Genomic prediction of hybrid performance: comparison of the efficiency of factorial and tester designs used as training sets in a multiparental connected reciprocal design for maize silage

Alizarine Lorenzi, Cyril Bauland, Tristan Mary-Huard, Sophie Pin, Carine Palaffre, Colin Guillaume, Christina Lehermeier, Alain Charcosset, Laurence Moreau

Genomic Prediction of Hybrid Performance: Comparison of the Efficiency of Factorial and Tester Designs Used as Training Sets in a Multiparental Connected Reciprocal Design for Maize Silage

Alizarine Lorenzi (alizarine.lorenzi@inrae.fr)
INRAE UMR0320: Genetique Quantitative et Evolution Le Moulon https://orcid.org/0000-0002-0214-4206
Cyril Bauland
INRAE UMR0320: Genetique Quantitative et Evolution Le Moulon
Tristan Mary-Huard
Mathématiques et Informatique Appliquées-Paris: Mathematiques et Informatique Appliquees-Paris
Sophie Pin
INRAE UMR0320: Genetique Quantitative et Evolution Le Moulon
Carine Palaffre
UE0394: Unite experimentale du mais
Colin Guillaume
MAS seeds
Christina Lehermeier
RAGT Seeds Ltd
Alain Charcosset
INRAE UMR0320: Genetique Quantitative et Evolution Le Moulon
Laurence Moreau
INRAE UMR0320: Genetique Quantitative et Evolution Le Moulon https://orcid.org/0000-0002-7195-1327

Research Article

Keywords: Hybrid breeding, Genomic selection, Factorial design, Tester design, General combining ability, Specific combining ability

Posted Date: February 2nd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1261394/v1

Genomic prediction of hybrid performance: comparison of the efficiency of factorial and tester designs used as training sets in a multiparental connected reciprocal design for maize silage.

A. Lorenzi[1], C. Bauland[1], T. Mary-Huard[1,2], S.Pin[1], C. Palaffre[3], C. Guillaume[4], C. Lehermeier[5], A. Charcosset[1], L. Moreau[1]

[1] Université Paris-Saclay, INRAE, CNRS, AgroParisTech, Génétique Quantitative et Evolution - Le Moulon, 91190 Gif-sur-Yvette, France

[2] MIA, INRAE, AgroParisTech, Université Paris-Saclay, 75005, Paris, France, 3 UE 0394 SMH, INRAE, 2297 Route de l'INRA, 40390, Saint-Martin-de-Hinx, France

[3] UE 0394 SMH, INRAE, 2297 Route de l'INRA, 40390, Saint-Martin-de-Hinx, Franc

[4] Maïsadour Semences SA, F-40001 Mont de Marsan Cedex, France

[5] RAGT2n, Genetics and Analytics Unit, 12510 Druelle, France

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

Abstract

In maize breeding, the selection of the candidate inbred lines is based on topcross evaluations using a limited number of testers. Then, a subset of single-crosses between these selected lines is evaluated to identify the best hybrid combinations. Genomic selection enables the prediction of all possible single-crosses between candidate lines but raises the question of defining the best training set design. Previous simulation results have shown the potential of using a sparse factorial design instead of tester designs as the training set. To validate this result, a 363 hybrid factorial design was obtained by crossing 90 dent and flint inbred lines from six segregating families. Two tester designs were also obtained by crossing the same inbred lines to two testers of the opposite group. These designs were evaluated for silage in eight environments and used to predict independent performances of a 951 hybrid factorial design. At a same number of hybrids and lines, the factorial design was as efficient as the tester designs, and, for some traits, outperformed them. All available designs were used as both training and validation set to evaluate their efficiency. When the objective was to predict single-crosses between untested lines, we showed an advantage of increasing the number of lines involved in the training set, by (i) allocating each of them to a different tester for the tester design, or (ii) reducing the number of hybrids per line for the factorial design. Our results confirm the potential of sparse factorial designs for genomic hybrid breeding.

Key words

Hybrid breeding, Genomic selection, Factorial design, Tester design, General combining ability, Specific combining ability

Key Message (30 words)

Calibrating a genomic selection model on a sparse factorial design rather than on tester designs is advantageous for some traits, and equivalent for others.

