Metabolomic Spectra For Phenotypic Prediction of Malting Quality In Spring Barley

Xiangyu Guo (Xiangyu.Guo@qgg.au.dk)
Aarhus University
Ahmed Jahoor
Nordic Seed A/S
Just Jensen
Aarhus University
Pernille Sarup
Nordic Seed A/S

Research Article

Keywords: barley, malting quality, metabolomic prediction, MBLUP, PLSR, training population size

Posted Date: December 3rd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1113863/v1

Metabolomic Spectra For Phenotypic Prediction Of Malting Quality In Spring Barley

Xiangyu Guo[1, 2 *], Ahmed Jahoor[3, 4], Just Jensen[1], Pernille Sarup[3]

[1] Center for Quantitative Genetics and Genomics, Aarhus University, 8830 Tjele, Denmark

[2] SEGES, Pig Research Centre, 8200, Aarhus, Denmark

[3] Nordic Seed A/S, 8300 Odder, Denmark

[4] Department of Plant Breeding, The Swedish University of Agricultural Sciences, 2353 Alnarp, Sweden

[*]Correspondance:

Dr. Xiangyu Guo

Xiangyu.Guo@qgg.au.dk

Abstract

The objectives were to investigate prediction of malting quality (MQ) phenotypes in different locations using information from metabolomic spectra, and compare the prediction ability using different models and different sizes of training population (TP).

A total of 2,667 plots of 564 malting spring barley lines from three years and two locations were included. Five MQ traits were measured in wort produced from each individual plot. Metabolomic features (MFs) used were 24,018 NMR intensities measured on each wort sample. Models involved in the statistical analyses were a metabolomic best linear unbiased prediction (MBLUP) model and a partial least squares regression (PLSR) model. Predictive ability within location and across locations were compared using cross-validation methods.

The proportion of variance in MQ traits that could be explained by effects of MFs was above 0.9 for all traits. The prediction accuracy increased with increasing TP size but when the TP size reached 1,000, the rate of increase was negligible. The number of components considered in the PLSR models can affect the performance of PLSR models and 20 components were optimal. The accuracy of individual plots and line means using leave-one-line-out cross-validation ranged from 0.722 to 0.865 and using leave-one-location-out cross-validation ranged from 0.517 to 0.817.

In conclusion, it is possible to carry out metabolomic prediction of MQ traits using MFs, the prediction accuracy is high and MBLUP is better than PLSR if the training population is larger than 100. The results have significant implications for practical barley breeding for malting quality.

Keywords

barley, malting quality, metabolomic prediction, MBLUP, PLSR, training population size

Introduction

Brewing for alcohol production is the major end use of malt, and barley is the primary cereal used in production of malt because of optimal content of carbohydrates, dietary fibers, protein, vitamins and minerals [1]. In the process of brewing, the cereal needs to be malted. Under specific controlled conditions, cereal grains are sprouted and the young seedlings are grown for four to six days in order to produce malt [2]. This process ensure a physical and biochemical transformation within the grain that is defined as malting [1]. During the process of malting, the cell-wall is degraded and protein is dissolved by hydrolytic enzymes such that the physical and the biochemical structure of the barley grain is modified in order to allow malt to be used in the subsequent stages of brewing [3].

The quality of malt can directly affect the quality and quantity of brewed beer, thus malting quality (MQ) traits are important traits in breeding of barley to be used for malting. A series of MQ traits are defined in the malting industry. These include traits as extract yield and grain protein; alpha-amylase, beta-glucanase, beta-glucan, soluble protein, and free amino nitrogen in wort; and some physical properties such as diastatic power, viscosity, taste, flavor, haze and foam head retention [4]. Measurement of MQ traits is expensive and labor-intensive and the MQ traits have been demonstrated to have complex inheritance [5,6]. A detailed analysis of genetic variation in MQ traits in spring barley was provided by a previous study [6], where a population of 1,329 spring barley lines from four breeding cycles were investigated and medium to high narrow sense heritabilities were found for the MQ traits included in this study.

The organic compounds in a plant are mostly produced by the plant itself so that the photosynthetic and metabolic capacity of a plant is the primary factor determining its growth potential [7]. Metabolites are typical intermediates of biochemical reactions during the growth and development at all stages of plant life [8]. A comprehensive view of cellular metabolites can be provided by metabolomics, which is an approach to quantify the endogenous metabolites in

cells and organisms. The development of metabolomics has contributed to the molecular and biological characterization of various organisms. Especially in the area of crops, compared with animals and microorganisms, metabolomics is of great importance since the crops produce very large array of metabolites collectively [9]. Omics technologies like genomics, transcriptomics, and metabolomics can be used in the investigation for the biological background in different organisms [10].

Nuclear magnetic resonance (NMR) spectroscopy is one of the technologies used to analyze many metabolites simultaneously [11]. NMR can produce signal intensities, which can be treated as an indicator of metabolites in a biological sample, were defined as metabolomic features (MFs) [12]. A total of 24,018 MFs from barley wort were investigated in a previous study where the genetic variation in the MFs was investigated using a univariate model and 8,604 MFs were found to be significantly heritable [13].

Partial least squares (PLS) was first developed for the modelling of information-scarce situations in social sciences by Wold [14]. It is a latent variable approach which has been used to find fundamental relationships between two matrices by modelling the inner covariance structures [15]. PLS regression (PLSR) is the simplest PLS approach, and is a dimension reduction method which has been widely used in chemometrics [16]. Similar with traditional regression, PLSR relates two matrices by a linear multivariate model, but compared with traditional regression, the structure of two matrices can also be modelled when using PLSR [16]. The use of PLSR in chemistry first started in 1980s and has increased steadily for about 40 years, due to its appealing mathematical properties [17]. PLSR is able to analyze data sets with a large number of explanatory variables compared to the number of observations, in cases of noisy data, multi-collinearity, and incomplete variables in both the matrix of dependent variables and the matrix of predictor variables [16,18].

Best linear unbiased prediction (BLUP), which is a method allowing prediction of random effects in a mixed model, was originally developed in animal breeding for prediction of breeding values (BVs) (Henderson, 1975). In the area of animal breeding, the selection of animals with highest BV was usually based on predicted/expected BVs (EBVs) derived from the records on the animals themselves and their relatives using BLUP. The use of BLUP is also widely studied in many other areas of research where the use of mixed linear models are relevant such as plant breeding [19].

BLUP can be used in general linear mixed models that include both fixed and random effects. The simplest case is BLUP without pedigree, where genotypic effect is treated as an independent unobservable normally distributed random variable and no relationships between individuals are considered [19]. Compared with a model based on individual performance, pedigree based BLUP leads to more accurate predictions and result in larger genetic gain because it efficiently uses information from all relatives by constructing an additive genetic relationship matrix (**A**), under the circumstances where genetic relationships between relatives exists (Falconer and Mackay, 1996). The higher the additive genetic relationship between the genotype of interest with its relatives, the more information can be gained from records of these related genotypes [19]. With the rapid development of biochip technology, genomic BLUP (GBLUP) has been developed and widely applied because it is easy and straightforward to be implemented since technically it just needed the replacement of the **A** matrix in pedigreed based BLUP by a genomic matrix (**G**) [20]. More recently, metabolomic BLUP (MBLUP) has been proposed by replacing the **A** or **G** matrix by a metabolomic similarity matrix (**M**) and MBLUP has been shown as an promising method [21].

In our previous study, around 36% of MFs were found having significant heritability and among which many were found to be correlated with MQ traits in spring barley [13]. With this information, it is worthwhile to investigate the role of MFs involved in the prediction of

109 phenotypes for MQ traits. PLSR is a popular method used in the studies of metabolic profiles
110 [22], and using metabolomic BLUP (MBLUP) model gave better prediction accuracies than the
111 BLUP model using genomic information for four of five quantitative traits investigated [23].
112 The objectives of this study were to: 1) investigate the possibility of prediction of phenotypes
113 of malting quality traits using metabolomic information; 2) compare the ability of predictions
114 using PLSR and MBLUP models; 3) study the effect of different training population size on
115 the accuracy of prediction; 4) explore the possibility of metabolomic prediction within and
116 across location; and 5) compare different number of components considered in the PLSR model.
117

118 ## Materials and methods

119 All the data used are available in a public accessible repository [24].

120 **Experiments**

121 In this study, a total of 2,667 plots of 564 spring barley malting lines were included. These
122 lines were part of the standard breeding program from Nordic Seed A/S. All experiments in
123 this study were conducted in land owned by Nordic Seed. There were no animal or human
124 experiments conducted for this research, the study also did not contain any GMO. Standard
125 farm operating procedure were used and therefore no ethical approval was needed for this study.
126 All the experiments involving plants adhered to plant ethics guidelines. Samples from two
127 locations in Denmark were used, and samples were taken from each plot individually and the
128 data covered three years from 2014 to 2016. In both locations, the fields were divided into trials,
129 which included 52 - 106 smaller plots (8.25 m^2). Each trial was designed as a randomized
130 complete block comprising 20 – 45 lines with three replicates of each line [25]. Each trial included
131 two control lines in three replications. As a consequence, testing was conducted in a number
132 of trials within each year-location combination. In total there were 139, 214 and 215 lines tested
133 in 2014, 2015, and 2016, respectively.

