australia, while its allelic form, termed denso, is frequently seen in European two-rowed germplasm. The recessive uzu allele is found in Japanese, Chinese and Korean cultivars [70,90]. Many QTL for PHT coincide with previously mapped QTL and genes (Table 8). The QTL4_PHT on chromosome 2H coincides with the mapping position of sdw3 which plays a major role in gibberellins-insensitive dwarfing barley [68]. Two allelic forms of the dwarfing gene denso/sdw1 map to the same genomic region as QTL8_PHT located on the long arm of chromosome 3H [70]. The QTL7_PHT is about 10 cM distant from the uzu locus based on the consensus map presented in grain genes database.

Thousand grain weight (TGW) is one of the major yield components having direct effect on the final yield. Altogether 21 QTL were found for this trait and some of them are in vicinity of row type genes. Some of the QTL were further confirmed by previously mapped QTL in same genomic regions (Table 9). QTL14_TGW on 5HL is observed to effect other traits like PHT, SC and CPC.

As outlined above, starch and protein content of the grain are major determinants of the end use quality. Several of the 25 QTL detected for starch content coincided with the previously identified QTL (Table 10). These include QTL for related traits like acid detergent fiber (ADF) content, starch granule size and granule shape [76]. QTL21_SC on 7H is located in the region of the waxy locus known to encode granule-bound starch synthase I (GBSS I), which catalyses the synthesis of amylose [91,92]. For the total grain crude protein content we identified 23 QTL, located on all the seven chromosomes. Eleven of these QTL regions co-localize with previously mapped QTL, while 12 QTL are novel (Table 11). Interestingly, the majority of QTL for traits SC and CPC are located on chromosome 7H. Some of the QTL identified for SC coincide with QTL for CPC e.g. chromosomes 1H (55 cM), 2H (33.74 cM), 3H (55 cM), 5H (110 cM) and 7H (12 cM and 121 cM) (Table 10 & 11). The coincidence of the QTL for these two traits can be expected due to their negative correlation (Table 2). On the other hand, we cannot rule out that some of the shared QTL are the result of linkage of underlying genes.

**GWA reveals small effects only**

Even the best associations observed in the present study showed only modest $R^2$ values (percentage of genetic trait variation explained) for the corresponding SNPs, implying low variance predicted by each SNP. This is exemplified by the QTL ‘Qsch7a’, which in a biparental QTL mapping study explained 47% of variation in SC [76]. In the present study, ‘QTL23_SC’ located in the same genomic region as ‘Qsch7a’ explains only 0.2% of the variation. Many GWAS in humans have reported low $R^2$ values and the rest of the unexplained variation is termed as ‘unexplained missing heritability’ [93]. Roy et al. [94], among others, reported $R^2$ values to range from 0.2% to 3.95% in GWAS for plants, which corresponds well with our present results. In a consorted study for the trait “body height”, an impressive number of 40 genotypic variants have been identified under a stringent threshold. Together these were only able to explain around 5% of the variation in human body height [95,96]. Possible explanations for this “missing heritability” include i) insufficient marker coverage, in cases where the causal polymorphism is not in perfect LD with the genotyped SNP reduces the power to detect associations and the variation explained by such a SNP marker. This has been demonstrated in the present study for the effect of the PpdH1 gene on HD; ii) rare alleles (MAF < 5%) with a major effect have been dropped from the analysis and will go undetected in cases where they are associated; iii) the expression of a character or trait depends on a large number of genes/QTL with small individual effects which escape statistical detection; iv) inadequacy of the statistical approaches available to detect epistatic interactions in GWAS and v) biased estimates of $R^2$ for individual SNPs due to the level of population stratification in
Conclusions

