Portability of genomic predictions trained on sparse factorial designs across two maize silage breeding cycles

Alizarine Lorenzi
INRAE UMR0320: Genetique Quantitative et Evolution Le Moulon
Cyril Bauland
INRAE UMR0320: Genetique Quantitative et Evolution Le Moulon
Sophie Pin
INRAE UMR0320: Genetique Quantitative et Evolution Le Moulon
Delphine Madur
INRAE UMR0320: Genetique Quantitative et Evolution Le Moulon
Valérie Combes
INRAE UMR0320: Genetique Quantitative et Evolution Le Moulon
Carine Palaffre
INRAE Nouvelle-Aquitaine Bordeaux: Institut National de Recherche pour l'Agriculture l'Alimentation et l'Environnement Centre Nouvelle-Aquitaine Bordeaux
Colin Guillaume
Maisadour
Gaëtan Touzy
RAGT2n
Tristan Mary-Huard
MIA-Paris: Mathematiques et Informatique Appliquees-Paris
Alain Charcosset
INRAE UMR0320: Genetique Quantitative et Evolution Le Moulon
Laurence Moreau
laurence.moreau@inrae.fr

INRAE: Institut National de Recherche pour l'Agriculture l'Alimentation et l'Environnement
https://orcid.org/0000-0002-7195-1327

Research Article

Keywords: Hybrid breeding, Genomic selection, Factorial design, Training set optimization, Inter-generation genomic prediction

Posted Date: September 6th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3286945/v1

Version of Record: A version of this preprint was published at Theoretical and Applied Genetics on March 7th, 2024. See the published version at https://doi.org/10.1007/s00122-024-04566-4.

Portability of genomic predictions trained on sparse factorial designs across two maize silage breeding cycles

Alizarine Lorenzi[1,4], Cyril Bauland[1], Sophie Pin[1], Delphine Madur[1], Valérie Combes[1], Carine Palaffre[2], Colin Guillaume[3], Gaëtan Touzy[4], Tristan Mary-Huard[1,5], Alain Charcosset[1], Laurence Moreau[1]

[1] Université Paris-Saclay, INRAE, CNRS, AgroParisTech, Génétique Quantitative et Evolution - Le Moulon, 91190 Gif-sur-Yvette, France
[2] UE 0394 SMH, INRAE, 2297 Route de l'INRA, 40390, Saint-Martin-de-Hinx, France
[3] Maïsadour Semences SA, F-40001 Mont-de-Marsan Cedex, France
[4] RAGT2n, Genetics and Analytics Unit, 12510 Druelle, France
[5] MIA, INRAE, AgroParisTech, Université Paris-Saclay, 75005, Paris, France

Corresponding author: Laurence Moreau *laurence.moreau@inrae.fr*. *ORCID:* 0000-0002-7195-1327

Abstract

Genomic selection offers new prospects for revisiting hybrid breeding schemes by replacing extensive phenotyping of individuals with genomic predictions. Finding the ideal design for training genomic prediction models is still an open question. Previous studies have shown promising predictive abilities using sparse factorial instead of tester-based training sets to predict single-cross hybrids from the same generation. This study aims to further investigate the use of factorials and their optimization to predict line general combining abilities (GCAs) and hybrid values across breeding cycles. It relies on two breeding cycles of a maize reciprocal genomic selection scheme involving multiparental connected reciprocal populations from flint and dent complementary heterotic groups selected for silage performances. Selection based on genomic predictions trained on a factorial design resulted in a significant genetic gain for dry matter yield in the new generation. Results confirmed the efficiency of sparse factorial training sets to predict candidate line GCAs and hybrid values across breeding cycles. Compared to a previous study based on the first generation, the advantage of factorial over tester training sets appeared lower across generations. Updating factorial training sets by adding single-cross hybrids between selected lines from the previous generation or a random subset of hybrids from the new generation both improved predictive abilities. The CDmean criterion helped determine the set of single-crosses to phenotype to update the training set efficiently. Our results validated the efficiency of sparse factorial designs for calibrating hybrid genomic prediction experimentally and showed the benefit of updating it along generations.

Keywords

Hybrid breeding, Genomic selection, Factorial design, Training set optimization, Inter-generation genomic prediction

Key message (30 words)

We validated the efficiency of genomic predictions calibrated on sparse factorial training sets to predict the next generation of hybrids and tested different strategies for updating predictions along generations.

Conflicts of interest

The authors declare no conflict of interest.

Acknowledgments

We thank Lidea, Limagrain Europe, Maïsadour Semences, Corteva, RAGT 2n, KWS and Syngenta Seeds grouped in the frame of the ProMais SAM-MCR project for the funding, inbred lines development, hybrid production, and phenotyping. We are also grateful to scientists from these companies and scientists of the INRAE-CIRAD "R2D2" network for helpful discussions on the results. A.L. PhD contract was funded by RAGT 2n and ANRT contract n° 2020/0032, with the contribution of all SAM-MCR project partners. GQE-Le Moulon benefits from the support of Saclay Plant Sciences-SPS (ANR-17-EUR-0007).

Introduction

Maize varieties are generally single-cross hybrids obtained by crossing two inbred lines that belong to complementary heterotic groups. The challenges for breeders are (i) selecting lines within each heterotic group that will be used as parents for the next generation and (ii) identifying the best single-cross hybrids among all possible ones in order to derive new varieties. The advent of Doubled-Haploid (DH) technology now enables the rapid production of numerous fully homozygous inbred lines each year. This large number of candidate lines produced each year in breeding programs makes generating and evaluating all potential single-cross hybrids practically undoable. To overcome this difficulty, conventional maize hybrid breeding schemes are typically divided into two stages. In the first stage (1), topcross hybrids are produced by crossing candidate lines from one heterotic group with a limited number of inbred lines from the complementary group, referred to as "testers". The performances of these topcross hybrids provide an estimation of the general combining abilities (GCA) of the candidate lines. In the second stage (2), the selected lines from stage 1 are crossed using a sparse factorial design to identify the best single-cross hybrid combinations. At this stage, the selection is performed on the GCA of the parental lines and the specific combining ability (SCA) of the pair of parental lines. Selecting lines based on a limited number of testers at stage 1 does not fully exploit the complementarity between the candidate lines from the two heterotic groups and can bias the line GCA estimation since the line GCA and its SCAs with the testers are confounded (Hallauer et al. 2010). Also, the two-stage process is time-consuming and requires extensive phenotyping (at least as many as the total number of candidate lines in both groups in stage 1).

Due to limited resources for phenotyping, predicting the performance of untested hybrids has been a critical objective in hybrid selection. Bernardo (1994) was the first to propose a marker-based model for plant hybrid performance prediction. He combined marker-based distances between parental lines of hybrids and the performances of a related set of single-crosses to predict GCAs and SCAs of non-phenotyped hybrids. This prediction model, which aims at predicting the value of unphenotyped individuals based on their marker-based relationship with a set of individuals both phenotyped and genotyped, is similar to the genomic best linear unbiased prediction (GBLUP) model (VanRaden 2008) that has been proposed more recently and is now widely used to perform genomic selection (GS) in plants and animal. Different other genomic prediction models have been proposed (see Meuwissen et al. 2001 seminal paper and Howard et al. 2022 for a review), all of them use molecular markers scored across the entire genome to predict the genetic values of genotyped individuals, referred to as the prediction set (PS), using individuals both phenotyped and genotyped, referred to as the training set (TRS). Since the pioneer work of Bernardo (1994), different prediction models adapted to hybrid value prediction have been

proposed considering non-additive effects, either by modeling the GCA and SCA effects or the additive, dominance, and epistasis effects (Vitezica et al. 2013, 2017; Varona et al. 2018; González-Diéguez et al. 2021). Even if several studies have confirmed the efficiency of these GS models for predicting single-cross hybrid values in maize (see review by Kadam and Lorenz 2018), the relative interest of the different prediction models is still unclear. Besides the statistical model, various factors are known to affect genomic prediction accuracies, such as trait heritability, the number of markers, the size of the TRS, and the relationship between the TRS and the PS (see reviews: Kadam and Lorenz 2018; Isidro y Sánchez and Akdemir 2021; Merrick and Carter 2021; Kadam et al. 2021). In the context of hybrid prediction, in addition to these factors, the crossing design used to produce the TRS hybrids also affects prediction accuracy (Technow et al. 2014; Seye et al. 2020; Lorenzi et al. 2022).

In most studies, GS for hybrid value prediction has been considered in the second stage of the hybrid breeding scheme, i.e., by using as TRS hybrids between lines that have already undergone a selection based on their testcross values. To improve the efficiency of hybrid breeding schemes, Kadam et al. (2016) and (Giraud 2016) proposed (1) to replace topcross evaluation in stage 1 with a sparse factorial design between unselected candidate lines from both groups and (2) to use GS to predict GCAs of all lines and SCA of all potential single-cross combinations. This makes it possible to perform selection in one stage instead of two. Both studies found good prediction accuracies for untested hybrids using factorial designs as TRS. Later, simulations and experimental studies have shown the potential of using sparse factorial instead of tester TRSs when predicting the same generation (Seye et al. 2020; Burdo et al. 2021; Lorenzi et al. 2022). Although a simulation work validated the advantage of factorial compared to tester TRSs to predict hybrid values across breeding cycles (Seye et al. 2020), further experimental validation is needed. From one cycle to the next, the average relatedness between the TRS and PS decreases and the joint effect of selection, drift, and recombination events change allele frequencies and the linkage disequilibrium between markers and QTLs, which decrease prediction accuracy if the TRS is not updated along cycles (Pszczola et al. 2012; Isidro y Sánchez and Akdemir 2021; Rio et al. 2022b). This raises questions about how to efficiently update the TRS to maximize prediction accuracy while minimizing phenotyping costs.

According to the literature, an ideal TRS should maximize the accuracy by maximizing the relationship between the TRS and PS (Zhong et al. 2009; Zhao et al. 2012; Technow et al. 2013) and minimizing the within TRS relationship to capture a large genetic variance (Pszczola et al. 2012; Isidro y Sánchez and Akdemir 2021). Different optimization criteria have been proposed to define the TRS (Rio et al. 2022b). Rincent et al. (2012) proposed optimizing the TRS by maximizing the mean of the coefficient of determination (CDmean) of contrasts

between each unphenotyped PS individual and the target population mean. Numerous studies have shown that building the TRS using the CDmean significantly increases the accuracy of GS models relative to random sampling (Isidro y Sánchez and Akdemir 2021; Rio et al., 2022; Fernández-González et al. 2023). In a breeding program, where genomic prediction is applied routinely, a large dataset from previous years of phenotyping is available for model training. One can wonder which phenotypic data from the previous generation(s) should be included in the TRS and which additional hybrids should be phenotyped to complete the existing TRS and achieve the highest prediction accuracy for the new generation with a given phenotyping effort. One idea could be to use the CDmean to optimize the choice of the individuals from the new generation to be phenotyped while considering the existing TRS comprising data from the previous generations. To our knowledge, this strategy has never been tested in this context.

The present study investigates the use and optimization of factorial TRS for genomic prediction of hybrid performance across breeding cycles. It relies on two breeding cycles of a reciprocal genomic selection scheme initiated from multiparental connected reciprocal populations generated in the flint and dent complementary heterotic groups. Data from the first cycle was already analyzed in previous studies (Giraud et al. 2017a, b; Seye et al. 2019) and have shown promising results in terms of genomic predictive abilities for replacing testcross evaluation by sparse factorial evaluations (Lorenzi et al. 2022). We present in study results from a new breeding cycle that was produced and evaluated in a factorial design to: (i) estimate the genetic gain achieved after selection based on genomic predictions calibrated on a sparse factorial, (ii) assess the predictive ability in the new breeding cycle and compare different GS models, (iii) evaluate the efficiency of training GS models on a factorial design for predictions across breeding cycles and compare it to tester designs, (iv) investigate the benefit of different strategies to update the factorial TRS across cycles and optimize it to predict the new generation.

Materials and Methods

This study relies on data from a reciprocal breeding experiment aiming at improving the silage performance of maize single-cross hybrids produced between the dent and flint heterotic groups, the two main heterotic groups used for silage maize hybrids in Northern Europe. The experimental data comprises two breeding cycles, further called G0 and G1. Inbred lines from the G0 cycle were evaluated for hybrid performances in three experimental designs already analyzed in previous publications (Giraud et al. 2017a, b; Seye et al. 2019; Lorenzi et al. 2022). A summary of the G0 cycle production is provided below. The best G0 lines in each group were selected based on

genomic predictions and intercrossed to produce the new breeding cycle (G1) we will focus on in this study. All experimental designs are described in **Table 1** and **Fig.1**.

Summary of the G0 plant material production and selection of the best candidate lines

Four founder lines were intercrossed in each group to derive six biparental families. In total, 822 flint lines and 802 dent lines were produced, further called G0 lines. The G0 lines were crossed to produce three experimental hybrid designs. The G0_F-1H was obtained by crossing the 822 flint lines to the 801 dent lines following a sparse factorial design, leading to 951 single-cross hybrids (on average, one line contributed to 1.2 hybrids). This experimental design was evaluated in France and Germany in 2013 and 2014 for silage performances (Giraud et al. 2017a, b; Seye et al. 2019). Then, 30 G0 lines were selected in each heterotic group based on genomic predictions trained on the G0_F-1H for an index combining silage yield, moisture content at harvest, and silage quality. Additionally, 60 G0 lines (10 lines per family) were chosen randomly. These lines were used to create two other experimental designs. The G0_F-4H factorial design composed of 363 hybrids (on average, one line contributed to four hybrids) was produced by randomly crossing (i) the 30 G0 selected flint lines to the 30 G0 selected dent lines to produce 131 hybrids (further called "G0S hybrids") and (ii) the 60 G0 random dent lines to the 60 G0 random flint lines leading to 232 hybrids (further called "G0R hybrids"). In parallel, the G0_T-F (and G0_T-D) tester design was produced by crossing the same 90 G0 flint (dent) lines from one group to two founder lines from the dent (flint) group used as testers. Together, the G0_T-F and G0_T-D tester designs were called G0 tester designs or G0_T. The G0_F-4H and the G0_T were evaluated jointly in eight trials in Northern France and Germany in 2016 and 1017 (Seye 2019; Lorenzi et al. 2022). In all trials, 18 hybrids were used as controls and evaluated twice: two commercial hybrids (LG30.275 and RONALDINIO) and 16 founder hybrids produced by crossing the four flint founder lines with the four dent founder lines.