Measurements of malting quality traits

The malt sample from each plot was milled and extracted in water in order to produce a wort as described in the previous study [6]. The wort was used to measure five MQ traits which included filtering speed (FS), extract yield (EY), wort color (WCO), beta glucan content (BG), and wort viscosity (WV). The wort samples needed to be filtered first, and 20 minutes after filtering begun, FS was scored by measuring the height of the liquid surface in the glass (cm flow-through in 20 min). EY was the percentage of dry matter in the filtered wort. Spectrophotometer was used to determine WCO following the method of European Brewery Convention (EBC) [26]. After the filtration, the wort samples were separated in two parts and all wort phenotypes were obtained according to the Analytica-EBC 2004 manual. Briefly, one sample of 25 ml of wort was used for WV (mPa/s, Analytical-EBC 8.4) and EY (Analytical-EBC 8.3). A second sample of 3 - 4 ml of wort was used for BG (mg/L, Analytical-EBC 8.13.1) and WCO (Analytical-EBC 8.5). Detailed description of MQ traits also can be found in a previous study by Sarup, et al. [6].

Metabolomic features and NMR intensities

The preparation of NMR analysis is described in detail in Guo, et al. [13]. MFs used in this study were 24,018 NMR intensities which obtained from one-dimensional (1D) ^{1}H NMR spectra. The NMR intensities were integrated over small chemical shift (δ) intervals and expressed in parts per million (ppm) in the frequency range of 0.00 ppm to 11.00 ppm. An in-house custom Matlab script was used to process the spectra [27]. First an exponential apodization function equivalent to 0.5 Hz line-broadening was used and then Fourier transformation was applied. Afterwards, all spectra were referenced to the DSS-d_6 signal, automatically phased, and baseline corrected. After visual inspection data below 0.70 ppm and above 9.00 ppm was removed as it did not contain any signal. The water peak which was in the range of 4.7 ppm to 4.9 ppm, and the region of the added standard which was -0.2 ppm to 0.2 ppm, were excluded.

159 The raw data was then normalized using the probabilistic quotient method [28], and the spectra
160 were aligned using icoshift [29,30]. Finally, the MFs were centered and standardized to a mean of
161 0 and standard deviation as 1 in order to refine variation that could be attributed to experimental
162 sources and signal intensities [12].

163 **Statistical models and methods**

164 Two models involved in the statistical analyses. The models were a metabolomic best linear
165 unbiased prediction (MBLUP) model and a partial least squares regression (PLSR) model.

166 **Metabolomic best linear unbiased prediction (MBLUP)**

167 MBLUP model was as follows:

168 $\boldsymbol{y} = \mathbf{1}\boldsymbol{\mu} + \boldsymbol{m} + \boldsymbol{e}$,

169 where $\boldsymbol{y}$ referred to the vector of each MQ trait, $\boldsymbol{\mu}$ was intercept, $\boldsymbol{m}$ was the vector metabolomic
170 effects, and $\boldsymbol{e}$ was a vector of residual terms that could not be explained by the other effects in
171 the model. In this model, $\boldsymbol{\mu}$ was a fixed parameter, $\boldsymbol{m}$ was a random parameter with
172 $\boldsymbol{m} \sim N(0, \boldsymbol{M}\sigma_m^2)$, and $\boldsymbol{e}$ was a random parameter with $\boldsymbol{e} \sim N(0, \boldsymbol{I}\sigma_e^2)$. The reason for only $\boldsymbol{\mu}$ taken
173 as fixed parameter, instead of following the model in our previous study to consider more fixed
174 parameters [13], was because metabolomic information include environmental factors in addition
175 to genomic information. **M** denoted the metabolomic similarity matrix built from MFs using
176 the method as for building a genomic relationship matrix (**G**) computed using VanRaden
177 method 1 [20]. Specifically, $\boldsymbol{M} = \frac{\boldsymbol{QQ'}}{m}$, where **Q** is a n × m matrix of adjusted, centered and
178 scaled NMR intensities with m = 24,018 (equal to number of MFs) and n = 2,667 (equal to
179 number of samples). Both the build of **M** and the MBLUP analysis were carried out by using
180 the "qgg" R-package [31].

181 The (co)variance components in the MBLUP model described in the previous section were
182 estimated by restricted maximum likelihood using "qgg" R-package [31].

The total variation of each MQ trait was calculated as the sum of variance components in MBLUP model:

$\sigma_P^2 = \bar{M}\sigma_m^2 + \sigma_e^2$,

where $\bar{M}$ was the average diagonal of **M**, which equal to 1. The relative variance component due to effects of $\boldsymbol{m}$ was $RVC_m = \frac{\bar{G}\sigma_m^2}{\sigma_P^2}$ which describe the proportion of total variation (across fixed effects) in MQ traits that can be described by the MFs.

Partial least squares regression (PLSR)

The PLSR model decompose **Q**, the matrix of MFs, into orthogonal scores **T** and loadings **P**:

$\boldsymbol{Q} = \boldsymbol{TP}$,

so that regressing $\boldsymbol{y}$ not on **Q** itself but on the first *t* columns of the scores $\boldsymbol{T}$, *t* is the number of components fitted in the model when using the package mentioned below. Detailed description of PLSR method can be found in previous study on the "pls" package [18], which was used in the current study to carry out the PLS analysis.

Cross-validation

Three different leave-set-out (LSO) cross-validation strategies, in which the whole dataset was divided into a training population (TP) and a validation population (VP), were investigated in this study based on three different hypotheses regarding factors that influence prediction accuracies. The first strategy, named SIZE, was to randomly leave out VP in order to create TP of different size; the second strategy, named LINE, was to leave out VP according to line i.e. all observations of a specific line; the third strategy, named LOC, was to leave out VP according to location.

In the SIZE strategy, since the TP samples and VP samples were randomly selected, the TP contained observations on the same lines, locations and years as in VP – although not the same combinations of lines, locations and years. In the LINE strategy, one out of 564 lines was left out so that the accuracy of predicting one line from all the other lines could be investigated,

this strategy is similar to prediction of new lines. In the LOC strategy, one out of two locations was left out, which means the accuracy of predicting one location based on data from the other one location could be investigated.

For each strategy, cross-validation was carried out to evaluate the accuracy of metabolomic prediction of five MQ traits using MFs. In order to study the effect of different size of TP in SIZE strategy, eight scenarios were investigated in this study. These scenarios varied on TP having the size of 50, 100, 200, 500, 1,000, 1,500, 2,000, and 2,500. Since the selection of TP was random, 15 replicates were carried out when selecting the TP. For the strategy of LINE, data from 564 lines were left out separately so that 564 replicates were carried out in this strategy. For the strategy of LOC, data from one of the two locations were left in turn so each location were predicted based on the other location.

In each round of the cross-validation, according to the setup of TP size, a certain number of the phenotypes were selected and then the rest of phenotypes were masked. The phenotypes of the masked samples were predicted based on the TP together with the metabolomic information. Thereafter, the correlation between phenotypes and the predicted values was calculated as the accuracy of prediction. The accuracies obtained from strategy LINE and LOC were computed based on both plot level and line mean level. This means, two accuracies were obtained for each model in LINE and LOC strategies. In each round of the prediction in LINE and LOC strategies, the predicted values for VP in this round were collected, and then when all the rounds of prediction completed, predicted values were collected for the whole population. Afterwards, the accuracy on plot level in LINE and LOC strategies was calculated as the correlation between observed phenotypes and the predicted values of each plot, and the accuracy of line mean was calculated as correlation between average observed phenotypes and the average predicted values of each line.

When applying PLSR model, compared with MBLUP model, a leave-one-sample-out (LOO) cross-validation was carried out within the TP to train the model first, and afterwards the LSO cross-validation was processed using the trained model from the preliminary LOO cross-validation. In this study, different number of components considered in the PLSR model was also compared, and the number of components were 5, 10, 20 and 50.

Software and setup

The cross-validation procedure using MBLUP was carried out by "qgg" package [31] and the procedure using PLSR model was carried out by "pls" package [18]. In all three strategies, both MBLUP and PLSR models were applied, and the datasets utilized in each round of the cross-validation were same for MBLUP and PLSR models to make sure they were comparable.