Overall, our results provide new details on the chances and pitfalls of GWAS in structured populations of inbreeding crops like barley. Results from the present study provide an insight into the genetic architecture of important agronomic traits for barley (HD, PHT, TGW, SC and CPC). In total, we identified 107 QTL for these traits. Some genomic regions harbor QTL for more than one trait and, based on map comparisons, 50 QTL have been found to concur with previously mapped QTL. For all traits together, 57 novel QTL have been detected. To mitigate the shortcomings of GWAS in inbreeding crops, future association studies might implement novel strategies such as joint linkage and LD mapping which were already successfully applied in various species [89,100-102]. Furthermore, to fine map and "mendelize" selected QTLs, staggered patterns of LD decay observed for different gene pools of barley (cultivars, landraces, wild barley) may be exploited in combination with biparental mapping and marker saturation strategies exploiting the ever increasing body of genomic sequence [30,103]. The feasibility of such an approach was recently demonstrated by identifying a candidate gene for the ANTHOCYANINLESS 2 locus using a combination of association mapping followed by a segregation analysis in a biparental population and a BAC contig analysis [104].

Additional material

Additional file 1: Table S1 Information of 957 mapped SNP markers from the IPK customized OPA that were successful in our panel.

Additional file 2: Table S2 Details of the 212 accessions used for GWAS. Name of the accession, row type, number of successful markers, Structure group, region of origin and country of origin.

Additional file 3: Figure S1 STRUCTURE results using DArT markers. Log probability data (LnP(D)) as function of $K$ (number of clusters) from the STRUCTURE run using 1088 DArT markers with the same association panel. The plateau of the graph at $K = 6$ indicates the minimum number of subgroups possible in the panel.

Additional file 4: Figure S2 Phenotypic distribution of 224 spring barley accessions for the traits heading date (HD), plant height (PHT), thousand grain weight (TGW), starch content (SC) and protein content (CPC).

Additional file 5: Table S3 Phenotypic variation among two-rowed and six-rowed groups. Estimation of means, standard deviation (SD), variation (VAR), standard error variation (SEVAR) and coefficient of variance (CV%) for each trait among two-rowed and six-rowed groups.

Additional file 6: Table S4 Estimation of means, SD, variation (VAR), standard error variation (SEVAR) and coefficient of variance (CV%) among all six subgroups in the panel.

Additional file 7: Figure S3 Comparison of BLUPs and BLUEs for starch content. The graph implies that there is not much difference between the BLUPs and BLUEs in our experiment.

Additional file 8: Table S5 Marker polymorphism information of the 918 SNP markers used in GWAS in the panel.

Additional file 9: Figure S4 Principal Co-ordinate analysis (PCoA) of the panel based on the first two components derived using 918 SNPs. The primary axis tend to separate into subgroups based on their spike morphology character (blue: six-rowed barley; red: two-rowed barley). Further clustering is based on origin of the accessions.

Additional file 10: Figure S5 LD plots for each chromosome in barley. The color of squares illustrate the strength of pairwise $r^2$ values on a black and white scale, where black indicates perfect LD ($r^2 = 1.00$) while white indicates perfect equilibrium ($r^2 = 0$). Failed and monomorphic SNPs as well as SNPs with MAF < 0.05 are not considered.

Additional file 11: Figure S6 GWAS whole genome scans for row type using different association models (naive, P, Q, GK, PK and K).

Additional file 12: Figure S7 GWAS for all traits. Localization of QTL and candidate genes for the traits row type (RT), heading date (HD), plant height (PHT), thousand grain weight (TGW), starch content (SC) and crude protein content (CPC) on the genetic map with 918 SNP markers.
Acknowledgements
We thank Kathrin Baake, Kerstin Wolf, Ute Krajewski, Jürgen Marlow and Peter Schreiber for their excellent technical assistance. Special thanks to Nils Stein and Kerstin Neumann for helpful discussions. We would like to thank all collaborating partners of the GABI-GENOBAR consortium and especially Stephanie Weise for bioinformatics support. We also acknowledge and thank the GABI-GENOPLANT partners for the support in field trials. This work was funded through the GABI program of the German Ministry of Education and Research (BMBF).

Author details
1Leibniz-Institute of Plant Genetics and Crop Plant Research (IPK), Correnstr. 3, 06466 Gatersleben, Germany. 2Biometris, Wageningen UR, P.O.Box 100, Wageningen, The Netherlands. 3Plant Breeding, Centre of Life and Food Sciences Weißenstephan, Technical University Munich, Freising, Germany.

