# Economical optimization of a breeding scheme by selective phenotyping of the calibration set in a multi-trait context: application to bread making quality

S. Ben-Sadoun, Renaud Rincent, J. Auzanneau, François-Xavier Oury, B. Rolland, E. Heumez, Catherine Ravel, Gilles Charmet, Sophie Bouchet

## ▶ To cite this version:

HAL Id: hal-02553534

https://hal.inrae.fr/hal-02553534v1

Submitted on 14 Dec 2020

# Economical optimization of a breeding scheme by selective phenotyping of the calibration set in a multi-trait context: application to bread making quality

S Ben-Sadoun[1], R Rincent[1], J Auzanneau[2], FX Oury[1], B Rolland[3], E Heumez[4], C Ravel[1], G Charmet[1], S Bouchet[1]

[1] INRAE- UCA UMR1095, Genetics Diversity and Ecophysiology of Cereals, 5 chemin de Beaulieu 63000 Clermont-Ferrand - France

[2] Agri-Obtentions, Ferme de Gauvilliers 78660 Orsonville - France

[3] INRAE – Agrocampus Ouest Rennes- Université Rennes 1 UMR 1349 IGEPP BP35327, 35653 Le Rheu Cedex, France

[4] INRAE UE Lille, 2 chaussée Brunehaut, Estrées-Mons, BP 50136, 80203 Peronne Cedex, France

**Corresponding author:**

Sophie Bouchet

INRAE-UCA UMR1095, Genetics Diversity and Ecophysiology of Cereals, 5 chemin de Beaulieu 63000 Clermont-Ferrand – France

sophie.bouchet@inra.fr

ORCID: 0000-0001-5868-3359

# Abstract

***Key message*** **Trait-assisted genomic prediction approach is a way to improve genetic gain by cost unit, by reducing budget allocated to phenotyping or by increasing the program's size for the same budget.**

This study compares different strategies of genomic prediction to optimize resource allocation in breeding schemes by using information from cheaper correlated traits to predict a more expensive trait of interest. We used bread wheat baking score (BMS) calculated for French registration as a case study. To conduct this project, 398 lines from a public breeding program were genotyped and phenotyped for BMS and correlated traits in 11 locations in France between 2000 and 2016.

Single trait (ST), multi-trait (MT) and trait-assisted (TA) strategies were compared in terms of predictive ability and cost. In MT and TA strategies, information from dough strength (W), a cheaper trait correlated to BMS ($r = 0.45$), was evaluated in the training population or in both the training and the validation sets, respectively. TA models allowed to reduce the budget allocated to phenotyping by up to 65% while maintaining the predictive ability of BMS. TA models also improved the predictive ability of BMS compared to ST models for a fixed budget (maximum gain: +0.14 in cross-validation and +0.21 in forward prediction).

We also demonstrated that the budget can be further reduced by approximately one fourth while maintaining the same predictive ability by reducing the number of phenotypic records to estimate BMS adjusted means. In addition, we showed that the choice of the lines to be phenotyped can be optimized to minimize cost or maximize predictive ability. To do so, we extended the mean of the generalized coefficient of determination (CDmean) criterion to the multi-trait context (CDmulti).

**Key words**

Multi-trait selection. Trait-assisted selection. Selective phenotyping. Resource allocation optimization. Wheat baking quality. Coefficient of determination (CD)

# Introduction

Thanks to exponential genotyping cost decrease, computing and data storage capacities improvement, methodological and statistical developments, genomic selection is becoming an essential tool to optimize breeding efficiency by predicting the genetic value of candidates for complex traits using molecular marker information (Whittaker et al. 2000; Meuwissen et al. 2001). The use of genomic predictions at key steps of a breeding program can improve genetic gain per unit of time and cost, by skipping some field evaluation steps or optimizing the experimental design, i.e. minimizing the number of lines or replicates evaluated in one environment (Heslot et al. 2017).

To optimize phenotyping cost, we can also take advantage of correlated (secondary) traits that are easier or cheaper to measure, or can be obtained at an earlier stage. It is possible to predict correlated traits simultaneously using multivariate best linear unbiased prediction (BLUP; Henderson and Quaas 1976). Multi-trait genomic prediction (MT) models benefit from information contained in both genetic correlation between traits and genetic relationship among individuals (Calus and Veerkamp 2011). Two strategies are possible: 1) either the training population is genotyped and phenotyped for both traits and the candidate population is genotyped but not phenotyped for any of the traits, or 2) the candidates can be partly phenotyped for the trait of interest or the secondary trait. The latter strategy is called trait-assisted genomic selection (TA) (Fernandes et al. 2018). Benefits of MT models over single-trait genomic prediction (ST) models were reported using simulated and empirical data (Calus and Veerkamp 2011; Jia and Jannink 2012; Hayashi and Iwata 2013; Guo et al. 2014). As for single trait prediction, under a major QTL genetic architecture, Jia and Jannink (2012) found that Bayesian multivariate models (BayesA or BayesCπ) performed better than multi-trait GBLUP model. But for traits with polygenic genetic architecture, multi-trait GBLUP model was equal to the Bayesian multivariate models. MT models can however suffer from a high computational demand, time and some convergence problems (Michel et al. 2018). Obviously, genetic correlation between traits is a key factor determining the MT advantage over ST methods (Calus and Veerkamp 2011; Jia and Jannink 2012; Hayashi and Iwata 2013; Guo et al. 2014). Although MT models improve the predictive ability when the targeted trait has a low heritability and the secondary trait has higher heritability, the advantage of MT models to predict high heritability traits is low (Jia and Jannink 2012; Hayashi and Iwata 2013; Guo et al. 2014). Other factors also affect the prediction accuracy, including the size and the composition of the training set. Assuming that the number of markers is sufficient, the accuracy of the calibration model strongly depends on congruency between the allelic composition represented in the training population to estimate marker effects, and the allelic composition of the candidates whose performance is to be predicted (Habier et al. 2007). When the prediction uses unrelated populations to train the prediction equations, prediction accuracy actually becomes negligible (Crossa et al. 2014). Methods based on minimizing the mean of the prediction error variance (PEVmean; Rincent et al. 2012; Akdemir et al. 2015; Isidro et al. 2015; Sarinelli et al. 2019) or maximizing the mean of the generalized coefficient of determination (CDmean) of the estimated contrast between candidates and the mean of the panel were proposed to optimize the sampling of key individuals to be phenotyped in unstructured (Rincent et al. 2012) and in structured multi-familial populations (Rincent et al. 2017). In any case, accurate predictions will not be obtained until a large enough pool of individuals that is relevant to the candidates is genotyped and phenotyped with a sufficient precision.

3

In bread wheat breeding programs, baking tests or dough rheology measurements are part of official variety evaluation for registration in many countries. Improving baking quality is a key aspect of wheat improvement but criteria may be specific to countries and targeted markets. In France, the most important criterion for baking quality is a normalized (NF V03-716) bread-making score (BMS). It is used to classify registered varieties into end-use quality classes. Varieties presenting high BMS scores are reserved for French baking; they must yield at least 2% higher than reference varieties in registration trials. Varieties with intermediate BMS scores are mainly meant to be exported or used in mixed flours and must yield 4% higher than reference varieties. Varieties with lower BMS can only be used as feed or fuel and must yield 7% more than reference varieties. This BMS score is a complex quantitative trait that integrates several physical measurements of dough and loaf. Its assessment requires a large amount of grains and flour which are not available in the first generations of the breeding program, i.e. in row-nursery or small plots. In addition, BMS evaluation is time-consuming and quite expensive. Therefore, this trait is only evaluated at the very last steps of wheat breeding programs after the number of candidate lines was largely reduced. In French breeding programs, BMS is often estimated at earlier stages using less expensive traits such as the Chopin alveograph parameters which describe the dough rheological characteristics. Dough strength (W), in particular has a correlation of approximately 0.4 to BMS (Oury et al. 2010).

Furthermore, dough strength was shown to be itself influenced by high-molecular-weight glutenin subunits (HMW-GS; reviewed by Shewry 2009) that are cheaper to type and necessitate just a few grains. HMW-GS are encoded by the *Glu-1* loci, named *Glu-A1*, *Glu-B1* and *Glu-D1*, located on the long arm of the homoeologous chromosomes of group 1 (Payne 1987). Each locus comprises the two tightly linked genes *Glu-1-1* and *Glu-1-2*. To predict the degree of elasticity or tenacity of the dough of candidates to selection, HMW-GS used to be typed by sodium-dodecyl-sulphate polyacrylamide gel electrophoresis (SDS-PAGE) of grain proteins. Several attempts were made to derive molecular markers from glutenin DNA sequences (reviewed by Liu et al. 2012) because they are cheaper and easier to implement routinely. Ravel et al (2020) recently developed KASP markers that predict the rheological properties of the dough slightly better than SDS-PAGE.

Previous studies concerning multi-trait genomic predictions of baking quality related parameters such as dough rheological traits (Hayes et al. 2017; Michel et al. 2018; Lado et al. 2018) did not focus on BMS. In addition, none of these approaches tested the economic gain achievable with MT and TA models.

In this study, we demonstrate how to decrease the cost of breeding an expensive trait by using the information of a cheap correlated trait. To do so, we used genomic prediction of BMS as a case study. We evaluated different strategies in terms of predictive ability and cost. We compared ST models, univariate genomic prediction models with molecular markers associated with HMW-GSs as fixed effects (ST-glu), MT and TA models using dough strength (W) as a correlated trait. Only individuals of the training set were phenotyped for the trait of interest in both MT and TA models. In MT model, only the individuals of the training set were phenotyped for the correlated trait W. In TA model, the training set and a variable proportion of the validation set were phenotyped for the correlated trait. In addition, we adapted the CDmean criterion (Rincent et al. 2012) to a multi-trait context (CDmulti) in order to determine which individuals should be phenotyped for the secondary and the trait of interest to optimize the predictive ability of the trait of interest.

# Materials and Methods

## *Plant material*

We analyzed a breeding population of $F_8$-$F_9$ winter-type bread wheat lines developed by the Institut National de la Recherche en Agriculture, Alimentation et Environnement (INRAE) and its subsidiary company Agri-Obtentions. Each of the 1912 lines were evaluated between 2000 and 2016. Each year, the lines were phenotyped in 7 to 9 locations in France (Supplementary Table S1). In each environment, some registered varieties were used as controls, (6.8 control lines on average for trials in which BMS was phenotyped). Each control line has been evaluated for BMS for several consecutive years, up to 8 years for some of them. Those controls are essential to estimate adjusted means in further analyses. Crop management methods corresponded to high yield objectives (optimized pesticide, fungicide and nitrogen amount) to avoid confusing effects of disease or N-supply limitation.

## *Genotyping data*

In total, 814 lines were genotyped using a 280K SNP array (Rimbert et al. 2018). As predictive ability was the same using a subset of 35K markers that minimize Linkage Desequilibrum (LD; result not shown), the 57 breeding lines that were produced after 2015 were genotyped with a 35K SNP array included in the 280K array. Finally, 21210 markers with minor allele frequency superior to 0.05 and less than 10% of missing data were kept for analyses. Missing data were imputed using a kinship-based EM algorithm (Poland et al. 2012), based on the assumption that marker genotypes follow a multivariate normal (MVN) distribution. A genomic relationship matrix K was computed using Endelman and Jannink equation (2012):

$$K = \frac{WW^T}{2\sum(p_k - 1)p_k} \quad (1)$$

where W is a centered N $\times$ M marker matrix whose i-th row and k-th column is $w_{ik} = x_{ik} + 1 - 2p_k$ with $X_{ik}$ the genotype of the *i*-th individual for the *k*-th marker as $\{-1,0,1\}$ and $p_k$ the allele frequency at the *k*-th marker. These algorithms were implemented in the A.mat function of the rrBLUP package (Endelman 2011).

