

Prediction of malting quality traits in barley based on genome-wide marker data to assess the potential of genomic selection

Malthe Schmidt¹ · Sonja Kollers¹ · Anja Maasberg-Prelle¹ · Jörg Großer¹ · Burkhard Schinkel¹ · Alexandra Tomerius^{1,2} · Andreas Graner³ · Viktor Korzun¹

Received: 6 August 2015 / Accepted: 11 November 2015 / Published online: 9 December 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract

Key message Genomic prediction of malting quality traits in barley shows the potential of applying genomic selection to improve selection for malting quality and speed up the breeding process.

Abstract Genomic selection has been applied to various plant species, mostly for yield or yield-related traits such as grain dry matter yield or thousand kernel weight, and improvement of resistances against diseases. Quality traits have not been the main scope of analysis for genomic selection, but have rather been addressed by marker-assisted selection. In this study, the potential to apply genomic selection to twelve malting quality traits in two commercial breeding programs of spring and winter barley (*Hordeum vulgare* L.) was assessed. Phenotypic means were calculated combining multilocational field trial data from 3 or 4 years, depending on the trait investigated. Three to five locations were available in each of these years. Heritabilities for malting traits ranged between 0.50 and 0.98. Predictive abilities (PA), as derived from cross validation,

ranged between 0.14 to 0.58 for spring barley and 0.40–0.80 for winter barley. Small training sets were shown to be sufficient to obtain useful PAs, possibly due to the narrow genetic base in this breeding material. Deployment of genomic selection in malting barley breeding clearly has the potential to reduce cost intensive phenotyping for quality traits, increase selection intensity and to shorten breeding cycles.

Introduction

Since its first introduction in 1998 (Haley and Visscher) and its extension with methods like ridge regression BLUP and Bayesian methods to calculate genomic estimated breeding values (Meuwissen et al. 2001), genomic selection (GS) has become a highly attractive complement or even alternative to conventional selection strategies employed in animal and plant breeding (Desta and Ortiz 2014; Henryon et al. 2014). The advantages of GS, in comparison to phenotypic selection, include the ability to predict breeding values within families (as opposed to only between families in pedigree based approaches), to generate and use genetic information for the planning of crosses, the provision of phenotypic values if they cannot be observed in a given season (e.g., no disease pressure present), and, in particular, the shortening of the breeding cycle (Hickey et al. 2014).

Owing to the limited availability of marker resources in the past, marker-assisted selection in barley was essentially restricted to monitor the inheritance of single QTL alleles with large effects, or of monogenic traits such as disease resistances (e.g., Graner et al. 1999; Hofmann et al. 2013). Regarding the latter, marker-assisted selection represents an instrumental tool for the rapid introgression of monogenic traits from unadapted germplasm into elite lines (König

Communicated by D. E. Mather.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-015-2639-1) contains supplementary material, which is available to authorized users.

✉ Viktor Korzun
viktor.korzun@kws.com

¹ KWS LOCHOW GMBH, Ferdinand-von-Lochow-Straße 5, 29303 Bergen, Germany

² AIB Dr. Alexandra Tomerius, Saffeweg 32, 38304 Wolfenbüttel, Germany

³ Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, 06466 Stadt Seeland, Germany

et al. 2012; Werner et al. 2007) and to minimize linkage drag during subsequent backcross cycles (König et al. 2012; Miedaner and Korzun 2012; Werner et al. 2007).

Regarding quantitative traits, which are characterized by a large number of small effect loci, marker-assisted selection can harness only a limited portion of the genetic variation. By contrast, GS captures a much larger spread of genetic variation: because a large number of markers cover the entire genome, all chromosome segments are simultaneously considered for analysis to predict the breeding value of a given progeny line or genotype (Meuwissen et al. 2001). GS captures all QTL effects irrespective of size by estimating allele effects of all markers simultaneously, using phenotypes and marker data of the training set (TS). These effects are then used to estimate the breeding value of selection candidates, based solely on their allelic profile. The accuracy of GS is crucial for its success and effectiveness in a practical breeding scheme. Important factors influencing accuracy include heritability of the trait under consideration, number of markers used, statistical models employed, linkage disequilibrium, effective population size, relationship between TS and selection candidates, population structure and size of the TS (Isidro et al. 2015).

In plant breeding, GS has been employed mostly to yield and yield-related traits, e.g., thousand kernel weight in barley (Schmid and Thorwarth 2014), grain moisture and grain yield in maize (Endelman et al. 2014; Zhao et al. 2012), grain yield, plant height, starch- and total pentosan content in rye (Wang et al. 2014), grain yield, plant height, flowering time in rice (Spindel et al. 2015) or grain yield in wheat (Poland et al. 2012).

Barley is a major crop with wide adaptability grown on all continents except for Antarctica. About 60–70 % of the global production is used for feed, while malt for brewing, distilling and food makes up for 30–40 % (Shewry and Ullrich 2014, pp. 1–9). The use of barley as a substrate of fermentable carbohydrates and a source of enzymes for beer production requires high quality seed as a raw material. Malting quality is a complex trait that is described by a series of trait components which follow complex inheritance. A series of Quantitative Trait Loci (QTL) have been identified over the past decade, shedding light on the genetic architecture of malting quality and functional genomics studies have been performed to identify candidate genes (for review, see Shewry and Ullrich (2014, pp. 293–309).

These trait components are expensive to measure and require relatively large amounts of seed. Therefore, in practical breeding, quality tests are restricted to advanced generations and marker-assisted selection during early generations is limited to few genes and major QTLs, e.g., β -amylase activity (*Bmy1*) (Chiapparino et al. 2006) or β -glucan content (Han et al. 1997). Often the first reliable

phenotypic quality results are available just before entering National List Trials. Against this backdrop, genomic-estimated breeding values (GEBVs) have the potential to (a) save costs if used to select candidates with high malting quality without phenotyping and (b) to allow for a (pre) selection of the target quality traits in early stages of the breeding program.

To explore the applicability of GS in barley for the complex trait malting quality, this study aimed at the establishment of training populations for two malting barley breeding programs and to assess the potential of GS to efficiently improve malting quality components in elite germplasm within the framework of a practical barley breeding program.

Materials and methods

Phenotypic data and analysis

Plant material

The plant material used for spring and winter barley consists of elite material, i.e., already registered lines and/or potential candidates for registration. They have been selected for yield, yield stability, resistances and quality scores if available. In winter barley, all genotypes are DH lines. In spring barley, 80 % of the materials are DH lines and 20 % have been derived from an SSD breeding scheme. The developed breeding material, excluding listed varieties, remains the property of KWS LOCHOW GMBH and cannot be provided without specific legal agreement.