Authors’ contributions
RKP carried out the study and performed data curing, association analysis, interpretation of the data and drafted the manuscript. RS participated in statistical analysis and improving the manuscript. MM and FAE advised on statistical analysis. GH generated the phenotypic data. BK coordinated the project, and assisted in GoldenGate data analysis and improving the manuscript. AG initiated the study, participated in data interpretation and in improvement of the manuscript. All the authors have read and approved the final manuscript.

Received: 14 July 2011 Accepted: 27 January 2012
Published: 27 January 2012

References
1. Flint-Garcia SA, Thornsberry JM, Buckler ES: Structure of linkage disequilibrium in plants. Annu Rev Plant Biol 2003, 54:357-374.
2. Zhu C, Gore M, Buckler ES, Yu J: Status and Prospects of Association Mapping in Plants. The Plant Genome Journal 2008, 1(1).
3. Hastbacka J, Delachapelle A, Kaitila I, Sistonen P, Weaver A, Lander E: Linkage Disequilibrium Mapping in Isolated Founder Populations - Diastrophic Dysplasia in Finland. Nature Genetics 1992, 2(3):204-211.
4. Lander ES, Schork NJ: Genetic dissection of complex traits. Science 1994, 265(5181):2037-2048.
5. Flint-Garcia SA, Thillault AC, Yu J, Pressoir G, Romero SM, Mitchell SE, Doebley J, Krevisch S, Goodman MM, Buckler ES: Maize association population: a high-resolution platform for quantitative trait locus dissection. Plant J 2005, 44(6):1054-1064.
6. Mackay I, Powell W: Methods for linkage disequilibrium mapping in crops. Trends in Plant Science 2007, 12(2):57-63.
7. Gupta PK, Rustgi S, Kulwal PL: Linkage disequilibrium and association studies in higher plants: Present status and future prospects. Plant Mol Biol Rev 2005, 57(4):461-485.
8. Tenailion MI, Samkis MC, Long AD, Gaut RL, Doebley JF, Gaut BS: Patterns of DNA sequence polymorphism along chromosome 1 of maize (Zea mays ssp. mays L). Proc Natl Acad Sci USA 2001, 98(16):9161-9166.
9. Caldwell KS, Russell J, Langridge P, Powell W: Extreme population-dependent linkage disequilibrium detected in an inbreeding plant species, Hordeum vulgare. Genetics 2006, 172(1):557-567.
10. Hirschhorn JN, Daly MJ: Genome-wide association studies for common diseases and complex traits. Nat Rev Genet 2005, 6(2):95-108.
11. Atwell S, Huang YS, Vilhjalmsson BJ, Willems G, Horton M, Li Y, Meng D, Piat A, Tarone AM, Hu TT, et al: Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature 2010, 465(7298):627-631.
12. Thornsberry JM, Goodman MM, Doebley J, Krevisch S, Nielsen D, Buckler ES: Dwarf8 polymorphisms associate with variation in flowering time. Nature Genetics 2001, 28(3):286-289.
13. Palaisa KA, Morgante M, Williams M, Rafalski A: Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci. Plant Cell 2003, 15(8):1795-1806.
14. Stadler S, Hasenjger G, Veyreras JB, Geiger HH, Sauer S, Gruber A, Piepho HP: Association mapping reveals gene action and interactions in the determination of flowering time in barley. Theor Appl Genet 2005, 118(2):259-273.
15. Singh A, Reimer S, Poznaino CJ, Clarke FR, Clarke JM, Knox RE, Singh AK: Allelic variation at Psyl1-A1 and association with yellow pigment in durum wheat grain. Theoretical and Applied Genetics 2009, 118(8):1539-1548.
16. Li YH, Zhang C, Gao ZS, Smulders MJM, Ma ZL, Liu ZX, Nan HY, Chang RZ, Qiu LJ: Development of SNP markers and haplotype analysis of the candidate gene for rhg1, which confers resistance to soybean cyst nematode in soybean. Mol Breeding 2009, 24(1):63-76.
17. Ehremsch H, Hanza W, Chou L, Roe JI, Kever PX, Purugian MD: Candidate Gene Association Mapping of Arabidopsis Flowering Time. Genetics 2009, 183(3):525-535.
18. Zhao KY, Aranazana MJ, Kim S, Lister C, Shindo C, Tang CL, Toomajian C, Zheng HG, Dean C, Marjoram P, et al: An Arabidopsis example of association mapping in structured samples. Plos Genetics 2007, 3(1).