Additionally, a subset of 200 lines was genotyped at 11 DNA-based KASP markers derived from the HMW-GS loci *Glu-A1*, *Glu-B1*, and *Glu-D1* (Ravel et al. 2020). We calculated Pearson's correlation between each of the KASP marker and the 21210 SNP markers. We identified 18 SNP markers with correlation higher than 0.9 with at least one of the HMW-GS markers. In the model using those KASP markers as fixed effects, these 18 SNP markers were filtered out to calculate the kinship matrix.

The datasets with the genotyping data are available in the INRAE Dataverse repository (https://data.inra.fr/). They can be accessed with the following link https://doi.org/10.15454/AABGO7.

5

## *Phenotypic data*

In total, 871 lines were both genotyped and evaluated for grain yield in 9.9 environments on average. Although yield, height, heading date and protein were recorded for all the $F_8$ lines, baking and rheological traits were only available in half of the lines that were selected for $F_9$ evaluation. Protein content was estimated by Near-Infrared-Reflectance. This test was realized according to American Association of Cereal Chemists (AACC 39-10) method, using a wholemeal flour produced on a Cyclotec mill with a 0.8 mm sieve. Chopin alveograph (method AACC 54-30A) was carried out on flour obtained with a Chopin mill or a Brabender Senior mill (the extraction rate of these two mills is about 70%). The dough strength (W), the tenacity (P) and the extensibility (L) of the dough were evaluated with a Chopin Alveolink apparatus (Tripette & Renaud, 92396 Villeneuve-la-Garenne, France). The bread-baking tests were realized according to the AFNOR method NF V03-716 (also called BIPEA test). This baking test gives a measure of loaf volume, a score for dough (out of 100 units), a score for crust (out of 100 units) and a score for crumb (out of 100 units); the sum of these 3 scores giving the final BIPEA score (out of 300 units). This final BIPEA score corresponds to the Bread Making Score (BMS). In total, 398 lines were both genotyped using an array and phenotyped for BIPEA BMS.

The dataset with the phenotypic data is available in the INRAE Dataverse repository (https://data.inra.fr/). It can be accessed with the following link https://doi.org/10.15454/AABGO7.

# Statistical analysis of phenotypic data

Adjusted means of the lines were estimated using Best Linear Unbiased Estimations (BLUEs) in ASReml-R library (Butler et al. 2009) using the following model for each trait:

$$y_{ijk} = \mu + g_i + e_j + g_i \times e_j + \varepsilon_{ijk} \qquad (2)$$

where $y_{ijk}$ is the phenotypic value of the i-th genotype in the j-th environment (combination of year and site) and for the k-th repetition, $\mu$ is the overall mean, $g_i$ is the fixed effect of the i-th genotype, $e_j$ is the effect of j-th environment, $g_i \times e_j$ is the effect of the interaction between the i-th genotype and the j-th environment, and $\varepsilon_{ijk}$ is the residual error for the i-th genotype in the j-th environment and for the k-th repetition. All effects except the genotype effect were considered as random. Correlation between traits was calculated as the Pearson's correlation between BLUEs. Note however that traits linked to baking quality were measured only once in each environment. Therefore, the $g_i \times e_j$ interaction was not included in the model for these traits.

The estimation of adjusted means can affect the predictive ability of the genomic prediction models. Therefore, we evaluated the impact of the number of phenotypic records used to calculate BLUEs on the predictive ability. We especially investigated the importance of control lines that were phenotyped in all locations during several

successive years in the estimation of BLUEs and the predictive ability, by comparing two different cases of BLUE calculations:

1) In the first case, all phenotypic records for control lines and one to four phenotypic records (in four different trials) for the other lines were used to estimate BLUEs for BMS and W;
2) In the second case, one to four phenotypic records for each line (the control and the other lines) were used to estimate BLUEs for BMS and W.

For each scenario, the phenotypic records were randomly sampled 10 times.

The variance components of the traits were assessed with ASReml-R using the model (2) but in this case all the effects (including genotypic effects) were considered as random.

Repeatability was estimated at the plot and design levels as:

$$\text{plot repeatability} = \frac{\sigma_G^2}{\sigma_G^2 + \sigma_{G \times E}^2 + \sigma_\varepsilon^2} \qquad (3)$$

$$\text{design repeatability} = \frac{\sigma_G^2}{\sigma_G^2 + \sigma_{G \times E}^2/t + \sigma_\varepsilon^2/(t \times r)} \qquad (4)$$

Where $\sigma_G^2$; $\sigma_{G \times E}^2$ and $\sigma_\varepsilon^2$ are the genotype, genotype $\times$ environment and residual variances respectively, t and r are the average number of trials per genotype and replicates in each trial per genotype, respectively. The variance components and repeatability were estimated using lines which were both phenotyped and genotyped. For the traits linked to rheological quality, $\sigma_{G \times E}^2$ term was removed from equations (3) and (4).

## *Genomic prediction models*

### *Single-trait GBLUP models*

First, a single-trait GBLUP model was used to obtain genomic predictions for different traits:

$$y = X\beta + Za + \varepsilon \qquad (5)$$

Where y is a N $\times$ 1 vector of BLUEs obtained in the first step with N the number of lines; $\beta$ is the vector of fixed effects; X is the corresponding design matrix for the fixed effect; Z is the design matrix for the random effects; a is the N x 1 vector of lines random effects with the additive genetic variance $\sigma_a^2$ and a~N(0,K$\sigma_a^2$); and $\varepsilon$ is the residual error with the residual variance $\sigma_\varepsilon^2$ and $\varepsilon$~N(0,I$\sigma_\varepsilon^2$).

Three types of single-trait models were performed. The first one was a simple ST model in which the vector of fixed effects β modelled the grand mean. The second model, called ST-glu, included the 11 HMW-GS markers as fixed effect. The third model (ST-glu_stepwise) included as fixed effect the HMW-GS markers that explained a significant genetic variance. They were selected using a stepwise regression linear model implemented in the stepAIC function from the MASS package (Venables and Ripley 2002). In this model, no line random effect was included.

### *Multi-trait GBLUP models*

The same genomic relationship matrix K was used for fitting the multi-trait model:

$$y = X\beta + Za + \varepsilon \qquad (6)$$

Where y corresponds to an $N \times 2$ vector of BLUEs for two traits, BMS which is the trait of interest and W which is a correlated trait less expensive to measure, β is the vector of fixed effects (the grand mean in this study); X is the corresponding design matrix for the fixed effect; a is the vector of $N \times 2$ line effects with the corresponding random effect design matrix Z and $a \sim MVN$ $(0, \Sigma_a \otimes K)$ with the variance–covariance matrix $\Sigma_a$ of the form:

$$\begin{pmatrix} \sigma_{a1}^2 & \sigma_{a12} \\ \sigma_{a12} & \sigma_{a2}^2 \end{pmatrix} \qquad (7)$$

where $\sigma_{a1}^2$ and $\sigma_{a2}^2$ are the genetic variance of the first and second trait, respectively, and $\sigma_{a12}$ is the genetic covariance between both traits. The variance of the residual effect followed $\varepsilon \sim MVN$ $(0, \Sigma_\varepsilon \otimes I)$ where I is the identity matrix of dimension $N \times N$ and $\Sigma_\varepsilon$ has the same form than $\Sigma_a$. $\otimes$ indicate the Kronecker product operator between matrices.

We investigated two ways of predicting the value of the trait of interest using multivariate models: either we predicted it for un-phenotyped individuals (MT model), or we predicted it for a set of individuals that were fully or partly phenotyped for the correlated trait (TA model). In both MT and TA scenarios, measurements of the targeted trait were only available for the training population.

## *Model validation*

To compare the different models, predictions were evaluated using 5-fold cross-validations. This method consists in randomly splitting the data set in 5 folds of equal size and using 4 folds as the training set in order to predict performance of the lines from the fifth fold. Predictive ability is calculated for each fold as the Pearson's correlation between predicted values and BLUEs. The same procedure is repeated 20 times. Predictive abilities and their

standard deviations were calculated as the mean of the 100 Pearson's correlations. Models were tested using the lines which were both genotyped and phenotyped and which were not considered as control lines.

## Cost evaluation

Budget allocated to phenotyping was obtained using the following equation:

$$B = Cp_{BMS} \times N_{BMS} + Cp_W \times N_W \qquad (8)$$

Where $N_{BMS}$ and $N_W$ are the number of lines phenotyped for BMS and for W respectively; and $Cp_{BMS}$ (150€) and $Cp_w$ (20€) are the cost of phenotyping one line for BMS and for W respectively.

In the equation (8) $Cp_{BMS}$ and $Cp_w$ are numeric constants but $N_{BMS}$ and $N_W$ are variable depending on the experimental design. It means that for a fixed budget there is several ways to allocate it between lines that are phenotyped for BMS or W, in the training set and in the validation set.

Phenotyping cost for W corresponds to a minimum cost when we do it internally at INRA. BMS must be evaluated by certified companies and the price can vary between providers and depending on the volume we phenotype each year. In order to evaluate the impact of the cost ratio between W and BMS on the predictive ability of BMS, we made this ratio vary (ratio = 1/2, 1/3, 1/4, 1/5, 1/10 and 1/20). We performed this analysis with 3 different budgets, with $Cp_{BMS}$=150€, with a MT approach and a TA approach where all lines are phenotyped for W (Supplementary Table S2).

## Multi-trait CDmean (CDmulti) and optimization algorithm

The objective was to optimize the choice of individuals from the training set that should be phenotyped for the trait of interest BMS and for the correlated trait W to predict the individuals from the validation set. For that purpose, we extended the CDmean criterion (Rincent et al, 2012) to a multi-trait CDmean criterion (CDmulti). We first computed the prediction error variance (PEV), using the following equations (Lynch and Walsh 1998):

$$\begin{bmatrix} X^T(\Sigma_\varepsilon^{-1} \otimes I)X & X^T(\Sigma_\varepsilon^{-1} \otimes I)Z \\ Z^T(\Sigma_\varepsilon^{-1} \otimes I)X & Z^T(\Sigma_\varepsilon^{-1} \otimes I)Z + (\Sigma_a^{-1} \otimes K^{-1}) \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{a} \end{bmatrix} = \begin{bmatrix} X^T(\Sigma_\varepsilon^{-1} \otimes I)y \\ Z^T(\Sigma_\varepsilon^{-1} \otimes I)y \end{bmatrix} \qquad (9)$$

where y is the vector of phenotypes, β is a vector of fixed effects (in our case it is the intercept), a is a vector of random genetic values, X and Z are the corresponding design matrices for the fixed and random effects respectively, K is the kinship matrix, $\Sigma_a$ is the genetic variance–covariance matrix between traits, and $\Sigma_\varepsilon$ is the residual variance-covariance matrix between traits.