Traits measured and structure of data

For spring barley, phenotypic data were available for 4 years (2011–2014) with three to five locations per year. All trials were laid out as alpha lattice designs in two or three replications. Phenotypic trial structure in the spring barley data was not completely orthogonal. The number of lines per location within a year ranged from 7 (BAI, UK) to 279 (WOH, GER) with a mean of 107 lines per location within a year and 116 (2012) to 294 (2014) unique lines per year over the duration of the study. There were 12 check lines and up to overlapping 37 lines between years.

In winter barley, orthogonal sets of lines were grown at four locations in each of the three seasons 2010/2011 (2011), 2011/2012 (2012) and 2012/2013 (2013). In those seasons, different combinations of locations were used. In 2011 and 2012, 64 lines were tested in each location and 54 lines in 2013. Trials were laid out as alpha lattice designs with two replications per location. At least two overlapping locations across all years were present. The overall number of lines,

including feed type varieties, in trials was 148 with 12 check lines and up to 17 lines overlapping between different years.

Field trials were conducted by KWS LOCHOW GmbH using standard growing practices including fungicide treatment. Malting quality analyses were carried out by VLB (Versuchs- und Lehranstalt für Brauerei, Berlin, Germany).

For both spring and winter barley, 12 malting traits were analyzed in this study. Descriptions of these traits have been given in Potokina et al. (2004).

AMA = alpha amylase activity (dry extract units/g dry weight), AMB = beta amylase activity (dry extract units/g dry weight), EXT = extract (% dry weight), FAN_L = free amino nitrogen (mg/l), Fin_At = final attenuation (%), FRI = friability (%), GLU = beta glucan content (mg/l), KOL = kolbach index (%) (percentage of denatured proteins), LOSS = malting loss (while malting, added water leads to growing of radicals, these are cleaned and the loss of mass during cleaning is measured, associated with early germination) (%), NIT = soluble nitrogen (mg/100 g dry weight), PRT = protein content (% dry weight), VIS = viscosity (connected to GLU, high molecular carbon hydrates, solubility of carbon hydrates) (8.6 %, mPas*s).

Calculation of BLUEs and heritabilities

Phenotypic data for spring barley were analyzed in a two-stage model. First, a within trial single location analysis was carried out using Plabstat (Utz 2011). The final adjusted means over locations, trials, and years were estimated with a mixed model using ASreml software (Butler 2009). In this model genotype was set as fixed term and all other factors as random terms [main factors being year, location and trial and interaction factors (genotype:year), (genotype:location), (genotype:year:location)]. The connection between the years was accomplished with at least 12 check lines overlapping between years. Broad sense heritabilities were estimated with the same formula as for winter barley (see below) but without replication because single location results provided by the breeder were already adjusted for replication.

The phenotypic data for the winter barley material were analyzed in a single-stage mixed model approach combining the 3 years with ASreml. In the mixed model equation, genotype was treated as fixed effect while year and location as well as all their interactions with genotypes were assumed to be random effects, resulting in BLUEs (best linear unbiased estimators) for the phenotypic means of each genotype and trait. Twelve lines were grown in all 3 years and up to 17 lines connect two consecutive years.

Broad sense heritabilities for winter barley were estimated as:

$$h^2 = \frac{\sigma_G^2}{\sigma_G^2 + \frac{\sigma_{GL}^2}{L} + \frac{\sigma_{GY}^2}{Y} + \frac{\sigma_{GLY}^2}{Y*L} + \frac{\sigma_\epsilon^2}{Y*L*R}}$$

where σ^2_G , σ^2_{GL} , σ^2_{GY} , σ^2_{GLY} and σ^2_ϵ are the genotype, genotype x location, genotype × year, genotype × location × year and residual variance, respectively, and L , Y and R are the number of locations, years, and replications. Since numbers of years and locations varied between trials, harmonic means were used as the denominator instead of absolute values.

For some traits (AMA, Fin_At, KOL, Loss, VIS), genotype x environment interaction components could not be separated from the residual. In these cases, a reduced model was used with the main effects genotype, location, year and the residual error (containing all the interaction terms with genotype), which was divided by the harmonic mean of locations, years and two replications in the resulting reduced formula for h^2 .

Correlations between phenotypic means in different years were calculated from checks and other lines present across years using adjusted single-year data.

Genotypic data and analysis

Genotyping and data cleaning

An Illumina 9 k array was employed for genotypic analysis (Comadran et al. 2012). Raw data were cleaned by removing markers with a minor allele frequency (maf) ≤ 0.01 , more than 20 % missing data points, and more than 20 % heterozygous allele calls. Lines with ≥ 20 % missing data points and ≥ 15 % heterozygous calls were also excluded. Missing values were imputed using Beagle (Browning and Browning 2007) with a confidence level of 0.95, using only information based on neighboring markers, as identified from the genetic map. Where genotypically identical individuals were present, all but one were excluded from the training set. Out of 7916 potential markers, 3958 (50 %) were mapped on the Morex × Barke genetic map from Comadran et al. (2012), with a total chromosome length of 2127.4 cM.

Analysis of population structure

Population structure was estimated with principal component analysis. Genetic distances for population analysis were based on modified Roger's Distance, as implemented in the R package SelectionTools (Frisch 2015). To estimate the probability for the number of groups for spring and winter barley, the Calinski criterion (Calińska and Harabasz 1974) from the vegan R-package (Oksanen et al. 2014) was used.

Analysis of linkage disequilibrium

The Synbreed R-package (Wimmer et al. 2012) was used for creating the Linkage Disequilibrium (LD)-plots. LD-decay was estimated using the 95 % percentile of all

pairwise LD values from all mapped markers between chromosomes, excluding LD results from markers on the same chromosome. This value was used as cutoff LD or intercept to estimate the average LD size of each chromosome for the LD decay smoothing curve. The average across all chromosomes was used as final LD estimate. This was done separately for spring and winter barley.