New breeding cycle (G1)

40 intragroup single-crosses were produced in each group by crossing the 30 selected G0 lines described above. 351 dent and 351 flint DH lines (G1) were derived from the 40 single-crosses in the dent and flint groups, respectively. The dent G1 lines were crossed with the flint G1 lines following a sparse factorial design to produce 442 G1 hybrids. Crosses were made at random with an average number of hybrids per line close to one. A new set of 47 G0S hybrids between the 30 G0 selected lines from each group was produced and evaluated jointly with the G1 hybrids, yielding a total of 489 hybrids further referred to as (G0S+G1)_F-1H. Hybrids were evaluated for two years in the North of France and Germany: three trials in 2019 and five in 2020. The same 18 control hybrids (two

commercial and 16 founder hybrids) as in the G0 experiments were evaluated twice in each trial. 15% of the experimental hybrids were also replicated once at each location. The field experiments were laid out as augmented partially replicated designs (p-rep) (Williams et al. 2011). Each trial comprised 512 to 520 elementary plots distributed in 26 incomplete blocks of 20 plots. Each genotype was evaluated in 7 trials across 2019 and 2020 and was replicated in at least one trial. For each trial, repetitions were allocated to blocks to form an efficient incomplete block design using the DiGGer R package (Coombes 2009).

Hybrids were evaluated for 11 traits, four agronomical traits: silage yield (DMY in tons of dry matter per ha), dry matter content at harvest (DMC in % of fresh weight), female flowering date (DtSilk in days after January the first) and plant height (PH in cm) and seven silage traits for digestibility: milk fodder unit per kilogram of dry matter (MFU in MFU per kg) (Andrieu 1995) (computed using model 4.2), cell wall content of the harvested dry matter measured by the neutral detergent fiber content (NDF in % of dry matter), cell wall in vitro digestibility of the non-starch and non-soluble carbohydrates part of silage (DINAG in %) and cell wall in vitro digestibility of the non-starch, non-soluble carbohydrates and non-crude protein part of silage (DINAGZ in %), lignin, cellulose and hemicellulose contents in the cell wall NDF evaluated with the Goering and Soest (1970) method (LIGN, CELL, and HCELL in % of NDF). The DINAG and DINAGZ are two digestibility criteria first proposed by Argillier et al. (1995). The silage traits were predicted using Near Infrared Reflectance Spectrometry (NIRS) measured in the lab on silage powders or directly on fields during the harvest, depending on the trial.

Outlier observations were detected by examining raw data and considering field observations. They were treated as missing data. Subsequently, filters were applied to identify plots with standing counts below 80% of the median, and DMC below 25% or above 45%, which were also considered as missing data. Values of DINAG, DINAGZ, and MFU measured in two trials were inconsistent with those of other trials and were excluded from further analyses. Following quality control and filters, the percentage of missing data across all traits was 8%.

Genotyping

The founder lines and the G0 parental lines were genotyped for 18,480 SNPs using a proprietary Affymetrix® array provided by Limagrain. The G1 parental lines were genotyped using a custom-made chip comprising a subset of 15,000 SNPs of the Illumina® MaizeSNP50 BeadChip (Ganal et al. 2011). Filters were applied for both G0 and G1 lines: markers with more than 20% of missing values within the dent and flint parental lines, markers with more than 5% of heterozygosity among the dent (flint) parental lines, and with Minor Allele Frequency (MAF)

inferior to 5% were discarded. After quality control, only markers common to the two arrays were considered. 4,812 SNP polymorphic markers (in at least the flint or dent population) were retained for further analyses.

Estimation of variance components and trait heritabilities

Variance components and trait heritabilities were estimated in the (G0S+G1)_F-1H design. Individual single-plot performances were corrected by the BLUPs of spatial effects predicted using the model defined in supplementary material **File S1**. Corrected data were then used to estimate variance components using the following model :

$$Y_{hii\prime jl} = \mu + \lambda_l + (\tau_h + \rho_{lh}) \times t_h + \left[\pi_j + H_{h(ii')j} + H\lambda_{lh(ii')j}\right] \times (1 - t_h) + E_{hii'jl}\ , (1)$$

where $Y_{hii\prime jl}$ is the phenotypic value corrected by spatial effects of hybrid h of generation j produced by crossing the flint parental line i and the dent parent line i' evaluated in trial l. μ is the intercept, λ_l is the fixed effect of trial l, t_h is an indicator function that distinguishes experimental hybrids (set to 0) from control hybrids (set to 1), τ_h is the fixed effect of control hybrids with 19 levels (2 for commercial hybrids + 16 for founder hybrids + one for non-control hybrids), ρ_{lh} is the effect of the interaction between trial l and control hybrid h, π_j is the fixed effect of the generation with two levels (G0S or G1 hybrids). $H_{h(ii')j}$ is the random genetic effect of experimental hybrid h of generation j, produced by crossing the flint line i and the dent line i'. $H_{h(ii')j}$ is decomposed into its GCA and SCA components as follows:

$$H_{h(ii')j} = U_{ij} + U'_{i\prime j} + S_{ii\prime j}$$

where U_{ij} (respectively $U'_{i\prime j}$) is the random GCA effect of the flint line i (respectively dent line i') of generation j. We assume that U_{ij} and $U'_{i\prime j}$ are independent and identically distributed (iid) within generation and follow a normal distribution: $U_{ij} \sim \mathcal{N}\left(0, \sigma^2_{GCA_f^j}\right)$ and $U'_{i\prime j} \sim \mathcal{N}\left(0, \sigma^2_{GCA_d^j}\right)$, respectively. $\sigma^2_{GCA_f^j}$ and $\sigma^2_{GCA_d^j}$ are the flint and dent GCA variances of generation j. $S_{kk\prime}$ is the random SCA effect of the interaction between the parental lines i and i', with $S_{ii\prime j} \sim \mathcal{N}\left(0, \sigma^2_{SCA^j}\right)$ ind with $\sigma^2_{SCA^j}$ being the SCA variance at generation j. $H\lambda_{lh(ii')j}$ is the genotype by trial interaction and is decomposed as follows:

$$H\lambda_{lh(ii')j} = (U\lambda)_{ilj} + (U'\lambda)_{i\prime lj} + (S\lambda)_{ii\prime lj},$$

where $(U\lambda)_{ilj}$ and $(U'\lambda)_{i\prime lj}$ are the random effects of the flint GCA effect by trial interaction, respectively dent GCA by trial interaction of generation j and $(S\lambda)_{ii\prime lj}$ is the random effect of the SCA by trial interaction of generation j. With $(U\lambda)_{ilj} \backsim \mathcal{N}\left(0, \sigma^2_{GCA\times E_f^j}\right)$, $(U'\lambda)_{i\prime lj} \backsim \mathcal{N}\left(0, \sigma^2_{GCA\times E_d^j}\right)$ and $(S\lambda)_{ii\prime lj} \backsim \mathcal{N}\left(0, \sigma^2_{SCA\times E^j}\right)$.

$\sigma^2_{GCA\times E_f^j}$, $\sigma^2_{GCA\times E_d^j}$ and $\sigma^2_{SCA\times E^j}$ are the flint GCA by trial interaction variance, the dent GCA by trial variance and the SCA by trial interaction variance of generation j, respectively. $E_{hii',jl}$ is the error term; we assume that the errors follow: $E_{hii',jl}\sim\mathcal{N}(0,\sigma^2_{E_l})$ and are iid within trial and independent between trials, $\sigma^2_{E_l}$ is the error variance of trial l. The different random effects of the model are assumed to be independent.

For each trait and each generation j (G0S or G1), the percentage of genetic variance due to SCA was estimated (%), and broad-sense heritability was computed as follows:

$$H_j^2 = \frac{\sigma^2_{H_j}}{\sigma^2_{H_j}+\frac{\sigma^2_{H\times E_j}}{n_{site}}+\frac{\sigma^2_{E_{moy}}}{n_{rep}\times n_{site}}},$$

where $\sigma^2_{H_j}$ is the hybrid genetic variance of generation j computed as $\sigma^2_{H_j}=\sigma^2_{GCA_f^j}+\sigma^2_{GCA_d^j}+\sigma^2_{SCA^j}$, $\sigma^2_{H\times E_j}$ is the total genotype by trial variance of generation j decomposed as: $\sigma^2_{H\times E_j}=\sigma^2_{GCA\times E_f^j}+\sigma^2_{GCA\times E_d^j}+\sigma^2_{SCA\times E^j}$, and $\sigma^2_{E_{moy}}$ is the mean residual variance across all trials, n_{site} is the average number of trials in which an hybrid has been evaluated and n_{rep} is the average number of within trial replicates per hybrid across trials.

Ls-means and genetic gain estimation

Least square-means (ls-means) of hybrids were computed over the eight trials. The model used was:

$$Y^*_{hrl}=\mu+\lambda_l+\gamma_h+E_{hrl} \quad (2)$$

In this model, experimental hybrids and founder hybrids were considered jointly. Y^*_{hrl} is the performance corrected by the spatial field effects of repetition r of hybrid h in trial l. μ is the intercept, λ_l is the fixed effect of trial l, γ_h is the fixed genetic effect of hybrid h. E_{hrl} is the error term of environment l, with $E_{hrl}\sim\mathcal{N}(0,\sigma^2_{E_l})$ iid within trial and independent between trials. All genomic predictions were performed on the ls-means thus obtained.

The founder hybrids were used as a reference for the initial unselected population. The observed genetic gain was computed as the difference between the performances of the founder hybrids and the experimental hybrids (either the G0S or the G1 hybrids). Then, we compared the observed to the predicted genetic gain estimated from the genomic predictions of hybrid values trained on the G0_F-1H.

Pedigree based Best Linear Unbiased Prediction (PBLUP) model

A prediction model based on the pedigree information (PBLUP) model was implemented and used as a benchmark compared to the GBLUP models. The model was:

$$\boldsymbol{y} = \mathbf{1_n}.\mu + \boldsymbol{Zg} + \boldsymbol{E}, \quad (3)$$

where **y** is the vector of ls-means of the n phenotyped hybrids, $\mathbf{1_n}$ is a vector of n ones and μ is the intercept. $\boldsymbol{g}$ is the vector of random hybrid effects, with $\boldsymbol{g} \sim \mathcal{N}(0, \boldsymbol{K}\sigma_h^2)$ where K is the pedigree kinship matrix computed on the hybrid population considering the founder lines as the base generation. The pedigree kinship matrix was computed with the recursive method presented in Mrode and Thompson (2005) using the AHGmatrix R-package (Amadeu et al. 2016). σ_h^2 is the hybrid variance. $\boldsymbol{Z}$ is the corresponding incidence matrix. $\boldsymbol{E}$ is the vector of error terms, with $\boldsymbol{E} \sim \mathcal{N}(0, \boldsymbol{I}\sigma_E^2)$. The random effects are assumed to be independent.

Genomic Best Linear Unbiased Prediction (GBLUP) models

Several GBLUP models were tested to evaluate the predictive ability within the G1 cycle. Two types of models can be distinguished: the GCA-models, which decompose the hybrid genetic effect into its parental GCAs and its SCA components and the G-models, which consider genetic effects defined based on the hybrid marker genotypes.

GCA.1-model. Two GBLUP models were implemented for genomic predictions depending on the TRS design (factorial or tester). The model implemented on the factorial designs including SCA effects was:

$$\boldsymbol{y} = \mathbf{1_n}.\mu + \boldsymbol{Z_d}\,\boldsymbol{g_{GCA_d}} + \boldsymbol{Z_f}\boldsymbol{g_{GCA_f}} + \boldsymbol{Z}\boldsymbol{g_{SCA_{df}}} + \boldsymbol{E}, \quad (4.1)$$

where $\boldsymbol{y}$ is the vector of ls-means of the n phenotyped hybrids, $\mathbf{1_n}$ is a vector of n ones and μ is the intercept. $\boldsymbol{g_{GCA_f}}$ (respectively $\boldsymbol{g_{GCA_d}}$) is the vector of random GCA effects of the n_f flint parental lines (respectively n_d dent lines), with $\boldsymbol{g_{GCA_f}} \sim \mathcal{N}\left(0, \boldsymbol{K_{GCA_f}}\sigma^2_{GCA_f}\right)$ (respectively $\boldsymbol{g_{GCA_d}} \sim \mathcal{N}\left(0, \boldsymbol{K_{GCA_d}}\sigma^2_{GCA_d}\right)$) where $\boldsymbol{K_{GCA_f}}$ (respectively $\boldsymbol{K_{GCA_d}}$) is the genomic relatedness matrix between the flint lines (respectively dent lines). The kinship matrix was computed for all the flint (dent) parental lines following method 1 from VanRaden (2008). $\sigma^2_{GCA_f}$ and $\sigma^2_{GCA_d}$ are the flint and dent GCA variances. $\boldsymbol{g_{SCA_{df}}}$ is the vector of SCA random effects of the n hybrids, accounting for the interactions between the flint and dent parental lines, with $\boldsymbol{g_{SCA_{df}}} \sim \mathcal{N}\left(0, \boldsymbol{K_{SCA_{df}}}\sigma^2_{SCA_{df}}\right)$ where $\boldsymbol{K_{SCA_{df}}}$ is the SCA kinship matrix of the hybrids (phenotyped or not) and $\sigma^2_{SCA_{df}}$ the SCA variance. The coefficient of the SCA kinship between two flint-dent hybrids produced from crossing parental lines *i* to *j* and parental lines *i'* to *j'* were computed as the product between the flint GCA kinship coefficient between lines *i* and *i'* and the dent GCA kinship coefficient between lines *j* and *j'* (Stuber and Cockerham 1966). $\boldsymbol{Z_d}$, $\boldsymbol{Z_f}$ and $\boldsymbol{Z}$ are the corresponding incidence matrices. $\boldsymbol{E}$ is the vector of error terms, with $\boldsymbol{E} \sim \mathcal{N}(0, \boldsymbol{I}\sigma_E^2)$. The different random effects are assumed to be independent.

The model implemented on the G0_T-F was:

$$\boldsymbol{y} = \mathbf{1}_{\boldsymbol{n}}.\mu + \boldsymbol{Xv} + \boldsymbol{Z_f g_{GCA_f}} + \boldsymbol{Z g_{SCAt}} + \boldsymbol{E}, \quad (4.2)$$

where $\boldsymbol{y}$ is the vector of ls-means of the n phenotyped hybrids, $\mathbf{1}_{\boldsymbol{n}}$ is a vector of n ones and μ is the intercept. $\boldsymbol{v}$ is the vector of fixed effects of the two dent testers. $\boldsymbol{g_{GCA_f}}$ is the vector of random GCA effects of the n_f flint parental lines, with $\boldsymbol{g_{GCA_f}} \sim \mathcal{N}\left(0, \boldsymbol{K_{GCA_f}}\sigma^2_{GCA_f}\right)$ where $\boldsymbol{K_{GCA_f}}$ is the genomic relatedness matrix between the flint lines and $\sigma^2_{GCA_f}$ is the flint GCA variance. $\boldsymbol{g_{SCAt}}$ is the vector of random effects of the interaction between the flint line and the dent testers, with $\boldsymbol{g_{SCAt}} \sim \mathcal{N}\left(0, \boldsymbol{I_2} \otimes \boldsymbol{K_{GCAf}}\,\sigma^2_{SCAt}\right)$ where σ^2_{SCAt} is the SCA variance. The kinship matrix was computed for all the flint parental lines following method 1 from VanRaden (2008). $\boldsymbol{X}$, $\boldsymbol{Z_f}$ and $\boldsymbol{Z}$ are the corresponding incidence matrices. $\boldsymbol{E}$ is the vector of error terms, with $\boldsymbol{E} \sim \mathcal{N}(0, \boldsymbol{I}\sigma^2_E)$. The different random effects are assumed to be independent. The same model was adapted and implemented on the G0_T-D.

