

## RESEARCH ARTICLE

## Can metabolic prediction be an alternative to genomic prediction in barley?

Mathias Ruben Gemmer<sup>1</sup>, Chris Richter<sup>2</sup>, Yong Jiang<sup>3</sup>, Thomas Schmutzer<sup>1</sup>, Manish L. Raorane<sup>2</sup>, Björn Junker<sup>2</sup>, Klaus Pillen<sup>1</sup>, Andreas Maurer<sup>1\*</sup>

**1** Institute of Agricultural and Nutritional Sciences, Chair of Plant Breeding, Martin Luther University Halle-Wittenberg, Halle, Germany, **2** Institute of Pharmacy, Martin Luther University Halle-Wittenberg, Halle, Germany, **3** Department of Breeding Research, Quantitative Genetics, Leibniz Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany

\* [andreas.maurer@landw.uni-halle.de](mailto:andreas.maurer@landw.uni-halle.de)

## Abstract

Like other crop species, barley, the fourth most important crop worldwide, suffers from the genetic bottleneck effect, where further improvements in performance through classical breeding methods become difficult. Therefore, indirect selection methods are of great interest. Here, genomic prediction (GP) based on 33,005 SNP markers and, alternatively, metabolic prediction (MP) based on 128 metabolites with sampling at two different time points in one year, were applied to predict multi-year agronomic traits in the nested association mapping (NAM) population HEB-25. We found prediction abilities of up to 0.93 for plant height with SNP markers and of up to 0.61 for flowering time with metabolites. Interestingly, prediction abilities in GP increased after reducing the number of incorporated SNP markers. The estimated effects of GP and MP were highly concordant, indicating MP as an interesting alternative to GP, being able to reflect a stable genotype-specific metabolite profile. In MP, sampling at an early developmental stage outperformed sampling at a later stage. The results confirm the value of GP for future breeding. With MP, an interesting alternative was also applied successfully. However, based on our results, usage of MP alone cannot be recommended in barley. Nevertheless, MP can assist in unravelling physiological pathways for the expression of agronomically important traits.

## OPEN ACCESS

**Citation:** Gemmer MR, Richter C, Jiang Y, Schmutzer T, Raorane ML, Junker B, et al. (2020) Can metabolic prediction be an alternative to genomic prediction in barley? PLoS ONE 15(6): e0234052. <https://doi.org/10.1371/journal.pone.0234052>

**Editor:** Aimin Zhang, Institute of Genetics and Developmental Biology Chinese Academy of Sciences, CHINA

**Received:** January 23, 2020

**Accepted:** May 17, 2020

**Published:** June 5, 2020

**Copyright:** © 2020 Gemmer et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the manuscript and its Supporting Information files.

**Funding:** This work was funded by the European Fund for Regional Development (EFRE, [https://ec.europa.eu/regional\\_policy/en/funding/erdf/](https://ec.europa.eu/regional_policy/en/funding/erdf/)) within the research focus “Molecular Biosciences as a Motor of Knowledge-Based Economics” (awarded to KP). The funders had no role in study design,

## Introduction

Barley (*Hordeum vulgare* L.) is the fourth most important crop worldwide after wheat, maize and rice, with an acreage of 48.1 m hectares in 2017/18 [1]. Approximately 10,000 years ago, barley was domesticated and is thus one of the oldest crop plants [2]. Domestication and breeding for yield performance in elite barley (*Hordeum vulgare* ssp. *vulgare*) led to a reduction of biodiversity through allele erosion, the so-called genetic bottleneck effect. This phenomenon also applies to most other crop species [3, 4]. Consequently, further improvement of the performance of barley becomes increasingly difficult. Moreover, classical selection methods with several years of field trials are expensive. On top of that, the current climate change

data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

**Abbreviations:** BLUEs, *Best linear unbiased estimates*; CV, *Coefficient of variation*; EAR, *Ears per m<sup>2</sup>*; GC-MS, *Gas chromatography-mass spectrometry*; GEBV, *Genomic estimated breeding value*; GNE, *Grain number per ear*; GP, *Genomic prediction*; GS, *Genomic selection*; GWAS, *Genome-wide association study*; HEA, *Heading, flowering time*; HEI, *Plant height*; LC-MS, *Liquid gas chromatography-mass spectrometry*; MAS, *Marker-assisted selection*; MAT, *Maturity*; MNI, *Mean imputation*; MP, *Metabolic prediction*; NAM, *Nested association mapping*; NMR, *Nuclear magnetic resonance-mass spectrometry*; QTL, *Quantitative trait locus/loci*;  $r_{ab}$ , *Prediction ability*;  $r_{ac}$ , *Prediction accuracy*; RR-BLUP, *Ridge regression best linear unbiased prediction*; SD, *Standard deviation*; SHO, *Shooting*; SNP, *Single nucleotide polymorphism*; TGW, *Thousand grain weight*; YLD, *Grain yield*.

scenarios and the increasing world population, pose difficult challenges for breeders to use effective breeding methods that could lead to yield increase and stability.

To accelerate the breeding progress, indirect selection methods are of great importance. The most common method is the single nucleotide polymorphism (SNP) based estimation of breeding values through genomic prediction (GP) [5]. The advantage of GP is the early estimation of agronomically relevant traits already at seedling stage of single plants, which accelerates the selection of the best plants during the breeding process. In contrast to classical methods like genome-wide association studies (GWAS) or linkage mapping to define trait-specific molecular markers for subsequent marker-assisted selection (MAS), the approach of GP (also called genomic selection—GS) is different: Rather than focusing on the single effect and the position of one marker, the entirety of all markers is taken into account in GP. Ordinarily, for each marker allele, an effect is estimated and with the combination of all marker effects, a genomic estimated breeding value (GEBV) is computed. Depending on the model, interactions between alleles may also be included in the calculation of GEBV. This requires a large number (tens or hundreds of thousands) of markers distributed over the whole genome [6, 7]. The modern methods of genome sequencing and the large number of genotyped SNPs allow a broad application of GP across different living systems—including animal and plant species as well as human genetics [8]. GP overcomes the disadvantages of MAS, which mainly relies on few selected quantitative trait loci (QTL), identified through linkage mapping and GWAS. Those methods have achieved great success, also in barley, for instance in the elucidation of genetic issues like disease resistance and flowering [9–11]. However, the classical methods have certain weaknesses in the quantification of some polygenic traits that are influenced by numerous minor QTL with small effects [12]. This circumstance is considered in GP by assigning effects to all markers tested.

Apart from GP, studies with different species (*Arabidopsis thaliana*, tomato, rice, potato, maize) confirmed that a reliable estimation of trait performance is also possible through MP with metabolite data [13–17]. Metabolites play an important role in all living organism, so in plants. Estimates for the total number of metabolites in plant kingdom vary from 200,000 to 1,000,000 [18]. A rough classification of metabolites is the differentiation between primary and secondary metabolites. While the primary metabolites are responsible for growth and development, the secondary ones are built in response to various biotic and abiotic stresses. These two classes are subject of different genetic control. Whereas primary metabolites are mainly controlled by many interacting genes with small effects, the secondary ones are determined by a small number of genes with large effects [19–22]. The use of metabolite profiling in plant breeding is interesting as it can provide helpful information about the system under study; metabolites play a key role in gene expression and help to elucidate the function of genes [23]. Furthermore, metabolites can be used as biomarkers (when no genomic information is available) or as an addition to SNP markers to predict phenotype expression [16]. With a combination of gas chromatography and mass spectrometry (GC-MS), a high-throughput method for untargeted metabolite screening is available. Other high-throughput methods such as liquid chromatography-mass spectrometry (LC-MS) and nuclear magnetic resonance-mass spectrometry (NMR-MS) have also been established for metabolite profiling of the experimental system [24].