Conflicts of interest

The authors declare no conflict of interest.

Acknowledgments

We thank Lidea, Limagrain Europe, Maïsadour Semences, Corteva, RAGT 2n, and Syngenta Seeds grouped in the frame of the ProMais SAM-MCR program for the funding, inbred lines development, hybrid production, and phenotyping. We are also grateful to scientists from these companies and to scientists of the INRAE “R2D2” network for helpful discussions on the results. A.L. phD contract was funded by RAGT 2n and ANRT contract n° 2020/0032 and receive support from the "Investissement d’Avenir" project "Amaizing" (Amaizing, ANR-10-BTBR-0001). GQE-Le Moulon benefits from the support of Saclay Plant Sciences-SPS (ANR-17-EUR-0007).

Introduction

Maize genetic diversity has been structured into complementary heterotic groups and varieties are mainly single-crosses between two inbred lines issued from different heterotic groups. In Northern Europe, the two main heterotic groups are the flint and the dent groups. The straightforward way to identify the best single-cross hybrids would be to cross all inbred lines from each heterotic group following a complete factorial design. In breeding programs, the number of candidate lines is large and increases every year. This makes it impossible to generate and evaluate all possible single-cross hybrids. The development of Doubled-Haploid (DH) technology reinforces this difficulty due to the fact that it allows the fast production of a large number of fully homozygous inbred lines, compared to the use of selfing generations in a single-seeds descent process.

An approach to manage this situation is to preselect lines within each heterotic group in early stages of selection and to evaluate single-crosses between selected lines of each group in advanced stages of the process. In 1942, Sprague and Tatum introduced the partition of the hybrid value into general and specific combining abilities (GCA and SCA). The GCA of a line is defined as the average performance of its progeny in hybrid combinations and the SCA of a hybrid is the deviation from the expected performance based on the GCA of the parental lines. They illustrated the interest of topcross tests in the early breeding stages, (i.e. the evaluation of hybrid progeny obtained by crossing candidates from one group with few individuals from the other) for the preliminary evaluation of inbred lines and of single-cross tests in later stages to identify the best hybrid combinations. Today, in most hybrid breeding programs, variations of this two-step approach are used to improve simultaneously two parental populations in a recurrent reciprocal way. The process is divided into two stages. First, the candidate lines of one heterotic group are crossed to a limited number of individuals, usually inbred lines (one or few lines) from the opposite group which are called “testers”. The topcross hybrid progeny are then evaluated in the field and the best lines of each heterotic group are selected based on their GCA. In the second stage, the selected lines of each group are crossed following an incomplete factorial design and the best hybrid combinations are identified. However, a selection based on a few testers in the early breeding stages does not fully exploit the complementarity between groups, specifically SCA, and can bias the GCA estimation. The GCA of the inbred line is then confounded with the SCA of the topcross hybrids, especially when using only one tester. This two-stage breeding process increases the time required for marketable hybrid development and requires the phenotyping of a large number of lines (at least as many hybrids as lines in each group in the first step).

Since the resources available for phenotyping are limited, a major goal in hybrid selection is the prediction of untested hybrid performance. Until the 1990s, selection was conducted without knowing genes or loci implied in GCA and SCA components. The development of markers enabled the identification of genes or loci implied in quantitative traits variations (QTL detection) and opened the way to performance prediction based on marker information (Lande and Thompson 1990). Several approaches using markers in selection were developed, one of which is genomic selection (GS or genomic prediction) (Meuwissen et al. 2001). Implementation of GS requires the development of a training set (TRS) consisting of individuals both phenotyped and genotyped. The TRS is used to calibrate a prediction model of the value of individuals that have only been genotyped. Among the proposed GS models, one consists in using markers to estimate additive genomic relationships (kinship matrix) between