Results

In this study, the proportion of variance in malting quality (MQ) traits that can be explained by effects of metabolomic features (MFs) was evaluated. Then the phenotype of MQ traits was predicted by using MFs through MBLUP and/or PLSR models. Different size of TP, and the different number of components considered in the PLSR models, prediction across line and location were also investigated.

Descriptive statistics for malting quality traits

Table 1 gives descriptive statistics of all the MQ traits analyzed in this study. There were 2,667 records analyzed for five MQ traits. The average of all the traits were 4.83 for FS, 82.66 for EY, 5.83 for WCO, 217.10 for BG, and 1.47 for WV. The coefficient of phenotypic variance ranged from 2.21% for EY to 53.07% for BG.

Estimates of total variance of malting quality traits explained by metabolomic features

A univariate MBLUP model was applied to estimate the proportion of the total variance including potential fixed effects in each MQ trait that can be explained by the MFs. The

257 estimates were indicators for the proportion of total variance in MQ traits that is associated
258 with metabolites.
259 Figure 1 shows the estimated relative amount of total variance explained by MFs (RVCm) and
260 error in five MQ traits. The RVCm ranged from 0.926 ± 0.011 in FS to 0.981 ± 0.002 in BG.
261 For all the MQ traits, the effect of MFs explained very large proportions of the total variance.
262 RVCm in WCO, BG and WV were similar and larger than RVCm in FS and EY.
263 **Metabolomic prediction using MBLUP model**
264 A univariate MBLUP model was used for metabolomic prediction of the five MQ traits. The
265 cross-validation results from MBLUP model at each SIZE scenario are shown in Figure 2. As
266 shown in Figure 2, averaged across 15 replicates in the strategy of SIZE, the maximum
267 prediction accuracies using MBLUP model were 0.757 ± 0.027 for FS, 0.736 ± 0.050 for EY,
268 0.838 ± 0.025 for WCO, 0.780 ± 0.033 for BG and 0.816 ± 0.028 for WV. In addition, for all
269 traits, the maximum accuracy were obtained in the scenarios with 2,500 as TP size.
270 **Metabolomic prediction using PLSR model**
271 There were four PLSR models compared regarding the number of components utilized in the
272 model. The number of components considered were 5, 10, 20 and 50. As shown in Figure 3,
273 averaged across 15 replicates, the maximum prediction accuracies using PLSR model were
274 0.741 ± 0.026 for FS, 0.730 ± 0.053 for EY, 0.826 ± 0.024 for WCO, 0.760 ± 0.032 for BG
275 and 0.805 ± 0.028 for WV. In addition, all the maximum accuracy were obtained when using
276 PLSR model with 20 components, except FS, for which the maximum accuracy was provided
277 by the PLSR model with 10 components, but it was very close to the accuracy provided by
278 PLSR model with 20 components.
279 When the PLSR model considered 5 components, the prediction accuracy was low for all the
280 MQ traits. With increase in the number of components considered in the PLSR model, the
281 accuracy also generally increased. The accuracy kept increasing until the number of

components reached 20. However, when the number of components increased further to 50, the accuracy decreased and in some cases even smaller than the accuracy from 5 components.

Comparison of MBLUP and PLSR models

The accuracy from MBLUP and the maximum accuracy among four PLSR models are plotted for each SIZE scenario in Figure 2. The accuracy obtained from MBLUP model was smaller than at the maximum accuracy from PLSR model when the TP size was small. When the TP size was 50, MBLUP yielded smaller accuracy in all the five MQ traits. With the increase of TP size, the accuracy from MBLUP increased rapidly and was larger than for the PLSR model. For example, in FS, MBLUP yielded higher accuracy than PLSR when the TP size just increased to 100. For all the traits, MBLUP yielded higher or same accuracy compared with all the PLSR models when the TP size reached 500.

Metabolomic prediction with different training population size

A total of eight sizes, including 50, 100, 200, 500, 1,000, 1,500, 2,000, and 2,500, of TP was compared. With the increase of TP size, the prediction accuracy increased regardless of the model used. When the TP size was small, PLSR could provide better predictions than MBLUP model, while along with the increase of TP size, the MBLUP outperformed PLSR models as soon as the data size reached 500 samples. As can be observed from Figure 2, though the general trend was higher accuracy obtained in the scenario of largest TP size. When the TP was increased beyond 1,000, the increase in accuracy was limited.

Metabolomic prediction of new lines

The second cross-validation strategy investigated in this study was LINE, in which the data from one line were masked as VP and the data from the other lines were treated as TP to predict the VP. This corresponds to predicting a new line based on metabolomic information only. There were 564 lines in the whole dataset, one line was left out and then predicted based on all other lines. This process was repeated until all lines were predicted. Since PLSR model with

20 components generally yielded highest accuracy, only this PLSR model was conducted and compared with MBLUP model in this strategy. As shown in Figure 4, when using MBLUP model, the accuracy of plot ranged from 0.722 ± 0.013 for EY to 0.832 ± 0.011 for WCO, and the accuracy of line mean ranged from 0.800 ± 0.025 for EY to 0.865 ± 0.021 for WV. The MBLUP surpassed PLSR model though the difference are small.

Metabolomic prediction across location

The third cross-validation strategy investigated in this study was LOC, in which the data from one location were masked as VP and the data from the other location were treated as TP to predict the VP. There were two locations in the whole dataset, therefore, this strategy had two rounds of prediction by treating each location as VP in each round. Same with LINE strategy, both MBLUP and the PLSR with 20 components were carried out in this strategy. As shown in Figure 5, the accuracy of plot ranged from. 0.517 ± 0.017 for EY to 0.684 ± 0.014 for FS, the accuracy of line mean ranged from 0.713 ± 0.030 for WCO to 0.817 ± 0.024 for BG, when using MBLUP model. The accuracy provided by PLSR model were similar with MBLUP.

Discussion

Metabolomic prediction using 24,018 metabolomic features (MFs) were carried for a total of 2,667 plots of 564 spring barley malting lines each phenotyped for five malting quality (MQ) traits. MBLUP and PLSR models were compared. Accuracy of cross-validation was investigated by varying size of training population, also using leave-one-line-out and leave-one-location-out strategies. In addition, considering number of components in the PLSR model was also studied.

Descriptive statistics for malting quality traits

The descriptive statistics for MQ traits in the current study were similar with the previous study though the number of observations in the previous study was around three times larger than in

the current study [6]. The standard deviation of most of the MQ traits in the current study were smaller than the previous study and it is expected because in the previous study, 1,329 spring malting barley lines were involved and the harvest was done in four different years and three locations [6], so that the samples from the current study was a subset of the previous study. The more variation in the year and location compared with the current led to the larger variation in phenotypes.

Estimates of total variance of malting quality traits explained by metabolomic features

The variance of five MQ traits explained by MFs was explored by using a univariate model integrating a metabolomic similarity matrix and the proportion of metabolomic effects were larger than 90% for all the five traits.

The utilization of metabolomic similarity matrix in the model aimed at dissection of the total variance into a metabolomic part and a random error. The proportion of the variance of MFs shows the extent that MFs can be used to predict total variance in MQ traits. In our previous study, WCO, BG and WV were found to have significant phenotypic and genetic correlation to a large proportion of the MFs [13], similarly to a large extend, their total variance could be explained by MFs. Potentially MQ traits of WCO, WV and BG can be predicted from the MFs because MFs explains almost all the variance in these MQ traits. While among the two traits (FS and EY) having relative lower proportion of correlation with MFs, they could not be explained to the same very high degree by variation in metabolites.

The direct link between metabolites and phenotypic records in biological system provide the potential of utilizing metabolomic features as an objective proxy for phenotype data [32]. The fact that almost all the variation in MQ can be explained by the MFs confirmed that the MFs could be used as the objective proxy for phenotype of interest and even more valuable and meaningful when the phenotype of interest is difficult or expensive to be obtain.