GCA.2-model. This GCA-model was defined following González-Diéguez et al. (2021), where the genetic effect is defined according to gamete origin. The fullest model for the factorial TRS was:

$$\boldsymbol{y} = \mathbf{1}_{\boldsymbol{n}}.\mu + \boldsymbol{Z_d g_{A_d}} + \boldsymbol{Z_f g_{A_f}} + \boldsymbol{Z g_D} + \boldsymbol{Z_d g_{AA_d}} + \boldsymbol{Z_f g_{AA_f}} + \boldsymbol{Z g_{AA_{df}}} + \boldsymbol{E}, \quad (6)$$

where $\boldsymbol{y}$ is the vector of ls-means of the n phenotyped hybrids, $\mathbf{1}_{\boldsymbol{n}}$ is a vector of n ones and μ is the intercept. $\boldsymbol{g_{A_f}}$ and $\boldsymbol{g_{A_d}}$ are the vectors of the random additive effect from the flint and dent parental lines with $\boldsymbol{g_{A_f}} \sim \mathcal{N}\left(0, \boldsymbol{K_{A_f}}\sigma^2_{A_f}\right)$ and $\boldsymbol{g_{A_d}} \sim \mathcal{N}\left(0, \boldsymbol{K_{A_d}}\sigma^2_{A_d}\right)$, respectively. $\boldsymbol{g_D}$ is the vector of random dominance effect with $\boldsymbol{g_D} \sim \mathcal{N}(0, \boldsymbol{K_D}\,\sigma^2_D)$, $\boldsymbol{g_{AA_f}}$ is the vector of random additive-by-additive epistatic effect within the flint (resp. dent) population with $\boldsymbol{g_{AA_f}} \sim \mathcal{N}\left(0, \boldsymbol{K_{AA_f}}\sigma^2_{AA_f}\right)$ (resp. $\boldsymbol{g_{AA_d}} \sim \mathcal{N}\left(0, \boldsymbol{K_{AA_d}}\sigma^2_{AA_d}\right)$) and $\boldsymbol{g_{AA_{df}}}$ is the vector of random additive-by-additive epistatic effect across the flint and dent populations $\boldsymbol{g_{AA_{df}}} \sim \mathcal{N}\left(0, \boldsymbol{K_{AA_{df}}}\sigma^2_{AA_{df}}\right)$. $\boldsymbol{K_{A_f}}$, $\boldsymbol{K_{A_d}}$, $\boldsymbol{K_D}$, $\boldsymbol{K_{AA_f}}$, $\boldsymbol{K_{AA_d}}$ and $\boldsymbol{K_{AA_{df}}}$ are respectively the flint additive, dent additive, dominance, additive-by-additive epistasis within the flint population, additive-by-additive epistasis within the dent population and the additive-by-additive epistasis across populations genomic relatedness matrices computed following González-Diéguez et al. 2021. $\sigma^2_{A_f}$, $\sigma^2_{A_d}$, σ^2_D, $\sigma^2_{AA_f}$, $\sigma^2_{AA_d}$ and $\sigma^2_{AA_{df}}$ are the corresponding variances. $\boldsymbol{Z_f}$, $\boldsymbol{Z_d}$ and $\boldsymbol{Z}$ are the incidence matrices. $\boldsymbol{E}$ is the vector of error terms, with $\boldsymbol{E} \sim \mathcal{N}(0, \boldsymbol{I}\sigma^2_E)$. The different random effects are assumed to be independent.

G-model. This model was defined by Vitezica et al. (2017). It is based on the hybrid genotypes and does not account for the gamete origin (flint and dent parental origins). The fullest model considered for the factorial TRS was:

$$\boldsymbol{y} = \mathbf{1}_{\boldsymbol{n}}.\mu + \boldsymbol{Zg_A} + \boldsymbol{Zg_D} + \boldsymbol{Zg_{AA}} + \boldsymbol{E}, \quad (5)$$

where $\boldsymbol{y}$ is the vector of ls-means of the n phenotyped hybrids, $\mathbf{1}_{\boldsymbol{n}}$ is a vector of n ones and μ is the intercept. $\boldsymbol{g_A}$ is the vector of the random additive effect with $\boldsymbol{g_A} \sim \mathcal{N}(0, \boldsymbol{K_A}\,\sigma_A^2)$, $\boldsymbol{g_D}$ is the vector of random dominance effect with $\boldsymbol{g_D} \sim \mathcal{N}(0, \boldsymbol{K_D}\,\sigma_D^2)$ and $\boldsymbol{g_{AA}}$ is the vector of random additive-by-additive epistasis effect with $\boldsymbol{g_{AA}} \sim \mathcal{N}(0, \boldsymbol{K_{AA}}\,\sigma_{AA}^2)$. $\boldsymbol{K_A}$, $\boldsymbol{K_D}$ and $\boldsymbol{K_{AA}}$ are respectively the additive, dominance and additive-by-additive epistasis genomic relatedness matrices computed following Vitezica et al. 2017. σ_A^2, σ_D^2 and σ_{AA}^2 are the corresponding variances and $\mathbf{Z}$ is the incidence matrix. $\boldsymbol{E}$ is the vector of error terms, with $\boldsymbol{E} \sim \mathcal{N}(0, \boldsymbol{I}\sigma_E^2)$. The different random effects are assumed to be independent.

Prediction scenarios

We defined three prediction scenarios to achieve three objectives: (i) assess the predictive ability of GS in the new generation and compare different GS models, (ii) evaluate the efficiency of a factorial design for predictions across breeding cycles and compare it to the tester designs, and n (iii) investigate the benefit of different strategies to update the factorial TRS across cycles and optimize it using the CDmean to predict the new generation.

In Scenario 1, we evaluated the predictive ability within the new generation (G1 hybrids) to serve as a reference and compared the efficiency of several GS models. Cross-validations within the G1 hybrids were performed by training the GS model on 354 G1 hybrids (four-fifth) to predict the remaining 88 G1 hybrids (one-fifth). This process was repeated a hundred times. We compared three types of GBLUP models, namely the GCA.1-model, G-model, and GCA.2-model, to a benchmark PBLUP model. The GCA.1-model involved two nested models, with or without the SCA effect. For the GCA.2- and G-models, several nested models were tested by adding successively dominance and additive-by-additive genetic effects to additive effects. See **Table 2** for the summary of all tested models.

Scenario 2 evaluated the efficiency of training a GBLUP model (GCA.1) on the G0 generation to predict the next one (G1). In Scenario 2a, we evaluated the efficiency of the incomplete factorial TRS (G0_F-1H) to predict G1 hybrids. We assessed the prediction stability across breeding cycles by comparing the predictive abilities obtained for the G1 hybrids to the one obtained for the G0S hybrids evaluated in the same experiments.

The GCA.1 model was used to perform predictions. In Scenario 2b, we compared the efficiency of factorial and tester TRSs from the G0 cycle to predict the G1 cycle. The GCA.1 models (4.1) or (4.2) were trained on the G0_F-4H (363 hybrids) or the tester designs (360 hybrids) to predict G1 hybrids. We investigated the impact of the TRS on hybrid selection through the correlation between the GCA BLUPs predicted using the factorial and the ones obtained using the tester designs. In addition, to compare the similarity of selection between the different approaches (based on phenotypic evaluations (ls-means) or genomic predictions (BLUPs) trained on the factorial or the tester designs), the coincidence of selection was computed for each trait. For each pair of approaches, it corresponds to the percentage of common hybrids that would be selected by the two approaches at a given selection rate (%). This coincidence of selection was computed for different selection rates. As in Lorenzi et al. 2022, we sampled hybrids in the tester designs to evaluate the impact of the number of testers used in the TRS. In this Scenario 2b', each tester TRS was composed of 180 hybrids produced by crossing in each group: (i) 90 lines to one tester (180 lines in total): since there were two testers in each group, there were four possible tester combinations, referred to as 1T-180H-180L- followed by the names of the testers, (ii) 45 lines to one tester and the 45 other lines to the other tester, referred to as 2T-180H-180L, (iii) the same 45 lines to two testers referred to as 2T-180H-90L. We compared these tester TRS to a factorial TRS by sampling 180 hybrids from the G0_F-4H in a random and balanced manner between families to maximize the number of lines. This factorial TRS comprised 180 hybrids representing 170 lines (one line contributed to 2.1 hybrids on average) and was called F-180H-170L. In Scenario 2b', all the TRSs were sampled ten times except for the one-tester designs that were sampled only once.

Scenario 3 investigated TRS optimization across breeding cycles. In Scenario 3a, we evaluated the benefit of updating the TRS across cycles by adding either G0S or/and G1 hybrids to the initial G0R TRS. Several TRSs were sampled and compared to cross-validations within the G1 hybrids. To assess the benefit of adding G0S hybrids to the initial G0R TRS, we compared TRSs only composed of G0R hybrids with the same TRSs to which 132 G0S hybrids from the G0_F-4H design were added. To evaluate the benefit of updating G0 TRSs with G1 hybrids, we added from 0 to 354 randomly sampled G1 hybrids to G0 TRSs. One-fifth of the G1 hybrids (88 hybrids) were predicted using the GCA.1 model. The mean predictive ability over 100 replicates was computed for each TRS. In Scenario 3b, our objective was to maximize the predictive ability of the G1 hybrids by optimizing *a priori* the G1 hybrid subset used to update the initial G0 TRS using only G1 line genotypes. We considered the CDmean proposed by Rincent et al. (2012). We used a heritability of 0.7, corresponding to the average heritability of our traits, to compute the value for the shrinkage parameter λ and the additive covariance kinship between

hybrids defined by Vitezica et al. (2017). Two optimization strategies were considered and compared to random sampling. For both strategies, we optimized the mean of the CD of contrasts between each non-phenotyped G1 hybrid (PS) and the mean of the G1 hybrids. In the first strategy (CDmean1), the G1 hybrid set was optimized without considering the marker information on the G0 hybrids. The additive kinship considered to compute expected CDmeans only included the 442 G1 hybrids. In the second strategy (CDmean2), the optimization of the G1 was performed by also considering information on the G0 hybrids: the additive kinship was computed for all 1802 hybrids from both generations (1360 G0+ 442 G1). The procedure was performed in both scenarios with four sampling sizes for the G1 hybrids (50,100, 200, and 300) and replicated a hundred times each. For each optimized set, all G0 hybrids plus the chosen G1 hybrids were used as TRS to predict the remaining G1 hybrids, used as VS. Predictions were performed using the GCA.1 model.

Predictive ability and statistical tests

In all scenarios, the predictive ability was computed as Pearson's correlation between predicted hybrid values and hybrid ls-means. Different statistical tests were performed depending on the scenario to test the significance of differences between predictive abilities. In Scenario 1, paired t-tests were performed with a risk level α=0.05, and a Bonferroni correction (multiple comparison correction) was applied per trait. In Scenario 2b, Williams tests (Williams 1959) were performed with a risk level α=0.05 using the "r.test" function of the psych R-package (Revelle 2021). In Scenario 2b', t-tests with a risk level α=0.05 were performed, and a Bonferroni correction was applied per trait. For all scenarios, computations were performed in the R statistical environment (R Core Team 2020), and models were fitted using the "MM4LMM" R-package (Laporte and Mary-Huard 2020; Laporte et al. 2022).

Results

For clarity purposes, results on the four main traits of interest (DMY, DMC, DtSilk, and MFU) are presented in the following. The results on the 11 studied traits are shown in supplementary materials.

Variance components and broad-sense heritability (H^2) at the phenotypic level without marker information

Broad-sense heritabilities (H^2) were medium to high (**Table 3**). They ranged from 0.56 (MFU) to 0.93 (DtSilk) for G0S hybrids and from 0.62 (MFU) to 0.94 (DtSilk) for G1 hybrids. Large and significant genetic variances were observed for all traits (**Table 3, Table S1**) with no clear differences between G0S and G1 hybrids. The main part of the genetic variance was due to GCA. The proportion of genetic variance due to SCA ranged from 0% (DMC)

to 30% (MFU) for the G0S and from 0% (DMY) to 10% (DMC) for the G1. $\sigma^2_{GCA_f}$ was always larger than $\sigma^2_{GCA_d}$ except for G0S hybrids for DMC. Non-null GCA by trial variances were observed for G0S and G1 hybrids, but were lower than the GCA variances. For the G0S, σ^2_{SCA} was larger than σ^2_{SCAxE} for all traits. For G1 hybrids, σ^2_{SCAxE} was larger than σ^2_{SCA}, except for DtSilk.

Ls-means and genetic gain

On average, G0S and G1 hybrids performed similarly (**Table 4**). Compared to the 16 founder hybrids, which are representative of the performance of the unselected G0 hybrids (G0R), G0S and G1 hybrids showed a gain in performance for DMY (+1.55 t/ha for G0S and +1.52 t/ha for G1). This gain was associated with a later DtSilk (+1.83 days for G0S and +1.90 days for G1), a lower DMC (-0.77% for G0S and -0.67% for G1), and a lower MFU (-2.11 MFUx10²/kg for G0S and -2.15 MFUx10²/kg for G1). The observed genetic gain for DMY was similar to the predicted one based on the genomic predictions trained on the G0_F-1H design. However, for DMC, DtSilk, and MFU, the observed response to selection was higher in absolute value than the predicted one.

Scenario 1- Predictive ability within the G1 cycle and GS model comparison

We assessed the predictive ability in the new breeding cycle using cross-validations among G1 hybrids (**Fig.2**). GBLUP predictive abilities of the new generation were high for all traits, ranging from 0.63 (DMY) to 0.76 (DtSilk) when considering the best GBLUP model. All GBLUP models significantly outperformed the PBLUP model (differences between the worst GBLUP model and the PBLUP ranged from 0.07 (DMY) to 0.11 (MFU)). Differences among GBLUP models were sometimes significant but minor (<0.01), showing that models were equivalent and that adding non-additive effects had little effect.