9

With the following notations: $\begin{bmatrix} X^T\left(\Sigma_\varepsilon^{-1} \otimes I\right)X & X^T\left(\Sigma_\varepsilon^{-1} \otimes I\right)Z \\ Z^T\left(\Sigma_\varepsilon^{-1} \otimes I\right)X & Z^T\left(\Sigma_\varepsilon^{-1} \otimes I\right)Z + \left(\Sigma_a^{-1} \otimes K^{-1}\right) \end{bmatrix}^{-1} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$

(10)

we have $PEV(â) = diag(C_{22})$  (11)

The multi-trait CD was defined using the following equation:

$$CD_{multi}(â) = 1 - \frac{PEV(â)}{diag(\Sigma_a \otimes K)} \qquad (12)$$

$CD_{multi}$ is the expected reliability (before phenotyping) for each trait and each individual. It takes values between 0 and 1. The closer to 1 the criterion is, the more reliable the predictions are expected to be. We optimized the sampling by maximizing the mean of the $CD_{multi}$ ($\overline{CD}_{multi}$) of individuals in the validation set for BMS, the trait of interest. For this a simple exchange algorithm was used. The initialization step consisted in randomly sampling calibration lines for BMS and/or W phenotyping and the corresponding $\overline{CD}_{multi}$ was calculated for the validation set for BMS. For each next step, one of the two traits (BMS or W) was randomly chosen. Then a random exchange between one individual in the calibration set for the chosen trait and one individual which was not in the calibration set was realized. If this new calibration set resulted in a higher $\overline{CD}_{multi}$, the new calibration set was accepted, otherwise the exchange was rejected. This procedure was repeated 2000 times to reach an optimum.

The genetic and residual variance-covariance matrices between traits ($\Sigma_a$ and $\Sigma_\varepsilon$) are estimated from phenotypic data. CDmulti is maximized before phenotyping. So those matrices have to be estimated using an independent dataset. To do so, we used 201 lines phenotyped for W and BMS by the French institute responsible for variety testing and registration (Groupement d'Etude et de contrôle des Variétés Et des Semences : GEVES) and genotyped using the same set of SNP markers.

We tested this optimization algorithm for several scenarios (Figure 1).

The optimization algorithm was performed with two different conditions:

1) in the first case, only individuals of the training set could be phenotyped for BMS and W (MT approach),

2) in the second case, only individuals from the training set could be phenotyped for BMS, while any of the individuals (from the training or the validation sets) could be phenotyped for the secondary trait W (TA approach).

In order to be able to compare GP models for different scenarios, lines from the validation set were never considered phenotyped for BMS, so that BMS predictive ability was always calculated on the same set of individuals.

We performed the optimization algorithm using different sample sizes that should be phenotyped for BMS and/or for W (Table 1).

We defined three budgets:

1) The most expensive budget allows to phenotype all the lines from the training set for BMS (47 700€),

2) The intermediate budget allows to phenotype all the lines from the training set for W and half of the lines for BMS (30 210€),

3) the less expensive budget allows to phenotype a quarter of the lines from the training set for BMS and three quarter of the lines from the training set for W (16 760€).

For a fixed budget, equation (8) has two unknowns ($N_{BMS}$, $N_W$), and there are many different possible combinations of individuals and traits to phenotype. We tested this optimization algorithm with 4 resource allocation strategies for each of the three budgets and under both MT and TA conditions. The resource allocation strategies correspond to a proportion (25, 50, 75 and 100%) of the training set (MT approach) or of the total population (TA approach) that is phenotyped for W. Then the $N_{BMS}$, the number of lines phenotyped for BMS, was calculated using the following equation obtained from equation (8):

$$Nb_{BMS} = \frac{B - (Cp_W \times Nb_W)}{Cp_{BMS}} \quad (13)$$

Training and validation sets were defined using 5-fold cross-validations. For each scenario (MT or TA approach, one budget and one resource allocation), we tested 20 folds (4 replicates x 5-fold cross-validations). In addition, for each fold and each scenario the $\overline{CD}_{multi}$ optimization algorithm was repeated 5 times, which means that for each scenario the procedure has been performed 100 times. The BMS BLUEs from the training set and W BLUEs from the training set (MT) and part of the validation set (TA) were used to predict BMS in the validation set. We compared predictive abilities obtained from optimized ($\overline{CD}_{multi}$) and random training sets (the mean predictability of 100 random sets for each scenario and each validation set).

Furthermore, the most expensive budget (47 700€) allowed to perform a ST approach in which all the individual from the training set were phenotyped for BMS. Therefore, we also compared predictive ability obtained for optimized training sets within each scenario with the predictive ability obtained with a ST model for BMS (calculated for each cross-validation that has been tested with the optimization algorithm).

In order to mimic real life situation, we also used a forward prediction strategy. In this case, the lines evaluated in 2003-2013 composed the training set and were used to predict the lines evaluated from 2014 to 2016.

The script of the optimization algorithm is available from the corresponding author on request. In addition, we extended the generalized CD of contrasts to the multi-trait context to adapt the optimization criterion to specific prediction objectives (see Appendix).

11

# Results

## *Trait variation and variance components*

Variance components and repeatabilities of yield, phenology and traits linked to bread making quality are presented in Table 2.

Plot repeatability ranged from 0.25 for crumb score to 0.81 for heading date. Design repeatability ranged from 0.66 for crumb score to 0.98 for heading date. Pearson's correlations between BLUEs of each trait are presented in Supplementary Figure S1. As expected, yield was negatively correlated with protein content (r = -0.63). The Chopin alveograph parameters were also correlated with each other. Indeed, Tenacity (P) was highly negatively correlated with Extensibility (L) (r = -0.6), while W and P were highly positively correlated (r = 0.63). Loaf volume, dough, crust and crumb scores are components of the BIPEA test which mechanically explains their correlation with BMS (the final BIPEA score).

As W is the trait that is the more correlated to BMS (r = 0.45) and its plot repeatability and design repeatability is higher (0.66 and 0.92 respectively) than for BMS (0.40 and 0.80 respectively), we used it as a secondary trait in MT and TA models in further analyses.

## *Contribution of glutenin (HMW-GS) markers to genomic prediction predictive ability*

The 11 KASP markers derived from HMW-GS loci *Glu-A1*, *Glu-B1*, and *Glu-D1* explained 31% and 8% of the phenotypic variation for W and BMS, respectively. The stepwise regression identified 3 markers explaining 30% of the W genetic variance and 4 markers explaining 8% of the BMS genetic variance, thus nearly as much as the 11 markers used together.

In order to compare ST, ST-glu and ST-glu_stepwise models, only the 200 lines that were phenotyped and genotyped with the 35K array and the KASP markers derived from HMW-GS loci were used for genomic predictions (Figure 2). For ST model with 200 lines, predictive ability was 0.49 for W and 0.28 for BMS. ST-glu and ST-glu_stepwise models had almost the same predictive ability for both traits (0.55 for W and 0.28 for BMS). They both improved the predictive ability of W but not that of BMS.

# Single-trait versus multi-trait and trait-assisted genomic prediction models

Three models (ST, MT and TA) were tested using the 398 lines which were both genotyped and phenotyped for W and BMS. For ST model, predictive ability was 0.57 for W and 0.38 for BMS (Figure 2). Although MT model did not improve the quality of BMS predictions, TA model improved it (+0.15). In addition, we predicted W using BMS as correlated trait. MT model did not improve the quality of W predictions, unlike TA model. Besides, the gain in predictive ability generated by TA model was lower for W (+0.10) than for BMS (+0.15).

We compared MT and TA approaches with 3 different budgets and 4 different resource allocation strategies (Figure 3). TA method performed better in terms of predictive ability than MT method in each scenario in cross-validation and in forward prediction. For each of the three budgets, the predictive ability for BMS was higher when all the lines (of the training set in MT case and of the total population in TA case) were phenotyped for the less expensive trait, W. The intermediate budget with at least 75 % of lines phenotyped for W was almost as accurate (0.5) than the expensive budget with 100% of lines phenotyped for W (0.51). This means that we can save 36% of the budget with this phenotyping strategy.

We used as a reference the predictive ability obtained with a ST model in which all the individuals from the training set were phenotyped for BMS (horizontal line in Figure 3). The cost of this ST approach is equal to the most expensive budget. For the intermediate budget, when all the lines from the training set were phenotyped for W, the predictive ability of BMS with MT model was close to the predictive ability obtained with ST model and the budget was reduced by 36%. In addition, for each budget, the predictive ability for BMS obtained using the TA model was higher than the predictive ability obtained with the ST model when at least three quarters of the lines were phenotyped for W. This means that TA approach improved the predictive ability compared to ST and MT approaches (maximum gain: +0.14 in cross-validation and +0.21 in forward prediction when all the lines were phenotyped for W) and allowed to reduce the cost by up to 65%.

In addition, the ST model to predict BMS for the lines evaluated between 2014 and 2016 using older lines (evaluated from 2003 to 2013) as training set, showed a lower predictive ability than in cross-validation scenarios (-0.2). MT approach slightly improved the predictive ability compared with ST model, but it remained under 0.22. TA approach greatly improved this predictive ability (maximum predictive ability with random sampling: 0.41) whichever the budget. In addition, the predictive ability reached with TA approach in forward prediction was closed to the predictive ability obtained with ST approach in cross-validation scenario.

# Selective phenotyping using Multi-trait CDmean criterion

To test whether we could minimize phenotyping or improve predictive ability by optimizing the choice of individuals to phenotype for BMS and/or W, we used a multi-trait CDmean criterion ($\overline{CD}_{multi}$) in an optimization algorithm. The multi-trait CD criterion allowed to improve slightly but systematically the predictive ability (+0.013

on average) for BMS compared to random sampling. The gain in terms of predictive ability allowed by the multi-trait CD criterion was slightly higher for TA approach in forward prediction strategy (+0.023 on average).

On average, the gain in predictive ability generated by TA model compared to MT model was +0.13 in cross-validation scenarios and +0.15 in forward prediction strategy. Note that for TA method using the multi-trait CDmean criterion, the number of lines phenotyped for W in the validation set was higher compared to random sampling (Supplementary Figure S2).

## *Impact of the cost ratio between W and BMS on BMS predictive ability*

We evaluated the impact of the cost ratio between W and BMS on BMS predictive ability for MT and TA approaches. To do so, we compared predictive ability of BMS in several scenarios (Supplementary Table S2) by varying the cost ratio (ratio = 1/2, 1/3, 1/4, 1/5, 1/10 and 1/20). We fixed the budget allocated to phenotyping (the three budgets described above), the cost of phenotyping one line for BMS cost for one line ($Cp_{BMS}=150€$) and the percentage of the lines phenotyped for W (100% of the training set in MT approach and 100% of the total population in TA approach).

When the cost difference between the two traits was small and the budget allocated to phenotyping was low, the budget was not sufficient to phenotype each line for W (Supplementary Table S2). Therefore, only the results obtained with the intermediate and the expensive budget were shown in the Figure 4.

When the cost difference between the two traits increased, the number of lines phenotyped for BMS increased as well, and consequently the predictive ability of BMS was improved. Indeed for the expensive budget (which corresponds to the budget needed to phenotype each line of the training set for BMS), the predictive ability of BMS ranged from 0.35 when the ratio was 1/2 to 0.38 when the ratio was 1/20 for MT approach and ranged from 0.48 when the ratio was 1/2 to 0.52 when the ratio was 1/20 for TA approach. For the intermediate budget, the same trends were found. We could not evaluate the predictive ability of BMS with the TA model when the ratio between the costs of phenotyping of each trait was 1/2 because $N_{BMS}$ was too small and it led to convergence model issues. Even if the interest of using a TA approach rather than a ST approach was slightly lower when the cost difference decreased, TA model still performed better than ST approach when the ratio was 1/2 and for a fixed budget (gain in predictive ability: +0.09).

## *Impact of the number of phenotypic records on predictive ability*

We evaluated the impact of the number of phenotypic records to estimate BMS and W BLUEs on BMS predictive ability for ST, MT and TA models (Figure 5). The number of phenotypic records for each line (except the control lines) ranged from 1 to 4. The three models were tested using the 322 lines which were both genotyped and phenotyped for W and for BMS in at least 4 environments.

We also evaluated the impact of control lines (evaluated 7 environments on average for BMS) to estimate BLUEs on BMS predictive ability. To do so, we compared two cases regarding controls: 1) all the phenotypic records of control lines were used to estimate BLUEs in order to have a more connected design; 2) we sampled the same number of control lines as candidate lines to calculate BLUEs.