Estimation of GEBVs

To estimate the genetic effects of the cleaned SNPs, the following linear model (RR-BLUP) was used, as described by Endelman (2011):

$$y = X\beta + Zu + \varepsilon$$

where y is a numeric vector of the adjusted means of the corresponding trait, β is a vector of fixed effects, X is the design matrix of fixed effects (here all = 1 because all effects are adjusted within the phenotypic analysis), Z is the marker matrix assigning marker genotypes to phenotypes (y) with dimension rows equal to the number of phenotypes and columns equal to the number of markers. Genetic effects are given in vector u and ε is the residual error. The latter were assumed to follow a normal distribution with mean 0 and variance σ^2u and $\sigma^2\varepsilon$, respectively. The sum of all allele effects is the GEBV of a line. Prediction error variances (PEV) to estimate the reliability of single GEBVs were calculated for all individuals and traits for spring and winter barley.

For estimating unshrunk GEBVs, a linear model (lm) was applied in which centered phenotypic and genotypic means were fitted using the following formula:

$$\begin{aligned} \text{GEBV}_{\text{us}} &= \bar{x}_{\text{Phenotypic means}} \\ &+ (\bar{x}_{\text{Centered GEBV}} - \text{Intercept (lm)}) / \\ &\text{Regression coefficient (lm)} \end{aligned}$$

Calculation of predictive abilities (PA) Predictive abilities were obtained separately for each trait from fivefold cross validation, replicated five times. Within one replication, the data set is randomly split into five sets of equal size, of which four are combined as “training set” to predict the fifth set. This is then repeated such that each of the sets is four times part of the training set and predicted once. The mean correlation of the predicted value to the observed phenotype over all folds and replications is the predictive ability.

Results

Traits and phenotypes

For ten of the twelve malting quality traits studied (AMA, AMB, EXT, FAN_L, Fin_At, FRI, KOL, LOSS, NIT and

PRT), the breeding goal is to increase the values in the breeding lines, whereas a reduction of the values is the goal for GLU and VIS. In the spring barley breeding material, the phenotypes were higher than in winter barley for six traits (AMA, EXT, FAN_L, Fin_At, FRI and KOL) and lower for GLU and VIS where those results are desired. The difference between spring and winter barley ranged from +19 % for FRI to –33 % for AMB. The mean standard deviation of the traits for spring barley was 31.74 and 34.66 for winter barley. In relation to the corresponding mean, spring barley showed a 12 % higher mean standard deviation of all traits on average (Supplementary Table S1).

Broad sense heritabilities for spring barley ranged between 0.50 for AMA and >0.80 for EXT, Fin_At, FRI, GLU, KOL and NIT. For winter barley, heritabilities ranged from 0.71 for KOL to 0.98 for FRI and GLU (Table 1).

Correlations of phenotypic mean values over years for winter barley, assessed from 12 resp. 17 overlapping lines, ranged between 0.727 and 0.997 (Supplementary Table S3).

Genotypic data

After data cleaning, 4095 markers remained for spring and 4359 for winter barley. For a total chromosome length of 2127.4 cM, this yielded on average 1.9 and 2 markers per cM, respectively.

Assessment of population structure by principal component analysis revealed that spring barley did not show any subgroups (Fig. 1). The first two principal components together explained 15 % of the variance; the Calinsky criterion had a maximum at one group. In winter barley, mild subpopulation structure was detected. The Calinsky criterion had a maximum for three groups and the first two components explained 31 % of the variance.

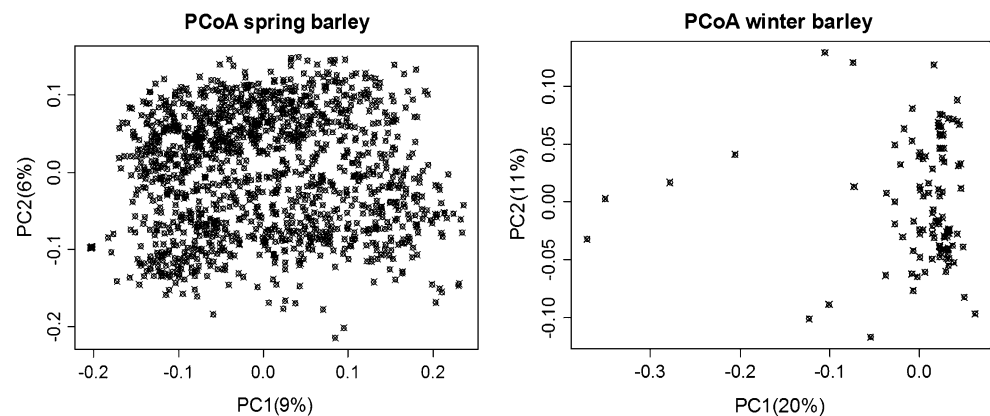
The family structure was different between both groups. In spring barley, there were 80 unique parents in the training population. These represent all parents for spring barley in this data set. Each of them was present with a maximum of 10 % of the 260 unique crosses. Sister plants have the same parents and were counted as a unique cross. 50 % of the crosses were single crosses, 42 % three-way crosses with three different parents, and 8 % were four-way crosses or a three-way cross with overlapping parents (AxB)xA which is similar to a back cross. In winter barley, 45 unique parents were used, resulting in 55 unique crosses. Some varieties occurred frequently in the pedigree of the analyzed lines. The most frequent line, Wintmalt, was present in 41 of the 102 pedigrees. From all crosses, 41 % were two-way crosses and 59 % were three-way crosses. The mean genetic distance of the winter barley lines in the trainings set was 0.18 ± 0.09 , whereas spring barley showed a mean distance of 0.35 ± 0.07 .

Table 1 Heritabilities for the twelve analyzed malting quality traits, with total number of locations (#loc), years (#yr) and harmonized number of locations (#loc_hm) and years (#yr_hm) for spring barley and winter barley

TRAIT	#loc	#loc_hm	#yr	#yr_hm	h^2
Spring barley					
AMA	2	1.70	3	2.87	0.50
AMB	2	1.64	4	3.63	0.77
EXT	7	5.46	4	3.31	0.84
FAN_L	2	1.81	3	2.88	0.79
Fin_At	7	5.46	4	3.31	0.82
FRI	7	5.46	4	3.31	0.81
GLU	7	5.47	4	3.32	0.85
KOL	7	5.46	4	3.31	0.83
LOSS	5	3.95	3	1.93	0.80
NIT	7	5.46	4	3.31	0.87
PRT	7	5.46	4	3.31	0.66
VIS	7	5.46	4	3.31	0.68
TRAIT	#loc	#loc_hm	#yr	#yr_hm	h^2
Winter barley					
AMA	6	5.50	3	2.95	0.75*
AMB	6	5.50	3	2.95	0.91
EXT	6	5.50	3	2.95	0.96
FAN_L	6	5.50	3	2.95	0.94
Fin_At	6	5.50	3	2.95	0.78*
FRI	6	5.50	3	2.95	0.98
GLU	6	5.50	2	1.92	0.98
KOL	6	5.50	3	2.95	0.71*
LOSS	6	5.50	3	2.95	0.74*
NIT	6	5.50	3	2.95	0.76*
PRT	6	5.50	3	2.95	0.80
VIS	6	5.50	3	2.95	0.79*