Scenario 2- Efficiency of a factorial TRS for predictions across breeding cycles and comparison with tester TRSs

In Scenario 2a, we compared the ability of the G0_F-1H TRS to predict the same generation (G0S hybrids) or the new generation (G1 hybrids). Predictive abilities were high for all traits (ranging from 0.56 for DMY to 0.67 for DtSilk for G1 hybrids and from 0.60 for DMY to 0.75 for MFU for G0S hybrids) (**Fig. 3**). As expected, predictive abilities were higher for G0S hybrids (hybrids from the same generation as the TRS hybrids) than for G1 hybrids for all traits. Lower predictive abilities were obtained when training on the G0_F-1H compared to those obtained by cross-validations within G1 hybrids (**Fig. 3).**

In Scenario 2b, we compared predictive abilities obtained using either the G0_F-4H (363 hybrids) or the G0 tester designs (360 hybrids) as TRS to predict all G1 hybrids (442 hybrids) (**Fig. 4**). They ranged from 0.59 (DMY and MFU) to 0.70 (DtSilk) when training on the G0_F-4H and from 0.60 (MFU) to 0.69 (DtSilk) when training on the G0 tester designs. Across the 11 traits, training on the G0_F-4H or the G0 tester designs gave equivalent predictive abilities except for four traits: the G0_F-4H design significantly outperformed the G0 tester designs for DMC and PH, and the G0 tester designs significantly outperformed the G0_F-4H design for DMY and CELL (**Fig. S1**). The GCA BLUPs of the G1 lines predicted using the G0_F-4H or the G0 tester designs as TRS were highly correlated. They ranged from 0.85 (DMC) to 0.94 (MFU) for the dent G1 lines and from 0.84 (DMY) to 0.94 (DtSilk) for the flint G1 lines, and from 0.87 (DMY, DMC) to 0.91 (DtSilk) for G1 hybrids (**Table S4**). The coincidence of selection for genomic predictions between the factorial and the tester TRS of the top 5% of hybrids was 52% for DMY, 61% for DMC, 65% for DtSilk, and 39% for MFU (**Fig. S2**), which indicates that the single-cross hybrid sets selected by the two approaches are not identical. To assess if one of the two approaches identified a higher proportion of the best-phenotyped hybrids, we compared the proportion of the top 5% hybrids identified based on the factorial or tester TRS to the top 5% phenotyped hybrids. For DMY, the major trait of interest in our study, the factorial design identified a higher proportion of the best-phenotyped hybrids compared to the tester designs.

In Scenario 2b', we investigated the efficiency of different G0 tester design compositions to predict G1 hybrids (442 hybrids) at the same number of hybrids (180) and compared them with a factorial design of same size (**Fig. 5**). Predictive abilities varied between the four one-tester TRS ranging from 0.005 (DMC) to 0.048 (DMY), and the best one-tester TRS depended on the trait. The best two-tester TRS maximized the number of evaluated candidate lines by crossing more lines each to a different tester (2T-180H-180L) and usually outperformed the worst one-tester TRS. The F-180H-170L factorial TRS was equivalent to or outperformed the tester TRS except for DMY. On average, over the 11 traits, the F-180H-170L TRS gave the highest predictive abilities (**Fig.5**, **Fig. S3**).

Scenario 3a- Benefit of updating the factorial TRS across breeding cycles

To evaluate the benefit of updating the TRS across breeding cycles, four TRS strategies were evaluated based on their ability to predict G1 hybrids: (i) training on G0R only (G0_F-1H, G0R_F-4H or G0_F-1H+G0R_F-4H), (ii) training on G0R plus 132 hybrids between G0 selected lines (G0S), (iii) training on G0R plus a subset *m* of hybrids from the new generation (G1 hybrids), and (iv) training on G0R plus 132 G0S hybrids and *m* G1 hybrids,

with *m* ranging from 1 to 354 (**Fig. 6**). The four TRS strategies were also compared to cross-validations within the G1 hybrids. The best G0R TRS (G0_F-1H+G0R_F-4H) was also the largest one (1183 hybrids) with predictive abilities ranging from 0.69 (MFU) to 0.76 (DMC and DtSilk), which were equivalent or higher than the ones obtained with a TRS composed of 354 G1 hybrids.

Adding 132 G0S hybrids to the initial G0R TRSs (G0_F-1H, G0R_F-4H, or G0_F-1H+G0R_F-4H) increased predictive abilities (with a gain on average of 0.10 for DMY, 0.14 for DMC, 0.16 for DtSilk and 0.05 for MFU). The largest gain in predictive ability was observed for the G0R_F-4H TRS, which was also the smallest G0R TRS (232 hybrids), with gains ranging from 0.08 (MFU) to 0.39 (DtSilk). There was always a gain in predictive ability when adding G1 hybrids to the TRS, whether composed of G0R or of G0R and G0S hybrids. As expected, the gain increased with the number of G1 hybrids included in TRS. Adding 354 G1 hybrids to G0R TRSs, increased predictive abilities on average by 0.13 for DMY, 0.20 for DMC, 0.21 DtSilk, and 0.12 for MFU. For TRSs comprising G0R and G0S hybrids, adding 354 G1 hybrids led to smaller gains (gain not exceeding 0.07 for MFU). The largest increase in predictive abilities when updating the TRS with G1 hybrids was obtained with the smallest initial G0R TRS (G0R_F-4H). It is interesting to note that TRSs composed of G0 and 354 G1 hybrids always outperformed prediction accuracies obtained with 354 G1 (cross-validations within G1), illustrating the benefit of keeping information from the previous generation in the TRS.

From **Fig 6**, it is possible to estimate the number of G1 hybrids to add to the initial G0R TRSs to achieve similar predictive abilities to the ones obtained when adding 132 G0S hybrids. For example, for DMY and the G0R_F-4H initial TRS, adding 132 G0S was equivalent to adding around 170 G1 hybrids. For all initial G0R TRSs, the number of G1 hybrids to include to be more efficient than 132 G0S hybrids was higher than 132 for all traits except MFU.

Scenario 3b- Optimization of the composition of the factorial TRS for G1 hybrid predictions

The G1 hybrid set to add to the existing G0 TRS (1360 G0 hybrids) was optimized using the CDmean following two strategies, and the results were compared to a TRS obtained from random sampling (**Fig. 7**). In the first strategy (CDmean1), the G1 hybrid set was optimized without considering the information from the G0 hybrids, whereas in the second strategy (CDmean2), the information from the G0 hybrids was considered. For all traits and all sampling sizes, the best CDmean strategy gave higher or at least equivalent predictive abilities compared to random sampling except for DMC for a sampling size of 300. The maximum gains were 0.03 for DMY, 0.01 for DMC, 0.03 for DtSilk, and 0.02 for MFU, depending on the sampling size. Across the 100 replicates, the variance of the

predictive abilities was always lower using the CDmean (1 or 2) than the random sampling. The CDmean1, which does not consider G0 hybrid information to optimize the G1 hybrid set included in the TRS, outperformed the CDmean2 except for small sampling sizes (size 50 and 100 for DMY, size 50 for DMC, DtSilk, and MFU).

Discussion

SCA variance and its importance in hybrid breeding

The SCA variance estimated in the G1 generation was small or equal to zero (**Table 3**, **Table S2**). Small SCA variance was expected in hybrids produced by crossing lines from divergent populations (Reif et al. 2007). The estimated SCA percentage decreased for all traits from G0 to G1 hybrids (**Table S3**). The precision of SCA variance estimation in our experiment is limited and does not allow us to draw a final conclusion on this evolution. However, one possible explanation for the decrease in SCA variance we observed is that the recurrent reciprocal selection increased the divergence between groups (as also observed by Gerke et al. (2015)) and, as a result, decreased the SCA variance in the flint-dent single-cross hybrids (consistent with theoretical expectations from Reif et al. (2007) and Legarra et al. (2023)).

Genetic gain after selection based on genomic predictions trained on a sparse factorial design

The population was selected for an index combining yield performance (DMY), dry matter content (DMC), and digestibility (MFU) based on genomic predictions. We successfully improved the mean performance of the new generation for DMY, but there was a decrease for MFU and DMC (**Table 4**), which was higher than expected. The negative correlation (-0.53) between DMY and MFU that was observed based on phenotypic data in the G0 generation (G0_F-1H) (**Fig. S4**) certainly explains the difficulty of improving both traits simultaneously. This was consistent with results found by Barrière and Emile (2000) and Surault et al. (2005), who also reported a negative correlation of -0.5 between these traits for maize silage. To maintain a stable level of DMC and improve MFU in the new generation, higher weights relative to DMY should have been put on these traits in the index calculation.

The genetic gain predicted by the GBLUP model trained on the G0_F-1H design was similar to the observed genetic gain for DMY. This illustrates the efficiency of GS models in predicting GCA values based on a sparse factorial TRS and confirms the results found by Seye et al. (2020) using simulations and Lorenzi et al. (2022) on the G0 generation.

Predictive ability in the new generation and comparison of different GS models

In Scenario 1, we evaluated the predictive ability within G1 hybrids and compared different prediction models. All models gave high predictive abilities, with the lowest reaching 0.66 (for DtSilk with the PBLUP model). The high predictive ability of the PBLUP model indicates that family structure alone could predict part of hybrid performances. However, GBLUP models always outperformed the PBLUP, confirming the efficiency of GBLUP to predict the mendelian sampling within a family, which is of main interest for breeding. Different GBLUP models were tested. Differences were sometimes significant but always small (<0.01). Including non-additive genetic effects had little or no effect on predictive abilities, which was also reported in studies using data from inter-heterotic group hybrids (Bernardo 1994; Schrag et al. 2006, 2018; Maenhout et al. 2010; Vitezica et al. 2017; González-Diéguez et al. 2021; Lorenzi et al. 2022). Note that the new SCA kinship formula proposed by González-Diéguez et al. (2021) used in model GCA.2 did not improve predictive abilities compared to the one used in the GCA.1 model. This was also observed by Lorenzi et al. (2022) for genomic predictions within the G0 generation. In their simulations, Seye et al. (2020) found an advantage of including SCA in prediction models when SCA explains about 23% of the genetic variance. The small SCA variances estimated in our experimental design are consistent with the fact that including non-additive effects did not improve prediction accuracies.

Assuming a single additive hybrid genetic effect (G models) or additive genetic effects defined according to the allele origin (GCA models) was equivalent in terms of quality of prediction for hybrid performance. This was surprising considering the large differences in GCA variances observed between the two groups and the detection of group-specific QTLs in the G0_F-1H design (Giraud et al. 2017b). The equivalence in terms of prediction accuracy between the G and GCA models was also shown in hybrid populations by Technow et al. (2014), González-Diéguez et al. (2021), and Alves et al. (2019). Even if the GCA model did not outperform the G model, it makes it possible to estimate parental line values and thus select the parental lines of the next cycle, which is less straightforward with a G model. We kept the GCA.1 model for the following genomic prediction scenarios for these reasons.

Portability of genomic predictions trained on a sparse factorial across breeding cycles

In Scenario 2a, we trained the GS model on the G0_F-1H to predict G0S and G1 hybrids, allowing us to evaluate the predictive ability across cycles and environments. We obtained high predictive abilities for G0S hybrids, which illustrates the ability of the GS model trained on the G0_F-1H design to predict the performances of a new set of hybrids between selected lines in new environments. This confirms previous results (Lorenzi et al.

2022), which considered another set of G0S hybrids evaluated in the 2016-2017 G0_F-4H trials. We observed lower predictive abilities for G1 compared to G0S hybrids. Note that G0S and G1 hybrids were evaluated in the same environments, therefore, the decrease in predictive ability is not attributable to an environmental effect. A decrease in prediction accuracy when generations differ between the TRS and PS was reported in simulations (Pszczola and Calus 2016; Seye et al. 2020) and experimental studies on hybrids (on sugar beet by Hofheinz et al. 2012; on barley Sallam et al. 2015 and Michel et al. 2016 and on maize by Wang et al. 2020). This decrease is expected as selection modifies allele frequencies along generations, and recombination events modify marker-QTL linkage disequilibrium. Allelic frequencies are identical in G0S and G1 hybrids since G1 lines are the unselected progeny of G0S lines. Thus, the observed decrease is due to the recombination events. Still, predictive abilities remained high, highlighting the efficiency of the GS model trained on the G0_F-1H design in decorrelating the contributions from each parental line to predict their GCAs, the GCAs of their progeny, and therefore the hybrid values across breeding cycles.

Efficiency of factorial compared to tester TRSs for predictions across the breeding cycle

In Scenario 2b, we compared, for the same number of hybrids and lines, the efficiency of the factorial and tester TRSs to predict hybrids across generations. A previous study using the same TRSs to predict the G0 generation showed slightly higher predictive abilities using the factorial compared to the tester TRSs (Lorenzi et al. 2022). This advantage decreased when predicting the new generation (G1). This is in accordance with results from simulations based on a similar design (Seye et al. 2020), which showed that the advantage of the factorial over the tester TRSs decreases across breeding cycles if the TRS was not updated. When investigating several tester designs composition (Scenario 2b'), we showed that the best strategy was always to use more testers while maximizing the number of candidate lines, a strategy comparable to using a sparse factorial design.

Benefit of updating the factorial TRS along breeding cycles

Once inbred lines from a new generation (G1) are available and can be genotyped, a key issue is to predict the best new hybrid combinations between them to prioritize hybrid production and evaluation. There are two possible situations, depending on the availability of phenotypes of a subset of hybrids from the new breeding cycle (G1 hybrids). When G1 phenotypes are available, they can be used to calibrate prediction equations. We showed the benefit of combining this information with historical data from G0 hybrids compared to using G1 phenotypes alone (**Fig.6**). Several studies also reported similar results (Jannink 2010; Denis and Bouvet 2013; Neyhart et al. 2017). Among the historical data, hybrids between the lines selected to generate the new generation (G0S) are the

most related to the G1 generation. We showed that even when G0S and G1 hybrids were already in the TRS, there was still a benefit of including hybrids between unselected lines from previous generations (G0R hybrids). This last result aligns with results found by Neyhart et al. (2017) and Brandariz and Bernardo (2018), showing that when constructing a TRS, one must consider keeping hybrids produced between unselected lines to maintain high prediction accuracy. Additionally, when including data from the two generations (G0 and G1) in the TRS, we also included TRS hybrids evaluated in different years and environments. This reduced the impact of genotype-by-environment interactions and, as a result, increased prediction accuracy. Similar results have been obtained by Auinger et al. (2016).

In the second situation, where G1 hybrids phenotypes are not yet available, we showed that using only historical data in the TRS can provide good prediction accuracies (**Fig.6**). We evaluated the benefit of producing and phenotyping additional data to update the historical (G0) TRS, particularly the benefit of adding G0S hybrids. G0S are single-crosses between the G0 lines selected to be the parents of the G1 generation, so including these hybrids increases the relationship between the TRS and the G1 PS. We compared G0S and G1 hybrids for their efficiency to update the TRS. Predictive abilities obtained with the 132 G0S hybrids were reached when adding a similar number or more G1 hybrids (**Fig.6**). This indicates that for a fixed number of hybrids, using G0S hybrids was equivalent to or slightly better than using G1 hybrids for updating the TRS. Once the best candidate lines are selected to become the parental lines (corresponding to G0S lines) for the subsequent breeding cycle, but hybrids from the new cycle (G1 hybrids) are not yet available, it is beneficial to phenotype new hybrid combinations between the selected lines to update the TRS. Several entangled factors can explain the result: (i) the increased TRS size, (ii) the increased relationship between the TRS and PS, and (iii) the increased number of years and environments in the data used as TRS (see reviews by Isidro y Sánchez and Akdemir 2021 and Rio et al. 2022). Adding G0S hybrids is a way to accumulate information on the hybrid values (GCAs) of the selected lines in different environments, which is helpful to predict the hybrid values of their progeny. Our results also show that even if G0S hybrids are added to the TRS, it is still interesting to add performances of G1 hybrids to the TRS when these become available, as it increases the genotypic relatedness between the TRS and the PS (**Fig.6**).