For the three models, the predictive ability increased with the number of phenotypic records when we used all controls. The highest gain in terms of predictive ability was when the traits were evaluated in two environments compared to one (+ 0.11, + 0.08 + 0.07 for ST, MT, TA respectively). In addition, the predictive ability for using three records was slightly higher compared to two (+ 0.05 on average) and almost similar using four records (+ 0.02 on average) whichever the model used. For each scenario, the predictive ability was higher when all the phenotypic records of control lines were used to calculate BLUEs (+0.04, +0.04 +0.03 for ST, MT, TA respectively for 2 records). Note that the budget allocated to phenotyping can be significantly lower (-50 or 25% for 2 or 3 records). In addition, when both traits were measured in more than one environment, MT did not allow to improve predictive ability of BMS compared to ST model. But TA model always performed better than MT and ST models whichever the number of phenotypic records.

# Discussion

This study focused on the comparison of different strategies to optimize resource allocation in breeding programs by using information from a cheap correlated trait to predict an expensive trait of interest. We used winter bread wheat baking quality (BMS) as a case study. Single-trait (ST), multi-trait (MT) and trait-associated (TA) strategies were compared in terms of predictive ability and cost. In multivariate approaches, information from dough strength (W) which is correlated ($r = 0.45$) to BMS and is less expensive to phenotype, was used. In MT analyses, only lines of the training set were phenotyped for W, whereas lines of the training and the validation sets were evaluated for W in TA analyses. In addition, we proposed and tested a multi-trait CDmean criterion (CDmulti) to optimize the choice of the set of lines to be phenotyped in the multi-trait context. To conduct this study, a public French breeding population of winter bread wheat lines was analyzed.

## *Ability of glutenin (HMW-GS) markers to predict BMS*

In order to complement scarce phenotyping or pedigree information, Lande and Thompson (1990) proposed to rank individuals using molecular scores calculated as the sum of the effects of their alleles at selected QTLs. In this strategy, a threshold has to be chosen to determine the list of loci that significantly contribute to the trait variation. As it has been shown that allele effects were often overestimated in QTL detection (Beavis 1998), and that the infinitesimal model (Fisher 1918) that considers many loci of small effects was the best model to explain the variation of many quantitative traits such as yield, Meuwissen et al. (2001) proposed instead to use all available independent markers. To do so, phenotypic values are regressed in a linear model on all markers simultaneously, considering them as random effects with the same small genetic variance. The genetic values of individuals are estimated by best linear unbiased prediction. The critical difference with the Lande and Thompson approach is that this method does not require setting a significance threshold for the loci selected for trait prediction since all the markers are used in the model. In some cases however, explicitly modeling major QTLs as fixed effects in genomic prediction models can improve prediction accuracy in simulated data (Bernardo 2014) and experimental bread wheat data (Arruda et al. 2016; Michel et al. 2018; Sarinelli et al. 2019). In bread wheat, HMW-GS are important determinants of rheological properties of dough (reviewed by Anjum et al. 2007). Ravel et al. (2020) developed a set of KASP markers derived from glutenin DNA sequences to identify HMW-GS alleles. They showed that these markers detected more alleles and explained more rheological traits variation than SDS-PAGE method, which was commonly used in plant breeding. In our study, we integrated 11 of these HMW-GS markers as fixed effects into ST models. It improved the predictive ability of W compared to ST model. These results are consistent with the conclusion of Michel et al. (2018) who found that prediction accuracy was increased for several dough rheological traits when *Glu-1* loci markers were including as fixed effects into the RR-BLUP model. But this model did not improve the predictive ability of BMS. This result might be due to the low percentage of genetic variance of BMS explained by the HMW-GS markers (8%). Bernardo (2014) actually demonstrated that only QTLs that explained at least 10% of the genetic variance could improve prediction accuracy.

## Prediction of unphenotyped individuals (MT)

Our results were consistent with some previous studies on experimental data in which MT models did not improve the predictive ability for individuals that were neither phenotyped for the trait of interest nor for the secondary trait, as in pine tree (Jia and Jannink 2012), soybean (Bao et al. 2015), rye (Schulthess et al. 2016), maize (Dos Santos et al. 2016), wheat (Schulthess et al. 2018; Michel et al. 2018; Lado et al. 2018) and sorghum (Fernandes et al. 2018).

Previous studies showed that MT could be beneficial in terms of accuracy compared to ST model when a part of the individuals of the training set were not phenotyped for the targeted trait but the correlated trait was evaluated for a larger number of individuals (Guo et al. 2014; Michel et al. 2018). We found a similar trend in our study. In fact, when only half of the lines in the training set were phenotyped for BMS but all the lines (training and validation sets)were phenotyped for W, the predictive ability was similar compared to a ST model in which all the lines from the training set were phenotyped for BMS, but the budget was reduced by one third.

In the INRA-Agri-Obtentions bread wheat breeding program, $F_8$ and $F_9$ lines are evaluated in 9 environments in France on average. BMS is evaluated in 7 environments on average for control lines and in 4 environments for most of the other lines. Those measurements are very expensive (150€) and in this study, we tested if we could decrease the number of records (or the number of environments) without affecting the predictive ability of BMS. As expected, predictive ability increased with the number of phenotypic records. The higher gap of predictive ability was observed when 2 phenotypic records instead of 1 were used to predict BMS with ST model (+0.08). However, with 3 phenotypic records instead of 4, the predictive ability was little affected (+0.02), but the budget allocated to phenotyping was reduced by 25%. The predictive ability was higher (+0.03) when all the phenotypic records of control lines were used to calculate BLUEs. Controls are essential in the breeding program in order to take into account G×E and it is important to evaluate them in a large number of environments. By constrast, the number of records for other lines could be decreased with little effect on the predictive ability. Although phenotypic records were selected randomly for each line, we could further optimize the experimental design (Heslot et al, 2017) and control lines distribution in particular.

## Prediction when (part of) the candidates are phenotyped for the correlated trait (TA)

TA method consists of predicting a trait using a multi-trait model when at least some of the candidates are phenotyped for one or several correlated traits. Several studies using experimental data demonstrated that TA models perform better than ST and MT models in terms of accuracy. TA models using high-throughput phenotyping for instance improved the prediction accuracy of bread wheat grain yield by up to 70% (Rutkoski et al. 2016; Sun et al. 2017; Crain et al. 2018). TA models also improved bread wheat baking quality related parameters using protein content (Michel et al. 2018) or dough rheological traits (Lado et al. 2018) as correlated traits. Predictive ability of *Fusarium* head blight severity in hybrid bread wheat was improved using plant height

and heading date as correlated traits (Schulthess et al. 2018). Wheat is not the only specie for which TA approach has been tested, Fernandes et al. (2018) showed that TA models increased prediction accuracy by up to 50% when using plant height as correlated trait to predict yield in sorghum.

Our results were consistent with these studies, since predictive ability using TA model was higher than ST and MT predictive abilities whichever the studied scenario.

For each studied budget (low, intermediate, high), the predictive ability of BMS was higher when all lines (training and validation) were phenotyped for the less expensive trait, W. It seems more important to phenotype a maximum of lines with a highly heritable trait than just a few lines for an expensive and less heritable trait. For similar budget, TA allowed higher predictive ability than ST (maximum gain: +0.14). For similar predictive ability (approximately 0.38), TA allowed reducing the budget by up to 65%. So TA is a way to improve genetic gain by cost unit, either by reducing phenotyping cost and budget or increasing the program's size for the same budget.

We showed that the gain in predictive ability increased with the difference of phenotyping cost between the expensive and the cheap trait. These results highlighted the advantage of using a correlated trait as cheap as possible in multi-variate genomic prediction models. Nevertheless, TA model was still interesting compared to ST model with the smallest cost difference (ratio = 1/2) tested in this study (gain in predictive ability: +0.09). In addition, even if it has no practical interest, we tried to predict W (plot repeatability = 0.66) using BMS (plot repeatability = 0.4) information and TA model, to compare the effect of the secondary trait's heritability (approximated by plot repeatability) on predictive ability. The gain was inferior to those when we predicted BMS using W. This is congruent with the hypothesis that the higher the heritability of the secondary trait compared to the trait of interest, the higher the gain in predictability (Jia and Jannink 2012, Hayashi and Iwata 2013, Guo et al, 2014).

## *Forward prediction*

In order to study a situation more similar to the one the breeders are dealing with, we investigated a forward prediction strategy. ST predictive ability was almost divided by two when the oldest lines were used to predict the most recent lines compared to a cross-validation scenario. This discrepancy in predictive ability cannot be attributed to differences in the size of training sets. Indeed, the size of training set in cross-validation scenario was 318 lines vs 309 in the case of forward prediction. Storlie and Charmet (2013) also reported that forward prediction in wheat is less accurate than cross-validation. Since the youngest lines are not differently (or less) genetically related to the oldest lines than the oldest lines with each other (results not shown), the most likely explanation is that new candidates are evaluated in new environments (years and locations) that have never been observed. Although BLUEs were used to correct the environmental main effect, they did not correctly (efficiently) account for G×E interactions. Using TA by phenotyping some candidates with a correlated trait, we could take into account G×E in the new environment and better predict the candidates. This result suggests that TA model are particularly interesting to predict lines evaluated in new environments.

Moreover, a genomic Bayesian multi-trait and multi-environment model was developed by Montesinos-Lopez et al (2016) to take into account trait x genotype x environment interaction that could be interesting to test in that context.

## *Optimization of the training set based on a multi-trait CD criterion (CDmulti)*

The size and the composition of the training population are key parameters of the accuracy of genomic prediction models. Optimizing the training set by selecting the most predictive individuals instead of using a random sampling leads to an increase of the predictive ability (Rincent et al. 2012; Akdemir et al. 2015; Isidro el al. 2015; Rincent et al. 2017; Sarinelli et al. 2019). In this study, we specifically developed a multi-trait CDmean criterion (CDmulti), which slightly but systematically improved the predictive ability for BMS. The gain of predictive ability was a bit higher for TA approach than MT approach, in particular with the forward prediction. This may be due to the fact that $\overline{CD}_{multi}$ leads to sample more lines to phenotype for W in the validation set because lines are more related to the candidates compared to the training set (Supplementary Figure S2). In this example the use of $\overline{CD}_{multi}$ resulted in small increase of predictive ability, but we can suppose that it could be more valuable in datasets with different genetic structure, or in the situation of lower budget in which the choice of the phenotyped lines would be essential. Note that the criterion CDmulti can be adapted to particular prediction objectives (eg: optimize predictive abilities within families as in Rincent et al. 2017) by using the corresponding contrasts (see Appendix for the generalized version of CDmulti).

## *Conclusion*

In this study, we addressed the question of resource allocation, which is hardly ever mentioned in multi-trait studies. We demonstrated how to decrease the breeding cost of an expensive trait by using the information of a cheap correlated trait using trait-assisted (TA) genomic predictions. TA models are useful tools for breeders to optimize resource allocation while maintaining the predictive ability for the traits of interest. In our study, it reduced the budget allocated to phenotyping by up to 65% for Bread Making Score (BMS). TA models can also be used to improve the predictive ability of the trait of interest (+0.14 using cross-validation and +0.21 using forward predictions) for a fixed budget. The less expensive and labor-intensive, and the more heritable the correlated trait is compared to the trait of interest, the more TA could be useful in a breeding program.

In addition, we highlighted the importance of control lines that are phenotyped in a large number of environments and which allow to decrease the number of phenotypic records (3 instead of 4) for the other lines with a small effect on BLUEs estimation.