Metabolomic prediction using MBLUP model

357 The MBLUP model used for estimation of variance components was then used for
358 metabolomic prediction of five MQ traits. The prediction accuracies using MBLUP model
359 were quite promising as the maximum accuracy were all above 0.7. This is higher than the
360 previous reported prediction accuracies for metabolomic prediction of plant phenotypes [7,33-
361 36]. The higher prediction accuracy in this study is probably due to a larger number of unique
362 genotypes in the study and the fact that the NMR was performed directly on wort and not on
363 e.g. leaves of the developing plant. When utilizing genomic information and fitting a
364 genomic BLUP (GBLUP) model, the comparable accuracies for MQ traits were reported as
365 from 0.28 to 0.68 [6]. The accuracy of GBLUP in the previous study was lower than the
366 accuracy of MBLUP in the current study can be due to metabolomic data included
367 information on both genetic factors as well as environmental factors. Thus the metabolomic
368 information was closer related with phenotype observations than the genomic information.
369 Though the spring barley lines involved in the current study were from the same breeding
370 program as the lines in the previous study [6], the number of lines studied in the current study
371 was a subset from the previous one, which can also lead to the difference in prediction
372 accuracy. However the accuracy of GBLUP is expected to be even lower if using the same
373 dataset as in the current study, which is smaller than and be a subset of the dataset in the
374 previous study [6].
375 One of the reasons for the high prediction accuracy using MBLUP could be because the total
376 variance explained by MFs were large in all the five MQ traits. The phenotypes of MQ traits
377 can be predicted very well and better than when using GBLUP, because most variation in MQ
378 is expected to be reflected in the NMR spectra. The high accuracy also shows that there is no
379 overfitting and MBLUP can explain and predict large variation in MQ. In addition, our
380 previous study on the genetic and phenotypic correlation between MFs and MQ traits also
381 showed a significant correlation between them, which can also be the reason for the high

382 prediction accuracy using metabolomic information [13]. A subset of MFs were detected as
383 significantly heritable, and a further subset of these had significant genetic correlation with
384 MQ traits in our previous study [13].Therefore, we carried out extra analysis in order to compare
385 the performance of MBLUP using different subsets of MFs. Three matrices were built
386 regarding to MFs included, the matrix using all the 24,018 MFs was M, the matrix using
387 significant heritable MFs was Ms, and the matrix using MFs which were significant heritable
388 and also had significant genetic correlation with each trait was Mgs (varied across traits). The
389 estimation of variance due to MFs were quite similar among the MBLUP models using M, Ms
390 and Mgs. The accuracy of prediction for using these three MBLUP models with different
391 training population size are shown in the Supplementary Figure S1. Very similar results
392 obtained from three MBLUP models (M, Mgs, Ms) indicated that selecting the significant
393 heritable MFs and/or MFs having significant genetic correlation with traits did not improve the
394 prediction accuracy of MBLUP. When applying MBLUP in the breeding system, breeders can
395 directly utilize all the MFs instead of filtering out some part of MFs which may involve more
396 work, cost, and potential for errors.
397 The performance of GBLUP and MBLUP was investigated in *Drosophila*, where the prediction
398 accuracy for two behavioral traits was below 0.1 when based on GBLUP and then increased to
399 above 0.4 when using MBLUP. Such an increase have also been found for two environmental
400 stress resistance traits in *Drosophila* [23]. In the plant field, metabolic information was introduced
401 into prediction of complex traits by Riedelsheimer, et al. [33], where the authors presented a
402 complementary approach to exploit large-scale genomic and metabolic information in hybrid
403 testcrosses. The MBLUP was also investigated in a previous study [37], where metabolomics data
404 were used to predict the performance agronomic traits in wheat, and metabolomic information
405 were found as providing strong predictive power for number of grains per spike and plant
406 height [37].

Metabolomic prediction using PLSR model

Four PLSR models were compared with different the number of components considered in the model. The maximum accuracies provided by PLSR models were smaller than for the MBLUP model. Increasing the number of components considered in the PLSR models generally led to the increase of prediction accuracy while when the number of components reached 50, the prediction accuracy was not larger than the one provided by the models only considered 20 components. These results indicated that the non-linear relationship between the number of components and the performance of the prediction. For all the MQ traits investigated in the current study, 20 components were already enough to provide good prediction accuracy though the exact number of components varied a bit from trait to trait. In a study of genomic selection for pork pH traits, 30 was found as the optimal number of components considered in the PLSR analysis [38].

PLSR has also been suggested as an efficient method to analyze genomic data, because of its ability to handle large data sets and its prediction ability, and the PLSR approach is particularly suitable to predict dependent variables from a very large number of predictors and especially the predictors might be highly correlated with each other [39]. The accuracy of prediction for yield traits in French dairy cattle were similar between PLSR and GBLUP models but in no case PLSR provided higher accuracy than GBLUP [39]. It was also reported that an increase in the number of relevant variables and observations contributed to the improvement in the precision of the model parameters, which was one desirable property of PLSR model [16].

Comparison of MBLUP and PLSR models

The comparison of MBLUP and PLSR models showed that MBLUP generally outperformed PLSR for all the traits, when the TP size larger than 500. PLSR could be a better choice than MBLUP only when the TP was small. In a previous study the Xu, et al. [21] analyzed a hybrid population of rice, and showed that the MBLUP model was superior to PLSR model [21]. A

similar situation was also found when utilizing genomic information instead of metabolomic information. For example, a study on rice also investigated the GBLUP model and PLSR using genomic information, showed that the GBLUP outperformed PLSR [21]. The superiority of BLUP model was also found in the study on genomic prediction in French Holstein and Montbéliarde breeds [40]. In addition to the better performance of MBLUP, it is also easy to implement, needs low demands regarding computation power, time and skill for the breeder, which makes MBLUP is more attractive in the practical breeding.

Metabolomic prediction with different training population size

A total of eight TP sizes from 50 to 2,500 were compared in this study. The results showed that the prediction accuracy generally increased with increasing TP size. Though the accuracy increased all the way from smallest TP until the largest dataset, the increase in accuracy was much smaller when the TP were larger than 1,000. The impact of TP size on the prediction accuracy had been demonstrated in the genomic prediction while rare in the metabolomic prediction using metabolomic information [41,42]. For example, the accuracy of genomic prediction in wheat has been investigated regarding to different population sizes and the results indicated that TP of around 700 lines were enough to yield the highest prediction accuracy. [43].

Metabolomic prediction across line/location

In addition to the first cross-validation strategy which selecting TP randomly within the whole population, two more strategies were investigated either predict the VP from different lines or growing in different locations. The accuracy of predicting plot MQ from these two strategies were smaller than when the TP randomly selected from the whole population. The reason is because when selecting TP from the whole population randomly, the observations on the same lines and/or locations were involved in TP and VP, which increased the degree of the metabolomic similarity between TP and VP.

In this study, 564 lines were harvested in three years separately, which means there was almost no lines involved in two or three years. This design created difficulty in the metabolomic prediction across year based on the current dataset. The across year metabolomic prediction could be better investigated when a dataset including overlap of lines been planted in different years is available.

Conclusion

Records of five malting quality (MQ) traits and metabolomic features (MFs) for 2,667 plots of 564 spring malting barley lines that were grown in two locations were studied. The ability of prediction based on metabolomic information was investigated.

The proportion of variance in MQ traits that can be explained by effects of MFs was above 0.9 for all traits when using all the records. The phenotype of MQ traits could be predicted by MFs through MBLUP and/or PLSR models. The prediction accuracy when using MBLUP was larger than 0.7 and generally surpassed PLSR models when size of training population (TP) larger than 500. When the size of TP smaller than 500, PLSR provided better accuracy than MBLUP. The prediction accuracy increased along with increasing TP size but when the population size reached 1,000, the rate of increase was very small. The number of components considered in the PLSR models can affect the performance of PLSR models and 20 was the optimal number. In addition, the prediction accuracy was also explored regarding to using the TP to predict the validation population (VP) in a different year or location. The results showed that it was possible carry out the prediction across line/location with the accuracy of plot ranged from 0.5 to 0.8, and the accuracy of line mean ranged from 0.7 to 0.9.

In conclusion, it is possible to carry out prediction for phenotypes of malting quality traits using metabolomic information. MBLUP is an ideal model for the prediction when TP size larger

480 than 500. The results from the current study indicate that barley breeders can predict MQ based
481 on MFs from the wort and have significant implications for the practical barley breeding.

References

1 Gupta, M., Abu-Ghannam, N. & Gallaghar, E. Barley for Brewing: Characteristic Changes during Malting, Brewing and Applications of its By-Products. *Comprehensive Reviews in Food Science and Food Safety* **9**, 318-328, doi:https://doi.org/10.1111/j.1541-4337.2010.00112.x (2010).

2 Burger, W. C. & LaBerge, D. E. Malting and Brewing Quality. *Barley*, 367-401, doi:https://doi.org/10.2134/agronmonogr26.c13 (1985).

3 MacLeod, L. & Evans, E. Malting. *Reference Module in Food Science*, doi:https://doi.org/10.1016/B978-0-08-100596-5.00153-0 (2016).

4 Li, C. D., Cakir, M. & Lance, R. in *Genetics and Improvement of Barley Malt Quality* (eds Guoping Zhang & Chengdao Li) 260-292 (Springer Berlin Heidelberg, 2010).

5 Gao, W. *et al.* Fine mapping of a malting-quality QTL complex near the chromosome 4H S telomere in barley. *Theor Appl Genet* **109**, 750-760, doi:10.1007/s00122-004-1688-7 (2004).

6 Sarup, P. *et al.* Genomic prediction for malting quality traits in practical barley breeding programs. *bioRxiv*, 2020.2007.2030.228007, doi:10.1101/2020.07.30.228007 (2020).