In this project we simultaneously characterize the multi-parental wild barley nested association mapping (NAM) population HEB-25 [11] with SNPs (50k SNP array [25]) and through metabolic profiling of 128 metabolites with sampling at two different developmental stages. We merge SNP, metabolite and phenotype data to alternatively predict phenotypes based on metabolites, SNPs or a combination of both and compare the prediction accuracies of the different methods.

## Materials and methods

### Plant material

The population HEB-25 is the worldwide first NAM population of barley. It was generated by crossing and subsequent backcrossing of 25 wild barley accessions (24 *Hordeum vulgare* ssp. *spontaneum* and one *Hordeum vulgare* ssp. *agriocrithon*) with the German elite spring barley cultivar Barke (*Hordeum vulgare* ssp. *vulgare*). The resulting BC<sub>1</sub>S<sub>3</sub> generation comprises 1,420 individual lines (whereof 1,307 were used in this study) subdivided into 25 families (for a detailed description see [11]).

### Genotypic evaluation

DNA of pooled BC<sub>1</sub>S<sub>3;8</sub> plants of each line was extracted according to the manufacturer's protocol, using the BioSprint 96 DNA Plant Kit and a BioSprint work station (Qiagen, Hilden, Germany), and finally dissolved in distilled water at approximately 50 ng/μl for genotyping with the recently developed barley Infinium iSelect 50K chip [25] at TraitGenetics, Gatersleben, Germany. SNP markers that did not meet the quality criteria (polymorphic in at least one HEB family, < 10% failure rate, < 12.5% heterozygous calls) were removed from the data set. Altogether, 33,005 SNPs met the quality criteria and were analysed in this study. Based on the Barke reference genotype, the wild barley allele can be specified in each segregating family. To set up the quantitative identity-by-state (IBS) matrix the state of the homozygous Barke allele was coded as 0, while HEB lines that showed a homozygous wild barley genotype were assigned a value of 2. Consequently, heterozygous HEB lines were assigned a value of 1. If a SNP was monomorphic in one HEB family but polymorphic in a second family, lines of the first HEB family were assigned a genotype value of 0, since their state is not different from the Barke allele. Gaps resulting from missing genotypes (0.84%) were estimated by applying the mean imputation (MNI) approach [26]. The genotype matrix is available at e!DAL [27, 28]. The markers are uniformly distributed over the whole genome with few gaps and decreasing density in the telomere regions (S1 Fig).

### Field trials

Between 2011 and 2018, eight field trials with HEB-25 were conducted at the Kühnfeld experimental station of the University Halle (51°29'45.72"N; 11°59'36.62"E) to gather phenotypic data. All field trials were sown in spring between March and April with fertilisation and pest management following local practice. Detailed information about field trials is given in S1 Table and S1 File.

The studies were conducted on land owned by the authors' institutions. The research conducted complied with all institutional and national guidelines.

### Phenotypic evaluation

The following traits were measured in the field trials: time to shooting (SHO), flowering (HEA) and maturity (MAT); plant height (HEI); number of ears per m<sup>2</sup> (EAR); grain number per ear (GNE); thousand grain weight (TGW); grain yield (YLD). Table 1 shows a detailed description of the trait assessment. Raw phenotype data is available as S2 File.

### Metabolic evaluation

A 2 cm tissue sample from the middle region of the last fully developed leaf of each HEB line was sampled on 22 May 2017 under a clear sky between nine and ten o'clock. This date represented the developmental stage BBCH 30–31 (beginning of shooting) for the majority of

**Table 1. List of evaluated traits for HEB-25 in eight-year field trials.**

Abbr. <sup>a</sup>	Trait	Unit	Method of measurement	Years studied
SHO	Time to shooting	days	Number of days from sowing until first node noticeable 1 cm above soil surface for 50% of all plants of a plot, BBCH 31 [29].	2011–2018
HEA	Time to heading (flowering)	days	Number of days from sowing until emergence of 50% of ears on main tillers of a plot, BBCH 49 [29].	2011–2018
MAT	Time to maturity	days	Number of days from sowing until hard dough: grain content firm and fingernail impression held, BBCH 87 [29].	2012–2018
HEI	Plant height	cm	Average plant height of all plants of a plot measured from soil surface to tip of the erected ear without awns at maturity.	2011–2018
EAR	Ears per m <sup>2</sup>	n	Number of ears per m <sup>2</sup> , counted in a representative 50 cm frame in the middle of a row and extrapolated to one m <sup>2</sup> .	2014–2018
GNE	Grain number per ear	n	Number of grains per ear, calculated by use of MARVIN seed analyser (GTA Sensorik GmbH, Neubrandenburg, Germany) based on a representative sample of 10 ears.	2014–2018
TGW	Thousand grain weight	g	Weight of 1000 grains, calculated after harvest by use of MARVIN seed analyser based on a 200 seeds sample of each plot (2011–2013). Before, seeds were cleaned and damaged seeds were sorted out.	2011–2018
YLD	Grain yield	dt/ha	Total grain yield determined after harvest of the whole plot (2014–2016) or derived from the three yield components EAR, GNE and TGW (2017/18) and extrapolated to dt/ha.	2014–2018

<sup>a</sup> Abbreviations of traits.

<https://doi.org/10.1371/journal.pone.0234052.t001>

plants. The leaf was cut approximately 1 cm from the stem and was put in an Eppendorf tube. The protruding leaf was cut off, the Eppendorf tube was closed and put instantly in liquid nitrogen to stop metabolic processes. All plots were sampled within one hour under constant weather conditions. In total, 29 people were involved to meet this schedule. Sampling was repeated under the same circumstances (constantly clear sky, equal time of day, equal sampling methods) on 22 June 2017. The plants were more heterogeneous at this time, representing developmental stages BBCH 59–69 (end of ear emergence to end of flowering).

The frozen leaf samples were pulverised using a Retsch-ball mill (MM 400, Retsch, Germany) for 2 minutes at 20 Hz. The homogenised leaf samples were then resuspended in 700 µl methanol:chloroform:water solution (3:2:4) containing 8 µg/ml <sup>13</sup>C-sorbitol as an internal quantitative standard. The mixture was shaken for 20 min at room temperature and at 500 rpm. The mixture was then centrifuged for 11,000 X g for 5 minutes at 4°C. After the extraction, 10 µl of the supernatant was dried in a vacuum concentrator without heating for 45 minutes. Online derivatization was performed using the Multi-Purpose Sampler (MPS, Gerstel, Germany) by adding 30 µl Methoxamine hydrochloride (20 mg/ml in Pyridine) to the samples and shaken for 30 min at 45°C. Furthermore, 45 µl N,O-Bis(trimethylsilyl)trifluoroacetamide and 5 µl Alkane-Standard (C10-C28; 6 mg/ml) were added and the samples were shaken again for 120 min at 45°C. As quality controls for the extraction procedure, leaf samples from 10 randomly chosen Barke reference lines were extracted and pooled together. All the samples along with 20% of quality controls were analysed with GC-MS (GC-qTOF system -7890B/7200, Agilent, Santa Clara, USA). One µl of the derivatized samples were injected at 250°C in a splitless mode with a helium gas flow set to 1 ml min<sup>-1</sup>. Chromatography was performed with a 30-m Zebron Capillary GC-Column (ZB-Semi Volatiles, 30 m, 0.25 mm, 0.25 µm). The Helium flow was constant at 1 ml/min. The temperature program was set to 60°C followed by a linear ramp of 10°C/min to 320°C and holding at this temperature for 3 minutes. Throughout the run, the transfer line, source and the quadrupole were set to 290°C, 230°C and 150°C respectively.