individuals. Then this matrix is used in a mixed model to predict the performance of unphenotyped individuals using the performance of phenotyped ones. An adaptation of this model for the prediction of hybrid values was first proposed by Bernardo (1994). For predicting the GCAs and SCAs of unphenotyped hybrids, he used marker-based distances between parental lines of hybrids and the performance of a related set of single-crosses. Several models adapted to the hybrid framework have been proposed more recently, modeling either the GCA and SCA components or the additive, dominance and epistasis effects (Vitezica et al. 2013, 2017; Varona et al. 2018; González-Diéguez et al. 2021). It has been shown that the prediction accuracy of genomic selection can be affected by various factors such as the trait heritability, the number of markers (Heslot et al. 2012), the statistical model, the calibration population size (Jannink et al. 2010; Technow et al. 2014; Seye et al. 2020), the relationship between the training set (TRS) and the prediction set (Saatchi et al. 2010; Albrecht et al. 2011; Pszczola et al. 2012; Technow et al. 2014; Kadam et al. 2016; Seye et al. 2020). In the particular framework of hybrid prediction, other factors affect prediction accuracies, such as including SCA in prediction models and the optimization of the TRS regarding the number of hybrids phenotyped and the number of parental lines contributing to these hybrids (Technow et al. 2014; Seye et al. 2020).

Studies confirmed the usefulness of genomic selection models to predict single-cross hybrid values in maize (see Kadam and Lorenz 2018 review), but most studies on the prediction of single-cross hybrids addressed only the last step of the selection process, i.e. the identification of the best hybrid combinations produced from crossings among the selected lines. The use of markers and especially the use of GS offers new prospects for improving the hybrid breeding scheme. A promising lead, first proposed by Giraud (2016) and Kadam et al. (2016), would be to replace topcross evaluations at an early stage by genomic predictions calibrated on a sparse factorial design between candidate lines. This would allow the identification of superior single crosses early in the hybrid breeding pipeline by (i) predicting all potential hybrid combinations using GS, (ii) exploiting the complementarity between the two heterotic groups in early stages and (iii) getting rid of the tester bias. This approach requires genotyping candidate lines and producing hybrids by hand-made pollination which is challenging and costly. Therefore, even if this approach is appealing (Kadam et al. 2016; Giraud et al. 2017a, b; Fristche-Neto et al. 2018; Kadam et al. 2021) and its efficiency compared to the tester approach was assessed by simulations (Seye et al. 2020), further investigations and experimental validation are needed. The first experimental study (Fristche-Neto et al. 2018) that investigated the influence of the mating design to build the TRS in maize breeding showed a clear advantage of a factorial or a diallel over a tester design. But the designs used as TRS had different sizes which might affect conclusions. Recently an experimental study compared the use of a topcross progeny to the use of randomly paired single-cross progeny as TRS for genomic predictions by cross-validations (Burdo et al. 2021). This study relied on two synthetic populations from Iodent and Stiff Stalk heterotic groups evaluated for grain yield performances.

In the present study, our objective is to further evaluate the interest of using a factorial design instead of a tester design as TRS in early stages of the breeding pipeline. Original factorial and tester experimental designs were produced to correspond to the population structure of the candidate lines generated by breeders in the early stages of the process (unselected lines): in each heterotic group (the flint and the dent), segregating lines were derived from biparental families, themselves issued from intercrossing four founder lines. Two factorial designs

differing in their composition (number of lines and number of hybrids per line) were generated as well as two tester designs derived from the same inbred line populations. All these designs were used as TRS or VS to evaluate their potential for calibrating prediction models for different prediction objectives. The aim of this study was to (1) assess the efficiency of using a sparse factorial design as calibration set to predict new hybrid combinations obtained with a tester versus a more complete factorial design, and compare its ability to predict hybrids derived from selected lines, (2) compare the efficiency of using a factorial or tester designs of equal size to calibrate GS models and evaluate different tester designs' compositions, and (3) assess the impact of the composition of the factorial design in terms of the number of hybrids per line on the prediction accuracy.

Materials and Methods

Genetic material

This study relies on four different experimental designs.

The first one, referred to as F-1H was already analyzed in previous studies (Giraud et al. 2017a, b; Seye et al. 2019). It is a factorial design derived from two multiparental populations, each corresponding to one of the major heterotic groups used for silage maize breeding in Northern Europe: the flint and the dent. In each heterotic group, three founder lines were chosen for their agronomical performances for silage production (F373, F03802, F02803 for the flint group and F98902, F1808, F04401 for the dent group) and one for its silage quality (F7088 for the flint group and F7082 for the dent group). Six biparental families were derived from the six F1 hybrids produced by intercrossing the four founder lines of each heterotic group. The dent biparental families were obtained by doubled haploidization and the flint ones were obtained by five to six generations of selfing using a single-seed descend process. A total of 822 flint lines and 801 dent lines were derived and crossed following a sparse factorial design to produce 951 flint-dent single-cross hybrids. Each parental line contributed to one or two hybrids (20% of the lines contributed to two hybrids), therefore this design will be referred to as F-**1H**. The F-1H is balanced between families: 22 to 35 hybrids were produced from each biparental family combination. For more details see Giraud et al. (2017a).