7 Meyer, R. C. *et al.* The metabolic signature related to high plant growth rate in Arabidopsis thaliana. *P Natl Acad Sci USA* **104**, 4759-4764, doi:10.1073/pnas.0609709104 (2007).

8 Jewett, M. C., Hofmann, G. & Nielsen, J. Fungal metabolite analysis in genomics and phenomics. *Current Opinion in Biotechnology* **17**, 191-197, doi:https://doi.org/10.1016/j.copbio.2006.02.001 (2006).

9 Saito, K. & Matsuda, F. Metabolomics for Functional Genomics, Systems Biology, and Biotechnology. *Annual Review of Plant Biology* **61**, 463-489, doi:10.1146/annurev.arplant.043008.092035 (2010).

10 Roessner, U. & Bowne, J. What is metabolomics all about? *BioTechniques* **46**, 363-365, doi:10.2144/000113133 (2009).

11 Lu, W. *et al.* Metabolite Measurement: Pitfalls to Avoid and Practices to Follow. *Annu Rev Biochem* **86**, 277-304, doi:10.1146/annurev-biochem-061516-044952 (2017).

12 Aliakbari, A. *et al.* Genetic variance of metabolomic features and their relationship with body weight and body weight gain in Holstein cattle1. *Journal of Animal Science* **97**, 3832-3844, doi:10.1093/jas/skz228 (2019).

13 Guo, X. *et al.* Genetic Variance of Metabolomic Features and Their Relationship With Malting Quality Traits in Spring Barley. *Frontiers in Plant Science* **11**, doi:10.3389/fpls.2020.575467 (2020).

14 Wold, H. Soft Modelling by Latent Variables: The Non-Linear Iterative Partial Least Squares (NIPALS) Approach. *Journal of Applied Probability* **12**, 117-142, doi:10.1017/S0021900200047604 (1975).

15 Xu, S. & Hu, Z. Methods of plant breeding in the genome era. *Genetics Research* **92**, 423-441, doi:10.1017/S0016672310000583 (2010).

16 Wold, S., Sjöström, M. & Eriksson, L. PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems* **58**, 109-130, doi:https://doi.org/10.1016/S0169-7439(01)00155-1 (2001).

17 Carrascal, L., Galván, I. & Gordo, O. Partial least squares regression as an alternative to current regression methods used in ecology. *Oikos* **118**, 681-690, doi:10.1111/j.1600-0706.2008.16881.x (2009).

18 Mevik, B.-H. & Wehrens, R. The pls Package: Principal Component and Partial Least Squares Regression in R. *2007* **18**, 23, doi:10.18637/jss.v018.i02 (2007).

19 Piepho, H. P., Möhring, J., Melchinger, A. E. & Büchse, A. BLUP for phenotypic selection in plant breeding and variety testing. *Euphytica* **161**, 209-228, doi:10.1007/s10681-007-9449-8 (2008).

20 VanRaden, P. M. Efficient methods to compute genomic predictions. *Journal of dairy science* **91**, 4414-4423, doi:10.3168/jds.2007-0980 (2008).

21 Xu, S., Xu, Y., Gong, L. & Zhang, Q. Metabolomic prediction of yield in hybrid rice. *The Plant Journal* **88**, 219-227, doi:https://doi.org/10.1111/tpj.13242 (2016).

22 Sarup, P., Pedersen, S. M. M., Nielsen, N. C., Malmendal, A. & Loeschcke, V. The Metabolic Profile of Long-Lived Drosophila melanogaster. *PLOS ONE* **7**, e47461, doi:10.1371/journal.pone.0047461 (2012).

23 Rohde, P. D., Kristensen, T. N., Sarup, P., Muñoz, J. & Malmendal, A. Prediction of complex phenotypes using the Drosophila metabolome. *bioRxiv*, 2020.2006.2011.145623, doi:10.1101/2020.06.11.145623 (2020).

24 Guo, X. (Mendeley Data, 2020).

25 Nielsen, N. H. *et al.* Genomic Prediction of Seed Quality Traits Using Advanced Barley Breeding Lines. *PloS one* **11**, e0164494-e0164494, doi:10.1371/journal.pone.0164494 (2016).

26 Bishop, L. R. EUROPEAN BREWERY CONVENTION TESTS OF THE E.B.C. COLOUR DISCS FOR WORT AND BEER. *Journal of the Institute of Brewing* **72**, 443-451, doi:10.1002/j.2050-0416.1966.tb02988.x (1966).

27 Haggart, G., Pearce, J. & Sands, C. *ghaggart*, <https://zenodo.org/record/3077413> (2019).

28 Dieterle, F., Ross, A., Schlotterbeck, G. & Senn, H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics. *Anal. Chem.* **78**, 4281-4290 (2006).

29 Savorani, F., Tomasi, G. & Engelsen, S. B. icoshift: A versatile tool for the rapid alignment of 1D NMR spectra. *J Magn Reson* **202**, 190-202, doi:10.1016/j.jmr.2009.11.012 (2010).

30 Vu, T. N. & Laukens, K. Getting your peaks in line: a review of alignment methods for NMR spectral data. *Metabolites* **3**, 259-276 (2013).

31 Rohde, P. D., Fourie Sørensen, I. & Sørensen, P. qgg: an R package for large-scale quantitative genetic analyses. *Bioinformatics* **36**, 2614-2615, doi:10.1093/bioinformatics/btz955 (2019).

32 Daygon, V. & Fitzgerald, M. Application of metabolomics for providing a new generation of selection tools for crop improvement. *Hot Topics in Metabolomics: Food and Nutrition*, 106, doi:10.4155/9781909453821 (2013).

33 Riedelsheimer, C. *et al.* Genomic and metabolic prediction of complex heterotic traits in hybrid maize. *Nature Genetics* **44**, 217-220, doi:10.1038/ng.1033 (2012).

34 Gärtner, T. *et al.* Improved heterosis prediction by combining information on DNA- and metabolic markers. *PLoS One* **4**, e5220, doi:10.1371/journal.pone.0005220 (2009).

35 Steinfath, M. *et al.* Discovering plant metabolic biomarkers for phenotype prediction using an untargeted approach. *Plant Biotechnol J* **8**, 900-911, doi:10.1111/j.1467-7652.2010.00516.x (2010).

36 Feher, K. *et al.* Deducing hybrid performance from parental metabolic profiles of young primary roots of maize by using a multivariate diallel approach. *PLoS One* **9**, e85435, doi:10.1371/journal.pone.0085435 (2014).

37 Shi, T. *et al.* Metabolomics analysis and metabolite-agronomic trait associations using kernels of wheat (Triticum aestivum) recombinant inbred lines. *The Plant Journal* **103**, 279-292, doi:https://doi.org/10.1111/tpj.14727 (2020).

38 Silveira, F. G. d. *et al.* The optimal number of partial least squares components in genomic selection for pork pH. *Cienc Rural* **47** (2017).

39 Colombani, C. *et al.* A comparison of partial least squares (PLS) and sparse PLS regressions in genomic selection in French dairy cattle. *Journal of dairy science* **95**, 2120-2131, doi:https://doi.org/10.3168/jds.2011-4647 (2012).

40 Colombani, C. *et al.* Application of Bayesian least absolute shrinkage and selection operator (LASSO) and BayesCπ methods for genomic selection in French Holstein and Montbéliarde breeds. *Journal of dairy science* **96**, 575-591, doi:https://doi.org/10.3168/jds.2011-5225 (2013).

41 Goddard, M. Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica* **136**, 245-257, doi:10.1007/s10709-008-9308-0 (2009).

42 Daetwyler, H. D., Villanueva, B. & Woolliams, J. A. Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS One* **3**, e3395, doi:10.1371/journal.pone.0003395 (2008).

43 Cericola, F. *et al.* Optimizing Training Population Size and Genotyping Strategy for Genomic Prediction Using Association Study Results and Pedigree Information. A Case of Study in Advanced Wheat Breeding Lines. *PloS one* **12**, e0169606-e0169606, doi:10.1371/journal.pone.0169606 (2017).

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. PS, AJ are employed by Nordic Seed A/S. XG is employed by both Aarhus University and SEGES. The funders had no influence on the study design or choice of analysis methods.

Author Contributions

XG implemented and carried out the statistical analysis, interpreted the results and drafted the manuscript. AJ, PS, JJ and XG contributed to the experimental design and PS, JJ and XG developed the statistical models. All authors participated in interpreting the results and all authors read and approved the final manuscript.

Funding

This project was funded by the Innovation Fund Denmark (Grant number: 5184-00032B) and by Green Development and Demonstration Programme (GUDP, Grant number: 34009-19-1586).