19. Kilian B, Oezkan H, Kohl J, von Haeseler A, Barale F, Deusch O, Brandolini A, Rucel C, Martin W, Salamin F: Haplosyme structure at seven barley genes: relevance to gene pool bottlenecks, phylogeography of ear type and site of barley domestication. Molecular Genetics and Genomics 2006, 276(3):230-241.
20. Morrell PL, Clegg MT: Genetic evidence for a second domestication of barley (Hordeum vulgare) east of the Fertile Crescent. Proceedings of the National Academy of Sciences 2007, 104(9):3289-3294.
21. Kilian B, Oezkan H, Pozzi C, Salamin F: Domestication of the Triticeae in the Fertile Crescent. In Genetics and Genomics of the Triticeae Volume 7: Edited by: Muehlbauer G, feuillet C. Springer New York, 2009:81-119.
22. Hayes P, Sauc P: Disequilibrium and association in barley: thinking outside the glass. Proc Natl Acad Sci USA 2006, 103(49):18385-18386.
23. Schulte D, Close TJ, Graner A, Langridge P, Matsumoto T, Muehlbauer G, Sato K, Schulman AH, Waughr R, Wise RP, et al: The International Barley Sequencing Consortium—At the Threshold of Efficient Access to the Barley Genome. Plant Physiology 2009, 149(1):142-147.
24. Rostoks N, Ramsay L, MacKenzie K, Cardle L, Bhat PR, Roose ML, Svensson JT, Stein N, Varshney RK, Marshall DF, et al: Recent history of artificial outcrossing facilitates whole-genome association mapping in elite inbred crop varieties. Proc Natl Acad Sci USA 2006, 103(49):18656-18661.
25. Wenf P, Li H, Carling J, Zhou M, Raman H, Paul E, Heamden P, Maier C, Xia L, Caig V, et al: A high-density consensus map of barley linking DAR markers to SSR RFLP and STS loci and agricultural traits. BMC Genetics 2006, 7:206.
26. Close TJBP, Lernardi S, Wu Y, Rostoks N, Ramsay L, Druka A, Stein N, Svensson JT, Wannamanuk S, Bozdag S, Roose ML, Moscou MJ, Chao S, Varshney RK, Sato K, Hayrer PM, Mathew DE, Kleinhoft A, Muehlbauer GJ, De’Young J, Marshall DF, Madihshetty K, Fenton RD, Condamin P, Graner A, Waughr R: Development and implementation of high-throughput SNP genotyping in barley. BMC Genetics 2009, 10:582.
27. Hasenjger G, Stracke S, Piepho HP, Sauver S, Geiger HH, Graner A: DNA polymorphisms and haplotype patterns of transcription factors involved in barley endosperm development are associated with key agronomic traits. BMC Plant Biology 2010, 10.
28. Cockram J, Leigh FJ, Lea VJ, Chiapparino E, Laurie DA, Mackay IJ, Powell W, O’Sullivan DM: Association mapping of partitioning loci in barley. BMC Genomics 2008, 9:16.
29. Ramsay L, Comadran J, Druka A, Marshall DF, Thomas WTB, Macaluy M, MacKenzie K, Simpson C, Fuller J, Bonton N, et al: INTERMEDIUM-C, a modifier of lateral spikelet fertility in barley, is an ortholog of the maize domestication gene TEOSINTE BRANCHED 1. Nat Genet 2011, 43(2):169-172.
30. Waughr R, Jannink JL, Muehlbauer GJ, Ramsay L: The emergence of whole genome association scans in barley. Curr Opin Plant Biol 2009, 12(2):218-222.
31. Michael Thiel, Graner Varshney. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (Hordeum vulgare L). TAG Theoretical and Applied Genetics 2003, 106(3):411-422.
32. Brescghello F, Sorrells ME: Association mapping of kernel size and milling quality in wheat (Triticum aestivum L.) cultivars. Genetics 2006, 172(2):1165-1177.
33. Miles S, Peiffer J, Brown PJ, Erooz ES, Zhang Z, Costich DE, Buckler ES: Association mapping: critical considerations shift from genotyping to experimental design. Plant Cell 2009, 21(8):2194-2202.
Malosetti M, van der Linden CG, Vosman B, van Eeuwijk FA: A Mixed-Model Approach to Association Mapping Using Pedigree Information With an Illustration of Resistance to Phytophthora infestans in Potato. Genetics 2007, 175(2):879-889.