The F-4H and the tester designs were produced from crossing a subset of the parental lines of the F-1H. In each heterotic group, 60 lines were chosen randomly in a balanced manner (10 lines per family) and 30 lines were selected based on genomic predictions obtained in the F-1H for an index combining silage yield, moisture content at harvest and silage quality. This index corresponds to the one used for silage hybrids registration in France (Seye 2019). Note that in each heterotic group only three families (out of six) were represented in the selected lines. An incomplete factorial design composed of 363 hybrids was produced by crossing randomly (i) the 30 flint selected lines to the 30 dent selected lines to produce 131 hybrids (further called "selected hybrids") and (ii) the 60 random dent lines to the 60 random flint lines leading to 232 hybrids (further called "random hybrids"). In this design each parental line contributed to produce generally four hybrids, therefore it will be referred to as F-**4H**. Note that the F-1H and the F-4H were issued from the same inbred line populations with the difference being their composition in terms of the number of lines and number of hybrids per line: the number of

hybrids per line was higher in the F-4H than in the F-1H. The same 90 dent lines were crossed to two flint testers to produce the 180 hybrids referred to as the T-**D** design and the 90 flint lines were crossed to two dent testers to produce the 180 hybrids of the T-**F** design. The testers used were two of the four founder lines from the opposite group (F1808 and F98902 for the dent testers and F373 and F02803 for the flint testers) that were chosen to be genetically distant and with good yield potential.

Field trials

The hybrids were evaluated in eight trials in the North of France and Germany. Hybrids from the F-1H were evaluated in four trials in 2013 and four in 2014, and hybrids from the F-4H and the tester designs were evaluated in three trials in 2016 and five in 2017. Trials were conducted by INRAE and seven private breeding companies (Lidea, Corteva, Maisadour, KWS, RAGT, Limagrain, Syngenta). The field experiments were laid out as augmented partially replicated designs (Williams et al. 2011). In each trial, two types of hybrids were used as controls: two commercial hybrids (LG30.275 and RONALDINIO) and 16 founder hybrids that were produced by crossing the founder lines of each heterotic group. In each trial, the controls were evaluated twice, as well as 20% of the experimental hybrids. The F-1H was evaluated in trials composed of 1,088 elementary plots distributed in 68 incomplete blocks. One block was composed of 16 plots, four to five of these were used for replicated genotypes. Among the 2013 and 2014 trials, on average each experimental hybrid was seen in seven trials and was replicated within at least one trial. See Giraud et al. (2017a) for more details on the F-1H. The F-4H and the tester designs were evaluated jointly in the same trials. Each trial was composed of 800 elementary plots laid out in 50 incomplete blocks. Among the 50 blocks, 26 were allocated to factorial hybrids and 24 to tester hybrids (six consecutive blocks per tester). Factorial blocks and tester blocks were grouped to limit potential competition between hybrids due to a design effect. One block was composed of 16 plots, four to five of these were used to replicate genotypes. Among the 2016 and 2017 trials, on average each experimental hybrid was seen in seven trials and was replicated within at least one trial.

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) (Andrieu 1995; Peyrat et al. 2016), cell wall content of the harvested dry matter measured by the neutral detergent fiber content (NDF in % of dry matter), 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), 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 %). The DINAG and DINAGZ are two digestibility criteria, first proposed by Argillier et al. (1995). The silage quality traits were predicted using Near-Infrared Reflectance Spectrometry (NIRS) equations on silage powders, or directly in the field at harvest, depending on the practices of each breeding company.

By inspection of raw data and field observations (from field trial visits), outliers were identified and considered as missing data. Then filters were applied, plots with abnormal standing counts (below 80% of the median), with DMC below 25% and above 45% were considered as missing data (NIRS predictions being

considered as unreliable for extreme moisture (Baker et al. 1994)). In total over the different traits, after inspection and filters, the percentage of missing data was equal to 11%.