Harmonic means were used to estimate heritabilities

* Genotype variance components were boundary with standard formula, reduced model was used

Fig. 1 Principle component analysis of spring barley (*left*) and winter barley (*right*). First two principle components explain 15- and 31 % of the variance

Spring barley showed an average LD block size of 60.46 ± 14.24 cM, with a minimum of 42.04 on chromosome 4H and a maximum of 77.68 on 2H. Compared to spring barley, winter barley had a smaller average LD block size of 16.45 ± 7.8 cM, with a minimum of 9.25 on

chromosome 4H and a maximum of 31.76 on 7H (Fig. 2a, b). In spring barley, the other chromosomes showed LD block sizes in cM of 63.7 (1H), 50.66 (3H), 46.4 (5H), 75.12 (6H), 68.33 (7H), and in winter barley 11.85 (1H), 20.86 (2H), 14.1 (3H), 16.87 (5H), 10.49 (6H) and

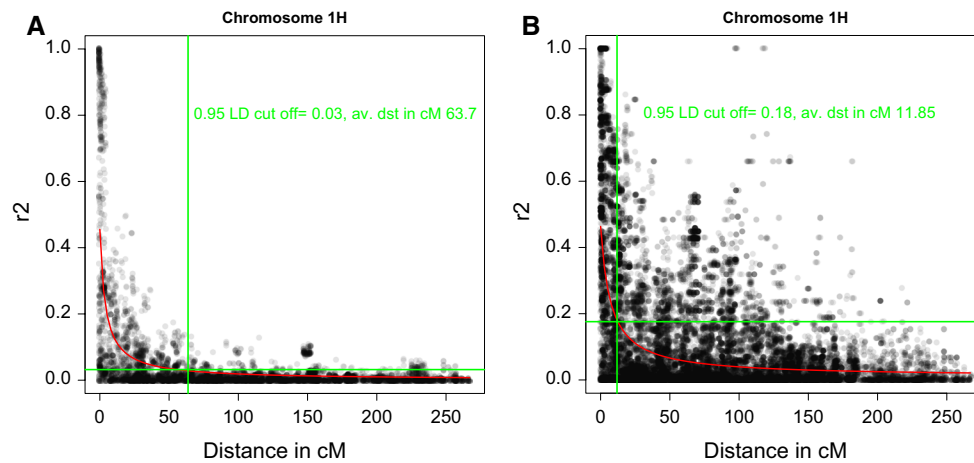


Fig. 2 LD decay on chromosome 1H in spring barley (**a**, left) and winter barley (**b**, right), with an LD decay of 63.7 and 11.85 cM, respectively

Table 2 Predictive abilities (PA), standard deviation of the predictive ability after fivefold cross validation and size of the available training sets (TS) for spring and winter barley

Trait	Spring barley			Winter barley		
	PA (mean)	PA (SD)	#TS	PA (mean)	PA (SD)	#TS
AMA	0.444	0.151	99	0.564	0.300	102
AMB	0.142	0.218	118	0.606	0.150	102
EXT	0.558	0.072	424	0.625	0.175	102
FAN_L	0.497	0.218	65	0.572	0.165	102
Fin_At	0.495	0.120	424	0.732	0.113	102
FRI	0.552	0.077	424	0.788	0.113	102
GLU	0.479	0.096	423	0.798	0.173	102
KOL	0.487	0.106	424	0.556	0.178	102
LOSS	0.419	0.101	322	0.652	0.165	102
NIT	0.583	0.074	424	0.551	0.150	102
PRT	0.521	0.064	424	0.399	0.158	102
VIS	0.449	0.111	424	0.744	0.116	102
mean	0.469	0.117		0.632	0.163	

31.76 cM (7H). Minor allele frequencies were unequally distributed between both groups (Supplementary Figure S1). Winter barley showed a higher proportion of extreme allele frequencies with 60 % of markers having a minor allele frequency lower than 0.01 compared to spring barley when compared to 40 % of markers with $maf < 0.01$.

GEBVs

All predictive abilities (PA) presented were estimated with RR-Blup (Endelman 2011). Other tested methods such as Elastic-Net and Lasso (Friedman et al. 2010) did not yield different GEBVs or predictive abilities (results not shown).

The training sets consisted of between 65 (FANL) and 424 (e.g., EXT) individuals for the different traits in spring barley and of 102 individuals in winter barley for all traits (Table 2).

In spring barley, the lowest PA was observed for AMB with 0.14 (Table 2). Medium PA of 0.42 to 0.5 were observed for AMA, FANL, Fin_At, GLU, KOL, LOSS and VIS. High PA of above 0.5 up to 0.58 were observed for EXT, FRI and NIT. In winter barley, the lowest PA of 0.4 were observed for PRT, whereas for all other traits medium to high PA were observed with the highest PA of 0.8 for GLU. The average PA for spring barley were 0.47 as compared to 0.63 for winter barley.

When training set size was small (e.g., 65), larger standard deviations were observed in spring barley. This amounted to an increase of the standard deviation in % of the PA value from 18 to 44 % when training set size decreased from 424 to 65, with a maximum of 153 % for AMB (stddev higher than PA, lowest PA of all traits) and a minimum of 12 % for PRT. In winter barley, training

Table 3 Correlations of GEBVs to adjusted phenotypes of different training set (TS) combinations for winter barley over different year combinations to a given prediction set (PS)