The raw data was processed by MassHunter Qualitative Analysis software (Agilent, B.07.00) and MassHunter Quantitative Analysis software for QTOF (Agilent, B.08.00). The mass spectra library NIST 14 (National Institute of Standards and Technology) and standard

compounds were used for identification and confirmation of the chromatographic peaks. Peak areas were normalized with the internal standard and fresh weight.

This resulted in data for 1,307 lines where 158 metabolites (alkanes, amino acids, organic acids, sugars and unknowns) could be defined (S3 File). Metabolites with > 10% missing values were removed from the data set so that 128 metabolites were used for prediction (S2 Table). Samples from the 2<sup>nd</sup> sample date resulted in data for 1,229 lines with 159 metabolites (one additional unknown metabolite). After data cleaning 122 metabolites remained for the subsequent analyses (S3 Table). Remaining missing values were replaced with the minimum value of the respective metabolite.

## Statistical analyses

All statistical analyses were performed with SAS 9.4 [30] and R [31]. Broad-sense heritabilities were computed using R software with the lmerTest package [32] across treatments and years as  $h^2 = \frac{V_G}{V_G + \frac{V_{GY}}{y} + \frac{V_R}{r}}$ , where  $V_G$ ,  $V_{GY}$  and  $V_R$  represent the genotype, genotype × year, and error variance components, respectively. The terms  $y$  and  $r$  indicate the number of years and replicates, respectively. To estimate variance components, all effects were assumed to be random. Best linear unbiased estimates (BLUEs) of all traits were calculated using the PROC HPMIXED procedure in SAS for each genotype assuming fixed genotype effects. Pearson's correlation coefficients were calculated with R software with the corrgram package [33]. The box-cox power transformation [34] was applied to metabolic data using SAS PROC TRANSREG with  $\lambda$  ranging from -3 to 3 by steps of 0.25. The genomic heritabilities of metabolites (also called SNP-based heritabilities, [35]) were estimated with the R package sommer [36] as

$h_{SNP}^2 = \frac{\sigma_A^2 + \sigma_D^2 + \sigma_I^2}{\sigma_A^2 + \sigma_D^2 + \sigma_I^2 + \sigma_R^2}$ , where  $\sigma_A^2$ ,  $\sigma_D^2$ ,  $\sigma_I^2$  and  $\sigma_R^2$  represent the additive, dominance, epistatic and residual variance components, respectively. Additionally, repeatability of metabolites was calculated as  $rep = \frac{V_G}{V_G + \frac{V_R}{r}}$  for the subset of 17 genotypes (elite cultivars, control lines) where multiple metabolite measurements were available. Euclidean distance matrices with SNP and metabolite data were calculated using R package stats. Subsets of SNPs or metabolites for GP and MP were created using R package dplyr [37]. Descriptive statistics for metabolites were calculated with R package psych [38]. Two-sided t-tests were carried out to detect significant differences between the models and datasets. The significance level was set to  $p < 0.01$ . All figures were created with R using the package ggplot2 [39].

## Genomic/metabolic prediction

Based on BLUEs of the 1,307 HEB genotypes (1,307 lines with complete datasets of SNP and metabolite data at the 1<sup>st</sup> sampling date, 1,229 lines at the 2<sup>nd</sup> sampling date), two approaches for genomic prediction were applied considering additive effects: ridge regression best linear unbiased prediction (RR-BLUP) [40] and BayesB [41]. All statistical procedures for genomic prediction approaches were executed using R. The R code for RR-BLUP was developed in-house [42]. For the BayesB model, the package BGLR [43] was used. The models are briefly described in the following.

Let  $n$  be the number of genotypes,  $m$  be the number of markers and  $l$  be the number of years. The RR-BLUP model has the form  $\mathbf{y} = \mathbf{1}_n \mu + \mathbf{X} \mathbf{g} + \mathbf{e}$ , where  $\mathbf{y}$  is the vector of BLUEs of the respective trait for all HEB genotypes across years,  $\mathbf{1}_n$  denotes the vector of 1's,  $\mu$  is the common intercept term,  $\mathbf{g} = (g_1, g_2, \dots, g_m)'$  is the vector of marker effects,  $\mathbf{X}$  is the matrix of marker information and  $\mathbf{e}$  is the residual term. In the model we assumed that  $\mathbf{g} \sim N(0, \sigma_g^2 \mathbf{I})$ ,  $\mathbf{e} \sim N(0, \sigma_e^2 \mathbf{I})$ , where  $\sigma_g^2 = \sigma_G^2 / m$  for SNP markers and  $\sigma_e^2 = \sigma_R^2 / l$ . Here  $\sigma_G^2$  and  $\sigma_R^2$  are the



genotypic and residual variance components obtained in the mixed model in the phenotypic data analysis. The penalty parameter is  $\lambda = (\sigma^2_R / l) / (\sigma^2_G / m)$ . The estimation of marker effects is then given by the mixed model equations [44]. The basic model of BayesB is the same as RR-BLUP. However, all parameters are treated as random variables in a Bayesian framework and we do not assume the same variance for all marker effects. More precisely, we defined the prior distributions as  $\mathbf{g} \sim N(0, \mathbf{D})$ ,  $\mathbf{e} \sim N(0, \sigma_e^2 \mathbf{I})$ , where  $\mathbf{D} = \text{diag}(\sigma_{g_1}^2, \sigma_{g_2}^2, \dots, \sigma_{g_p}^2)$ . For the intercept term  $\mu$  we assume a flat prior. For each  $i$ , the prior distribution of  $\sigma_{g_i}^2$  is assumed to be zero with probability  $\pi$  and a scaled inverse chi-squared distribution with probability  $(1-\pi)$ . The prior of  $\pi$  is a beta distribution. The prior of  $\sigma_e^2$  is also a scaled inverse chi-squared distribution. A Gibbs sampler algorithm was then applied to infer all the parameters in the model.

The accuracy of the prediction by the models was evaluated using five-fold cross-validation [45]. In each run of cross-validation, the training set included 80% of HEB lines, randomly selected per HEB family, while the remaining 20% of HEB lines were assigned to build the test set. The prediction ability ( $r_{ab}$ ) is the correlation between observed and predicted values, averaged over all 100 cross-validation runs. Prediction accuracy ( $r_{ac}$ ) is defined as  $r_{ac} = \frac{r_{ab}}{\sqrt{h^2}}$  [17].

Pairwise t-tests were carried out in R to determine significant differences in prediction accuracy between models and prediction methods. The significance level was set to  $p < 0.01$ .

Genomic prediction was realised for the agronomic traits measured in the field with 33,005 SNPs coded as 0,1,2 in the RR-BLUP model and -1,0,1 in the BayesB model to meet the specific requirements of the applied R packages.

For metabolic prediction (MP) of the agronomic traits measured in the field, the values of 128 metabolites (first sampling date) or 122 metabolites (second sampling date) were used in both models. In the combined approach, all 33,005 SNPs and 128 metabolites (or 122) were included in the prediction model.