Pritchard JK, Stephens M, Rosenberg NA, Donnelly P: Association mapping in structured populations. American Journal of Human Genetics 2000, 67(1):170-181.

Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamazaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, et al: A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 2006, 38(2):203-208.

Stich B, Melchinger AE: Comparison of mixed-model approaches for association mapping in rapeseed, potato, sugar beet, maize, and Arabidopsis. BMC Genomics 2009, 10.

Haseneyer G, Stracke S, Paul C, Einfelder C, Broda A, Piepho HP, Graner A, Geiger HH: Population structure and phenotypic variation of a spring barley world collection set up for association studies. Plant Breeding 2010, 129(3):271-279.

Knupffer H, van Hintum T: Chapter 13 Summarised diversity—the Barley Core Collection. In Developments in Plant Genetics and Breeding. Volume 7. Edited by: von Bothmer R, van Hintum T; Knupffer H, Kauzhuoro S; Elsevier; 2003:259-267.

Naumann C, Basler R: VDLUFA-Methodenbuch III: Die Chemische Untersuchung von Futtermitteln, 5. Ergänzungslieferung, Darmstadt, VDLUFA; 2004.

Payne RW, Murray DA, Harding SA, Baird DB, Soutar DW: GenStat for Windows (9th Edition) Introduction. VSN International, Hemel Hempstead; 2008.

Fan JB, Gunderson KL, Bibikova M, Yeakley JM, Chen J, Wickham Garcia E, Lebrato LL, Laurent M, Shen R, Barker D: Illumina Universal Bead Arrays. In Methods in Enzymology. Volume 470. Edited by: Alan K, Brian O: Academic Press; 2006:57-73.

Fan JB, Oliphant A, Shen R, Kermani BG, Garcia F, Gunderson KL, Hansen M, Steemers F, Butler SL, Deloukas P, et al: Highly parallel SNP genotyping. Cold Spring Harbor Symp Quant Biol 2003, 68:69-78.

Liu K, Muze S: PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 2005, 21(9):2128-2129.

Nie M: Genetic Distance between Populations. Am Nat 1972, 106(949):283.

Falush D, Stephens M, Pritchard JK: Inference of Population Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele Frequencies. Genetics 2003, 164(4):1567-1587.

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

Perrier X, Jacquemond-Collet JP: DA Twinn Software. 2006 [http://darwin.cirad.fr/dartwin].

Genten JC, Fry P, Maller J, Daly MJ: Haleypewe: analysis and visualization of LD and haplotype maps. Bioinformatics 2005, 21(2):263-265.

Weir BS: Genetic Data Analysis II: Methods for Discrete Population Genetics. Sinauer Associates, Sunderland, Massachusetts; 1996.

Bradbury PJ, Zhang Z, Kroon DA, Casstevens TM, Buckler ES: TASSEL: software for association mapping of complex traits in diverse populations. Bioinformatics 2007, 23(19):2639-2645.

Kang HM, Zaitlen NA, Wade CM, Kirby A,-Yassin N, et al: Mixed model association scans of multi-environmental trial data reveal major loci controlling yield and yield related traits in Hordeum vulgare in Mediterranean environments. Theoretical and Applied Genetics 2011, 122(7):1363-1373.

Griffiths S, Dunford RP, Coupal G, LaBerge DE: The evolution of CONSTANS-like gene families in barley, rice, and Arabidopsis. Plant Physiol 2003, 131(4):1855-1867.

Qi X, Nikes RE, Stam P, Lindhout P: Identification of QTLs for partial resistance to leaf rust (Puccinia hordei) in barley. Theoretical and Applied Genetics 1996, 93(12):1205-1215.