Acknowledgments

The authors are grateful to Vahid Edriss, Nanna Hellum Kristensen, Jens Due Jensen, and Jette Andersen from Nordic Seed for field and laboratory work, Frans A. A. Mulder, Lars Alf Jensen and Benjamin Wahlqvist from Aarhus University for sample preparation and performing NMR measurements, and Palle Duun Rohde from Aalborg University for very helpful discussion.

Abbreviations

BG: beta glucan

BLUP: best linear unbiased prediction

BV: breeding value

CV: coefficient of phenotypic variance

GBLUP: genomic best linear unbiased prediction

G: genomic relationship matrix

EBC: European Brewery Convention

EBV: expected breeding value

EY: extract yield

FS: filtering speed

LOO: leave-one-sample-out

LSO: leave-set-out

M: metabolomic similarity matrix built from all metabolomic features

MBLUP: metabolomic best linear unbiased prediction

MFs: metabolomic features

Mgs: metabolomic similarity matrix built from significant heritable metabolomic features having significant genetic correlation with metabolomic traits

MQ: malting quality

Ms: metabolomic similarity matrix built from significant heritable metabolomic features

NMR: nuclear magnetic resonance

PLS: partial least squares

PLSR: Partial least squares regression

ppm: parts per million

RVC: relative variance component

TP: training population

VP: validation population

WCO: wort color

WV: wort viscosity

Tables

Table 1. Descriptive statistics for malting quality traits.

Trait	No. of records	Unit	Average	S.D.	Min	Max	CV
FS	2,667	cm/20 min	4.83	0.61	2.30	6.30	12.72%
EY	2,667	%	82.66	1.82	70.38	92.39	2.21%
WCO	2,667	EBC units	5.83	0.83	3.59	8.99	14.23%
BG	2,667	mg/L	217.10	115.23	70.00	751.19	53.07%
WV	2,667	mPa·s	1.47	0.06	1.29	1.73	4.27%

Trait: FS = filtering speed, EY = extract yield, WCO = wort color, BG = beta glucan, WV = wort viscosity; CV is phenotypic coefficient of variance.

650 **Figures**

651 **Figure 1** Proportion of total variance explained by metabolomic features and error in malting
652 quality traits
653 Trait: FS = filtering speed, EY = extract yield, WCO = wort color, BG = beta glucan, WV = wort viscosity; y-
654 axis is relative variance component; m is relative variance of metabolomic effects and e is relative variance of
655 residuals.

656
657 **Figure 2** Accuracy of prediction for malting quality traits using MBLUP and PLSR models
658 with different training population size
659 Trait: FS = filtering speed, EY = extract yield, WCO = wort color, BG = beta glucan, WV = wort viscosity; x-
660 axis is training population size, y-axis is accuracy of prediction which is the correlation between observed and
661 predicted phenotypes; MBLUP is metabolomic best linear unbiased prediction model, PLSR is partial least
662 squares regression model; PLSR at each point are the results from PLSR model with best number of components.
663

664
665 **Figure 3** Accuracy of prediction for malting quality traits using PLSR models with different
666 training population size
667 Trait: FS = filtering speed, EY = extract yield, WCO = wort color, BG = beta glucan, WV = wort viscosity; x-
668 axis is training population size, y-axis is accuracy of prediction which is the correlation between observed and
669 predicted phenotypes; PLSR is partial least squares regression model; PLSR_05 – PLSR_50 are partial least
670 squares regression models with different number of components (5, 10, 20, 50).
671

672
673 **Figure 4** Accuracy of prediction for malting quality traits across line using MBLUP and PLSR
674 models
675 Trait: FS = filtering speed, EY = extract yield, WCO = wort color, BG = beta glucan, WV = wort viscosity; y-
676 axis is accuracy of prediction which is the correlation between observed and predicted phenotypes; MBLUP is
677 metabolomic best linear unbiased prediction model; PLSR is partial least squares regression model with 20
678 components; plot is accuracy of plot, mean is accuracy of line mean.
679

680
681
682 **Figure 5** Accuracy of prediction for malting quality traits across location using MBLUP and
683 PLSR models
684 Trait: FS = filtering speed, EY = extract yield, WCO = wort color, BG = beta glucan, WV = wort viscosity; y-
685 axis is accuracy of prediction which is the correlation between observed and predicted phenotypes; MBLUP is
686 metabolomic best linear unbiased prediction model; PLSR is partial least squares regression model with 20
687 components; plot is accuracy of plot, mean is accuracy of line mean.
688

689 **Supplementary Material**
690
691 **Figure S1** Accuracy of prediction for malting quality traits using MBLUP models with
692 different sets of metabolomic features
693 Trait: FS = filtering speed, EY = extract yield, WCO = wort color, BG = beta glucan, WV = wort viscosity; x-
694 axis is training population size, y-axis is accuracy of prediction which is the correlation between observed and
695 predicted phenotypes; MBLUP is metabolomic best linear unbiased prediction model, M is MBLUP using all
696 metabolomic features, Mgs is MBLUP using metabolomic features having significant genetic correlation with
697 each trait and significantly heritable, Ms is MBLUP using metabolomic features significant heritable.

Figures

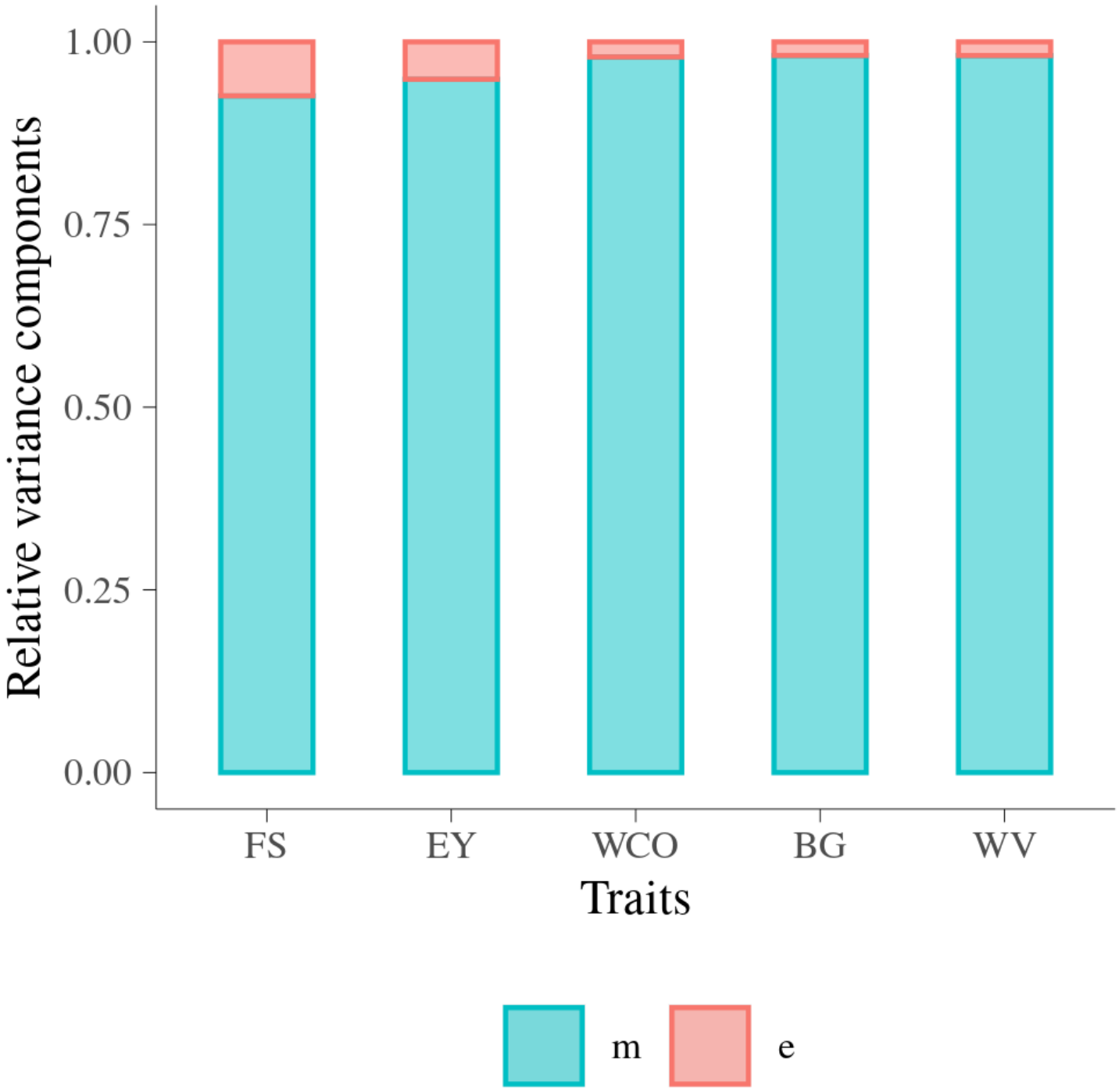

Figure 1

Proportion of total variance explained by metabolomic features and error in malting quality traits Trait: FS = filtering speed, EY = extract yield, WCO = wort color, BG = beta glucan, WV = wort viscosity; y-axis is relative variance component; m is relative variance of metabolomic effects and e is relative variance of residuals.