TS	PS	AMA	AMB	EXT	FAN_L	Fin_At	FRI	GLU	KOL	LOSS	NIT	PRT	VIS	\bar{x}
All	11	0.91	0.97	0.89	0.97	0.95	1.00	0.96	0.98	0.95	0.98	0.87	0.91	0.95*
All	12	0.98	0.98	0.94	0.99	0.99	1.00	0.98	0.99	0.99	0.99	0.90	0.99	0.98*
All	13	0.94	0.93	0.95	0.97	0.89	1.00	0.99	0.98	0.97	0.98	0.93	0.97	0.96*
11	12	-0.10	0.48	0.35	0.37	0.53	0.63	0.44	0.07	0.14	0.29	0.53	0.14	0.34
11	13	-0.18	0.25	0.39	0.25	0.66	0.74	0.80	0.22	0.51	0.33	0.32	0.35	0.42
12	11	0.15	0.60	0.38	0.55	0.48	0.47	0.30	0.30	0.48	0.68	0.56	0.00	0.41
12	13	0.61	0.46	0.56	0.29	0.62	0.88	0.88	0.54	0.52	0.71	0.67	0.87	0.63
13	11	-0.04	0.28	0.28	0.55	0.52	0.41	0.42	0.40	0.21	0.69	0.48	0.29	0.38
13	12	0.47	0.63	0.53	0.62	0.57	0.77	0.72	0.42	0.15	0.66	0.73	0.71	0.58
11 + 12	13	0.59	0.45	0.58	0.61	0.67	0.87	0.91	0.42	0.57	0.68	0.55	0.87	0.65
11 + 13	12	0.52	0.70	0.58	0.63	0.62	0.77	0.76	0.32	0.25	0.70	0.69	0.73	0.61
12 + 13	11	0.06	0.38	0.34	0.22	0.62	0.50	0.41	0.33	0.27	0.69	0.50	0.05	0.36
	\bar{x}	0.23	0.47	0.44	0.45	0.59	0.67	0.63	0.34	0.34	0.60	0.56	0.45	0.49

Auto correlations indicated by * were not used to estimate mean correlations for the traits over different training set combinations

set size was constant with 102 individuals for all traits. The standard deviations ranged from 14 % (FRI) to 53 % (AMA). The proportions of the standard deviations, compared to the PA, were slightly higher for spring barley with 33 % as compared to winter barley with 27 %.

Reliabilities based on PEV were medium to high for both spring and winter barley with 0.54–0.74 for spring barley and 0.63–0.77 for winter barley. The prediction error variance showed values in a wide range for both spring and winter barley (Supplementary Table S1). In spring barley, it ranged from very low values of 0–0.17 (VIS, PRT, LOSS and EXT) to extremely high values of 7340.68 (GLU) to 7518.41 (AMB). In winter barley, PEV values ranged from 0.01 to 0.32 (LOSS, EXT, PRT, VIS) and up to 15473.62 (GLU) to 16296.76 (AMB). The same three traits (AMB, GLU and NIT) showed the highest PEV values in both spring and winter barley.

The correlations of GEBVs to phenotypes deliberately assumed as non-observed (Table 3) had a range from 0.23 (AMA) to 0.67 (FRI). The correlation to the PA was 0.35 (excluding auto correlation) and 0.53 (mean correlation if all lines were in the training set to phenotypes in 2011, 2012 and 2013). In consecutive years, correlations increased, e.g., the lowest mean of all tested correlations with 0.36 was observed when 2012 and 2013 were used as training sets. Strong variation was observed within the traits, e.g., VIS had a minimum of 0 (2012 only to predict 2011) and a maximum of 0.87.

Discussion

Over the past decade, genome analysis of barley has greatly benefitted from the augmenting sequence information

which provided the basis for the establishment of high-throughput marker systems (Close et al. 2009; Comadran et al. 2012). These have paved the way for the development and implementation of novel approaches for trait mapping, such as genome-wide association analysis and, at the applied level, genomic selection (GS). Regarding the latter, we have investigated in the present study the potential and limitations of GS to predict the expression of malting quality in the two major gene pools, spring- and winter barley, which have evolved their genetic distinctness as a result of directed breeding and selection.

The major findings of this study were (1) in narrow genetic material “small” training sets of around 100 individuals were sufficient to achieve PA up to 0.798, depending on the trait and the corresponding heritability, (2) increasing the training population size leads to a higher stability of PA and (3) reliabilities of the GEBVs indicate that for many of the traits, the chances of applying GS as selection tool are good, especially in winter barley.

Phenotypic data

Spring barley heritabilities were on average 0.07 lower than those of winter barley, which can be partly explained by the reduced balance of the phenotypic data and also in some cases by the lower number of locations and years. In general, the heritabilities for all traits exceeded 0.5, which was expected with the available number of years and locations. The observed heritabilities and corresponding PA indicate that GS can be applied. Field trial designs might have to be adjusted for the requirements of calculating BLUEs across different years by, e.g., improving connectivity over years by a sufficient number of checks, increasing the number of

plots per entry or increasing the number of entries, as indicated by Endelman et al. (2014). When GS is fully implemented into a breeding scheme, possibilities arise to divide the breeding population into a carefully managed training set that is phenotyped more extensively (more locations) to deliver reliable BLUEs and consequently reliable GEBVs, and the rest of the breeding population for which phenotyping is reduced to a minimum and selection mainly based on GEBV instead.

In comparison to winter barley, spring barley lines were characterized by a 33 % lower Beta amylase activity (AMB). One explanation could be that the priorities of other traits in spring barley were higher in the last selection cycles and that the current level was sufficient for current malting quality requirements.

Relatedness, LD, population structure, genetic distances

The background noise of LD was estimated by the LD of markers between chromosomes, using the 95 % percentile as cutoff to estimate the average length of markers in LD. The cutoff of spring barley (0.032) was 5.5 times smaller than the winter barley cutoff (0.176), resulting in average LD block length difference of 44.1 cM. This mathematical translation into the average length of haplotype blocks may not be exact because the cutoff value is crucial and only estimated by this elite material. If only the dots are compared, the decay between spring- and winter barley is more similar. A recombination analysis may lead to different results. This is thought to be due to longer and more intensive selection of spring barley for malting quality. The concentration of elite breeding material should increase average LD block sizes in comparison to more unselected material, wide crosses, exotic or diversity material. The large difference can partly be explained by the lower background noise of spring barley, which was used as the basis to estimate the cutoff value. Since only inter chromosomal distances were used, the employed map, whose mapping parents (Barke \times Morex) were two- and six rowed but both spring barley varieties, should not have an impact. Comparable LD sizes in barley were also observed by Rostoks et al. (2006) who observed LD over distances up to 60 cM in controlled crosses. In conventional breeding programs, crosses are usually genetically narrow, especially if the breeding program has limited genetic resources and traits have limiting thresholds to get registered as official varieties. Both breeding programs were affected by these limitations, avoiding wide crosses and limited introduction of new resources. This in sum can explain the observed extent of LD. The family structure explains the larger mean genetic distances within the spring barley breeding program. In spring barley, more unique parents were used,

resulting in more divergent crosses relative to winter barley. The number of markers used should sufficiently capture all chromosome segments, especially in spring- but also winter barley.