Marquez-Cedillo LA, Hayes PM, Kleinrohs A, Legge WG, Rossnagel BG, Sato K, Ullrich SE, Wesenberg DM, Proj NABQAT: QTL analysis of agronomic traits in barley based on the doubled haploid progeny of two elite North American varieties representing different germplasm groups. Theoretical and Applied Genetics 2001, 103(4):625-637.

Gottwald S, Stein N, Borner A, Sasaki T, Graner A: The gibberellin-ac insensitive dwarfin gene sdw1 of barley is located on chromosome 2HS in a region that shows high colinearity with rice chromosome 7L. Mol Genet Genomics 2004, 271(4):426-436.

Hayes PM, Liu BH, Knapp SJ, Chen J, Jones B, Blake T, Franckowiak J, Rasmusson D, Sorells M, Ullrich SE, et al: Quantitative Trait Locus Effects and Environmental Interaction in a Sample of North-American Barley Germplasm. Theoretical and Applied Genetics 1993, 87(3):392-401.

Jia QJ, Zhang XQ, Westcott S, Broughton S, Cakir M, Yang J, Lance R, Li CD: Expression level of a gibberellin 20-oxidase gene is associated with multiple agronomic and quality traits in barley. Theoretical and Applied Genetics 2011, 122(8):1451-1462.

Yin X, Stam P, Ourlieren CJ, Kooppa A: ARL mapping of quantitative trait loci for yield-determining physiological characters in spring barley. Theoretical and Applied Genetics 1999, 99(1-2):244-253.

Pillen K, Zacharias A, Leon J: Advanced backcross QTL analysis in barley (Hordeum vulgare L). Theoretical and Applied Genetics 2007, 110(7):1350-1359.

von Korff M, Wang H, Leon J, Pillen K: ABQTL analysis in spring barley: II. Detection of favourable exotic alleles for agronomic traits introgressed from wild barley (H-vulgare ssp spontaneum). Theoretical and Applied Genetics 2006, 112(7):1221-1231.

von Korff M, Grando S, Del Greco A, This D, Baum M, Ceccarelli S: Quantitative trait loci associated with adaptation to Mediterranean dryland conditions in barley. Theoretical and Applied Genetics 2008, 117(5):633-649.

Szczesny P, Blake YC, Bhat PR, Chao S, Close TJ, Cuesta-Marcos A, Muehlbauer GJ, Ramsay L, Waugh R, Hayes PM: An Integrated Resource for Barley Linkage Map and Malting Quality QTL Alignment. Plant Gen 2009, 2(1):134-140.

Abdel-Haleem H, Bowman J, Giros M, Karzian V, Talbert H, Surber L, Blake T: Quantitative trait loci of acid detergent fiber and grain chemical composition in hulled x hull-less barley population. Euphytica 2010, 172(3):405-418.

Mather DE, Tinker NA, LaBerge DE, Edney M, Jones BL, Rossnagel BG, Legge WG, Briggs KG, Irvine RB, Falk DE, et al: Regions of the genome that affect grain and malt quality in a North American two-row barley cross. Crop Sci 1997, 37(2):544-554.
Hints of hidden heritability in GWAS.

Using association mapping to dissect the genetic basis of complex traits in plants. Brief Funct Genomics 2010, 9(2):157-165.

Molecular Dissection of a Quantitative Trait Locus in Mice Using Heterogeneous Stocks. Genetics 2002, 160(4):1609-1618.

Unlocking the Barley Genome by Chromosomal and Comparative Genomics. Plant Cell 2011, 23(4):1249-1263.

Genome-wide association mapping to candidate polymorphism resolution in the unsequenced barley genome. Proc Natl Acad Sci USA 2010, 107(50):21611-21616.

doi:10.1186/1471-2229-12-16

Cite this article as: Pasam et al.: Genome-wide association studies for agronomical traits in a world wide spring barley collection. BMC Plant Biology 2012 12:16.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Direct link to your published research
-publication without charge
- Open access (articles freely available online to all
researchers)
- Approximately 5633 full text views in the first year
- High visibility and impact

Submit your manuscript at www.biomedcentral.com/submit