The multi-trait CD criterion (CDmulti) specifically developed in this study allowed a slight but systematic improvement of BMS predictive ability. It would be interesting to evaluate this criterion in a larger and more

diverse population. The tools and approaches developed in this study and exemplified using BMS prediction of bread wheat can be used in any species when relevant cheap correlated traits are available.

# Author contribution statement

JA, FXO, BR and EH designed the field trials and collected the phenotypic data. RR supported in statistical analysis and in developing the multi-trait CDmean algorithm. CR developed KASP markers derived from HMW-GS loci. SBS analyzed the data and wrote the manuscript. SB and GG guided through the study and helped improving the manuscript. All authors approved the final manuscript.

# Acknowledgements

# Compliance with ethical standards

## *Conflict of interest*

The authors declare no conflict of interest

# References

Akdemir D, Sanchez JI, Jannink J-L (2015) Optimization of genomic selection training populations with a genetic algorithm. Genet Sel Evol 47:38. https://doi.org/10.1186/s12711-015-0116-6

Anjum FM, Khan MR, Din A, et al (2007) Wheat Gluten: High Molecular Weight Glutenin Subunits—Structure, Genetics, and Relation to Dough Elasticity. J Food Sci 72:R56–R63. https://doi.org/10.1111/j.1750-3841.2007.00292.x

Arruda MP, Lipka AE, Brown PJ, et al (2016) Comparing genomic selection and marker-assisted selection for Fusarium head blight resistance in wheat (*Triticum aestivum* L.). Mol Breeding 36:84. https://doi.org/10.1007/s11032-016-0508-5

Bao Y, Kurle JE, Anderson G, et al (2015) Association mapping and genomic prediction for resistance to sudden death syndrome in early maturing soybean germplasm. Mol Breeding, 35(6), 128.

Beavis WD (1998) QTL analyses: power, precision, and accuracy, pp. 145–162 inMolecular Dissection of Complex Traits, edited by Paterson A. H. CRC Press, New York.

Bernardo R (2014) Genomewide Selection when Major Genes Are Known. Crop Sci 54:68–75. https://doi.org/10.2135/cropsci2013.05.0315

Butler DG, Cullis BR, Gilmour AR and Gogel BJ (2009). ASReml-R reference manual. The State of Queensland, Department of Primary Industries and Fisheries, Brisbane.

Calus MP, Veerkamp RF (2011) Accuracy of multi-trait genomic selection using different methods. Genet Sel Evol 43:26. https://doi.org/10.1186/1297-9686-43-26

Crain J, Mondal S, Rutkoski J, et al (2018) Combining High-Throughput Phenotyping and Genomic Information to Increase Prediction and Selection Accuracy in Wheat Breeding. Plant Genome 11. https://doi.org/10.3835/plantgenome2017.05.0043

Crossa J, Pérez P, Hickey J, et al (2014) Genomic prediction in CIMMYT maize and wheat breeding programs. Heredity 112:48–60. https://doi.org/10.1038/hdy.2013.16

Endelman JB (2011) Ridge regression and other kernels for genomic selection with R package rrBLUP. Plant Genome 4:250-255. doi: 10.3835/plantgenome2011.08.0024

Endelman JB and Jannink JL (2012) Shrinkage estimation of the realized relationship matrix. G3 2:1405-1413. doi: 10.1534/g3.112.004259

Fernandes SB, Dias KOG, Ferreira DF, Brown PJ (2018) Efficiency of multi-trait, indirect, and trait-assisted genomic selection for improvement of biomass sorghum. Theor Appl Genet 131:747–755. https://doi.org/10.1007/s00122-017-3033-y

Fisher RA (1919) XV.—The Correlation between Relatives on the Supposition of Mendelian Inheritance. Earth Environ Sci Trans R Soc Edinb 52:399–433. https://doi.org/10.1017/S0080456800012163

Guo G, Zhao F, Wang Y, et al (2014) Comparison of single-trait and multiple-trait genomic prediction models. BMC Genet 15:30. https://doi.org/10.1186/1471-2156-15-30

Habier D, Fernando RL, Dekkers JCM (2007) The impact of Genetic Relationship Information on Genome-Assisted Breeding Values. Genetics 177:2389–2397. https://doi.org/10.1534/genetics.107.081190

Hayashi T, Iwata H (2013) A Bayesian method and its variational approximation for prediction of genomic breeding values in multiple traits. BMC Bioinformatics 14:34. https://doi.org/10.1186/1471-2105-14-34

Hayes BJ, Panozzo J, Walker CK, et al (2017) Accelerating wheat breeding for end-use quality with multi-trait genomic predictions incorporating near infrared and nuclear magnetic resonance-derived phenotypes. Theor Appl Genet 130:2505–2519. https://doi.org/10.1007/s00122-017-2972-7

Heffner EL, Sorrells ME, Jannink J-L (2009) Genomic Selection for Crop Improvement. Crop Sci 49:1–12. https://doi.org/10.2135/cropsci2008.08.0512

Henderson CR, Quaas RL (1976) Multiple Trait Evaluation Using Relatives' Records. J Anim Sci 43:1188–1197. https://doi.org/10.2527/jas1976.4361188x

Heslot N, Jannink J-L, Sorrells ME (2015) Perspectives for Genomic Selection Applications and Research in Plants. Crop Sci 55:1–12. https://doi.org/10.2135/cropsci2014.03.0249

Heslot, N and Feoktistov V (2017) Optimization of selective phenotyping and population design for genomic prediction. BioRxiv, 172064.

Isidro J, Jannink J-L, Akdemir D, et al (2015) Training set optimization under population structure in genomic selection. Theor Appl Genet 128:145–158. https://doi.org/10.1007/s00122-014-2418-4

Jia Y, Jannink J-L (2012) Multiple-Trait Genomic Selection Methods Increase Genetic Value Prediction Accuracy. Genetics 192:1513–1522. https://doi.org/10.1534/genetics.112.144246

Lado B, Vázquez D, Quincke M, et al (2018) Resource allocation optimization with multi-trait genomic prediction for bread wheat (*Triticum aestivum* L.) baking quality. Theor Appl Genet 131:2719–2731. https://doi.org/10.1007/s00122-018-3186-3

Laloë D (1993) Precision and information in linear models of genetic evaluation. Genet Sel Evol 25(6), 557.

Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124:743–756

Liu Y, He Z, Appels R, Xia X (2012) Functional markers in wheat: current status and future prospects. Theor Appl Genet 125:1–10. https://doi.org/10.1007/s00122-012-1829-3

23

Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits. Sunderland, MA: Sinauer.

Meuwissen TH, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829

Michel S, Kummer C, Gallee M, et al (2018) Improving the baking quality of bread wheat by genomic selection in early generations. Theor Appl Genet 131:477–493. https://doi.org/10.1007/s00122-017-2998-x

Montesinos-López OA, Montesinos-López A, Crossa J, et al (2016) A genomic Bayesian multi-trait and multi-environment model. G3 6(9), 2725-2744.

Oury F-X, Chiron H, Faye A, et al (2009) The prediction of bread wheat quality: joint use of the phenotypic information brought by technological tests and the genetic information brought by HMW and LMW glutenin subunits. Euphytica 171:87. https://doi.org/10.1007/s10681-009-9997-1

Payne PI (1987) Genetics of Wheat Storage Proteins and the Effect of Allelic Variation on Bread-Making Quality. Ann Rev Plant Physiol 38:141–153. https://doi.org/10.1146/annurev.pp.38.060187.001041

Poland J, Endelman J et al. (2012) Genomic selection in wheat breeding using genotyping-by-sequencing. Plant Genome 5:103-113. doi: 10.3835/plantgenome2012.06.0006

Ravel C, Faye A, Ben-Sadoun S, et al (2020) SNP markers for early identification of high molecular weight glutenin subunits (HMW-GSs) in bread wheat. Theor Appl Genet. https://doi.org/10.1007/s00122-019-03505-y

Rimbert H, Darrier B, Navarro J, et al (2018) High throughput SNP discovery and genotyping in hexaploid wheat. PLoS ONE 13:e0186329. https://doi.org/10.1371/journal.pone.0186329

Rincent R, Charcosset A, Moreau L (2017) Predicting genomic selection efficiency to optimize calibration set and to assess prediction accuracy in highly structured populations. Theor Appl Genet 130:2231–2247. https://doi.org/10.1007/s00122-017-2956-7

Rincent R, Laloë D, Nicolas S, et al (2012) Maximizing the Reliability of Genomic Selection by Optimizing the Calibration Set of Reference Individuals: Comparison of Methods in Two Diverse Groups of Maize Inbreds (*Zea mays* L.). Genetics 192:715–728. https://doi.org/10.1534/genetics.112.141473

Rutkoski J, Poland J, Mondal S, et al (2016) Canopy Temperature and Vegetation Indices from High-Throughput Phenotyping Improve Accuracy of Pedigree and Genomic Selection for Grain Yield in Wheat. G3 6:2799–2808. https://doi.org/10.1534/g3.116.032888

Santos JPR dos, Vasconcellos RC de C, Pires LPM, et al (2016) Inclusion of Dominance Effects in the Multivariate GBLUP Model. PLoS ONE 11:e0152045. https://doi.org/10.1371/journal.pone.0152045

Sarinelli JM, Murphy JP, Tyagi P, et al (2019) Training population selection and use of fixed effects to optimize genomic predictions in a historical USA winter wheat panel. Theor Appl Genet 132:1247–1261.

https://doi.org/10.1007/s00122-019-03276-6

Schulthess AW, Wang Y, Miedaner T, et al (2016) Multiple-trait- and selection indices-genomic predictions for grain yield and protein content in rye for feeding purposes. Theor Appl Genet 129:273–287. https://doi.org/10.1007/s00122-015-2626-6

Schulthess AW, Zhao Y, Longin CFH, Reif JC (2018) Advantages and limitations of multiple-trait genomic prediction for Fusarium head blight severity in hybrid wheat (*Triticum aestivum* L.). Theor Appl Genet 131:685–701. https://doi.org/10.1007/s00122-017-3029-7

Shewry PR (2009) Wheat. Journal of Experimental Botany 60:1537–1553. https://doi.org/10.1093/jxb/erp058

Storlie E, Charmet G (2013) Genomic Selection Accuracy using Historical Data Generated in a Wheat Breeding Program. Plant Genome 6. https://doi.org/10.3835/plantgenome2013.01.0001

Sun J, Rutkoski JE, Poland JA, et al (2017) Multitrait, Random Regression, or Simple Repeatability Model in High-Throughput Phenotyping Data Improve Genomic Prediction for Wheat Grain Yield. Plant Genome 10:. https://doi.org/10.3835/plantgenome2016.11.0111

Venables, W. N. & Ripley, B. D. (2002) Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0

# List of tables

**Table 1** Number of selected individuals in each scenario

**Table 2** Summary statistics, variance components and repeatabilities for the main traits and traits linked to bread making quality

| Budget | % of lines phenotyped for W (*In the training set for MT and in the total population for TA*) | MT | | TA | |
|---|---|---|---|---|---|
| | | $N_W$ | $N_{BMS}$ | $N_W$ | $N_{BMS}$ |
| **47 700€** | 100 | 318 | 275 | 398 | 264 |
| | 75 | 238 | 286 | 298 | 278 |
| | 50 | 159 | 296 | 199 | 291 |
| | 25 | 79 | 307 | 99 | 304 |
| **30 210€** | 100 | 318 | 159 | 398 | 148 |
| | 75 | 238 | 169 | 298 | 161 |
| | 50 | 159 | 180 | 199 | 174 |
| | 25 | 79 | 190 | 99 | 188 |
| **16 760€** | 100 | 318 | 69 | 398 | 58 |
| | 75 | 238 | 80 | 298 | 72 |
| | 50 | 159 | 90 | 199 | 85 |
| | 25 | 79 | 100 | 99 | 98 |