Reliability and accuracy of GEBVs

The accuracy of GS usually increases with the relatedness of training and prediction sets (Zhong et al. 2009). Genetic diversity analysis showed that the mean Roger's distance was considerably lower for winter barley and, therefore, the reduced size of the training sets did not result in a lower PA when compared to spring barley (Table 2). The predictive abilities for spring barley mostly had a lower standard deviation than winter barley, because the size of the training sets was larger for most traits. Over all traits, the PA of winter barley were on average 0.163 above spring barley, e.g., in AMB and GLU the PA of winter barley were 0.465 and 0.319 above spring barley. Only for NIT and PRT, spring barley showed a PA above winter barley (0.032 and 0.122). The pedigree structure of winter barley, where the frequently used parent Wintmalt was included in the training set, and the smaller mean genetic distance in comparison to spring barley are two explanations for increased mean PA. The size of the training sets and the standard deviations of the PA were negatively correlated (Table 2).

Reliabilities of the GEBV indicate that for many of the traits, the prospects of applying GS as selection tool are good, especially in winter barley. The higher the PA and the reliability, the more reliable the GS will be and the sharper GS selection can be.

The knowledge of increased standard deviations of the PA is especially important in small training sets, where few individuals with rare alleles and haplotype blocks could have a greater influence and, therefore, can contribute more to the PA. This can also happen if a frequently used parent or a major progenitor is excluded from a small training set, e.g., Wintmalt in winter barley. Nevertheless, Wintmalt was used in many crosses as parent, the remaining material would still have a good connectivity and, therefore, the PA should not drop.

Correlation of GEBVs and phenotypes

In general, all observed correlations of the traits (phenotypic and GEBVs) corresponded to the expectations (personal communication with breeders). VIS and GLU were negatively correlated to other traits because they were selected to have low values (Supplementary Table S2). The higher phenotypic variance (mean of the standard deviations) of the winter barley material compared to spring barley may explain the higher correlations. The observed correlations to phenotypes deliberately assumed

as non-observed were strongly influenced by the training set combination. This can be explained by the small number of individuals in the training set in general. If closely linked markers become monomorphic, the trait cannot be estimated. This can be compensated for if the size and genetic diversity of the training population are increased or by increasing marker density. The reduced size of the training set also affects the variability of phenotypic observations. In trials including around 100 individuals, inclusion of certain individuals in the training set may be important to minimize the occurrence of rare alleles and to maximize phenotypic variance. The distribution of the minor allele frequency for winter barley supports this assumption. As expected, a replacement of a single-year by a two-year training set increased the correlations (by 0.09 on average). This indicates that increasing the training set size increases the reliability to estimate unobserved years. Also relatively small training sets seem to be sufficient in the case of malting barley—in the current study, single-year correlations over the traits were 0.45 on average which is an acceptable figure. Nevertheless, increasing the size of the TS leads to an increased stability of PA and probably to more confident results to future years. For spring barley, the number of lines within the years was not as evenly distributed. In correlation results, the year effect and the size of the training set were thus confounded. Therefore, no correlations to artificially non-observed phenotypes were calculated.

The correlation of overlapping phenotypes over years for winter barley (Supplementary table S3) was at least 0.727, which shows that (1) genotype by environment interaction is generally low for the traits under consideration and (2) the variance between GEBVs and observed phenotypes of different training sets is probably not caused by large genotype by environment interaction but by the small size of the TS. Correlations for spring barley were not derived due to few overlapping phenotypes.

Marker application and molecular breeding in barley

Reports on improvement of quality traits in barley have so far been based on marker-assisted selection (MAS) (Heffner et al. 2011). MAS employs large effect markers and is a powerful tool (Lande and Thompson 1990), but studies indicated that MAS has serious limitations and that GS can be useful in combination with marker-assisted selection [reviewed by Schmid and Thorwarth (2014) or Heslot et al. (2015)]. In several QTL studies, QTL explained a large proportion of genetic variance for malting quality traits. Islamovic et al. (2014) reported that 56 % of the variance for free amino nitrogen was explained by applying 180 SNPs and 165 DArTs markers to an F6:8 bi-parental RIL population. In theory, a bi-parental population is the most effective way to demonstrate the applicability of GS

(Lorenzana and Bernardo 2009), but in a bi-parental population QTL may not segregate in some backgrounds or effects may be overestimated (Gutiérrez et al. 2011), where GS models correct for overestimation. Genomic selection for grain quality traits in wheat (Heffner et al. 2011) outperformed marker-assisted selection for all traits using less than 1000 markers, reaching a plateau at 256 markers. This might be due to the fact that two bi-parental populations were used in that study. In the current study, more diverse lines than a bi-parental population, i.e., elite lines from two breeding programs of spring and winter barley, were used as training populations. Therefore, more than 256 markers were required.

Implementation of GS in a breeding scheme

The point where GS is implemented into the breeding scheme is crucial. The implementation of GS in an early stage is most promising because phenotyping for yield and especially quality traits is expensive. Malting trait analysis requires approximately 100–1000 g of grains, and this amount of seed is not available at early stages of breeding. However, half or one seed is sufficient for DNA extraction for genotyping. The implementation of GS at an early stage of the breeding cycle bears the highest potential for a breeding program because (1) the selection intensity can be increased when more candidates are grown initially and GS is used as an additional tool for selection in a two-stage approach, (2) the effect of saving costs of phenotyping is maximized and (3) the potential for shortening the breeding cycle is highest. By GS the selection of further candidates to be phenotyped can be optimized, which saves costs due to the reduced number of new phenotypes in the future. In the breeder's equation for prediction of the expected selection gain, the cycle length is the denominator, so shortening the breeding cycle has the most impact on selection gain per unit time. Malting quality is a complex of multiple inter-related traits of which 12 were analyzed in this study. For the breeder, this poses the problem of how to handle those individual traits and how to define and select for the overall complex trait “malting quality”. Traits can either be selected for separately or combined in an arbitrary index—which may differ between countries or markets—or both. In practice, often one or several of the individual traits need to be improved while others need to be kept within a certain acceptable range. Very often, the major breeder's goal is to improve grain yield while maintaining a good malting quality. This problem affects phenotypic and GS in the same way. So currently, the GEBVs are treated in the same manner when applying genomic selection.