Variance decomposition on single-plot performances (without marker information)

The individual single plot performance was corrected by the BLUPs of spatial effects predicted using models described in detail in ***File S1***. Variance components were estimated on the single plot performance corrected by spatial effects independently for each design using Model (1.1) for the factorial designs and Model (1.2) for the tester designs.

The model implemented on the factorial designs (F-1H and F-4H) was:

$$Y_{hkk'lxyz} = \mu + \lambda_l + (\tau_h + \rho_{lh}) \times t_h + (H_{h(kk')} + H\lambda_{lh(kk')}) \times (1 - t_h) + E_{hkk'lxyz}, \quad (1.1)$$

where $Y_{hkk'lxyz}$ is the phenotypic value of hybrid h produced by crossing the parental lines k and k', evaluated in environment l, located at row x, column y and in block z. μ is the intercept, λ_l is the fixed effect of environment l, t_h distinguishes the type of hybrid, it is set to 0 for the experimental hybrids and set to 1 for the control hybrids (commercial or founder hybrids), τ_h is the fixed factor with 18 levels corresponding to the control hybrids, ρ_{lh} is the effect of the interaction between environment l and control hybrid h. $H_{h(kk')}$ is the random genetic effect of experimental hybrid h produced by crossing the flint line k and the dent line k'. $H_{h(kk')}$ is decomposed into its GCA and SCA components as follows:

$$H_{h(kk')} = U_k + U'_{k'} + S_{kk'},$$

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

$$H\lambda_{lh(kk')} = (U\lambda)_{kl} + (U'\lambda)_{k'l} + (S\lambda)_{kk'l},$$

where $(U\lambda)_{kl}$ and $(U'\lambda)_{k'l}$ are the random effects of the flint GCA effect by trial interaction, respectively dent GCA by trial interaction and $(S\lambda)_{kk'l}$ is the random effect of the SCA by trial interaction. With $(U\lambda)_{kl} \backsim \mathcal{N}(0, \sigma^2_{GCA_f \times E})$iid, $(U'\lambda)_{k'l} \backsim \mathcal{N}(0, \sigma^2_{GCA_d \times E})$iid and $(S\lambda)_{kk'l} \backsim \mathcal{N}(0, \sigma^2_{SCA \times E})$iid. $\sigma^2_{GCA_f \times E}$, $\sigma^2_{GCA_d \times E}$ and $\sigma^2_{SCA \times E}$ are the flint GCA by trial interaction variance, the dent GCA by trial variance and the SCA by trial interaction variance, respectively. $E_{hkk'lxyz}$ is the error term; we assume that the errors follow: $E_{hkk'lxyz} \sim \mathcal{N}(0, \sigma^2_{El})$ and are independent and identically distributed within trial and independent between trials, σ^2_{El} is the error variance of environment l. The different random effects of the model are assumed to be independent.

The model implemented on the T-F was:

$$Y_{hkmlxyz} = \mu + \lambda_l + (\tau_h + \rho_{lh}) \times t_h + (\gamma_m + H_{h(km)} + H\lambda_{lh(km)}) \times (1 - t_h) + E_{hkmlxyz}, \quad (1.2)$$

where λ_l, t_h, τ_h and ρ_{lh} are defined as in Model (1.1). Y_{hlmxyz} is the phenotypic value of hybrid h produced by crossing the dent founder line m used as tester and the flint parental line k, evaluated in environment l located at row x, colomn y and in block z. γ_m is the fixed effect of line m used as tester. $H_{h(km)}$ is the random genetic effect of hybrid h produced by crossing the dent founder line m used as tester and the flint parental line k evaluated for its GCA. The genetic value of hybrid h, $H_{h(km)}$, is decomposed as follows: $H_{h(km)} = U_k + S_{km}$, where U_k is the random GCA effect of the flint line k, with $U_k \sim \mathcal{N}(0, \sigma^2_{GCA_f})$ iid and S_{km} is the random effect of the interaction (SCA) between the flint line k and the founder dent line m used as tester, with $S_{km} \sim \mathcal{N}(0, \sigma^2_{SCAt})$ iid. $H\lambda_{lh(kk')}$ is the genotype by trial interaction, decomposed as follows: $H\lambda_{lh(kk')} = (U\lambda)_{kl} + (S\lambda)_{kml}$ where $(U\lambda)_{kl} \backsim \mathcal{N}(0, \sigma^2_{GCA_f \times E})$ iid and $(S\lambda)_{kml} \backsim \mathcal{N}(0, \sigma^2_{SCAt \times E})$ iid. $E_{hkmlxyz}$ is the error term. We assume that the errors follow $E_{hkmlxyz} \sim \mathcal{N}(0, \sigma^2_{E_l})$ iid within trial and independent between trials. The different random effects of the model are assumed to be independent. The same model was adapted and implemented on T-D.