Optimization of the G1 hybrid set to phenotype to update the TRS

In Scenario 3b, we optimized the G1 hybrid set used to update the initial G0 TRS. The G1 hybrid set was optimized based on the CDmean computed considering (CDmean 2) or not (CDmean 1) the information from the initial G0 TRS (G0R+G0S hybrids) (**Fig.7**). As expected, optimizing the TRS using the CDmean (CDmean 1 or CDmean 2)

instead of random sampling increased our predictive abilities in most of the cases. This was also reported in numerous other studies (Rincent et al. 2012, 2017; Isidro et al. 2015; Akdemir et al. 2015; Mangin et al. 2019; Isidro y Sánchez and Akdemir 2021; Kadam et al. 2021). Interestingly, we observed more stable predictions abilities across replicates using the CDmean, than with random sampling. It has to be noted that in this optimization process, we used CDs computed assuming a single additive genetic effect despite using a GCA/SCA prediction model for our predictions. We could have included non-additive effects in the computation of the CDmean, as done by Momen and Morota (2018) and Fristche-Neto et al. (2018). However, these authors did not find a clear benefit of accounting for dominance in the CDmean computation, and we did not see any advantage of including the dominance effect in our prediction models. For these reasons, we do not expect that adding dominance in the CDmean computation would have had a positive impact.

It was surprising to us that the CDmean 1 (which does not consider information from the G0 TRS hybrids) outperformed the CDmean 2 (which considers the G0 information). The CDmean 1 likely selected G1 hybrids that were representative of the whole range of G1 hybrids. In contrast, since the hybrids between the G0S parental lines of the G1 were already in the TRS, the CDmean 2 likely maximized the diversity of the TRS by favoring G1 hybrids genetically distant from the G0 hybrids. The CDmean 2 assumed that G0 and G1 hybrids were evaluated in the same environments, which was not true. As a consequence, some of the G0 hybrids may not have been as informative to predict the G1 hybrids as they seemed, based on the genomic relationship matrix. This may explain why CDmean 2 did not outperform CDmean 1. To compute the CDmean, we could have considered each environment as a different trait and used the correlation value between the two environments, as suggested by Ben-Sadoun et al. (2020). Rio et al. (2022a) showed the benefit of using such multi-environmental CDs to optimize the allocation of individuals in trial networks, and this could have been extended to multigeneration TRS optimization. In practice, one cannot know in advance the correlation between the environments where the previous generation was evaluated and those where the new generation will be evaluated. One solution might be to use historical data to estimate the magnitude of correlations that can be expected between years and use this value when computing the expected multi-environment CD.

Conclusions

Our study confirms the efficiency of combining genomic predictions and sparse factorial TRS to predict candidate lines GCAs and hybrid values across breeding cycles. Genomic prediction accuracy was high and increased when updating the TRS by incorporating performances of hybrids between selected lines from the previous generation

607 and potentially hybrids from the new generation. When incorporating hybrids from the new generation, choosing
608 them based on a criterion such as the CDmean was beneficial.

609 **References**

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

612 Alves FC, Granato ÍSC, Galli G, et al (2019) Bayesian analysis and prediction of hybrid performance. Plant
613 Methods 15:14. https://doi.org/10.1186/s13007-019-0388-x

614 Amadeu RR, Cellon C, Olmstead JW, et al (2016) AGHmatrix: R Package to Construct Relationship Matrices for
615 Autotetraploid and Diploid Species: A Blueberry Example. The Plant Genome
616 9:plantgenome2016.01.0009. https://doi.org/10.3835/plantgenome2016.01.0009

617 Andrieu J (1995) Prévision de la digestibilité et de la valeur énergétique du maïs fourrage à l'état frais. INRA Prod
618 Anim 8:273–274. https://doi.org/10.20870/productions-animales.1995.8.4.4136

619 Argillier O, Barrière Y, Hébert Y (1995) Genetic variation and selection criterion for digestibility traits of forage
620 maize. Euphytica 82:175–184. https://doi.org/10.1007/BF00027064

621 Auinger H-J, Schönleben M, Lehermeier C, et al (2016) Model training across multiple breeding cycles
622 significantly improves genomic prediction accuracy in rye (Secale cereale L.). Theor Appl Genet
623 129:2043–2053. https://doi.org/10.1007/s00122-016-2756-5

624 Barrière Y, Emile JC (2000) Le maïs fourrage. III - Evaluation et perspectives de progrès génétiques sur les
625 caractères de valeur alimentaire

626 Ben-Sadoun S, Rincent R, Auzanneau J, et al (2020) Economical optimization of a breeding scheme by selective
627 phenotyping of the calibration set in a multi-trait context: application to bread making quality. Theor Appl
628 Genet 133:2197–2212. https://doi.org/10.1007/s00122-020-03590-4

629 Bernardo (1994) Prediction of Maize Single-Cross Performance Using RFLPs and Information from Related
630 Hybrids. Crop Science 34:20. https://doi.org/10.2135/cropsci1994.0011183X003400010003x

631 Brandariz SP, Bernardo R (2018) Maintaining the Accuracy of Genomewide Predictions when Selection Has
632 Occurred in the Training Population. Crop Science 58:1226–1231.
633 https://doi.org/10.2135/cropsci2017.11.0682

634 Burdo B, Leon N, Kaeppler SM (2021) Testcross vs. randomly paired single-cross progeny tests for genomic
635 prediction of new inbreds and hybrids derived from multiparent maize populations. Crop Sci csc2.20545.
636 https://doi.org/10.1002/csc2.20545

637 Coombes NE (2009) DiGGer, a spatial design program. Biometric Bulletin NSW Department of Primary
638 Industries, Orange, NSW

639 Denis M, Bouvet J-M (2013) Efficiency of genomic selection with models including dominance effect in the
640 context of Eucalyptus breeding. Tree Genetics & Genomes 9:37–51. https://doi.org/10.1007/s11295-012-
641 0528-1

642 Fernández-González J, Akdemir D, Isidro y Sánchez J (2023) A comparison of methods for training population
643 optimization in genomic selection. Theor Appl Genet 136:30. https://doi.org/10.1007/s00122-023-04265-
644 6

645 Fristche-Neto R, Akdemir D, Jannink J-L (2018) Accuracy of genomic selection to predict maize single-crosses
646 obtained through different mating designs. Theor Appl Genet 131:1153–1162.
647 https://doi.org/10.1007/s00122-018-3068-8

Ganal MW, Durstewitz G, Polley A, et al (2011) A Large Maize (Zea mays L.) SNP Genotyping Array: Development and Germplasm Genotyping, and Genetic Mapping to Compare with the B73 Reference Genome. PLOS ONE 6:e28334. https://doi.org/10.1371/journal.pone.0028334

Gerke JP, Edwards JW, Guill KE, et al (2015) The Genomic Impacts of Drift and Selection for Hybrid Performance in Maize. Genetics 201:1201–1211. https://doi.org/10.1534/genetics.115.182410

Giraud H (2016) Genetic analysis of hybrid value for silage maize in multiparental designs: QTL detection and genomic selection. Thesis, Paris-Saclay

Giraud H, Bauland C, Falque M, et al (2017a) Linkage Analysis and Association Mapping QTL Detection Models for Hybrids Between Multiparental Populations from Two Heterotic Groups: Application to Biomass Production in Maize (*Zea mays* L.). G3 Genes|Genomes|Genetics 7:3649–3657. https://doi.org/10.1534/g3.117.300121

Giraud H, Bauland C, Falque M, et al (2017b) Reciprocal Genetics: Identifying QTL for General and Specific Combining Abilities in Hybrids Between Multiparental Populations from Two Maize (*Zea mays* L.) Heterotic Groups. Genetics 207:1167–1180. https://doi.org/10.1534/genetics.117.300305

Goering HK, Soest PJV (1970) Forage Fiber Analyses (apparatus, Reagents, Procedures, and Some Applications). U.S. Agricultural Research Service

González-Diéguez D, Legarra A, Charcosset A, et al (2021) Genomic prediction of hybrid crops allows disentangling dominance and epistasis. Genetics. https://doi.org/10.1093/genetics/iyab026

Hallauer AR, Carena MJ, Filho JBM (2010) Quantitative Genetics in Maize Breeding. Springer Science & Business Media

Henderson CR (1976) A Simple Method for Computing the Inverse of a Numerator Relationship Matrix Used in Prediction of Breeding Values. Biometrics 32:69. https://doi.org/10.2307/2529339

Hofheinz N, Borchardt D, Weissleder K, Frisch M (2012) Genome-based prediction of test cross performance in two subsequent breeding cycles. Theor Appl Genet 125:1639–1645. https://doi.org/10.1007/s00122-012-1940-5

Howard R, Jarquin D, Crossa J (2022) Overview of Genomic PredictionGenomic predictions (GP) Methods and the Associated Assumptions on the Variance of Marker Effect, and on the Architecture of the Target Trait. In: Ahmadi N, Bartholomé J (eds) Genomic Prediction of Complex Traits: Methods and Protocols. Springer US, New York, NY, pp 139–156

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

Isidro y Sánchez J, Akdemir D (2021) Training Set Optimization for Sparse Phenotyping in Genomic Selection: A Conceptual Overview. Frontiers in Plant Science 12:. https://doi.org/10.3389/fpls.2021.715910

Jannink J-L (2010) Dynamics of long-term genomic selection. Genetics Selection Evolution 42:35. https://doi.org/10.1186/1297-9686-42-35

Kadam DC, Lorenz AJ (2018) Toward Redesigning Hybrid Maize Breeding Through Genomics-Assisted Breeding. In: Bennetzen J, Flint-Garcia S, Hirsch C, Tuberosa R (eds) The Maize Genome. Springer International Publishing, Cham, pp 367–388

Kadam DC, Potts SM, Bohn MO, et al (2016) Genomic Prediction of Single Crosses in the Early Stages of a Maize Hybrid Breeding Pipeline. G3 Genes|Genomes|Genetics 6:3443–3453. https://doi.org/10.1534/g3.116.031286

Kadam DC, Rodriguez OR, Lorenz AJ (2021) Optimization of training sets for genomic prediction of early-stage single crosses in maize. Theor Appl Genet 134:687–699. https://doi.org/10.1007/s00122-020-03722-w

691 Laporte F, Charcosset A, Mary-Huard T (2022) Efficient ReML inference in variance component mixed models
692 using a Min-Max algorithm. PLOS Computational Biology 18:e1009659.
693 https://doi.org/10.1371/journal.pcbi.1009659

694 Laporte F, Mary-Huard T (2020) MM4LMM: Inference of Linear Mixed Models Through MM Algorithm

695 Legarra A, Gonzalez-Dieguez DO, Charcosset A, Vitezica ZG (2023) Impact of interpopulation distance on
696 dominance variance and average heterosis in hybrid populations within species. Genetics 224:iyad059.
697 https://doi.org/10.1093/genetics/iyad059

698 Lorenzi A, Bauland C, Mary-Huard T, et al (2022) Genomic prediction of hybrid performance: comparison of the
699 efficiency of factorial and tester designs used as training sets in a multiparental connected reciprocal
700 design for maize silage. Theor Appl Genet. https://doi.org/10.1007/s00122-022-04176-y

701 Maenhout S, De Baets B, Haesaert G (2010) Prediction of maize single-cross hybrid performance: support vector
702 machine regression versus best linear prediction. Theoretical and applied genetics 120:415–427

703 Mangin B, Rincent R, Rabier C-E, et al (2019) Training set optimization of genomic prediction by means of
704 EthAcc. PLoS ONE 14:e0205629. https://doi.org/10.1371/journal.pone.0205629

705 Merrick LF, Carter AH (2021) Comparison of genomic selection models for exploring predictive ability of
706 complex traits in breeding programs. The Plant Genome 14:e20158. https://doi.org/10.1002/tpg2.20158

707 Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of Total Genetic Value Using Genome-Wide Dense
708 Marker Maps. Genetics 157:1819–1829

709 Michel S, Ametz C, Gungor H, et al (2016) Genomic selection across multiple breeding cycles in applied bread
710 wheat breeding. Theor Appl Genet 129:1179–1189. https://doi.org/10.1007/s00122-016-2694-2

711 Momen M, Morota G (2018) Quantifying genomic connectedness and prediction accuracy from additive and non-
712 additive gene actions. Genetics Selection Evolution 50:45. https://doi.org/10.1186/s12711-018-0415-9

713 Mrode RA, Thompson R (2005) Linear models for the prediction of animal breeding values, 2nd ed. CABI Pub,
714 Wallingford, UK ; Cambridge, MA

715 Neyhart JL, Tiede T, Lorenz AJ, Smith KP (2017) Evaluating Methods of Updating Training Data in Long-Term
716 Genomewide Selection. G3 Genes|Genomes|Genetics 7:1499–1510.
717 https://doi.org/10.1534/g3.117.040550

718 Pszczola M, Calus MPL (2016) Updating the reference population to achieve constant genomic prediction
719 reliability across generations. animal 10:1018–1024. https://doi.org/10.1017/S1751731115002785

720 Pszczola M, Strabel T, Mulder HA, Calus MPL (2012) Reliability of direct genomic values for animals with
721 different relationships within and to the reference population. Journal of Dairy Science 95:389–400.
722 https://doi.org/10.3168/jds.2011-4338

723 R Core Team (2020) R: A Language and Environment for Statistical Computing. R Foundation for Statistical
724 Computing, Vienna, Austria

725 Reif JC, Gumpert F-M, Fischer S, Melchinger AE (2007) Impact of Interpopulation Divergence on Additive and
726 Dominance Variance in Hybrid Populations. Genetics 176:1931.
727 https://doi.org/10.1534/genetics.107.074146

728 Revelle W (2021) psych: Procedures for Psychological, Psychometric, and Personality Research. Northwestern
729 University, Evanston, Illinois

730 Rincent R, Charcosset A, Moreau L (2017) Predicting genomic selection efficiency to optimize calibration set and
731 to assess prediction accuracy in highly structured populations. Theor Appl Genet 130:2231–2247.
732 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

Rio S, Akdemir D, Carvalho T, Sánchez JI y. (2022a) Assessment of genomic prediction reliability and optimization of experimental designs in multi-environment trials. Theor Appl Genet 135:405–419. https://doi.org/10.1007/s00122-021-03972-2