## Results and discussion

### Phenotypic data

Descriptive analysis of the phenotypic data showed a high variation between lines and between years, resulting in high coefficients of variation (S4 Table). For instance, the difference for the trait HEA was 71 days between the minimum and maximum value. This reflects the high diversity of the HEB-25 population within and across years (S2 Fig). Heritabilities for all traits calculated over 4–8 years were  $> 0.8$  with the exception of EAR (0.41) and YLD (0.58, Table 2). In summary, this reflects the high quality of phenotypic data and the genotype impact on traits, underlining the suitability for genetic analyses such as GP and MP.

### Genomic and metabolic prediction

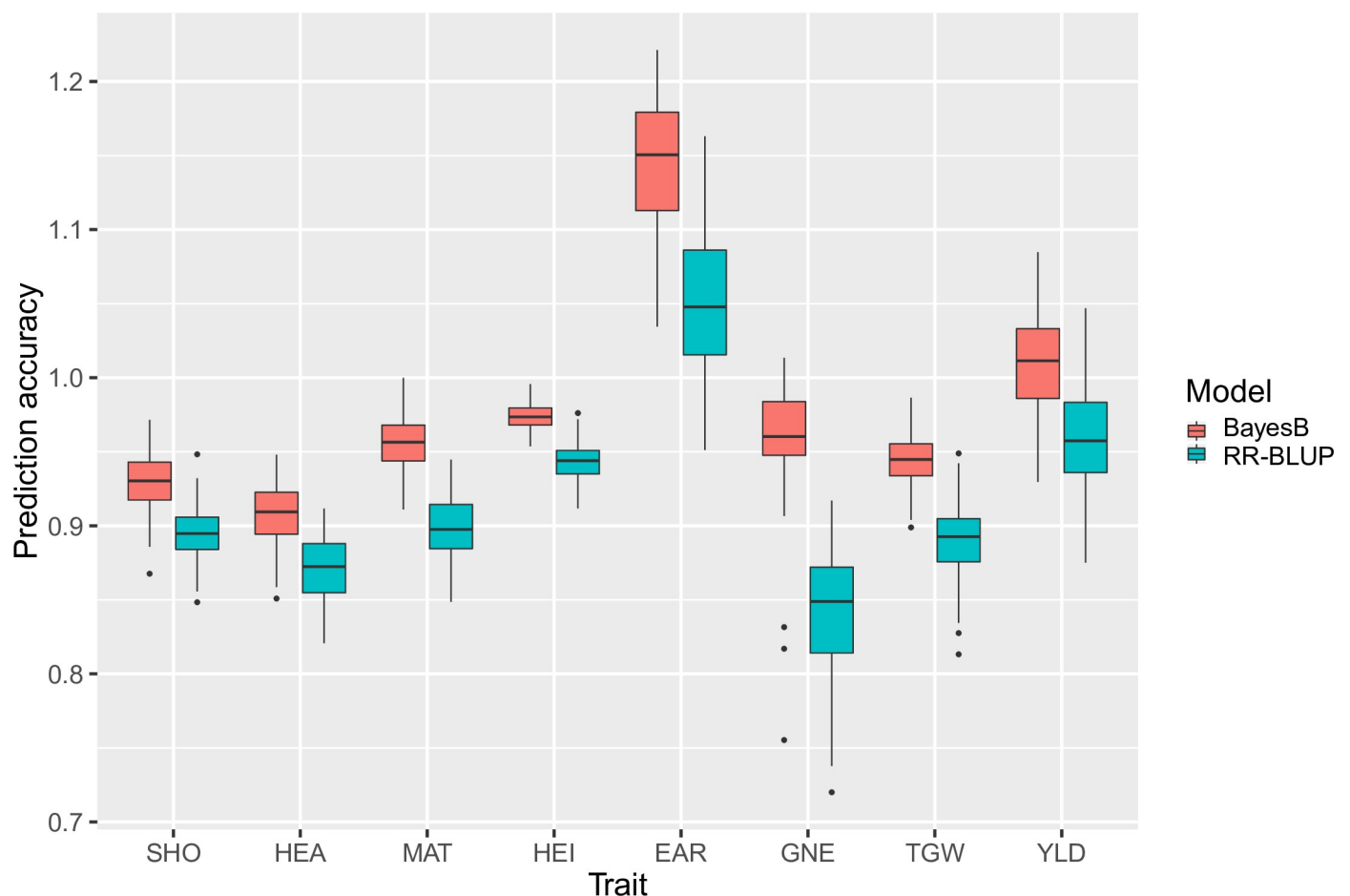
All results described below (including figures, tables and supplementary files) refer to the metabolite set of the first sampling date unless it is mentioned otherwise. Generally, in genomic prediction with SNP data, we observed a slight advantage of BayesB over RR-BLUP regarding prediction performance, which was significant for all traits (Fig 1). With metabolite data both models performed almost equal (S5 Table, S3 Fig). With the exception of EAR (better performance of RR-BLUP) and YLD (better performance of BayesB), no significant differences were detected. The better performance of BayesB depends on the genetic architecture of the target trait [46]. It is superior to RR-BLUP when the trait is controlled by few large QTL effects, which is true and well-studied for HEA [11] as well as for GNE and TGW [47] in the HEB-25 population. With SNP data high prediction accuracies ( $\geq 0.91$ ) for all traits were reached with BayesB (Table 2). It is noticeable that the accuracies for the traits EAR and YLD

**Table 2. Summary of genomic and metabolic prediction, BayesB model.**

Trait	$h^2$	SNPs			Metabolites			SNPs + Metabolites			Sig.
		$r_{ab}$	$r_{ac}$	SD	$r_{ab}$	$r_{ac}$	SD	$r_{ab}$	$r_{ac}$	SD	
SHO	0.91	0.88	0.93	0.02	0.57	0.59	0.05	0.89	0.93	0.02	-
HEA	0.93	0.87	0.91	0.02	0.59	0.61	0.05	0.88	0.91	0.02	-
MAT	0.83	0.87	0.95	0.02	0.55	0.61	0.06	0.88	0.96	0.01	*
HEI	0.91	0.93	0.97	0.01	0.37	0.39	0.05	0.93	0.97	0.01	-
EAR	0.41	0.74	1.14	0.04	0.38	0.59	0.09	0.74	1.15	0.04	-
GNE	0.84	0.88	0.96	0.04	0.27	0.29	0.07	0.88	0.96	0.03	-
TGW	0.83	0.86	0.94	0.02	0.26	0.28	0.06	0.86	0.94	0.02	-
YLD	0.58	0.77	1.01	0.03	0.35	0.46	0.08	0.77	1.01	0.03	-

BayesB model was applied using SNP and/or metabolite data. Prediction abilities ( $r_{ab}$ ) and prediction accuracies ( $r_{ac}$ ) for selected traits, averaged over 100 cross-validation runs and the standard deviation of  $r_{ac}$  (SD) are shown. In addition, heritabilities ( $h^2$ ) for the traits are given. Sign. indicates the significance (\*:  $p < 0.01$ , -: not significant) of a t-test between  $r_{ac}$  of SNPs and  $r_{ac}$  of SNPs + metabolites for the respective traits. Not indicated here: SNPs and SNPs + metabolites generally performed significantly better than metabolites alone for all traits. Trait abbreviations are given in Table 1.