Variance components were estimated with each model and a likelihood ratio test was performed to test their significance with adjusted p-values corresponding to mixed chi-square distributions (Self and Liang 1987; Molenberghs and Verbeke 2007) using the "lrt.asreml" function of the ASReml-R package (setting the parameter "boundary" to TRUE for the mixed chi-square distributions).

The percentage of genetic variance due to SCA was estimated (%) and broad-sense heritability was computed as follows:

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

where σ^2_H is the hybrid genetic variance. For the factorial designs it is computed as: $\sigma^2_H = \sigma^2_{GCA_f} + \sigma^2_{GCA_d} + \sigma^2_{SCA}$ and for the T-F (respectively T-D) it corresponds to the $\sigma^2_{GCA_f}$ (respectively $\sigma^2_{GCA_d}$). $\sigma^2_{H \times E}$ is the total genotype by trial variance decomposed as: $\sigma^2_{H \times E} = \sigma^2_{GCA_f \times E} + \sigma^2_{GCA_d \times E} + \sigma^2_{SCA \times E}$ for the factorial designs and as: $\sigma^2_{H \times E} = \sigma^2_{GCA_f \times E}$ for the T-F (respectively $\sigma^2_{H \times E} = \sigma^2_{GCA_d \times E}$ for the T-D). $\sigma^2_{E_{moy}}$ is the mean residual variance across all trials, n_{site} is the number of trials and n_{rep} is the mean number of within trial replicates across trials.

Adjusted means

For each trait and each design, least square-means (ls-means or adjusted means) of the hybrids were computed over trials, using a model considering the hybrid genetic effect as fixed:

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

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 environment l. μ is the intercept, λ_l is the fixed effect of environment l, γ_h is the fixed genetic effect of hybrid h. E_{hrl} is the error term of environment l, with $E_{hrl} \sim \mathcal{N}(0, \sigma^2_{El})$ iid within trial and independent between trials. All genomic predictions were performed on the ls-means thus obtained.

Genotyping and kinship estimation

The founder lines as well as the parental lines were genotyped for 18,480 SNPs using an Affymetrix® array provided by Limagrain. After quality control, 9,548 SNP markers were conserved and mapped on a consensus map (Giraud et al. 2017a).

Kinship matrices for the flint and dent GCA ($\boldsymbol{K_{GCA_f}}$ and $\boldsymbol{K_{GCA_d}}$) were computed for all parental lines following method 1 from VanRaden (2008). The coefficient of the flint GCA kinship between individuals i and i' was estimated as follows:

$$K_{GCA_f(i,i')} = \frac{\sum_{m=1}^{M}(G_{im}-f_m)(G_{i'm}-f_m)}{\sum_{m=1}^{M} f_m(1-f_m)}, \quad (3)$$

where G_{im} is the genotype of the flint line i at locus m (coded 0, 0.5 and 1) and f_m is the allele frequency of allele "1" at locus m estimated on the whole dataset. The kinship matrix $\boldsymbol{K_{GCA_d}}$ was computed similarly. The coefficient of the SCA kinship matrix ($\boldsymbol{K_{SCA}}$) between two flint-dent hybrids, produced from the crossings of parental lines i to j and parental lines i' to j', was computed as follows (Stuber and Cockerham 1966):

$$K_{SCA_{(ij,i'j')}} = K_{GCA_f(i,i')} * K_{GCA_d(j,j')}. \quad (4)$$

Genomic Best Linear Unbiased Prediction (GBLUP) models

Two GBLUP models were implemented for genomic predictions depending on the design used as TRS (factorial or tester).