Rio S, Charcosset A, Mary-Huard T, et al (2022b) Building a Calibration Set for Genomic Prediction, Characteristics to Be Considered, and Optimization Approaches. In: Ahmadi N, Bartholomé J (eds) Genomic Prediction of Complex Traits. Springer US, New York, NY, pp 77–112

Sallam AH, Endelman JB, Jannink J-L, Smith KP (2015) Assessing Genomic Selection Prediction Accuracy in a Dynamic Barley Breeding Population. The Plant Genome 8:plantgenome2014.05.0020. https://doi.org/10.3835/plantgenome2014.05.0020

Schrag TA, Melchinger AE, Sørensen AP, Frisch M (2006) Prediction of single-cross hybrid performance for grain yield and grain dry matter content in maize using AFLP markers associated with QTL. Theoretical and Applied Genetics 113:1037–1047. https://doi.org/10.1007/s00122-006-0363-6

Schrag TA, Westhues M, Schipprack W, et al (2018) Beyond genomic prediction: combining different types of omics data can improve prediction of hybrid performance in maize. Genetics 208:1373–1385

Seye AI (2019) Prédiction assistée par marqueurs de la performance hybride dans un schéma de sélection réciproque : simulations et évaluation expérimentale pour le maïs ensilage. Thesis, Paris Saclay

Seye AI, Bauland C, Charcosset A, Moreau L (2020) Revisiting hybrid breeding designs using genomic predictions: simulations highlight the superiority of incomplete factorials between segregating families over topcross designs. Theor Appl Genet 133:1995–2010. https://doi.org/10.1007/s00122-020-03573-5

Seye AI, Bauland C, Giraud H, et al (2019) Quantitative trait loci mapping in hybrids between Dent and Flint maize multiparental populations reveals group-specific QTL for silage quality traits with variable pleiotropic effects on yield. Theor Appl Genet 132:1523–1542. https://doi.org/10.1007/s00122-019-03296-2

Stuber CW, Cockerham CC (1966) GENE EFFECTS AND VARIANCES IN HYBRID POPULATIONS. Genetics 54:1279–1286. https://doi.org/10.1093/genetics/54.6.1279

Surault F, Emile JC, Briand M, et al (2005) Variabilité génétique de la digestibilité in vivo d'hybrides de maïs. Bilan de 34 années de mesures. Fourrages 183:459

Technow F, Bürger A, Melchinger AE (2013) Genomic Prediction of Northern Corn Leaf Blight Resistance in Maize with Combined or Separated Training Sets for Heterotic Groups. G3 3:197–203. https://doi.org/10.1534/g3.112.004630

Technow F, Schrag TA, Schipprack W, et al (2014) Genome Properties and Prospects of Genomic Prediction of Hybrid Performance in a Breeding Program of Maize. Genetics 197:1343–1355. https://doi.org/10.1534/genetics.114.165860

VanRaden PM (2008) Efficient Methods to Compute Genomic Predictions. Journal of Dairy Science 91:4414–4423. https://doi.org/10.3168/jds.2007-0980

Varona L, Legarra A, Toro MA, Vitezica ZG (2018) Non-additive Effects in Genomic Selection. Front Genet 9:78. https://doi.org/10.3389/fgene.2018.00078

Vitezica ZG, Legarra A, Toro MA, Varona L (2017) Orthogonal Estimates of Variances for Additive, Dominance, and Epistatic Effects in Populations. Genetics 206:1297–1307. https://doi.org/10.1534/genetics.116.199406

Vitezica ZG, Varona L, Legarra A (2013) On the Additive and Dominant Variance and Covariance of Individuals Within the Genomic Selection Scope. Genetics 195:1223–1230. https://doi.org/10.1534/genetics.113.155176

Wang N, Wang H, Zhang A, et al (2020) Genomic prediction across years in a maize doubled haploid breeding program to accelerate early-stage testcross testing. Theor Appl Genet 133:2869–2879. https://doi.org/10.1007/s00122-020-03638-5

Williams E, Piepho H-P, Whitaker D (2011) Augmented p-rep designs. Biom J 53:19–27. https://doi.org/10.1002/bimj.201000102

Williams EJ (1959) 136. Query: Significance of Difference between Two Non-Independent Correlation Coefficients. Biometrics 15:135. https://doi.org/10.2307/2527608

Zhao Y, Gowda M, Liu W, et al (2012) Accuracy of genomic selection in European maize elite breeding populations. Theor Appl Genet 124:769–776. https://doi.org/10.1007/s00122-011-1745-y

Zhong S, Dekkers JCM, Fernando RL, Jannink J-L (2009) Factors Affecting Accuracy From Genomic Selection in Populations Derived From Multiple Inbred Lines: A Barley Case Study. Genetics 182:355–364. https://doi.org/10.1534/genetics.108.098277

Statements and declarations

Funding

Lidea, Limagrain Europe, Maïsadour Semences, Corteva, RAGT 2n, KWS and Syngenta Seeds grouped in the frame of the ProMais association funded the SAM-MCR project. A.L. PhD contract was funded by RAGT 2n and ANRT contract n°660 2020/0032, with contributions of all SAM-MCR project partners. GQE-Le Moulon benefits from the support of Saclay Plant Sciences-SPS (ANR-17-EUR-0007).

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Author's contributions

CB, AC and LM initiated this project. LM and CB coordinated it with the help of GT and CG. AC and LM supervised this work with contributions of TMH. SP, and CP contributed to the multiplication of DH lines and hybrid seeds production. DM and VC prepared DNA samples for the genotyping of the inbred lines and assembled the genotypic data. AL analyzed the results and prepared the manuscript. All authors discussed the results and contributed to the final manuscript. All authors revised and approved the manuscript.

Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Ethics approval

All usual standards have been respected.

Consent to participate

Not relevant for this study on plants.

Consent to publish

Not relevant for this study on plants.

Figure legends

Fig. 1 Hybrid experimental designs produced by crossing inbred lines from the initial generation (G0) and the inbred lines obtained after one cycle of selection (G1).

Fig. 2 Predictive abilities obtained by cross-validations within the 442 G1 hybrids using different prediction models (PBLUP, GCA.1, GCA.2 or G models) in Scenario 1. The mean predictive ability over the 100 replicates is represented by a white cross. Significant differences (as obtained by paired t-tests at a level risk α=0.05) are indicated with letters: two different letters indicate a significant difference and at least one common letters indicate no significant difference.

Fig. 3 Predictive abilities obtained in Scenario 2a when training the GS model on the G0_F-1H design (951 hybrids) to predict the 47 G0S hybrids or the 442 G1 hybrids. The dotted line corresponds to the mean predictive ability over 100 replicates of cross-validations within the 442 G1 hybrids.

Fig. 4 Predictive abilities obtained for the G1 hybrids (442) by training the GS model on the G0_F-4H (363) or the G0_T (360) TRSs in Scenario 2b. Williams tests were performed (α=0.05) and significant differences were indicated with letters: two different letters indicate a significant difference and at least one common letters indicate no significant difference.

Fig. 5 Predictive abilities obtained in Scenario 2b' by training the GS model on 180 hybrids issued from tester-based or factorial TRSs to predict the G1 hybrids (442). The different tester-based TRSs correspond to: 90 lines crossed to one tester (1T-180H-180L-A, 1T-180H-180L-B, 1T-180H-180L-C, 1T-180H-180L-D), 90 lines crossed to two testers (2T-180H-180L), 45 lines crossed to two testers (2T-180H-90L). The factorial design (F-180H-152L) corresponds to the crosses of 76 flint lines with 76 dent lines. The sampling was repeated 10 times and t-tests (α=0.05) were performed for the F-180H-170L, 2T-180H-180L and 2T-180H-90L. Significant differences were indicated with letters: two different letters indicate a significant difference and at least one common letters indicate no significant difference.

Fig. 6 Predictive abilities obtained in Scenario 3a when predicting one-fifth of the G1 hybrids (88) using different TRSs: G0R hybrids (in solid colored lines) completed by 132 G0S hybrids (in doted colored lines) and *m* G1 hybrids (with *m* ranging from 0 to 354 from the left to the right of each graph). The mean predictive ability over 100 replicates is represented by a dot for each TRS. The number of hybrids in the initial G0R TRSs is indicated between brackets in the figure legend.

Fig. 7 Predictive abilities obtained with TRSs composed of an initial G0 set (1360 hybrids) completed by a CDmean optimized G1 hybrid set of different sizes (50, 100, 200 and 300). The G1 hybrid set is optimized considering only G1 information (CDmean 1) or considering G1 and G0 information (CDmean 2) in the calculation of the CDmean and compared to a randomly sampled TRS (Random). The white cross represents the mean predictive ability over the 100 replicates.

Table 1 Description of all experimental designs used in this study.

Years of phenotyping	Breeding cycle	Design	Name	Hybrids within the design [a]	Reference [c]
2013, 2014	G0	Factorial	G0_F-1H	G0R[a]	Giraud et al. 2017a, b; Seye et al. 2019; Lorenzi et al. 2022
2016, 2017	G0	Factorial	G0_F-4H	G0R + G0S[b]	Seye 2019; Lorenzi et al. 2022
		Tester	G0_T		
2019, 2020	G0+G1	Factorial	(G0S+G1)_F-1H	G0S + G1	Current study

[a] G0R hybrids were produced by crossing two random lines from the G0 cycle
[b] G0S hybrids were produced by crossing two selected lines from the G0 cycle
[c] A reference was indicated for data that was already analyzed in previous studies

Table 2 Definition of the genomic prediction models tested in Scenario 1.

Models	Model code	Random genetic effects[a]	Reference
PBLUP		g	Henderson 1976
GCA.1	GCA	$g_{GCA_f} + g_{GCA_d}$	VanRaden 2008
	GCA_SCA	$g_{GCA_f} + g_{GCA_d} + g_{SCA}$	
GCA.2	GCA:A	$g_{A_f} + g_{A_d} + r$	González-Diéguez et al. 2021
	GCA:AD	$g_{A_f} + g_{A_d} + g_D$	
	GCA:A(AAdf)	$g_{A_f} + g_{A_d} + g_{AA_{df}}$	
	GCA:AD(AAdf)	$g_{A_f} + g_{A_d} + g_D + g_{AA_{df}}$	
	GCA:AD(AAf)(AAd)(AAdf)	$g_{A_f} + g_{A_d} + g_D + g_{AA_f} + g_{AA_d} + g_{AA_{df}}$	
G	G:A	g_A	Vitezica et al. 2017
	G:AD	$g_A + g_D$	
	G:A(AA)	$g_A + g_{AA}$	
	G:AD(AA)	$g_A + g_D + g_{AA}$	

[a] The list of the random genetic effects considered in the GCA models correspond to: dent GCA (g_{GCA_d}), flint GCA (g_{GCA_f}), SCA (g_{SCA}), intragroup additive-by-additive epistasis for the dent (g_{AA_d}) and flint group (g_{AA_f}), and intergroup additive-by-additive epistasis ($g_{AA_{df}}$) effects. In the G models, random genetic effects correspond to: additive (g_A), dominance (g_D) and additive-by-additive epistasis (g_{AA}) effects.

Table 3 Broad-sense heritability (H^2), percentage of genetic variance assigned to SCA variance (%SCA) and variance components estimated on phenotypic data corrected for spatial effects for the (G0S+G1)_F-1H without marker information.

	DMY (t/ha)		DMC (%)		DtSilk (days)		MFU (MFUx10²/kg)	
	G0S	**G1**	**G0S**	**G1**	**G0S**	**G1**	**G0S**	**G1**
$\sigma^2_{GCA_f}$	0.50(0.24)[d]	0.31(0.05)	0.46(0.28)	1.47(0.21)	1.63(0.71)	2.06(0.29)	0.44(0.24)	0.49(0.09)
$\sigma^2_{GCA_d}$	0.13(0.18)	0.25(0.05)	1.42(0.51)	0.73(0.21)	1.57(0.65)	1.41(0.27)	0.00	0.36(0.09)
σ^2_{SCA}	0.14(0.18)	0.00	0.00	0.25(0.20)	0.03(0.27)	0.05(0.22)	0.19(0.13)	0.00(0.09)
$\sigma^2_{GCA_f \times E}$	0.07(0.08)	0.05(0.05)	0.00	0.11(0.05)	0.41(0.16)	0.17(0.07)	0.27(0.17)	0.20(0.07)
$\sigma^2_{GCA_d \times E}$	0.12(0.09)	0.07(0.04)	0.39(0.13)	0.31(0.05)	0.05(0.12)	0.09(0.06)	0.06(0.12)	0.13(0.07)
$\sigma^2_{SCA \times E}$	0.00	0.23(0.07)	0.00	0.00	0.00	0.24(0.11)	0.00	0.11(0.09)
σ^2_E [a]	0.31(0.05)-1.40(0.12)		0.57(0.07)-3.62(0.27)		0.64(0.09)-2.33(0.20)		0.12(0.02)-7.97(0.55)	
%SCA[b]	19	0	0	10	1	1	30	0
H^2 [c]	0.87	0.80	0.90	0.92	0.93	0.94	0.56	0.62

[a] Minimum and maximum residual variance across all environments

[b] Percentage of SCA variance computed as $\frac{\sigma^2_{SCA}}{\sigma^2_{GCA_d}+\sigma^2_{GCA_f}+\sigma^2_{SCA}} \times 100$

[c] Broad-sense heritability

[d] Standard error in brackets

Table 4 Performances (ls-means) of commercial, founder and experimental hybrids (G0S and G1 hybrids) and genetic gain of the experimental hybrids compared to the founder hybrids corresponding to the (G0S+G1)_F-1H design.