<https://doi.org/10.1371/journal.pone.0234052.t002>



**Fig 1. Cross-validated prediction accuracies of traits with SNP data using RR-BLUP and BayesB, respectively.** Boxplots contain all 100 prediction values of the cross-validation runs. Red boxes show results of BayesB, while blue boxes show results of RR-BLUP. BayesB performed better for all traits. Prediction accuracies with BayesB were significantly better than with RR-BLUP for all traits.

<https://doi.org/10.1371/journal.pone.0234052.g001>

were  $> 1$ , which is caused by the low  $h^2$  estimates of these traits. Nevertheless, the usage of  $r_{ac}$  is common in GP, as it corrects  $r_{ab}$  for nongenetic effects of the target trait [17]. The correlation between  $h^2$  and  $r_{ab}$  was highly positive ( $r = 0.95$ ) and, consequently, the correlation between  $h^2$  and  $r_{ac}$  was highly negative ( $r = -0.94$ ). This underlines the importance of high-quality phenotypic data, resulting in high prediction performance. The observed prediction accuracies are comparable to other studies in wheat, maize and barley [8, 17, 48].

The concept of estimating SNP-based heritability [35], also called genomic heritability, was applied to the metabolite data resulting in values of up to 0.50 with a mean value of 0.10 (S6 and S7 Tables, S4 Fig). Repeatabilities of metabolite measurements showed high variation across metabolites (0.00–0.87) with a mean value of 0.26 (S6 and S7 Tables), hinting on limited data quality for several metabolites that may affect metabolic prediction.

Prediction accuracies with metabolite data instead of SNPs were generally lower. The highest accuracies were observed for the developmental parameters ( $r_{ac}$  up to 0.61 for HEA and MAT), while for HEI and especially the yield parameters GNE and TGW low accuracies of no more than 0.29 were obtained (Table 2). The decay of  $r_{ac}$  for yield parameters seems logical since sampling took place early on during the shooting phase of plants. The assumption is that metabolites which are involved in plant development are more reflected in the early metabolite profile than the ones responsible for grain filling and yield formation and vice versa. To pursue this question, it is worth to compare  $r_{ac}$  of the first sampling with  $r_{ac}$  of the second sampling (S8 Table). Actually, based on the second metabolite sampling the prediction accuracies for developmental traits were worse (ca. 0.10 less for SHO and HEA), but also for yield parameters no notable improvements could be achieved. Metabolic prediction with data from the first sampling date performed significantly better for the traits SHO, HEA and HEI. MAT and EAR showed no significant differences. Slight, but significant improvements at the second sampling date could be achieved for the yield parameters GNE, TGW and YLD. In conclusion, sampling during a young and more homogeneous plant stage seems more effective, also in terms of time management.

To our knowledge, there exists no study on MP in barley. Prediction accuracies of MP were, depending on the trait, below the accuracies reported in studies with other species [15–17]. However, the comparability of different studies on MP is difficult, since metabolite determination is highly sensitive. Steinfath et al. [16] predicted blackspot susceptibility of potatoes with correlations between observed and predicted values ranging from 0.68 to 0.82. Riedelsheimer et al. [17] reached accuracies of up to 0.80 for female flowering in maize. The use of both SNPs and metabolites in the combined approach did not lead to an improvement in prediction compared to the sole use of SNPs. This applies to our study as well as to Riedelsheimer et al. [17].

To gain insights which metabolites are decisive for different trait predictions, Pearson's correlations between metabolite measurements and agronomic traits across all lines were calculated. As expected, correlations were comparably low ( $-0.36 < r < 0.30$ , S9 Table), showing that single metabolites generally exert only a moderate impact on trait expression. Interestingly, one of the strongest negative correlations was observed for TMET101 and HEA ( $r = -0.35$ ), indicating that this unknown metabolite might be directly involved in flowering time regulation. This is confirmed by the high effect estimation for TMET101 in the MP model for HEA (S9 Table). In general, there was the trend that metabolites with a high effect estimated in MP also had a higher correlation with the respective agronomic trait, as exemplified for HEA (S5 Fig). Similar observations could be made in the metabolite set of the second sampling (S10 Table). This indicates that MP effect estimates can give hints to metabolites that are involved in trait expression and thus might be worth further investigation for instance to deepen the understanding of molecular pathways.

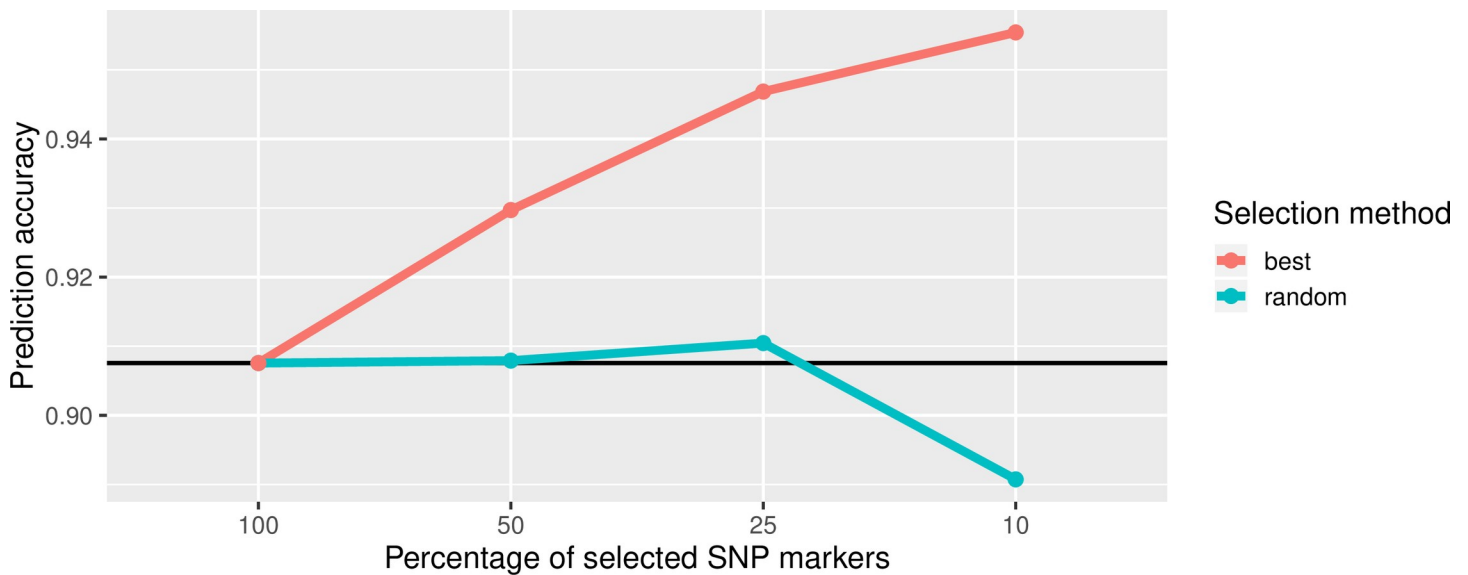


The accuracies with metabolite data seem to be low compared to the accuracies with SNP data. However, it is important to remember that 128 metabolites face 33,005 SNPs (approximately 260 times more SNPs). Moreover, metabolites were sampled in an early developmental stage of the plants, reflecting just a snapshot in the highly dynamic system of plant metabolism, and used for prediction of eight-year phenotypic data. This raises the question of whether the metabolites are used to predict something they cannot provide. Therefore, the MP model was run again, restricting the phenotypic data to the season 2017, the year in which also the metabolite samples were collected. Surprisingly, this resulted in almost equal or even slightly lower prediction accuracies compared to eight-year phenotypes (S11 Table). With  $r_{ac} = 0.47$  for MAT, the prediction accuracy was even worse. However, the metabolite-trait correlations were quite similar to the complete set (S12 Table). Like SNPs, metabolites seem to fix information about the underlying genotype, which seems to be environmentally stable. Our results support the assumption that a prediction of phenotypic traits is possible even with metabolite data from one year at one sampling date.