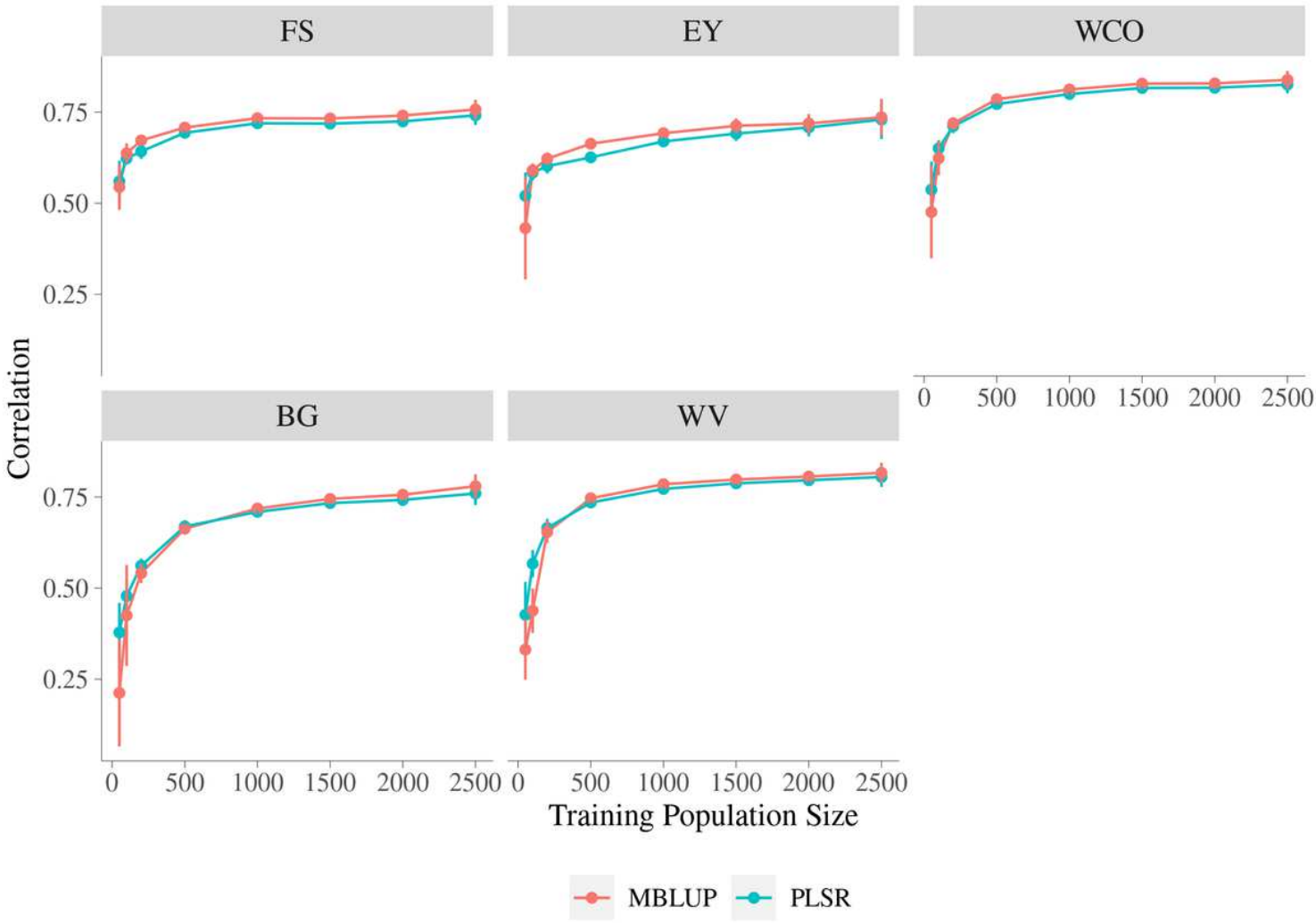

Figure 2

Accuracy of prediction for malting quality traits using MBLUP and PLSR models with different training population size Trait: FS = filtering speed, EY = extract yield, WCO = wort color, BG = beta glucan, WV = wort viscosity; x-axis is training population size, y-axis is accuracy of prediction which is the correlation between observed and predicted phenotypes; MBLUP is metabolomic best linear unbiased prediction model, PLSR is partial least squares regression model; PLSR at each point are the results from PLSR model with best number of components.

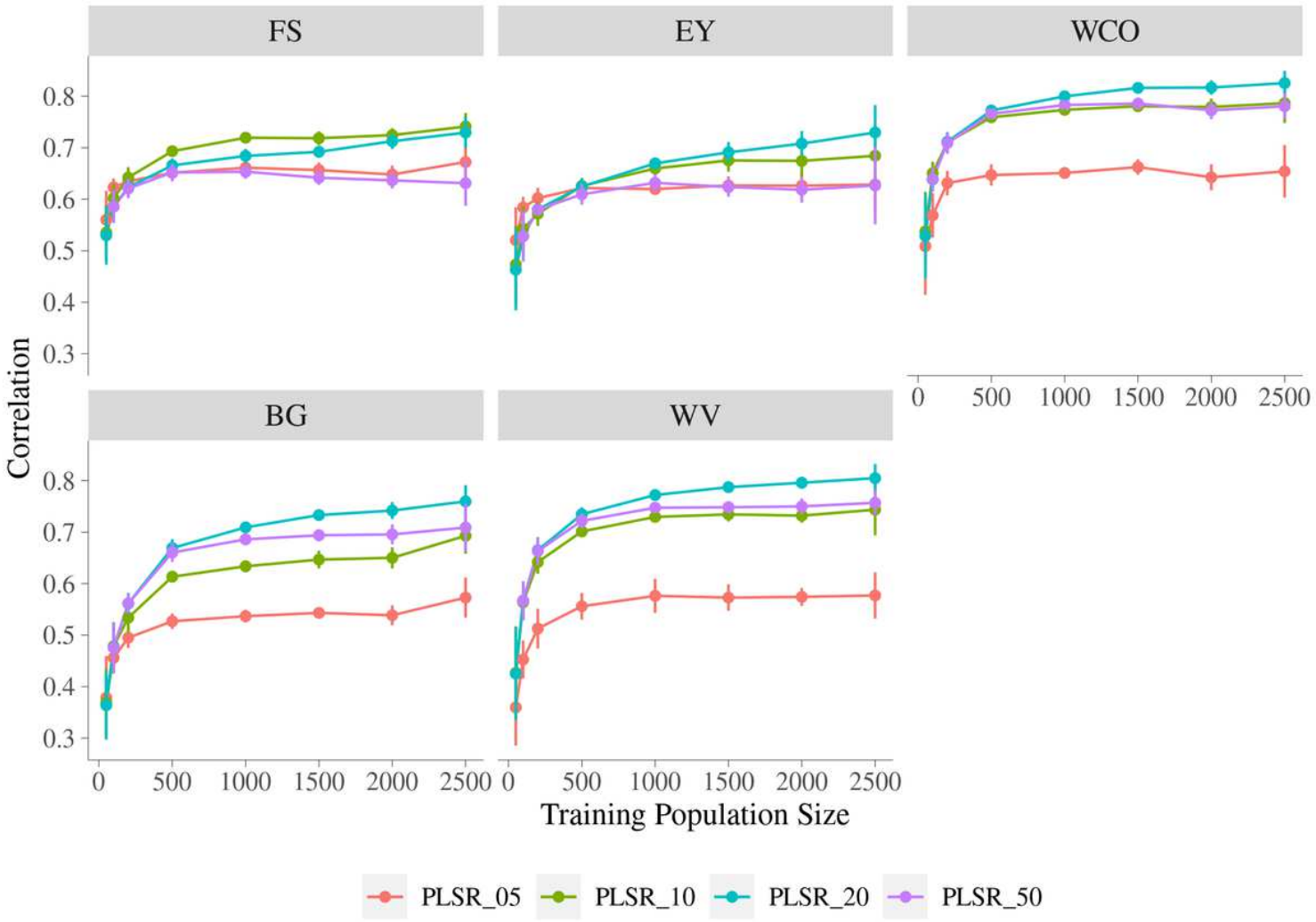

Figure 3

Accuracy of prediction for malting quality traits using PLSR models with different training population size Trait: FS = filtering speed, EY = extract yield, WCO = wort color, BG = beta glucan, WV = wort viscosity; x-axis is training population size, y-axis is accuracy of prediction which is the correlation between observed and predicted phenotypes; PLSR is partial least squares regression model; PLSR_05 – PLSR_50 are partial least squares regression models with different number of components (5, 10, 20, 50).