	Hybrid type	Component	DMY (t/ha)	DMC (%)	DtSilk (days)	MFU (MFUx10²/kg)
Ls-means	**Commercial**	**Mean**	**17.96** (17.18-18.74)[b]	**34.69** (34.37-35.00)	**201.89** (200.27-203.50)	**95.54** (95.34-95.74)
		Sd[c]	1.10	0.45	2.28	0.28
	Founder	**Mean**	**15.80** (14.27-17.47)	**34.10** (30.53-37.48)	**203.14** (199.80-206.16)	**95.28** (91.13-98.14)
		Sd	0.84	1.94	1.50	2.05
	G0S	**Mean**	**17.35** (14.82-18.89)	**33.33** (30.54-37.13)	**204.97** (201.80-208.79)	**93.17** (89.04-97.06)
		Sd	0.92	1.40	1.82	1.56
	G1	**Mean**	**17.33** (14.33-19.72)	**33.43** (29.39-39.18)	**205.04** (197.98-211.40)	**93.13** (87.59-101.36)
		Sd	0.85	1.63	1.97	1.71
Genetic gain[a]	**G0S**		1.55	-0.77	1.83	-2.11
	G1		1.52	-0.67	1.90	-2.15
Predicted genetic gain[d]	**G0S**		1.45	-0.33	1.18	-1.35
	G1		1.41	-0.28	0.91	-1.14

[a] Genetic gain computed as the difference between the mean performance of the experimental hybrids and the founder hybrids

[b] Minimum and maximum mean performance in brackets

[c] Standard deviation of the ls-means of the experimental hybrid performances

[d] Predicted genetic gain based on genomic predictions trained on the G0_F-1H

Figures

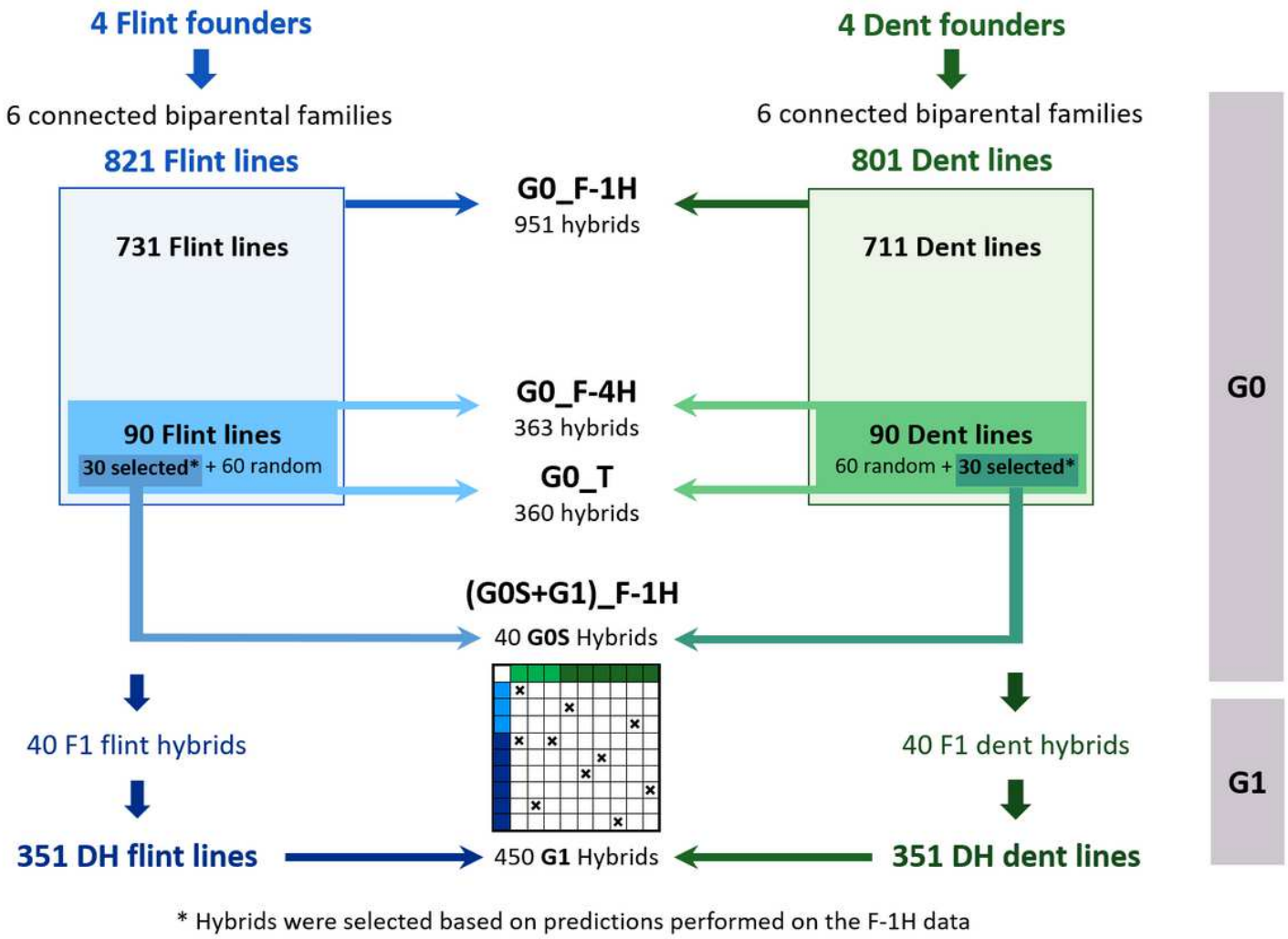

Figure 1

Hybrid experimental designs produced by crossing inbred lines from the initial generation (G0) and the inbred lines obtained after one cycle of selection (G1).

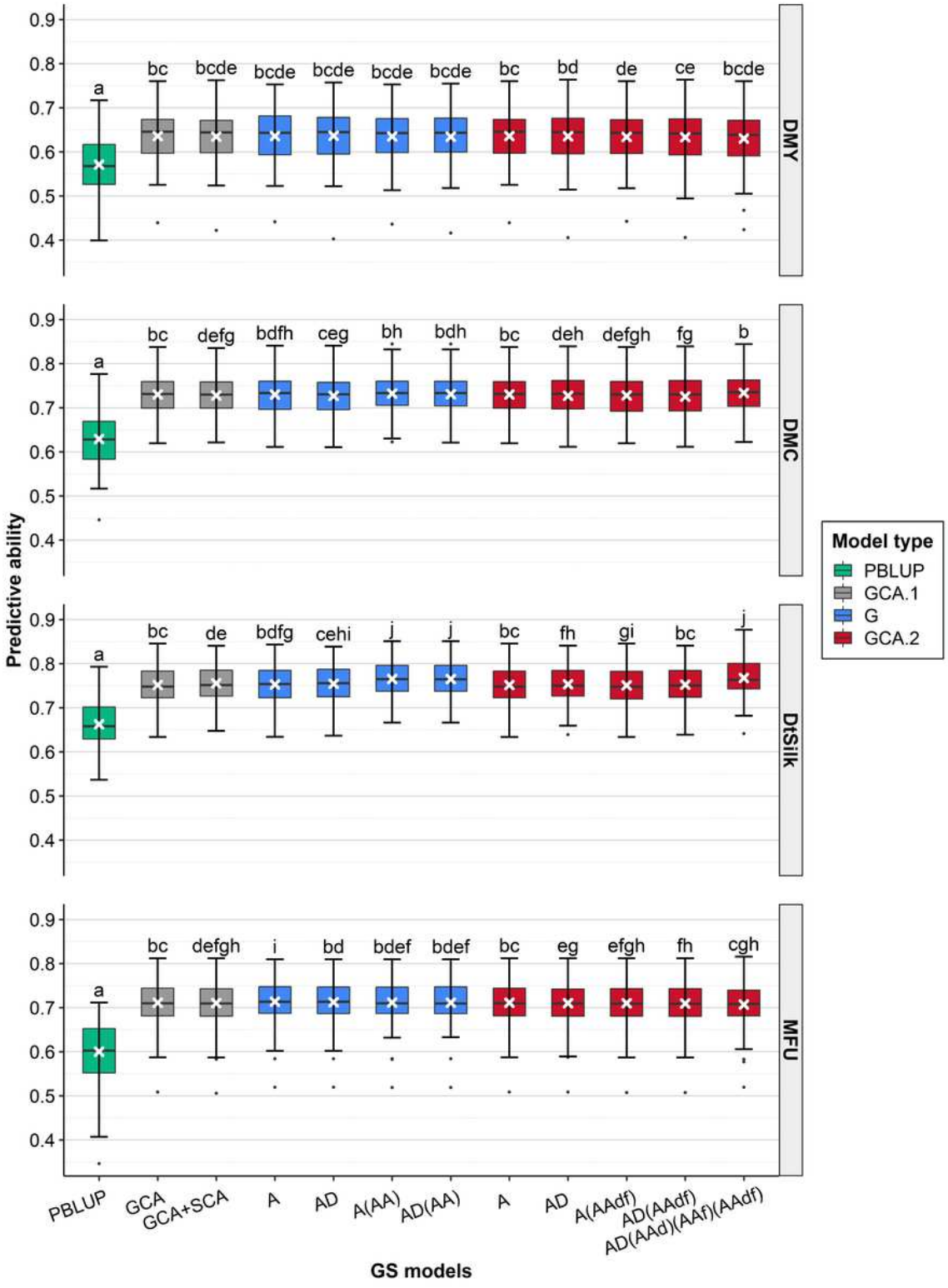

Figure 2

Predictive abilities obtained by cross-validations within the 442 G1 hybrids using different prediction models (PBLUP, GCA.1, GCA.2 or G models) in Scenario 1. The mean predictive ability over the 100 replicates is represented by a white cross. Significant differences (as obtained by paired t-tests at a level risk α=0.05) are indicated with letters: two different letters indicate a significant difference and at least one common letters indicate no significant difference.

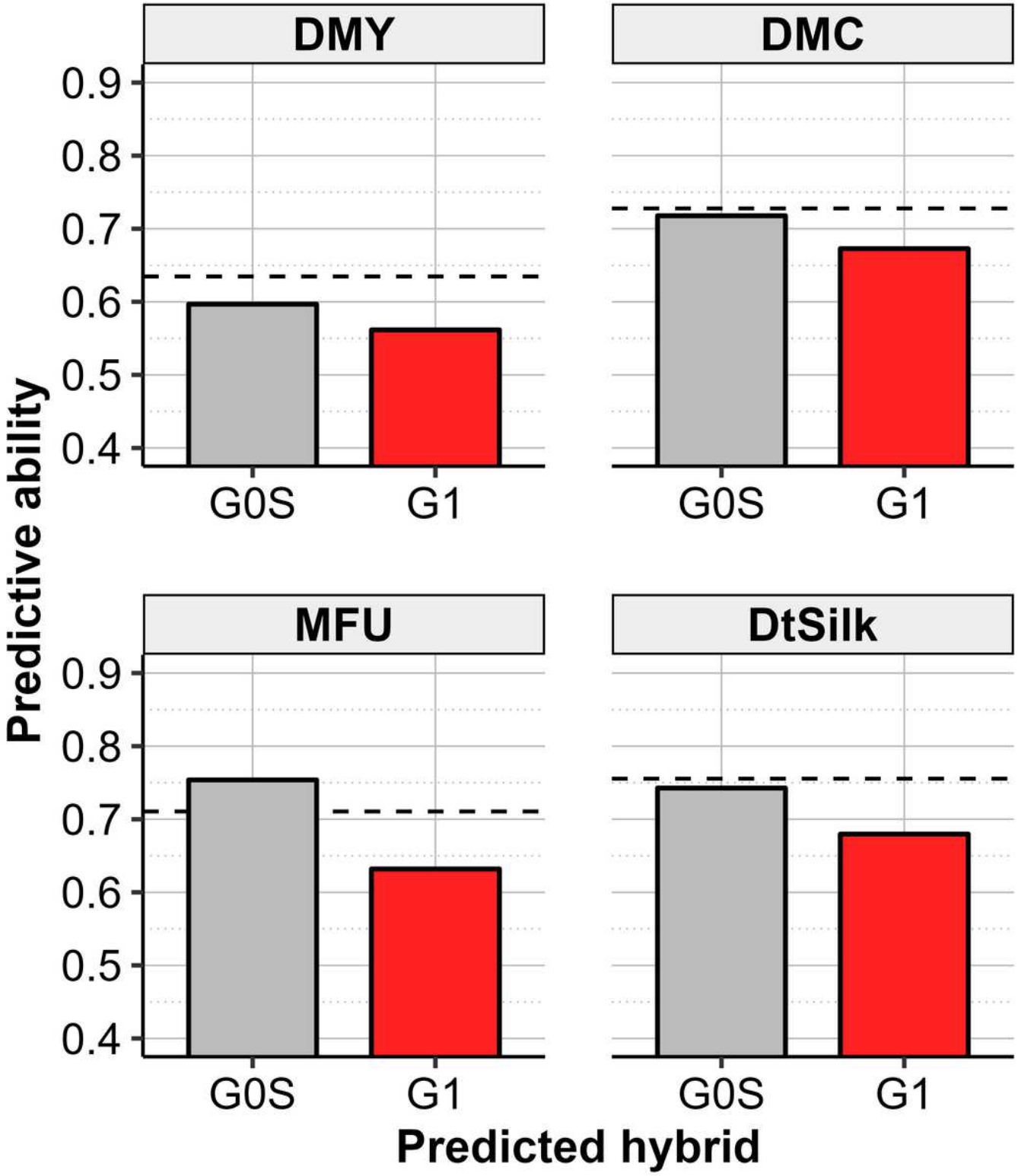

Figure 3

Predictive abilities obtained in Scenario 2a when training the GS model on the G0_F-1H design (951 hybrids) to predict the 47 G0S hybrids or the 442 G1 hybrids. The dotted line corresponds to the mean predictive ability over 100 replicates of cross-validations within the 442 G1 hybrids.

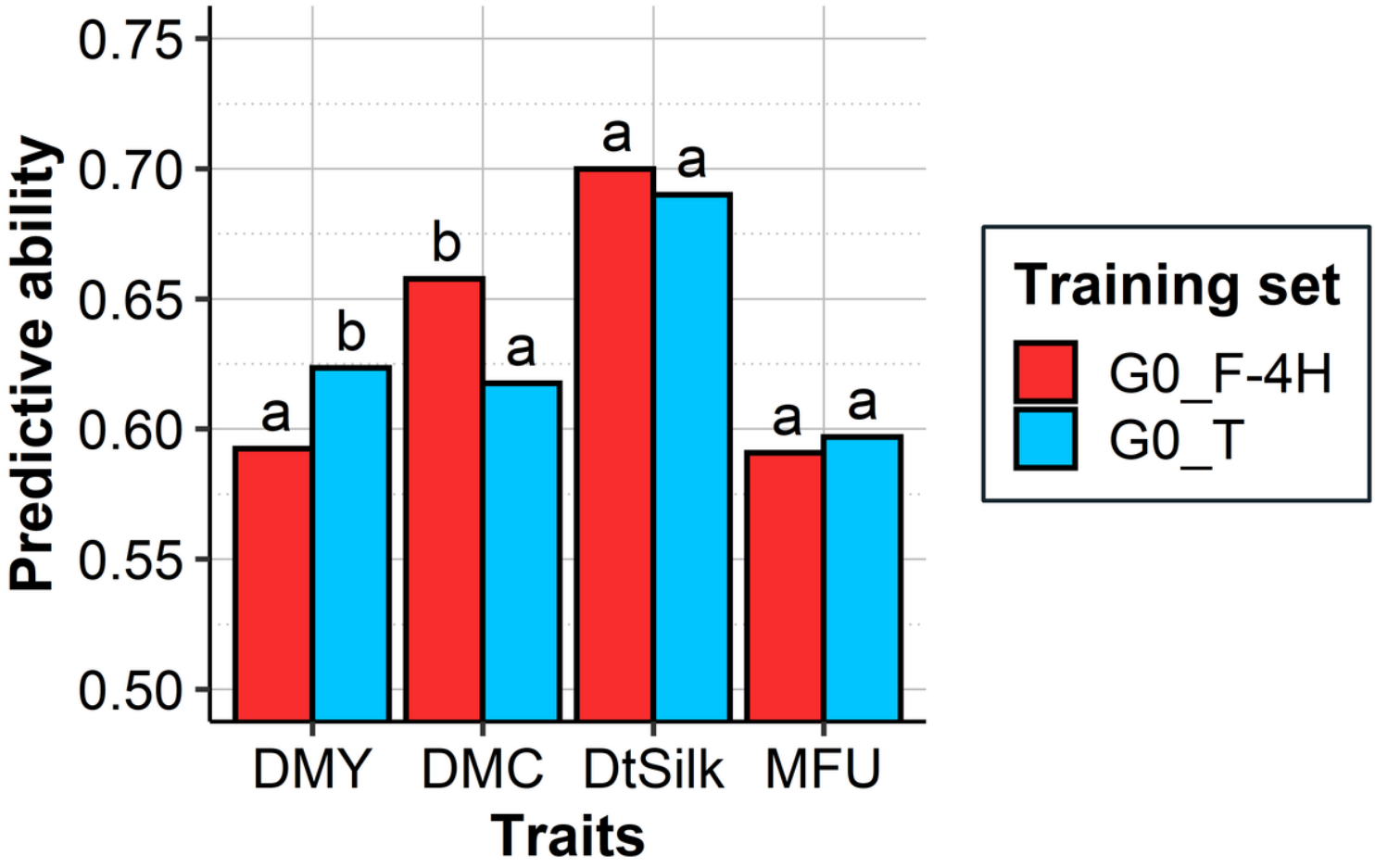

Figure 4

Predictive abilities obtained for the G1 hybrids (442) by training the GS model on the G0_F-4H (363) or the G0_T (360) TRSs in Scenario 2b. Williams tests were performed (α=0.05) and significant differences were indicated with letters: two different letters indicate a significant difference and at least one common letters indicate no significant difference.