A closer look on the estimated effects in GP and MP showed that there was a clear correlation pattern between the estimated effects of different traits (S6A and S6B Fig). Both in GP and MP, the marker and metabolite effects for SHO, HEA and MAT were highly correlated ( $0.88 < r < 0.95$ ), indicating that the same genes and metabolites are responsible for the expression of these traits. Interestingly, the correlation plot of the phenotypic traits (S6C Fig) reflected the same patterns like the plots for the estimated effects of GP and MP. For instance, the negative correlations between TGW and the developmental parameters ( $-0.22 < r < -0.37$ ) were quite close to the correlations of their estimated effects, the same applies to the correlations among developmental parameters. Apparently, the GP and MP models were able to quantify these phenotypic connections in their estimation of effects with high precision and therefore they reflected the underlying genetic and metabolic mechanisms. Remarkably, the genetic and metabolic distance matrices were not correlated ( $r = 0.04$ , S7 Fig). It seems that they contain similar information, though based on different backgrounds.

Interestingly, a reduction of used SNPs and metabolites in the prediction model can lead to an improvement or at least to no decay in prediction accuracy. For instance, the prediction accuracy for HEA was steadily increased when reducing the number of SNP markers to subsets of 50%, 25% and 10%, provided that the markers with the biggest effects in GP from the model with the whole marker set were selected. But even with 25% randomly selected markers (8,251 SNP markers) of the complete set a small increase in  $r_{ac}$  was observed (Fig 2). Selecting the best markers increased the  $r_{ac}$  for all investigated traits whereas random selection, especially by selecting only 10%, clearly reduced the accuracy (S8 Fig). The reason for enhanced prediction accuracy with best markers may be the reduction of SNPs causing background noise in the model. But even random selection did not worsen the accuracy up to a certain point suggesting that fewer markers are sufficient for a reliable coverage of genome information.

For MP, randomly selected metabolites reduced  $r_{ac}$  but when selecting those 50% of metabolites with the highest effects in MP using the whole metabolite set, the accuracy increased to up to 0.65 for HEA (Fig 3). This trend applied to most of the traits (S9 Fig). Traits with a generally weaker  $r_{ac}$  in the MP based on all metabolites (EAR and TGW) even increased their prediction accuracy when only 10% of the most impactful metabolites were selected (S9 Fig). The model was not as robust against reduction when using metabolites instead of SNPs. This may be due to the fact that much less metabolites than SNPs are available and thus a further reduction has a stronger impact on accuracy of the model, especially with random selection. The reason for the enhancement in  $r_{ac}$  by selecting 50% of the best metabolites is probably due to the reduction of noise in the model resulting from metabolites with questionable determination quality. A study in rapeseed also showed that high prediction accuracies are possible with

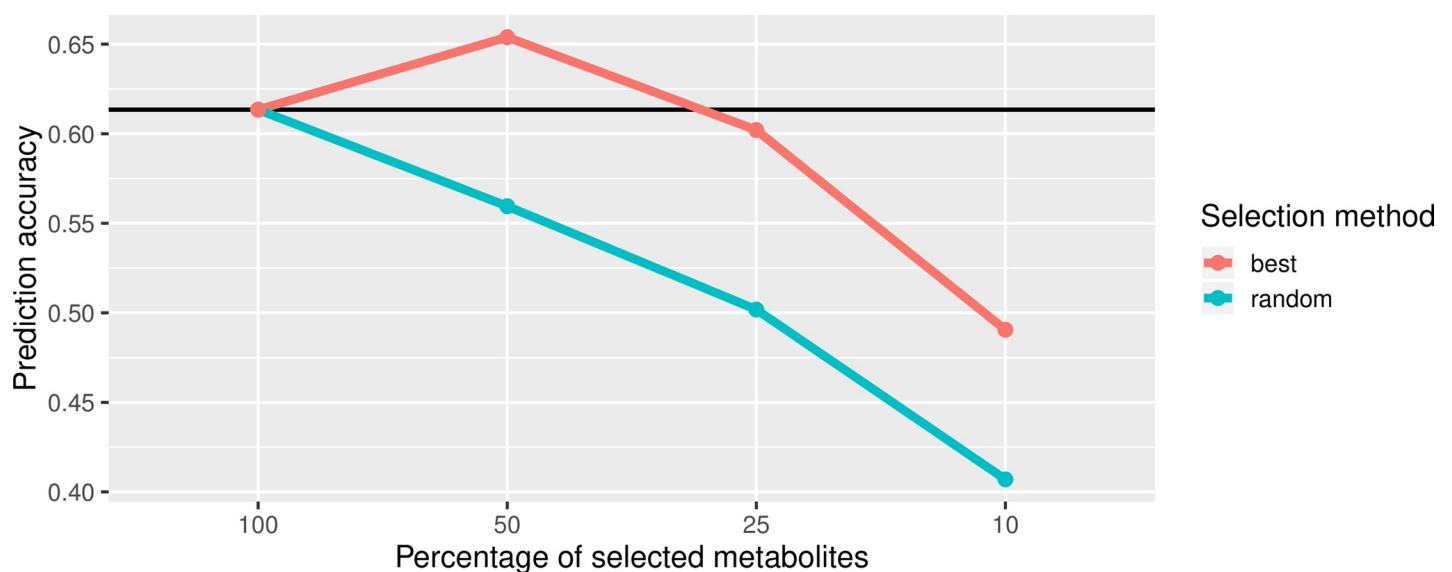


**Fig 2. Variation of prediction accuracy for HEA with BayesB after reduction of the number of SNP markers.** The black reference line indicates the prediction accuracy using all SNP markers in the model. The red line indicates the trend of prediction accuracy by selecting the best markers (markers with the highest effects in BayesB model), the blue line indicates the trend of prediction accuracy by selecting random markers.

<https://doi.org/10.1371/journal.pone.0234052.g002>

a reduced marker set [49]. These findings allow the consideration of using reduced and selected marker sets for GP, this way reducing computation time and costs as fewer markers have to be evaluated.

The high accuracies, especially in GP, may partly be attributed to the population design of HEB-25, which is genetically highly diverse due to the crossings with 25 different wild barley accessions. Breeding populations usually have a much smaller genetic variability [17]. Moreover, the large sample size influences the accuracies [50]. Nevertheless, the high accuracies in



**Fig 3. Variation of prediction accuracy for HEA with BayesB after reduction of the number of metabolites.** The black reference line indicates the prediction accuracy using all metabolites in the model. The red line indicates the trend of prediction accuracy by selecting the best metabolites (metabolites with the highest effects in BayesB model), the blue line indicates the trend of prediction accuracy by selecting random metabolites.

<https://doi.org/10.1371/journal.pone.0234052.g003>

this study confirmed the value of applying GP in barley breeding, especially the time and cost savings are mentioned here. Results of MP indicate it as an interesting alternative to GP under certain circumstances, but according to the current status, its practical use in barley breeding is not recommendable. Metabolites as predictor variables are an attractive alternative to SNPs when no genotypic data is available, as it is the case in many orphan crop species [16]. Moreover, MP has the potential to detect metabolites involved in the expression of important agronomic traits, which might assist in unravelling the involved molecular pathways. Further research in HEB-25, like GWAS on metabolite expression, to investigate metabolite-trait associations is in progress. This promises to achieve a deeper knowledge of the complex interaction between genes, metabolites and plant physiology.