**Table 1** Number of selected individuals in each scenario
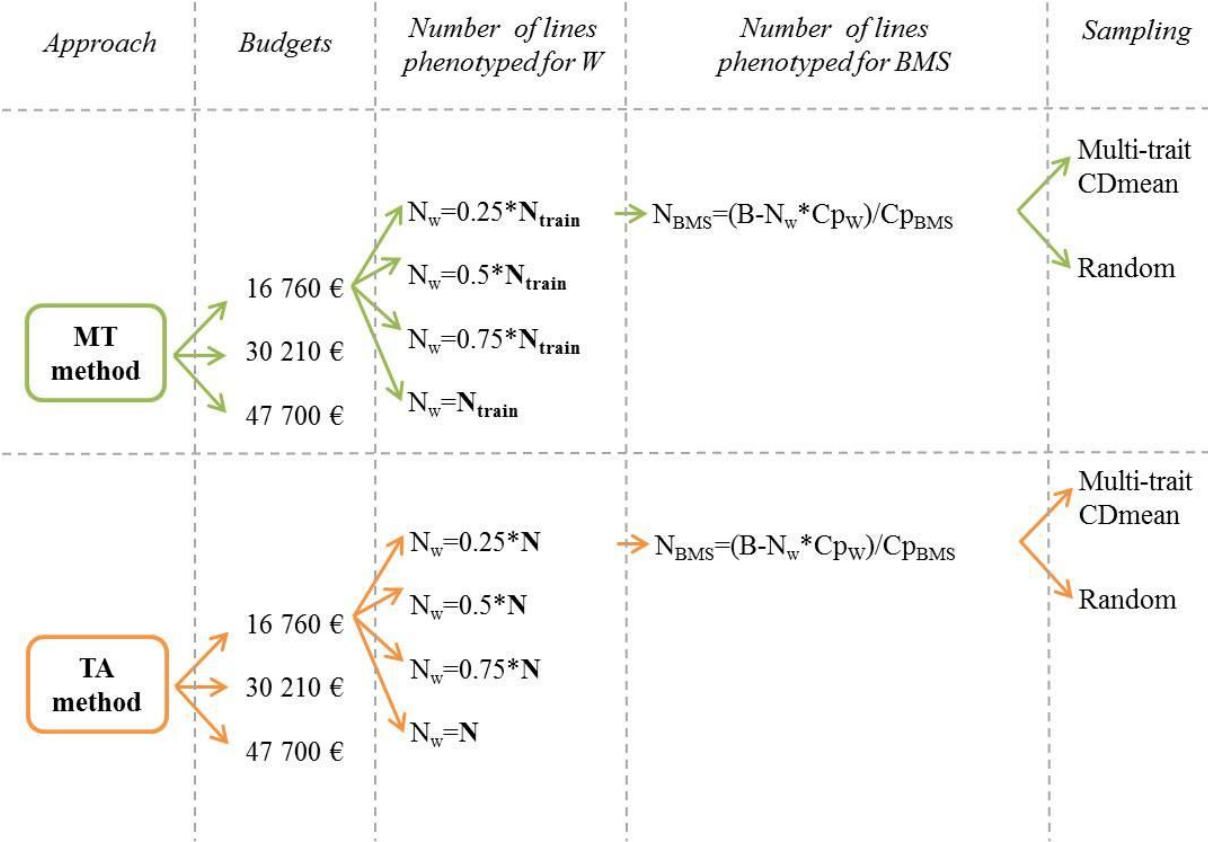
*$N_W$, $N_{BMS}$: Number of lines phenotyped for W and BMS, respectively*

| | Trait | Mean | Var | $\sigma_E^2$ | $\sigma_G^2$ | $\sigma_{GxE}^2$ | $\sigma_e^2$ | Plot repeatability | Design repeatability |
|---|---|---|---|---|---|---|---|---|---|
| | **Grain yield** | 91.34 | 269.67 | 236.06 | 14.05 | 26.55 | 9.27 | 0.28 | 0.84 |
| | **Plant height** | 87.85 | 94.27 | 50.85 | 30.31 | 5.87 | 4.94 | 0.74 | 0.95 |
| | **Heading date** | 140.98 | 63.04 | 45.47 | 8.08 | $7{,}72 \times 10^{-4}$ | 1.93 | 0.81 | 0.98 |
| | **Protein content** | 11.17 | 1.89 | 1.4 | 0.26 | 0.21 | 0.1 | 0.46 | 0.91 |
| **Chopin alveograph Parameters** | **Dough strength:W** | 205.63 | 3844.01 | 788.98 | 2030.03 | --- | 1045.85 | 0.66 | 0.92 |
| | **Tenacity: P** | 77.46 | 695.31 | 166.1 | 413.3 | --- | 117.67 | 0.78 | 0.95 |
| | **Extensibility: L** | 85.05 | 936.17 | 254.77 | 465.46 | --- | 282.5 | 0.62 | 0.9 |
| | $P/_L$**ratio** | 1.16 | 0.74 | 0.14 | 0.34 | --- | 0.24 | 0.59 | 0.89 |
| **BIPEA Test** | **Dough score** | 79.24 | 215.21 | 27.78 | 69.5 | --- | 120.13 | 0.37 | 0.77 |
| | **Crust score** | 50.18 | 483.88 | 96.94 | 145.1 | --- | 249.32 | 0.37 | 0.77 |
| | **Crumb score** | 93.98 | 61.16 | 15.13 | 12.16 | --- | 36.53 | 0.25 | 0.66 |
| | **Loaf volume** | 1482.44 | 48629.99 | 10647.39 | 15815.57 | --- | 23975.46 | 0.4 | 0.79 |
| | **BMS** | 223.4 | 1276.6 | 141.13 | 466.23 | --- | 688.96 | 0.4 | 0.8 |

**Table 2** Summary statistics, variance components and repeatabilities for the main traits and traits linked to bread making quality

$\sigma^2_G$; $\sigma^2_{G \times E}$ and $\sigma^2_\varepsilon$ are the genotype, genotype x environment and residual variances respectively

# Figure



**Fig. 1** Definition of each scenario

*$N_W$, $N_{BMS}$: Number of lines phenotyped for W and BMS, respectively. N: Total number of lines (number of lines in the training and the validation sets). $N_{train}$: Number of lines in the training set. $Cp_{BMS}$, $Cp_w$: the cost of phenotyping for one line for BMS and for W respectively. First, we defined $Cp_{BMS} = 150€$ and $Cp_w = 20€$. Then we made the cost ratio between W and BMS vary (ratio = 1/2, 1/3, 1/4, 1/5, 1/10 and 1/20).*
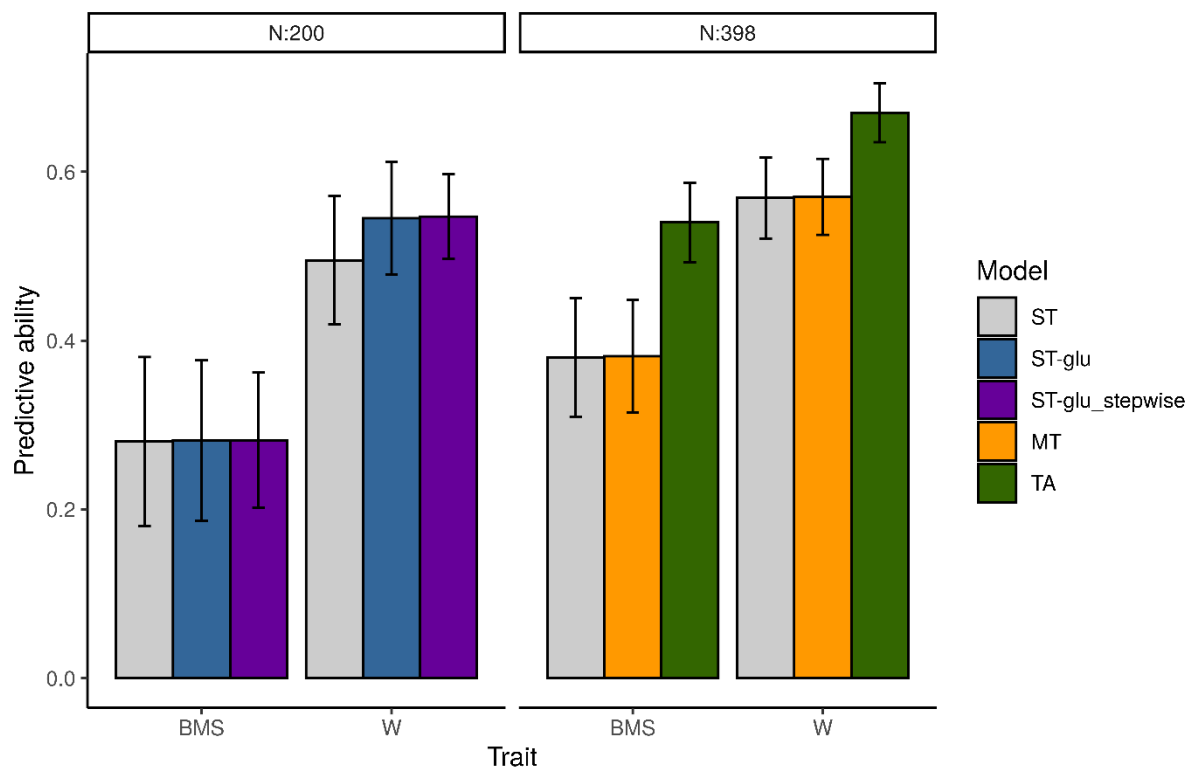
**Fig.2** Comparison of predictive ability using single-trait and glutenin diagnostic markers (*ST-glu and ST-glu_stepwise)*, single-trait (ST), multi-trait (MT) and trait assisted (TA) genomic prediction models