Real multivariate genomic selection on multiple traits has rarely been explored in plant breeding to date (Jia and Jannink 2012), but might offer chances for the complex

trait “malting quality”, e.g., by making use of genetic correlations between individual quality traits.

Training set optimization

In accordance with the findings of Endelman et al. (2014), increasing training population sizes in this study for spring barley resulted in increased predictive abilities with a lower standard deviation in the cross validation (Table 2). The training population sizes used in this study, especially for winter barley, yielded standard deviations of the cross-validated predicted abilities (Albrecht et al. 2011) and predicted error variances (Rincent et al. 2012) comparable to other published results. Several combinations of training sets were analyzed for this study (data not shown), but these resulted mostly in reduced PA because the total size of the training sets became very small and the standard error of the PA increased accordingly.

Increasing training set size by combining spring and winter barley was of no avail. If both spring- and winter barley were combined, the first principal component explained more than 60 % of the genetic variance (result not shown), which would violate the assumption that the material is related. The correlation between GEBVs and phenotypic results would then be artificially increased owing to the population structure caused by the two subgroups. The minor subpopulation structure in winter barley can partly explain the increased PA but since the total size of the training set was 102, the effect of auto correlation should be minor. In addition, individuals distant to the main group do not show extreme phenotypic values.

Next to increasing training population sizes, predictive abilities can be increased by the selection of particular individuals to maximize the relationship between training and prediction population. This can be done on the basis of the individuals’ prediction error variance (PEV) or its reliability (Isidro et al. 2015; Rincent et al. 2012). These values were also estimated for all individuals forming the training sets in this study. If the data set is split on the basis of PEV or reliability, an optimized training set of the remaining lines can be created to achieve highest correlations in predicting the other set (Akdemir et al. 2015). Optimization within the existing training set does not necessarily improve the correlation of GEBVs to non-phenotyped lines, but it can be used together with genetic distance information to select a stable and genetically representative sample of genotyped lines chosen for phenotyping.

Conclusions

In this study, GS training populations for two malting barley breeding programs were established for which GEBV and predictive abilities for twelve malting quality traits

were derived. The results suggest a high potential of genomic selection to improve malting barley breeding: in the early generations of a breeding program, a combined phenotypic selection (for easy to select agronomic traits) and genomic selection (for malting quality traits) could be employed, which would (1) reduce the costs to assess malting quality and (2) shorten the breeding cycle, since selection for quality is brought forward.

Author contribution statement Schmidt M contributed to phenotypic data analysis, genotypic data analysis and writing; Kollers S contributed to genotypic data analysis and writing; Maasberg-Prelle A contributed to the raw phenotypic data preparation for winter barley; Großer J contributed the winter barley breeding material; Schinkel B contributed the spring barley breeding material; Tomerius A contributed to writing; Graner A contributed to the preparation of concept InnoGRAIN-MALT project; Korzun V contributed conceptualization of the project, initialization and final writing of this manuscript.

Acknowledgments The authors greatly thank Professor Mark E. Sorrells (Cornell University, USA) and Dr. Edward Henry Byrne (KWS UK Ltd., UK) for careful reading and providing useful comments for this manuscript. This research was supported by Grant (FKZ 0315960) from the Bundesministerium Bildung und Forschung (BMBF) within the framework of the PLANT 2030 program.

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interest.

References

- Akdemir D, Sanchez JJ, Jannink JL (2015) Optimization of genomic selection training populations with a genetic algorithm. *Genet Sel Evol* 47:38–47
- Albrecht T, Wimmer V, Auinger H-J, Erbe M, Knaak C, Ouzunova M, Simianer H, Schön C-C (2011) Genome-based prediction of test-cross values in maize. *Theor Appl Genet* 123:339–350
- Browning SR, Browning BL (2007) Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet* 81:1084–1097
- Butler D (2009) ASReml R package version 3.0. <https://www.vsnl.co.uk/de/software/asreml/>
- Calinski T, Harabasz J (1974) A dendrite method for cluster analysis. *Commun Stat* 3:1–27
- Chiapparino E, Donini P, Reeves J, Tuberosa R, O’Sullivan D (2006) Distribution of β -amylase I haplotypes among European cultivated barleys. *Mol Breed* 18:341–354
- Close TJ, Bhat PR, Lonardi S, Wu Y, Rostoks N, Ramsay L, Druka A, Stein N, Svensson JT, Wanamaker S, Bozdag S, Roose ML, Moscou MJ, Chao S, Varshney RK, Szucs P, Sato K, Hayes PM, Matthews DE, Kleinhofs A, Muehlbauer GJ, DeYoung J, Marshall DF, Madishetty K, Fenton RD, Condamine P, Graner A, Waugh R (2009) Development and implementation of high-throughput SNP genotyping in barley. *BMC Genomics* 10:852