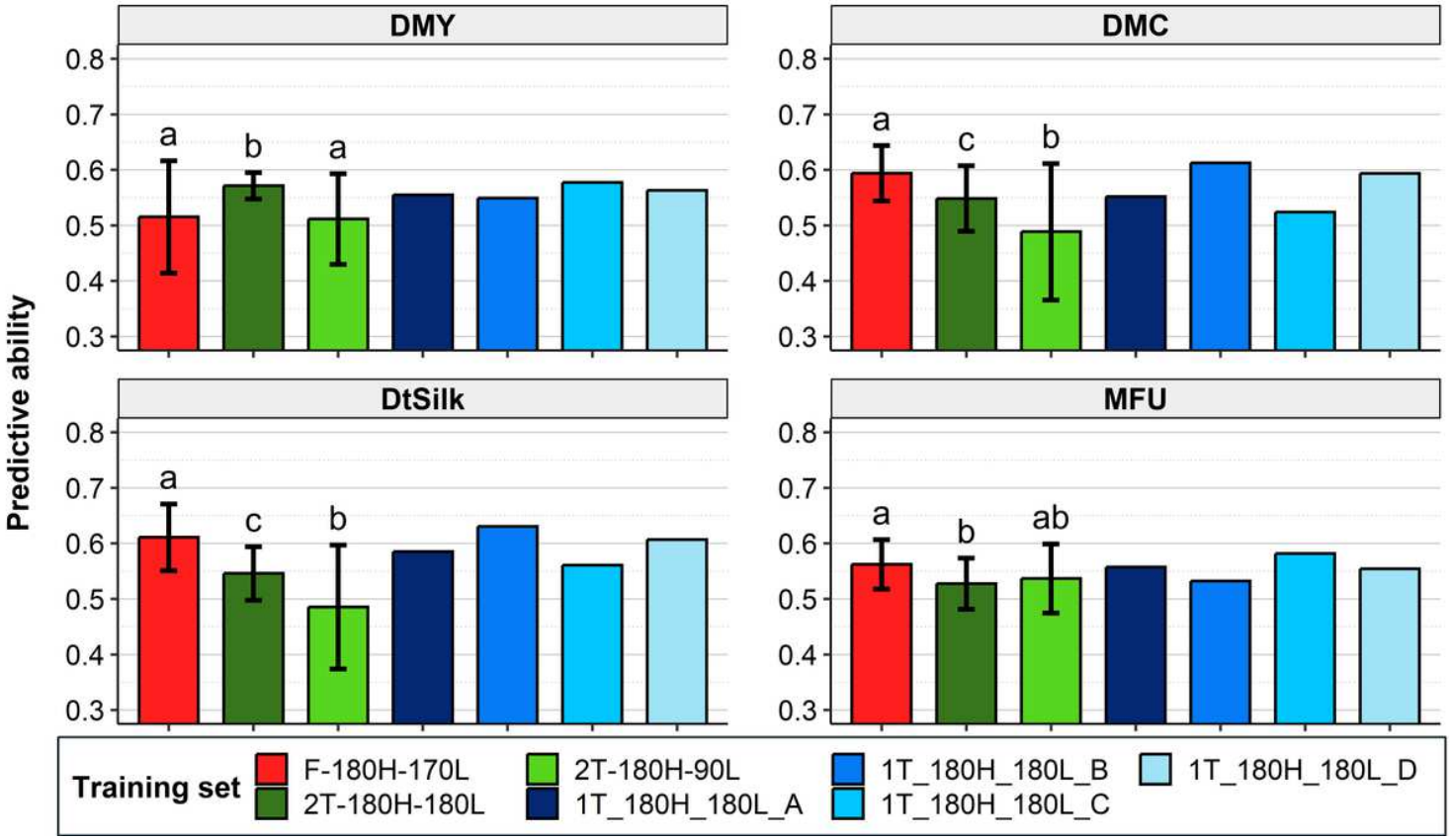

Figure 5

Predictive abilities obtained in Scenario 2b' by training the GS model on 180 hybrids issued from tester-based or factorial TRSs to predict the G1 hybrids (442). The different tester-based TRSs correspond to: 90 lines crossed to one tester (1T-180H-180L-A, 1T-180H-180L-B, 1T-180H-180L-C, 1T-180H-180L-D), 90 lines crossed to two testers (2T-180H-180L), 45 lines crossed to two testers (2T-180H-90L). The factorial design (F-180H-152L) corresponds to the crosses of 76 flint lines with 76 dent lines. The sampling was repeated 10 times and t-tests (α=0.05) were performed for the F-180H-170L, 2T-180H-180L and 2T-180H-90L. Significant differences were indicated with letters: two different letters indicate a significant difference and at least one common letters indicate no significant difference.

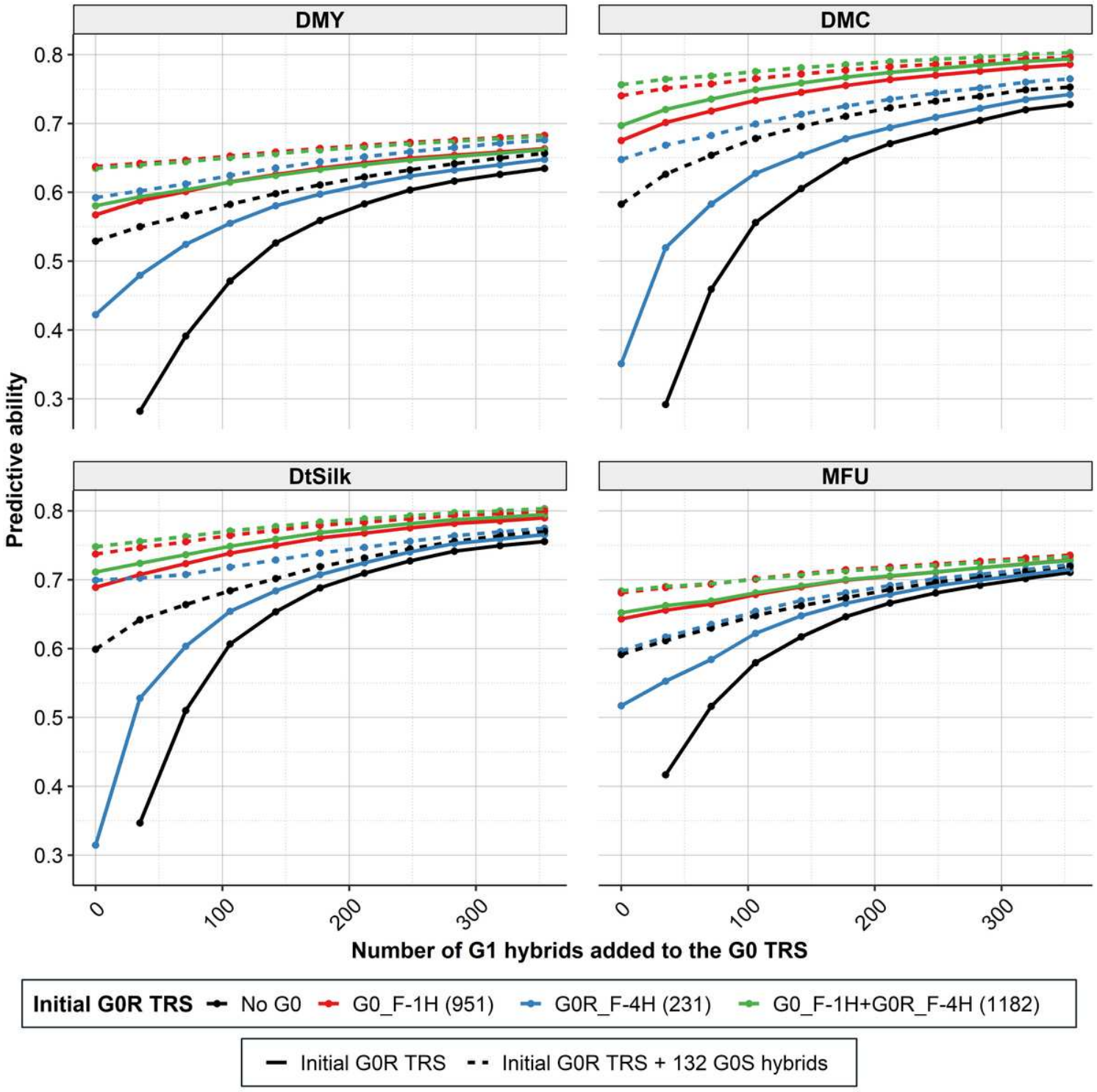

Figure 6

Predictive abilities obtained in Scenario 3a when predicting one-fifth of the G1 hybrids (88) using different TRSs: G0R hybrids (in solid colored lines) completed by 132 G0S hybrids (in doted colored lines) and *m* G1 hybrids (with *m* ranging from 0 to 354 from the left to the right of each graph). The mean predictive ability over 100 replicates is represented by a dot for each TRS. The number of hybrids in the initial G0R TRSs is indicated between brackets in the figure legend.

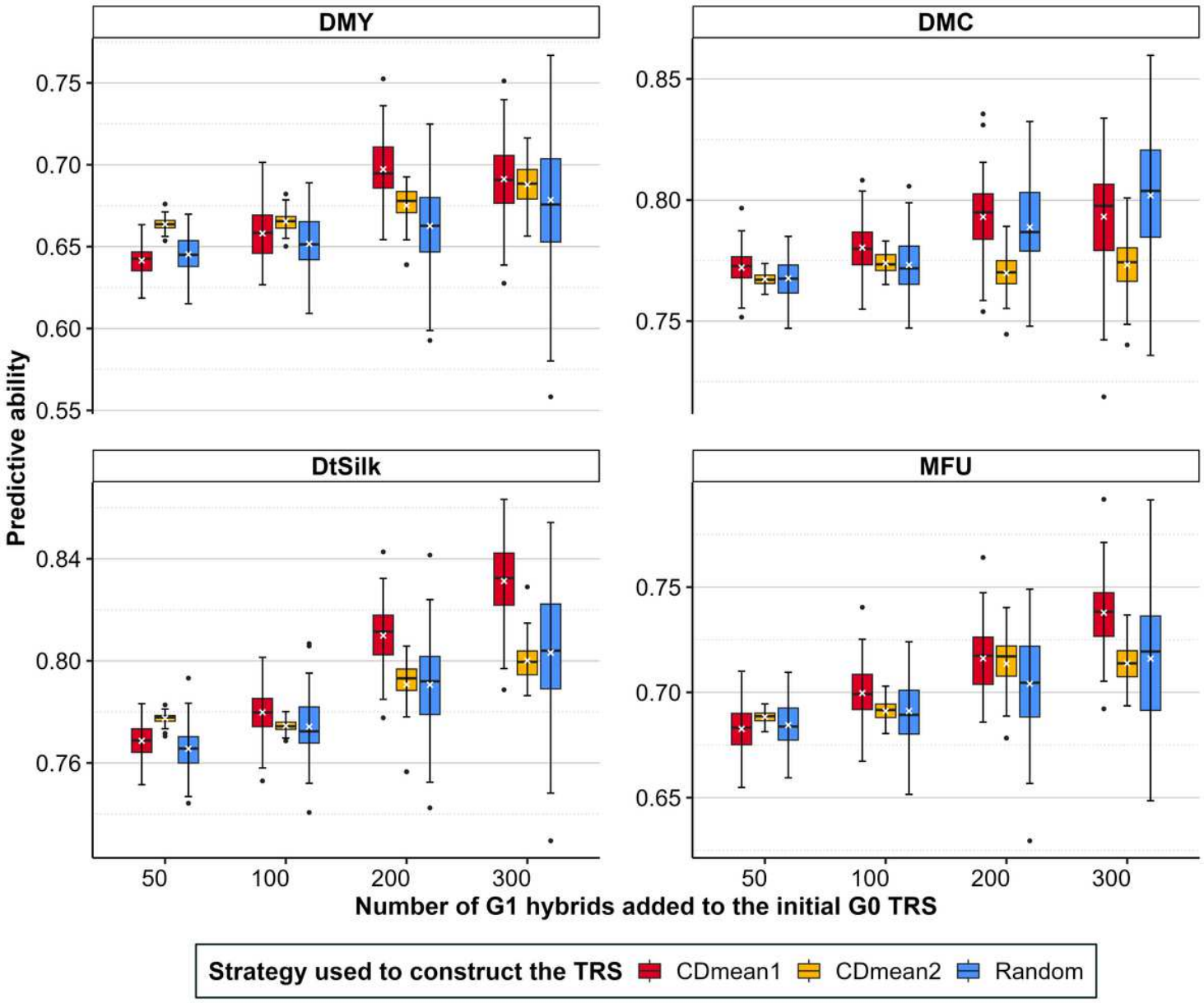

Figure 7

Predictive abilities obtained with TRSs composed of an initial G0 set (1360 hybrids) completed by a CDmean optimized G1 hybrid set of different sizes (50, 100, 200 and 300). The G1 hybrid set is optimized considering only G1 information (CDmean 1) or considering G1 and G0 information (CDmean 2) in the calculation of the CDmean and compared to a randomly sampled TRS (Random). The white cross represents the mean predictive ability over the 100 replicates.

Supplementary Files

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

- 20230529SupplementaryMaterial.pdf