*N: Number of lines. On the left: comparison of predictive ability for each trait using three single-trait models (ST, ST-glu and ST-glu_stepwise). On the right: comparison of predictive ability using single-trait and multi-trait models. For the prediction of BMS, W was used as the correlated traits in bivariate models (MT and TA). For the prediction of W, BMS was used as the correlated traits in bivariate models (MT and TA). Error bars stand for standard deviation.*
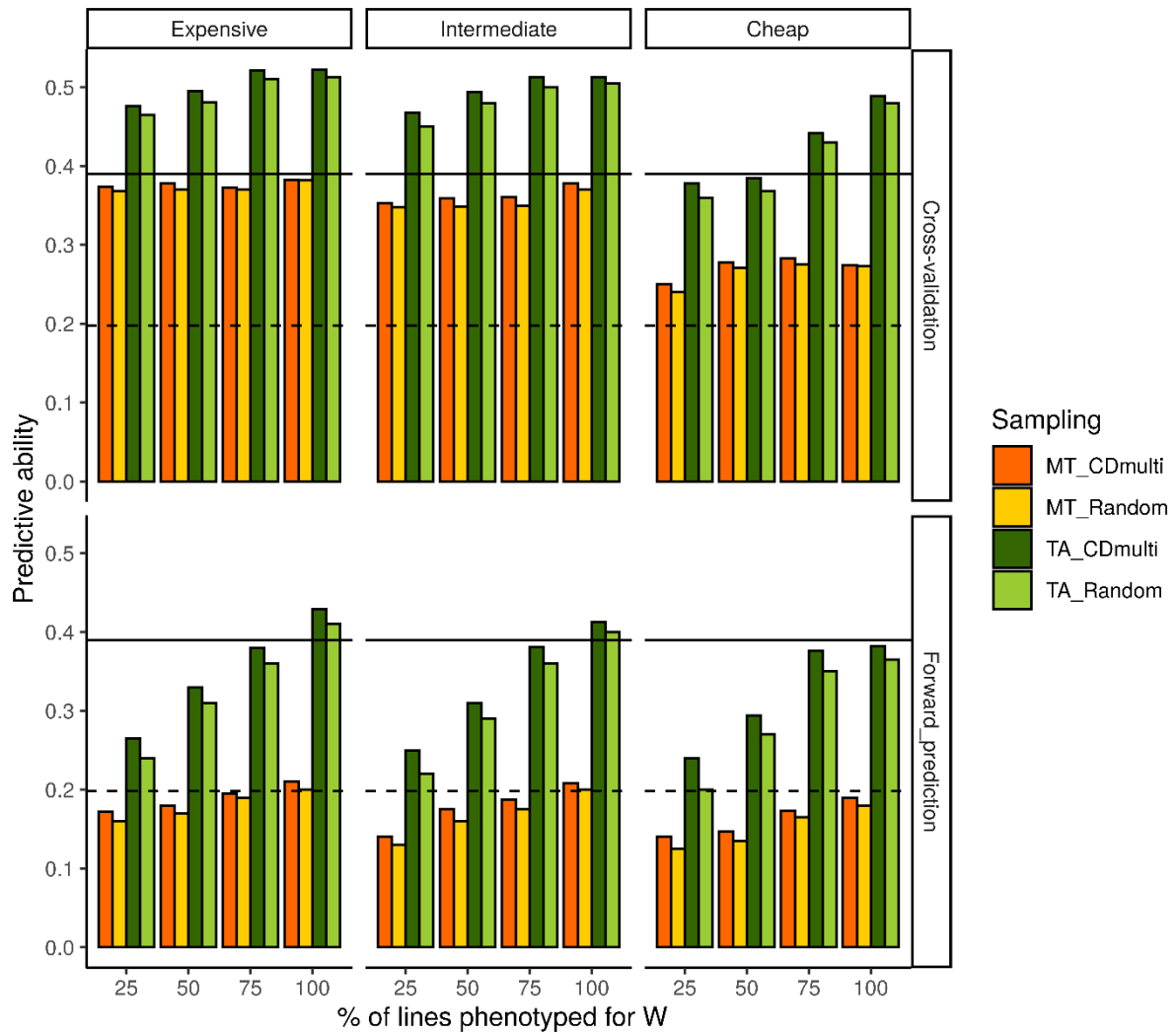
**Fig. 3** Comparison of predictive ability for BMS in each scenario

*Results obtained with the 5-fold cross-validation method on the top. Results obtained when only the most recent lines are predicted on the bottom (Forward prediction). Solid lines correspond to predictive ability obtained using the ST model in cross-validation scenario. Dashed lines correspond to predictive ability obtained using the ST model in forward prediction scenario. Expensive, intermediate and cheap correspond to the three budgets. MT_CDmulti, TA_CDmulti: MT and TA approaches using the CDmulti to sample lines to phenotype, respectively. MT_Random, TA_Random: MT and TA approaches using a random sampling, respectively.*
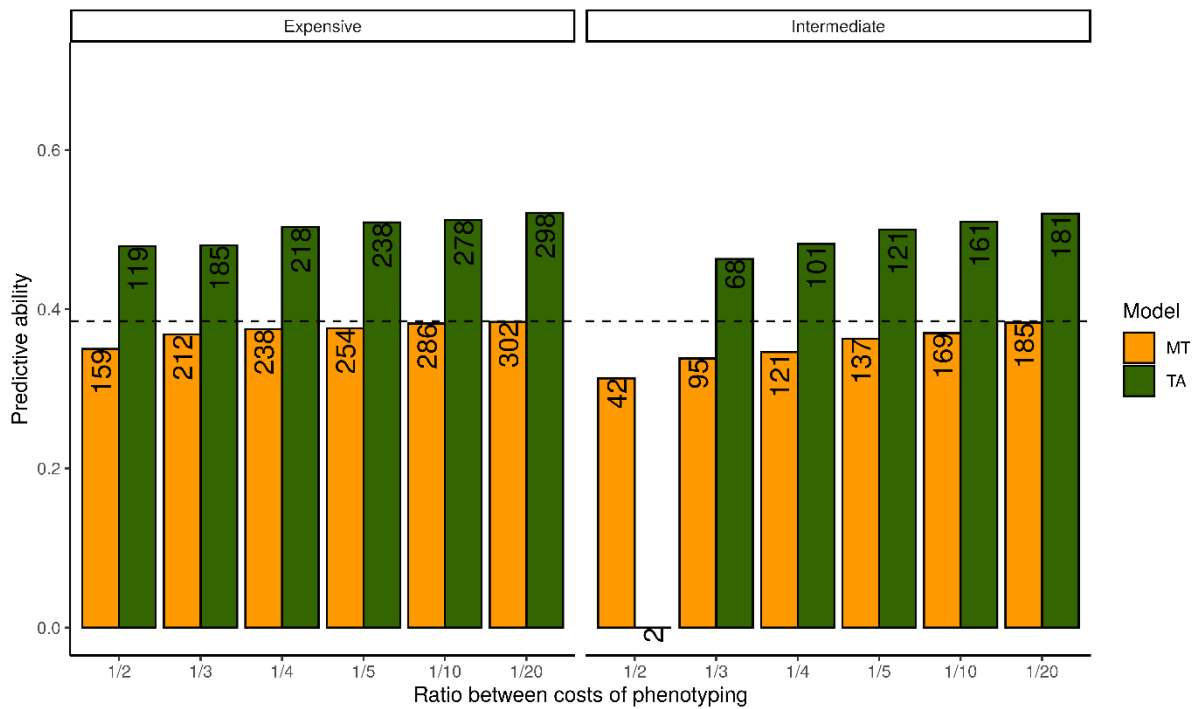
**Fig. 4** Impact of cost ratio between W (cheap) and BMS (expensive) traits on the predictive ability of BMS using multi-trait (MT) and trait assisted (TA) genomic prediction models

*The dashed line corresponds to predictive ability obtained using ST model. Ratio between phenotyping costs: $\frac{Cp_W}{Cp_{BMS}}$ with $Cp_W$ W cost for one line and $Cp_{BMS}$ BMS cost for one line. $Cp_{BMS}$ is a constant (150€) and $Cp_W$ depends on the cost ratio. "Expensive" and "intermediate" correspond to the two budgets. The number on each barplot corresponds to the number of lines phenotyped for BMS. In each scenario, 100% of the lines of the training set (MT) or of the total population (TA) were phenotyped for W.*
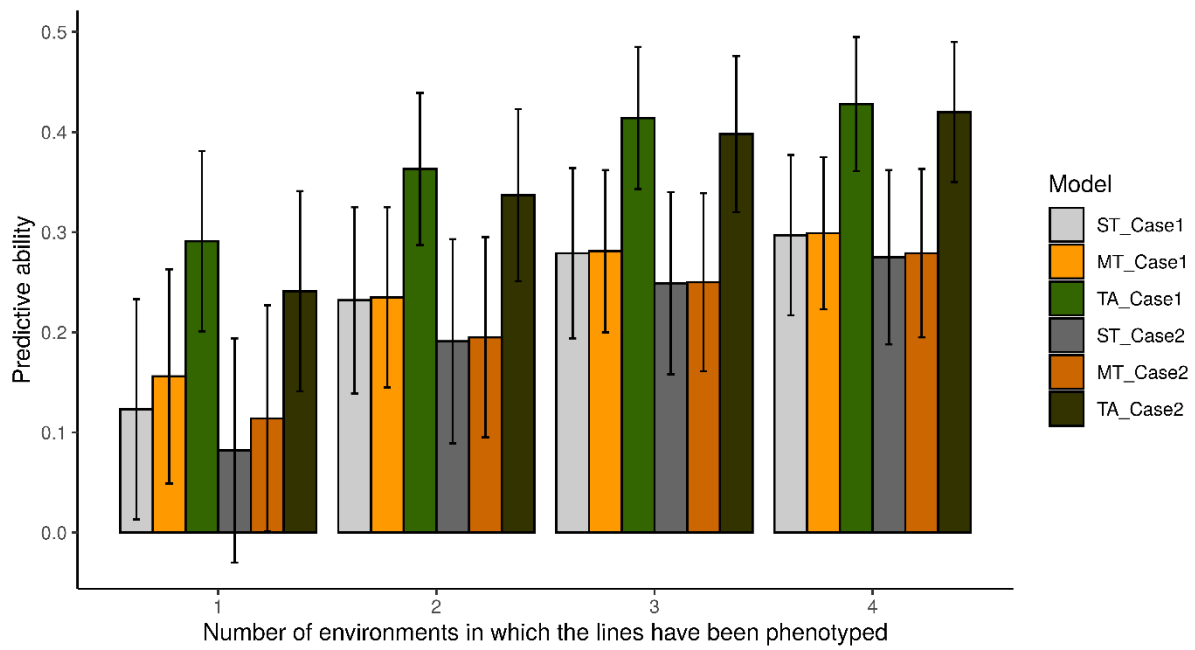
**Fig. 5** Impact of the number of phenotypic records on predictive ability of BMS

*Case 1: all the phenotypic records of control lines were used to estimate BLUEs in order to have a more connected design. Case 2: phenotypic records of control lines were sampled in the same way as candidate lines to calculate BLUEs. In both cases, the total population (training and validation sets) includes 398 lines. Error bars stand for standard deviations.*

Graphics programs used to create the figures:

- Microsoft PowerPoint: Figure 1
- R environment : Figures 2-5 (ggplot2 package)

# Supplementary material

| Name of trial | GPS coordinates |
| --- | --- |
| Champagne-Céréales | 48°52'N/4°09'E |
| Clermont-Ferrand | 45°46'N/3°04'E |
| Colmar | 48°03'N/7°20'E |
| Dijon | 47°19'N/5°04'E |
| Epi Centre | 47°05'N/2°25'E |
| Estrées-Mons | 49°52'N/3°00'E |
| Le Moulon | 48°42'N/2°09'E |
| Lusignan | 46°25'N/0°11'E |
| Orgeval | 48°55'N/1°58'E |
| Orsonville | 48°27'N/1°52'E |
| Rennes | 48°06'N/1°40'W |