The model implemented on the factorial designs (F-1H and F-4H) including SCA effects was:

$$\boldsymbol{y} = \boldsymbol{1_n}.\mu + \boldsymbol{Z_d}\,\boldsymbol{g_{GCA_d}} + \boldsymbol{Z_f}\boldsymbol{g_{GCA_f}} + \boldsymbol{Z_{df}}\boldsymbol{g_{SCA_{df}}} + \boldsymbol{E}, \quad (5.1)$$

where $\boldsymbol{y}$ is the vector of ls-means of the n phenotyped hybrids, $\boldsymbol{1_n}$ 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). $\sigma^2_{GCA_f}$ 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. $\boldsymbol{Z_d}$, $\boldsymbol{Z_f}$ and $\boldsymbol{Z_{df}}$ are the corresponding incidence matrices of dimensions $[n \times n_d]$, $[n \times n_f]$, and $[n \times n]$ respectively, that relate the observations to the GCA and SCA effects of lines and single-cross hybrids considered in the model. $\boldsymbol{E}$ 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. A model without SCA effects was also considered.

The model implemented on the T-F (the same model was adapted and implemented on the T-D) was:

$$\boldsymbol{y} = \boldsymbol{1_n}.\mu + \boldsymbol{Xv} + \boldsymbol{Z_f g_{GCA_f}} + \boldsymbol{Z g_{SCAt}} + \boldsymbol{E}, \quad (5.2)$$

where $\boldsymbol{y}$ is the vector of ls-means of the n phenotyped hybrids, $\boldsymbol{1_n}$ is a vector of n ones and μ is the intercept. $\boldsymbol{v}$ is the vector of fixed effects of the n_t testers. $\boldsymbol{g_{GCA_f}}$ is the vector of random GCA effects of the n_f flint parental lines, with $\boldsymbol{g_{GCA_f}} \sim \mathcal{N}(0, \boldsymbol{K_{GCA_f}}\sigma^2_{GCA_f})$ 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 tester, with $\boldsymbol{g_{SCAt}} \sim \mathcal{N}(0, \boldsymbol{I_2} \otimes \boldsymbol{K_{GCAf}}\,\sigma^2_{SCAt})$ where σ^2_{SCAt} is the SCA variance. $\boldsymbol{X}, \boldsymbol{Z_f}$ and $\boldsymbol{Z}$ are the corresponding incidence matrices of dimensions $[n \times n_t]$, $[n \times n_f]$, and $[n \times n_t n_f]$ respectively. $\boldsymbol{E}$ 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.

Predictive ability with different scenarios

Three objectives were investigated using three different scenarios.

In Scenario 1, the aim was to evaluate the efficiency of using a sparse factorial design mostly composed of one hybrid per line to predict new hybrid combinations evaluated in new environments. The F-1H, was used as TRS to predict the F-4H, T-F and T-D. The predictive ability was computed as the correlation between the observed phenotypes (ls-means) and the predicted hybrid values (sum of the predicted GCA and SCA BLUPs when included in the model). In the tester designs, for a given line, two ls-means values were available (one ls-mean per tester). Therefore, the predictive ability was computed as the correlation between the mean of the two ls-means and the predicted GCA BLUP of the same line. Predictive abilities were also computed separately for the selected and the random hybrids to evaluate the quality of prediction among hybrids derived from selected lines. In this scenario, we evaluated the impact of the level of relationship between the TRS and the validation set (VS) on predictions. To vary the level of relationship between the TRS and the VS, we used four different TRS constituted of hybrids sampled within the F-1H to predict the same VS (F-4H) ***(Fig. 2)***. Hybrids included in the TRS were sampled to select those issued or not from parental lines contributing to the VS plus others to reach 742 hybrids and to preserve the balance between families in each TRS. The four TRS led us to consider the prediction of: T0 hybrids where none of the VS parental lines contributed to the TRS, T1 hybrids where only one of the VS parental lines contributed to the TRS and T2 hybrids where both VS parental lines contributed to the TRS. Each TRS was sampled 10 times and the mean of the quality of prediction over the 10 repetitions was computed.