## Supporting information

**S1 Table. Detailed information about field trials 2011 to 2018.**

(XLSX)

**S2 Table. List of metabolites 1st sampling date.**

(XLSX)

**S3 Table. List of metabolites 2nd sampling date.**

(XLSX)

**S4 Table. Descriptive statistics of agronomic traits across the years 2011–2018.**

(XLSX)

**S5 Table. Results for genomic and metabolic prediction using the RR-BLUP model and metabolites from 1st sampling date.**

(XLSX)

**S6 Table. Descriptive statistics for the metabolites from 1st sampling date.**

(XLSX)

**S7 Table. Descriptive statistics for the metabolites from 2nd sampling date.**

(XLSX)

**S8 Table. Results for genomic and metabolic prediction using the BayesB model, including results from 2nd sampling date.**

(XLSX)

**S9 Table. Pearson's correlation coefficients between traits and metabolites and estimated metabolite effects in BayesB model, 1st sampling date.**

(XLSX)

**S10 Table. Pearson's correlation coefficients between traits and metabolites and estimated metabolite effects in BayesB model, 2nd sampling date.**

(XLSX)

**S11 Table. Results for metabolic prediction using BayesB model, phenotypic data 2017 and metabolites of 1st sampling date.**

(XLSX)

**S12 Table. Pearson's correlation coefficients between traits (phenotypic data only from 2017) and metabolites of 1st sampling date.**

(XLSX)

**S1 Fig. Distribution of markers on chromosomes.**  
(PDF)

**S2 Fig. Boxplots of all traits over years and across treatments.**  
(PDF)

**S3 Fig. Cross-validated prediction accuracies of traits with metabolite data using RR-BLUP and BayesB, respectively.**  
(PDF)

**S4 Fig. Genomic heritability of single metabolites.**  
(PDF)

**S5 Fig. Estimated effects of metabolites in BayesB model plotted against Pearson's correlation coefficients of metabolite measurements with the agronomic trait, exemplified for HEA.**  
(PDF)

**S6 Fig.** Pearson's correlations of SNP effects (a) or metabolite effects (b) estimated for respective traits in BayesB model, in comparison to correlations of trait BLUEs (c).  
(PDF)

**S7 Fig. Scatter plot of Euclidean distances estimated with SNPs and metabolites, respectively.**  
(PDF)

**S8 Fig. Variation of prediction accuracy for selected traits in BayesB through reduction of used SNP markers.**  
(PDF)

**S9 Fig. Variation of prediction accuracy for selected traits in BayesB through reduction of used metabolites.**  
(PDF)

**S1 File. Detailed description of field trials.**  
(PDF)

**S2 File. Raw phenotype data.**  
(XLSX)

**S3 File. Raw metabolite data.**  
(XLSX)

## Acknowledgments

We thank all cooperation partners and employees which contributed to this study.

## Author Contributions

**Conceptualization:** Björn Junker, Klaus Pillen.

**Data curation:** Mathias Ruben Gemmer.

**Formal analysis:** Mathias Ruben Gemmer, Chris Richter, Thomas Schmutzer, Manish L. Raorane.

**Funding acquisition:** Klaus Pillen.

**Investigation:** Mathias Ruben Gemmer, Andreas Maurer.

**Methodology:** Chris Richter, Yong Jiang, Manish L. Raorane.

**Project administration:** Björn Junker, Klaus Pillen.

**Supervision:** Manish L. Raorane, Björn Junker, Klaus Pillen, Andreas Maurer.

**Visualization:** Mathias Ruben Gemmer, Thomas Schmutzer.

**Writing – original draft:** Mathias Ruben Gemmer, Chris Richter, Yong Jiang, Manish L. Raorane, Björn Junker, Klaus Pillen, Andreas Maurer.

**Writing – review & editing:** Mathias Ruben Gemmer, Chris Richter, Yong Jiang, Manish L. Raorane, Björn Junker, Klaus Pillen, Andreas Maurer.

## References

1. Statista. Anbaufläche der wichtigsten Getreidearten weltweit in den Jahren 2010/11 bis 2018/19 2019. Available from: <https://de.statista.com/statistik/daten/studie/28883/umfrage/anbauflaeche-von-getreide-weltweit/>.
2. Sakuma S, Salomon B, Komatsuda T. The domestication syndrome genes responsible for the major changes in plant form in the Triticeae crops. *Plant and cell physiology*. 2011; 52(5):738–49. <https://doi.org/10.1093/pcp/pcr025> PMID: 21389058
3. Zamir D. Improving plant breeding with exotic genetic libraries. *Nature reviews genetics*. 2001; 2(12):983. <https://doi.org/10.1038/35103590> PMID: 11733751
4. Tanksley SD, McCouch SR. Seed banks and molecular maps: unlocking genetic potential from the wild. *Science*. 1997; 277(5329):1063–6. <https://doi.org/10.1126/science.277.5329.1063> PMID: 9262467
5. Heffner EL, Sorrells ME, Jannink J-L. Genomic selection for crop improvement. *Crop Science*. 2009; 49(1):1–12.
6. Goddard M, Hayes B. Genomic selection. *Journal of Animal breeding and Genetics*. 2007; 124(6):323–30. <https://doi.org/10.1111/j.1439-0388.2007.00702.x> PMID: 18076469
7. Meuwissen T, Hayes B, Goddard M. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*. 2001; 157(4):1819–29. PMID: 11290733
8. Crossa J, Perez P, Hickey J, Burgueno J, Ornella L, Cerón-Rojas J, et al. Genomic prediction in CIM-MYT maize and wheat breeding programs. *Heredity*. 2014; 112(1):48. <https://doi.org/10.1038/hdy.2013.16> PMID: 23572121
9. Aghnoum R, Marcel TC, Johrde A, Pecchioni N, Schweizer P, Niks RE. Basal host resistance of barley to powdery mildew: connecting quantitative trait loci and candidate genes. *Molecular plant-microbe interactions*. 2010; 23(1):91–102. <https://doi.org/10.1094/MPMI-23-1-0091> PMID: 19958142
10. Grewal TS, Rosnagel BG, Scoles GJ. Mapping quantitative trait loci associated with spot blotch and net blotch resistance in a doubled-haploid barley population. *Molecular breeding*. 2012; 30(1):267–79.
11. Maurer A, Draba V, Jiang Y, Schnaithmann F, Sharma R, Schumann E, et al. Modelling the genetic architecture of flowering time control in barley through nested association mapping. *Bmc Genomics*. 2015; 16(1):290.
12. Bernardo R. Molecular markers and selection for complex traits in plants: learning from the last 20 years. *Crop science*. 2008; 48(5):1649–64.
13. Meyer RC, Steinfath M, Lisek J, Becher M, Witucka-Wall H, Törjék O, et al. The metabolic signature related to high plant growth rate in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences*. 2007; 104(11):4759–64.
14. Schauer N, Semel Y, Roessner U, Gur A, Balbo I, Carrari F, et al. Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. *Nature biotechnology*. 2006; 24(4):447. <https://doi.org/10.1038/nbt1192> PMID: 16531992
15. Dan Z, Hu J, Zhou W, Yao G, Zhu R, Zhu Y, et al. Metabolic prediction of important agronomic traits in hybrid rice (*Oryza sativa* L.). *Scientific reports*. 2016; 6:21732. <https://doi.org/10.1038/srep21732> PMID: 26907211
16. Steinfath M, Strehmel N, Peters R, Schauer N, Groth D, Hummel J, et al. Discovering plant metabolic biomarkers for phenotype prediction using an untargeted approach. *Plant Biotechnology Journal*. 2010; 8(8):900–11. <https://doi.org/10.1111/j.1467-7652.2010.00516.x> PMID: 20353402