- Comadran J, Kilian B, Russell J, Ramsay L, Stein N, Ganai M, Shaw P, Bayer M, Thomas W, Marshall D, Hedley P, Tondelli A, Pecchioni N, Francia E, Korzun V, Walther A, Waugh R (2012) Natural variation in a homolog of *Antirrhinum CENTRORADIALIS* contributed to spring growth habit and environmental adaptation in cultivated barley. *Nat Genet* 44:1388–1392
- Destá ZA, Ortiz R (2014) Genomic selection: genome-wide prediction in plant improvement. *Trends Plant Sci* 19:592–601
- Endelman JB (2011) Ridge regression and other kernels for genomic selection with R package rBLUP. *Plant Genome* 4:250–255
- Endelman JB, Atlin GN, Beyene Y, Semagn K, Zhang X, Sorrells ME, Jannink J-L (2014) Optimal design of preliminary yield trials with genome-wide markers. *Crop Sci* 54:48–59
- Friedman J, Hastie T, Tibshirani R (2010) Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* 33:1–22
- Frisch M (2015) SelectionTools. <http://www.fb09-pg-s207agraruni-giessende/~frisch/m/>. R-Library 15.1.1
- Graner A, Streng S, Kellermann A, Schiemann A, Bauer E, Waugh R, Pello B, Ordon F (1999) Molecular mapping and genetic fine-structure of the *rym5* locus encoding resistance to different strains of the Barley Yellow Mosaic Virus complex. *Theor Appl Genet* 98:285–290
- Gutiérrez L, Cuesta-Marcos A, Castro AJ, von Zitzewitz J, Schmitt M, Hayes PM (2011) Association mapping of malting quality quantitative trait loci in winter Barley: positive signals from small germplasm arrays. *Plant Genome* 4:256–272
- Haley CS, Visscher PM (1998) Strategies to utilize marker-quantitative trait loci associations. *J Dairy Sci* 81:85–97
- Han F, Romagosa I, Ullrich SE, Jones BL, Hayes PM, Wesenberg DM (1997) Molecular marker-assisted selection for malting quality traits in barley. *Mol Breed* 3:427–437
- Heffner EL, Jannink J-L, Iwata H, Souza E, Sorrells ME (2011) Genomic selection accuracy for grain quality traits in biparental wheat populations. *Crop Sci* 51:2597–2606
- Henryon M, Berg P, Sørensen AC (2014) Animal-breeding schemes using genomic information need breeding plans designed to maximise long-term genetic gains. *Livest Sci* 166:38–47
- Heslot N, Jannink J-L, Sorrells ME (2015) Perspectives for genomic selection applications and research in plants. *Crop Sci* 55:1–12
- Hickey JM, Dreisigacker S, Crossa J, Hearne S, Babu R, Prasanna BM, Grondona M, Zambelli A, Windhausen VS, Mathews K, Gorjanc G (2014) Evaluation of genomic selection training population designs and genotyping strategies in plant breeding programs using simulation. *Crop Sci* 54:1476–1488
- Hofmann K, Silvar C, Casas AM, Herz M, Buttner B, Gracia MP, Contreras-Moreira B, Wallwork H, Igartua E, Schweizer G (2013) Fine mapping of the *Rrs1* resistance locus against scald in two large populations derived from Spanish barley landraces. *Theor Appl Genet* 126:3091–3102
- Isidro J, Jannink J-L, Akdemir D, Poland J, Heslot N, Sorrells ME (2015) Training set optimization under population structure in genomic selection. *Theor Appl Genet* 128:145–158
- Islamovic E, Obert D, Budde A, Schmitt M, Brunick R II, Kilian A, Chao S, Lazo G, Marshall J, Jellen E, Maughan P, Hu G, Klos K, Brown R, Jackson E (2014) Quantitative trait loci of barley malting quality trait components in the Stellar/01Ab8219 mapping population. *Mol Breed* 34:59–73
- Jia Y, Jannink J-L (2012) Multiple-trait genomic selection methods increase genetic value prediction accuracy. *Genetics* 192:1513–1522
- König J, Kopahnke D, Steffenson BJ, Przulj N, Romeis T, Röder MS, Ordon F, Perovic D (2012) Genetic mapping of a leaf rust resistance gene in the former Yugoslavian barley landrace MBR1012. *Mol Breed* 30:1253–1264
- Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* 124:743–756
- Lorenzana R, Bernardo R (2009) Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. *Theor Appl Genet* 120:151–161
- Meuwissen TH, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829
- Miedaner T, Korzun V (2012) Marker-assisted selection for disease resistance in wheat and barley breeding. *Phytopathology* 102:560–566
- Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Henry M, Stevens H, Wagner H (2014) Vegan: community ecology package. <http://www.CRANR-project.org/package=vegan>. R package version 2.2-1
- Poland J, Endelman J, Dawson J, Rutkoski J, Wu S, Manes Y, Dreisigacker S, Crossa J, Sánchez-Villeda H, Sorrells M, Jannink J-L (2012) Genomic selection in wheat breeding using genotyping-by-sequencing. *Plant Genome* 5:103–113
- Potokina E, Caspers M, Prasad M, Kota R, Zhang H, Sreenivasulu N, Wang M, Graner A (2004) Functional association between malting quality trait components and cDNA array based expression patterns in barley (*Hordeum vulgare* L.). *Mol Breed* 14:153–170
- Rincint R, Laloë D, Nicolas S, Altmann T, Brunel D, Revilla P, Rodríguez VM, Moreno-Gonzalez J, Melchinger A, Bauer E, Schoen CC, Meyer N, Giauffret C, Bauland C, Jamin P, Laborde J, Monod H, Flament P, Charcosset A, Moreau L (2012) Maximizing the reliability of genomic selection by optimizing the calibration set of reference individuals: comparison of methods in two diverse groups of maize inbreds (*Zea mays* L.). *Genetics* 192:715–728
- Rostoks N, Ramsay L, MacKenzie K, Cardle L, Bhat PR, Roose ML, Svensson JT, Stein N, Varshney RK, Marshall DF, Graner A, Close TJ, Waugh R (2006) Recent history of artificial outcrossing facilitates whole-genome association mapping in elite inbred crop varieties. *Proc Natl Acad Sci* 103:18656–18661
- Schmid KJ, Thorwarth P (2014) Genomic Selection in Barley Breeding. In: Kümlehn J, Stein N (eds) *Biotechnological approaches to barley improvement*. Springer, Berlin, pp 367–378
- Shewry PR, Ullrich SE (2014) *Barley: Chemistry and Technology*, 2nd edn. AACC International
- Spindel J, Begum H, Akdemir D, Virk P, Collard B, Redoña E, Atlin G, Jannink JL, McCouch SR (2015) Genomic selection and association mapping in rice *Oryza sativa*: effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. *PLoS Genet* 11:e1004982
- Utz HF (2011) PLABSTAT - Ein Computerprogramm zur statistischen Analyse von pflanzenzüchterischen Experimenten. *Saatgutforschung und Populationsgenetik*, Universität Hohenheim, Stuttgart, Institut für Pflanzenzüchtung
- Wang Y, Mette MF, Miedaner T, Gottwald M, Wilde P, Reif JC, Zhao Y (2014) The accuracy of prediction of genomic selection in elite hybrid rye populations surpasses the accuracy of marker-assisted selection and is equally augmented by multiple field evaluation locations and test years. *BMC Genom* 15:556
- Werner K, Friedt W, Ordon F (2007) Localisation and combination of resistance genes against soil-borne viruses of barley (BaMMV, BaYMV) using doubled haploids and molecular markers. *Euphytica* 158:323–329
- Wimmer V, Albrecht T, Auinger H-J, Schoen C-C (2012) synbreed: a framework for the analysis of genomic prediction data using R. *Bioinformatics* 28:2086–2087
- Zhao Y, Gowda M, Liu W, Würschum T, Maurer H, Longin F, Ranc N, Reif J (2012) Accuracy of genomic selection in European maize elite breeding populations. *Theor Appl Genet* 124:769–776
- Zhong S, Dekkers JCM, Fernando RL, Jannink J-L (2009) Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a Barley case study. *Genetics* 182:355–364