**Table S1** Location of trials

| Budget | Method | W cost per line (€) | BMS cost per line (€) | Ratio between W and BMS costs | $N_W$ | $N_{BMS}$ | Predictive ability of BMS and sd |
|---|---|---|---|---|---|---|---|
| 47700 | TA | 75 | 150 | 1/2 | 398 | 119 | 0.48 ± 0.06 |
| | | 50 | 150 | 1/3 | 398 | 185 | 0.48 ± 0.06 |
| | | 37.5 | 150 | 1/4 | 398 | 218 | 0.50 ± 0.06 |
| | | 30 | 150 | 1/5 | 398 | 238 | 0.51 ± 0.05 |
| | | 15 | 150 | 1/10 | 398 | 278 | 0.51 ± 0.05 |
| | | 7.5 | 150 | 1/20 | 398 | 298 | 0.52 ± 0.04 |
| | MT | 75 | 150 | 1/2 | 318 | 159 | 0.35 ± 0.12 |
| | | 50 | 150 | 1/3 | 318 | 212 | 0.37 ± 0.11 |
| | | 37.5 | 150 | 1/4 | 318 | 238 | 0.37 ± 0.11 |
| | | 30 | 150 | 1/5 | 318 | 254 | 0.38 ± 0.11 |
| | | 15 | 150 | 1/10 | 318 | 286 | 0.38 ± 0.10 |
| | | 7.5 | 150 | 1/20 | 318 | 302 | 0.38 ± 0.10 |
| 30210 | TA | 75 | 150 | 1/2 | 398 | 2 | NA |
| | | 50 | 150 | 1/3 | 398 | 68 | 0.46 ± 0.06 |
| | | 37.5 | 150 | 1/4 | 398 | 101 | 0.48 ± 0.06 |
| | | 30 | 150 | 1/5 | 398 | 121 | 0.50 ± 0.06 |
| | | 15 | 150 | 1/10 | 398 | 161 | 0.51 ± 0.05 |
| | | 7.5 | 150 | 1/20 | 398 | 181 | 0.52 ± 0.05 |
| | MT | 75 | 150 | 1/2 | 318 | 42 | 0.31 ± 0.12 |
| | | 50 | 150 | 1/3 | 318 | 95 | 0.34 ± 0.12 |
| | | 37.5 | 150 | 1/4 | 318 | 121 | 0.36 ± 0.12 |
| | | 30 | 150 | 1/5 | 318 | 137 | 0.36 ± 0.11 |
| | | 15 | 150 | 1/10 | 318 | 169 | 0.37 ± 0.11 |
| | | 7.5 | 150 | 1/20 | 318 | 185 | 0.38 ± 0.11 |
| 16760 | TA | 75 | 150 | 1/2 | 223 | NA | NA |
| | | 50 | 150 | 1/3 | 335 | NA | NA |
| | | 37.5 | 150 | 1/4 | 398 | 12 | NA |
| | | 30 | 150 | 1/5 | 398 | 32 | 0.45 ± 0.07 |
| | | 15 | 150 | 1/10 | 398 | 71 | 0.48 ± 0.07 |
| | | 7.5 | 150 | 1/20 | 398 | 91 | 0.50 ± 0.06 |
| | MT | 75 | 150 | 1/2 | 223 | NA | NA |
| | | 50 | 150 | 1/3 | 318 | 5 | NA |
| | | 37.5 | 150 | 1/4 | 318 | 32 | 0.23 ± 0.16 |
| | | 30 | 150 | 1/5 | 318 | 48 | 0.26 ± 0.15 |
| | | 15 | 150 | 1/10 | 318 | 79 | 0.31 ± 0.12 |
| | | 7.5 | 150 | 1/20 | 318 | 95 | 0.32 ± 0.12 |

**Table S2** Variation of cost ratio between W (cheap) and BMS (expensive) and its impact on the predictive ability of BMS

*$N_W$, $N_{BMS}$: Number of lines phenotyped for W and BMS, respectively. Sd: Standard deviation. $N_{BMS}$ is not available when the budget is not enough to phenotype all the lines for W. Predictive ability is not available (NA) when $N_{BMS}$ is a negative number and when $N_{BMS}$ is small because of convergence model issues.*
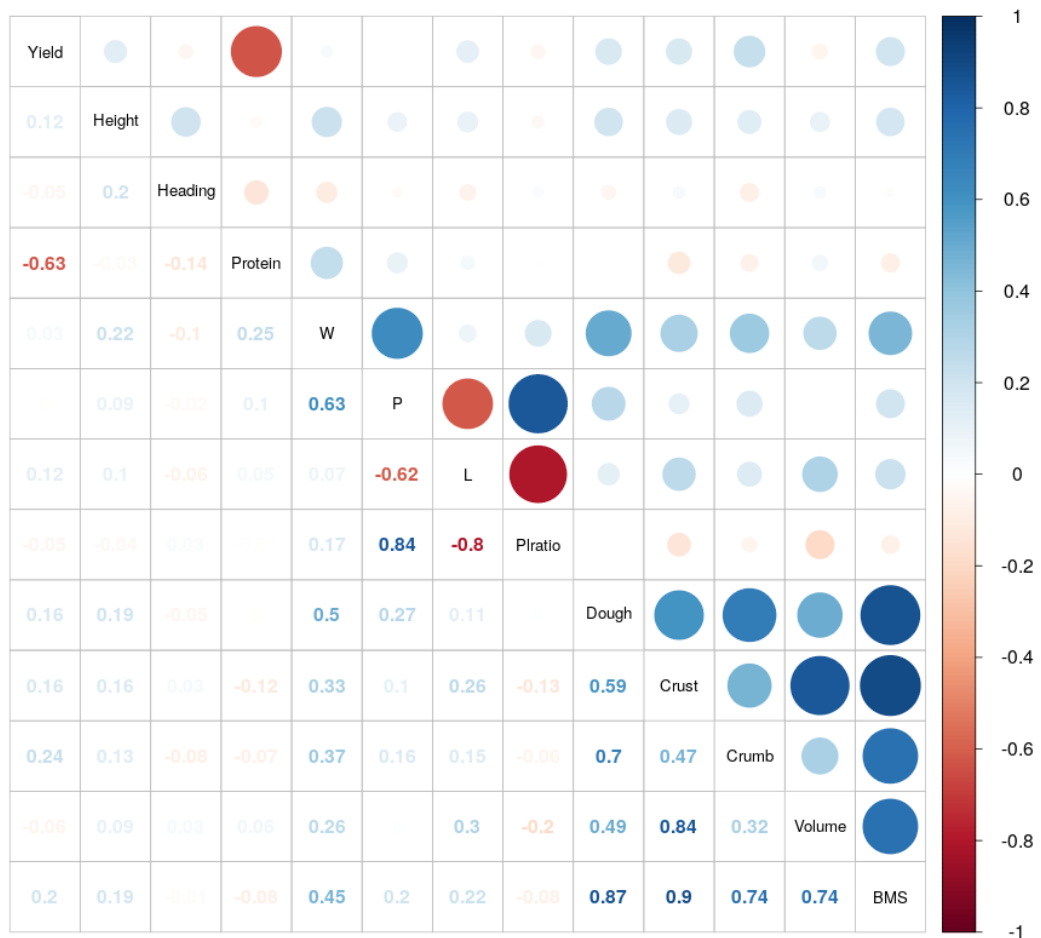
**Fig.S1** Pearson correlations between main traits and traits linked to bread making quality

*Yield: grain yield. Height: plant height. Heading: heading date. Protein: protein content. W: dough strength. P: Tenacity. L: Extensibility. Plratio: ratio Tenacity / Extensibility (P/L). Dough, Crust and Crumb: scores for dough, crust and crumb, respectively (obtained with the BIPEA test). Volume: dough volume. BMS: Bread Making Score (final BIPEA score)*
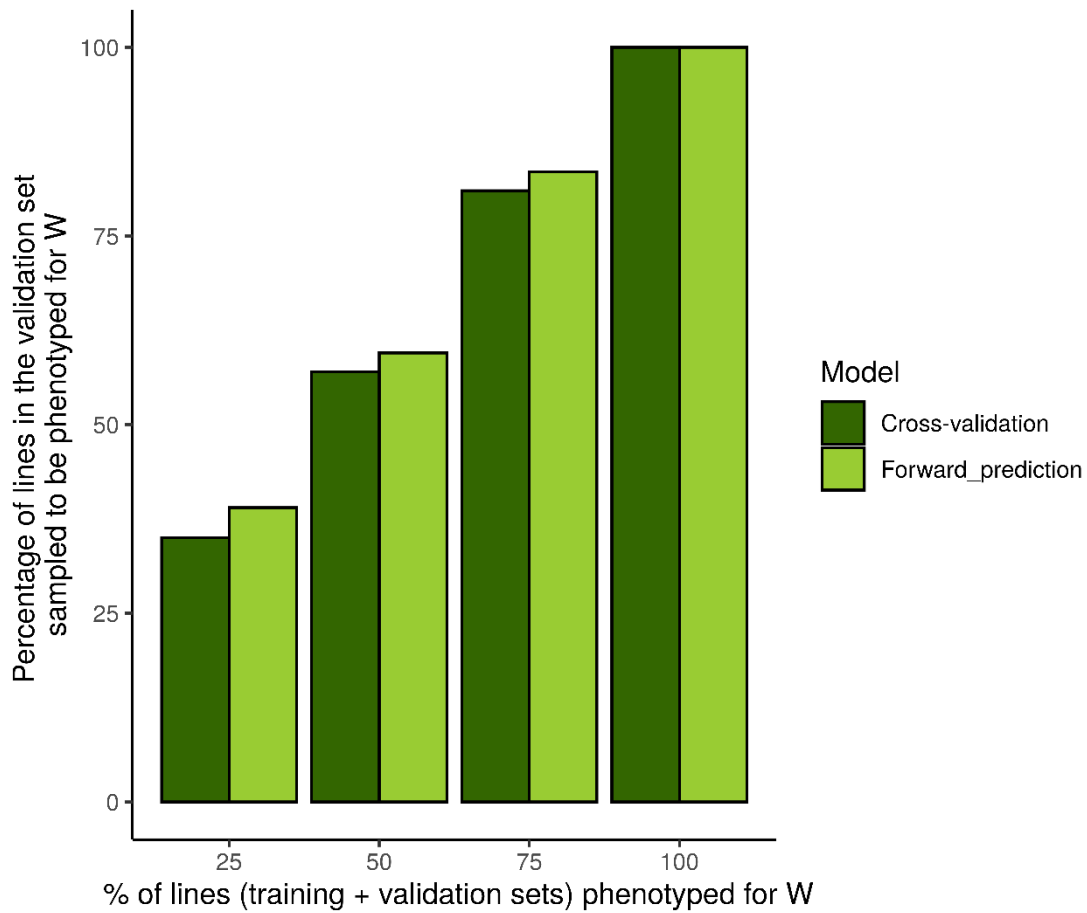
**Fig. S2** Percentage of lines phenotyped for W in the validation set for each resource allocation scenario

Program used to create the figure S1: R environment with corrplot package.

Graphics program used to create the figure S2: R environment with the ggplot2 package.

# Appendix

We extended here the generalized CD (Laloë 1993) to the multi-trait context. The objective is to compute the expected reliability (before phenotyping) in a multi-trait context for contrasts corresponding to the prediction objectives. To give an example, it is clear that if the objective is to accurately predict the difference (contrasts) between families, or to focus on predictions within families, the contrasts to be considered will be different and the optimal calibration sets as well. By defining contrasts corresponding to the prediction objectives, one can adapt the CDmulti criterion to the specific prediction objectives (see Rincent et al. 2017 for more details in the single trait context).

For a given contrast c, the generalized Prediction Error Variance (PEV(c)) in the multi-trait context is equal to:

$$PEV(c) = c^T (Z^T M Z + (\Sigma_a^{-1} \otimes K^{-1}))^{-1} c \qquad (14)$$

With $M = (\Sigma_\varepsilon^{-1} \otimes I) - (\Sigma_\varepsilon^{-1} \otimes I) X (X^T (\Sigma_\varepsilon^{-1} \otimes I) X)^{-1} X^T (\Sigma_\varepsilon^{-1} \otimes I)$     (15)

where X and Z are the design matrices for the fixed and random effects respectively, K is the kinship matrix, $\Sigma_a$ is the genetic variance–covariance matrix between traits, and $\Sigma_\varepsilon$ is the residual variance-covariance matrix between traits. $\otimes$ indicate the Kronecker product operator between matrices.

The generalized multi-trait CD for a given contrast c is equal to:

$$CDmulti\ (c) = \frac{c^T ((\Sigma_a \otimes K) - (Z^T M Z + (\Sigma_a^{-1} \otimes K^{-1}))^{-1}) c}{c^T (\Sigma_a \otimes K) c} \qquad (16)$$

As a reminder, a contrast is a vector whose elements sum to 0 and indicating the difference in which we are interested. For example, if we are interested in accurately predicting the difference between individual 1 and individual 2, the contrast to consider will be: $c^T = [1, -1, 0, 0, 0, 0, \dots]$.