17. Riedelsheimer C, Czedik-Eysenberg A, Grieder C, Lisec J, Technow F, Sulpice R, et al. Genomic and metabolic prediction of complex heterotic traits in hybrid maize. *Nature genetics*. 2012; 44(2):217. <https://doi.org/10.1038/ng.1033> PMID: 22246502
18. Dixon RA. Phytochemistry meets genome analysis, and beyond. *Phytochemistry*. 2003; 62:815–6. [https://doi.org/10.1016/s0031-9422\(02\)00712-4](https://doi.org/10.1016/s0031-9422(02)00712-4) PMID: 12590109
19. Joseph B, Corwin JA, Li B, Atwell S, Kliebenstein DJ. Cytoplasmic genetic variation and extensive cyto-nuclear interactions influence natural variation in the metabolome. *Elife*. 2013; 2:e00776. <https://doi.org/10.7554/eLife.00776> PMID: 24150750
20. Chan EK, Rowe HC, Hansen BG, Kliebenstein DJ. The complex genetic architecture of the metabo-lome. *PLoS genetics*. 2010; 6(11):e1001198. <https://doi.org/10.1371/journal.pgen.1001198> PMID: 21079692
21. Rowe HC, Hansen BG, Halkier BA, Kliebenstein DJ. Biochemical networks and epistasis shape the Ara-bidopsis thaliana metabolome. *The Plant Cell*. 2008; 20(5):1199–216. <https://doi.org/10.1105/tpc.108.058131> PMID: 18515501
22. Chen W, Gao Y, Xie W, Gong L, Lu K, Wang W, et al. Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. *Nature genetics*. 2014; 46(7):714. <https://doi.org/10.1038/ng.3007> PMID: 24908251
23. Luo J. Metabolite-based genome-wide association studies in plants. *Current opinion in plant biology*. 2015; 24:31–8. <https://doi.org/10.1016/j.pbi.2015.01.006> PMID: 25637954
24. Fernie AR, Schauer N. Metabolomics-assisted breeding: a viable option for crop improvement? *Trends in genetics*. 2009; 25(1):39–48. <https://doi.org/10.1016/j.tig.2008.10.010> PMID: 19027981
25. Bayer MM, Rapazote-Flores P, Ganal M, Hedley PE, Macaulay M, Plieske J, et al. Development and evaluation of a barley 50k iSelect SNP array. *Frontiers in plant science*. 2017; 8:1792. <https://doi.org/10.3389/fpls.2017.01792> PMID: 29089957
26. Rutkoski JE, Poland J, Jannink J-L, Sorrells ME. Imputation of unordered markers and the impact on genomic selection accuracy. *G3: Genes, Genomes, Genetics*. 2013; 3(3):427–39.
27. Maurer A, Pillen K. 50k Illumina Infinium iSelect SNP Array data for the wild barley NAM population HEB-25 e!DAL—Plant Genomics and Phenomics Research Data Repository (PGP). 2019. <https://doi.org/10.5447/ipk/2019/20>
28. Arend D, Lange M, Chen J, Colmsee C, Flemming S, Hecht D, et al. e! DAL—a framework to store, share and publish research data. *BMC bioinformatics*. 2014; 15(1):214.
29. Lancashire PD, Bleiholder H, Boom Tvd, Langelüddeke P, Stauss R, WEBER E, et al. A uniform deci-mal code for growth stages of crops and weeds. *Annals of applied Biology*. 1991; 119(3):561–601.
30. SAS SiC, North Carolina, USA. 2013.
31. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation forStatistical Computing; 2018.
32. Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest: Tests in Linear Mixed Effects Models. R package version 2.0–33. 2016.
33. Friendly M. Corrgrams: Exploratory displays for correlation matrices. *The American Statistician*. 2002; 56(4):316–24.
34. Box GE, Cox DR. An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1964; 26(2):211–43.
35. Yang J, Zeng J, Goddard ME, Wray NR, Visscher PM. Concepts, estimation and interpretation of SNP-based heritability. *Nature genetics*. 2017; 49(9):1304. <https://doi.org/10.1038/ng.3941> PMID: 28854176
36. Covarrubias-Pazarán G. Genome-assisted prediction of quantitative traits using the R package som-mer. *PLoS one*. 2016; 11(6):e0156744. <https://doi.org/10.1371/journal.pone.0156744> PMID: 27271781
37. Hadley Wickham RF, Henry Lionel and Müller Kirill. A Grammar of Data Manipulation. R package ver-sion 075. 2018.
38. Revelle W. psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA. 2018.
39. Wickham H. ggplot2: elegant graphics for data analysis: Springer; 2016.
40. Whittaker JC, Thompson R, Denham MC. Marker-assisted selection using ridge regression. *Genetics Research*. 2000; 75(2):249–52.
41. Meuwissen TH, Solberg TR, Shepherd R, Woolliams JA. A fast algorithm for BayesB type of prediction of genome-wide estimates of genetic value. *Genetics Selection Evolution*. 2009; 41(1):2.



42. Zhao Y, Gowda M, Liu W, Würschum T, Maurer HP, Longin FH, et al. Accuracy of genomic selection in European maize elite breeding populations. *Theoretical and Applied Genetics*. 2012; 124(4):769–76. <https://doi.org/10.1007/s00122-011-1745-y> PMID: 22075809
43. Pérez P, de Los Campos G. Genome-wide regression and prediction with the BGLR statistical package. *Genetics*. 2014; 198(2):483–95. <https://doi.org/10.1534/genetics.114.164442> PMID: 25009151
44. Henderson CR. *Applications of linear models in animal breeding*: University of Guelph Guelph; 1984.
45. Hjorth JU. *Computer intensive statistical methods: Validation, model selection, and bootstrap*: Routledge; 1993.
46. Clark SA, Hickey JM, Van der Werf JH. Different models of genetic variation and their effect on genomic evaluation. *Genetics Selection Evolution*. 2011; 43(1):18.
47. Herzig P, Backhaus A, Seiffert U, von Wirén N, Pillen K, Maurer A. Genetic dissection of grain elements predicted by hyperspectral imaging associated with yield-related traits in a wild barley NAM population. *Plant Science*. 2019; 285:151–64. <https://doi.org/10.1016/j.plantsci.2019.05.008> PMID: 31203880
48. Sallam A, Endelman J, Jannink J-L, Smith K. Assessing genomic selection prediction accuracy in a dynamic barley breeding population. *The Plant Genome*. 2015; 8(1).
49. Werner CR, Voss-Fels KP, Miller CN, Qian W, Hua W, Guan C-Y, et al. Effective Genomic Selection in a Narrow-Genepool Crop with Low-Density Markers: Asian Rapeseed as an Example. *The plant genome*. 2018.
50. Desta ZA, Ortiz R. Genomic selection: genome-wide prediction in plant improvement. *Trends in plant science*. 2014; 19(9):592–601. <https://doi.org/10.1016/j.tplants.2014.05.006> PMID: 24970707