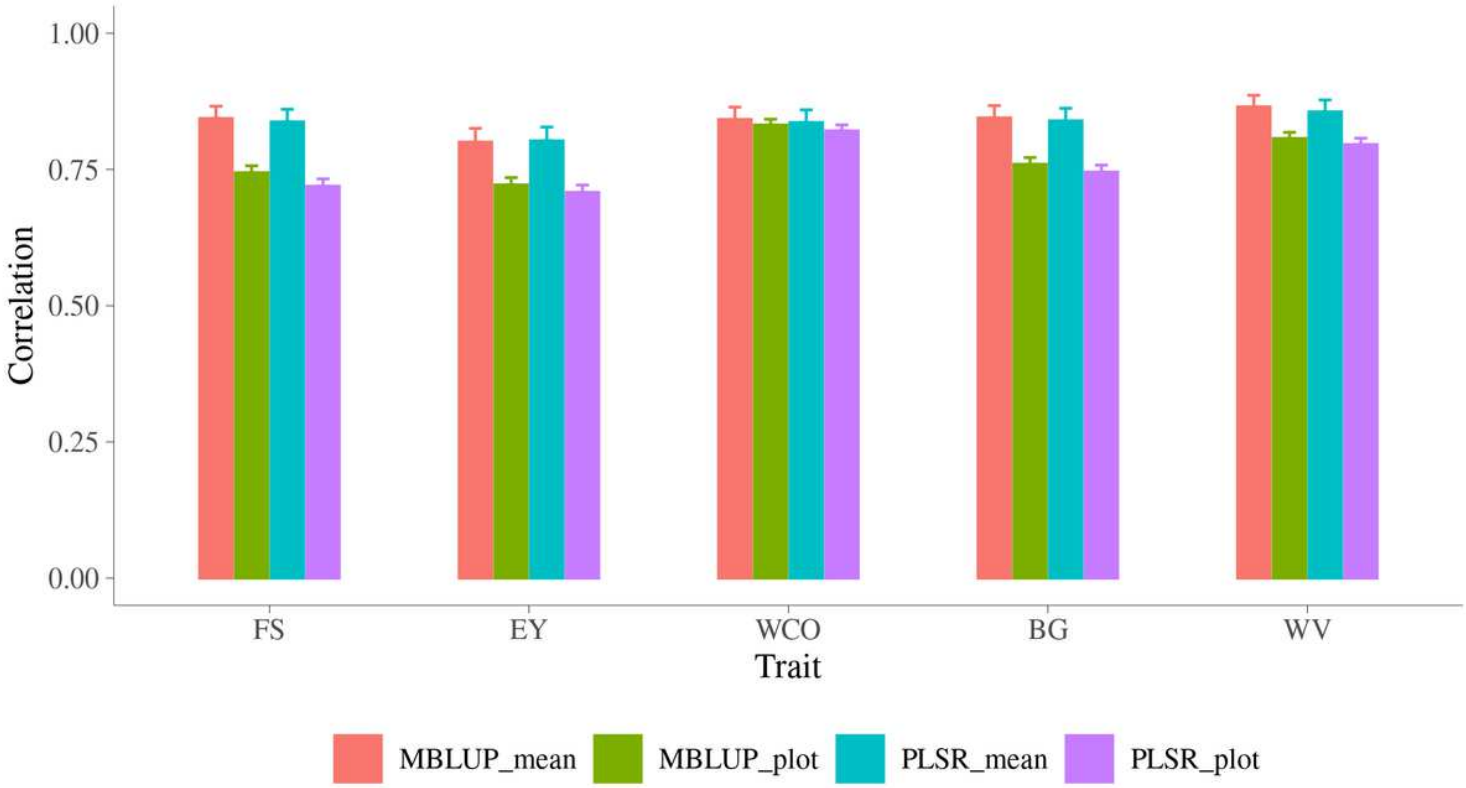

Figure 4

Accuracy of prediction for malting quality traits across line using MBLUP and PLSR models Trait: FS = filtering speed, EY = extract yield, WCO = wort color, BG = beta glucan, WV = wort viscosity; y-axis is accuracy of prediction which is the correlation between observed and predicted phenotypes; MBLUP is metabolomic best linear unbiased prediction model; PLSR is partial least squares regression model with 20 components; plot is accuracy of plot, mean is accuracy of line mean.

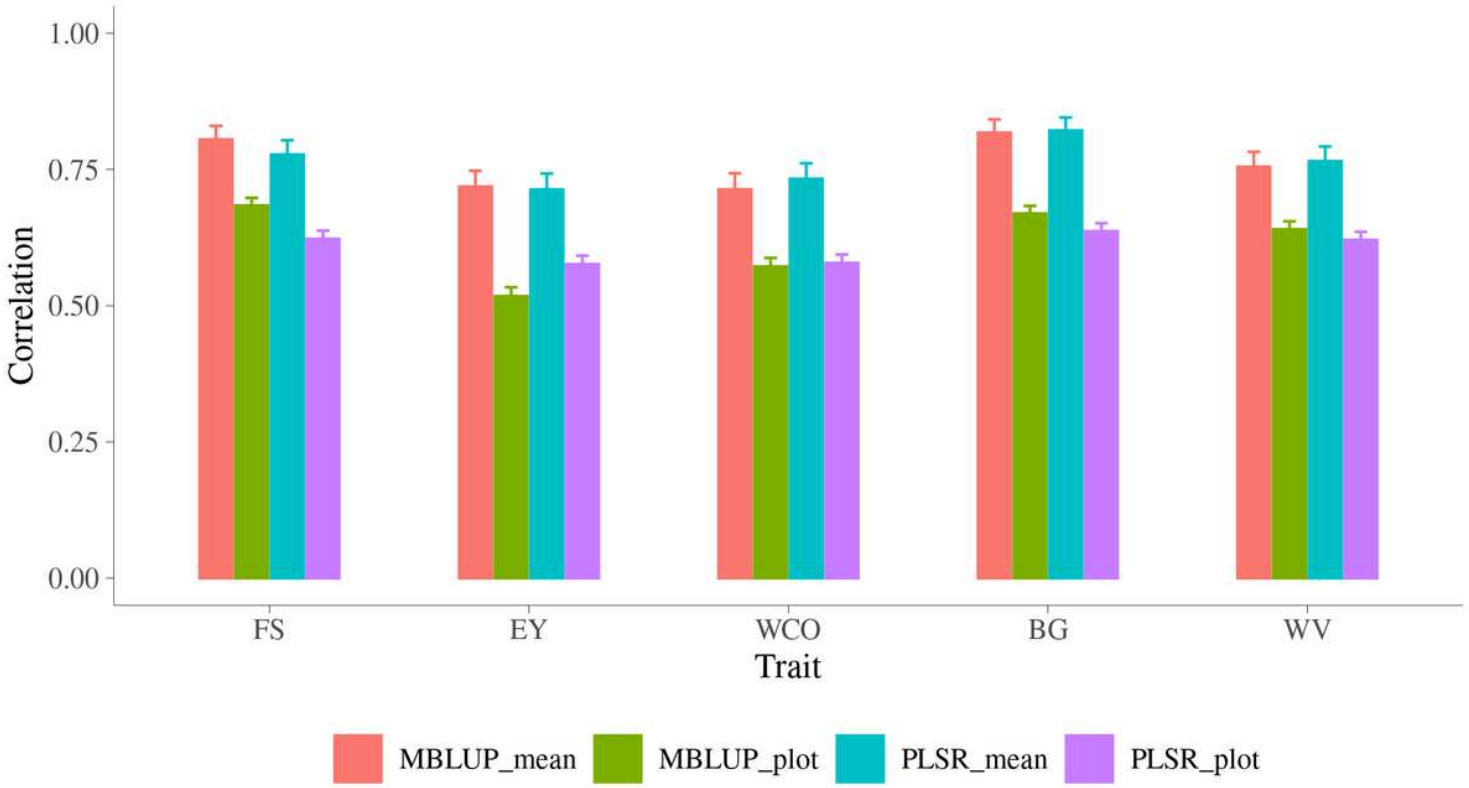

Figure 5

Accuracy of prediction for malting quality traits across location using MBLUP and PLSR models Trait: FS = filtering speed, EY = extract yield, WCO = wort color, BG = beta glucan, WV = wort viscosity; y-axis is accuracy of prediction which is the correlation between observed and predicted phenotypes; MBLUP is metabolomic best linear unbiased prediction model; PLSR is partial least squares regression model with 20 components; plot is accuracy of plot, mean is accuracy of line mean.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- phenometaxiangyusup.docx