In Scenario 2, we compared the efficiency of calibrating a GBLUP model with a factorial or a tester design. The F-4H (composed of 363 hybrids) or the tester designs (360 hybrids) were used to calibrate respectively models (5.1) or (5.2) to predict the F-1H design. The predictive ability was computed as the correlation between the observed phenotypes (ls-means) and the predicted hybrid value (sum of the predicted GCA and SCA BLUPs when included in the model). The GCA BLUPs predicted when calibrating on the factorial design were correlated to the ones predicted with each of the tester designs. To compare the similarity of selection between the different approaches (based on phenotypic evaluations (ls-means) or genomic predictions (BLUPs) calibrated on the factorial or the tester designs) the coincidence of selection was computed for yield (DMY). For each pair of approaches, it corresponds to the percentage of common hybrids that would be selected by the two approaches for a given selection rate (%). The coincidences of selection were computed for different selection rates. We also investigated the impact of the composition of the tester designs by considering designs composed of one or two testers. Different TRS were considered, each composed of 180 hybrids: (i) 180 hybrids produced by crossing 90 lines to one tester in each group (180 lines in total), since there were two testers in each group, there were four possible tester combinations to predict an hybrid, referred to as 1T-180H-180L-A, 1T-180H-180L-B, 1T-180H-180L-C and 1T-180H-180L-D, (ii) 180 hybrids produced by crossing 45 lines to one tester and the 45 other lines to the other tester in each group, referred to as 2T-180H-180L, (iii) 180 hybrids produced by crossing 45 lines to the first tester and the same 45 lines to the second tester in each group referred to as 2T-180H-90L. To make it comparable 180 hybrids were sampled within the F-4H in a random and balanced manner between families, with the objective of maximizing the number of lines. This led to sample 152 lines (76 dent and 76 flint), and on average one line contributed to 2.4 hybrids. This scenario was called the F-180H-152L.

In Scenario 3, the difference of composition of the two factorials was exploited to investigate its impact on predictions. At the same number of hybrids, there are more lines and fewer hybrids per line in the F-1H (1.2 hybrids per line on average) compared to the F-4H (3.5 hybrids per line on average). The F-1H and the F-4H were used in turn as TRS and VS. The two factorial designs were sampled to the same number of hybrids (216 hybrids) but represented a different number of lines (207 dent lines and 209 flint lines for the F-1H and 60 dent lines and 60 flint lines for the F-4H). Sampling was done such that hybrids of the two sets had no common parental lines (only hybrids produced by crossing lines that did not contribute to TRS were predicted), and were balanced between families. For the F-4H, sampling was done only among the random hybrids. 100 samples of 216 hybrids in each factorial design were considered and the mean of the 100 predictive abilities was computed.

To test the significance of the differences between predictive abilities, William's tests (Williams 1959) were performed using the "r.test" function of the *psych* R-package (Revelle 2021) for dependent correlations with a common variable. All models were implemented using the ASReml-R package (version 4) (Butler 2019; R Core Team 2020).

Results

For clarity purpose, results on only four traits (DMY, DMC, DtSilk and MFU) are presented in the following. The results on the 11 studied traits are presented in supplementary materials.

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

Broad-sense heritabilities (H^2) at the design level were high for all traits and all designs ***(Table 1 and Table S1)***. For the four main traits, they ranged from 0.79 (MFU) to 0.91 (DMC and DtSilk) for the F-1H, from 0.86 (MFU) to 0.93 (DMC and DtSilk) for the F-4H, from 0.85 (MFU) to 0.89 (DMC) for the T-D and from 0.78 (MFU) to 0.94 (DMC) for the T-F. For a given trait, heritabilities were similar in both factorial designs. Variance decomposition at the phenotypic level without marker information showed large and significant genetic variances for all designs and all traits (***Table 1, Table S1***). For all traits and all designs, the SCA variance was lower than the sum of the GCA variances. The percentage of genetic variance due to SCA ranged from 10% (MFU) to 20% (DMY) for the F-1H, from 3% (MFU) to 9% (DMY) for the F-4H design, from 11% (DtSilk and MFU) to 21% (DMY) for the T-D and from 13% (DMC) to 20% (DMY) for the T-F.

Variance components obtained using marker information

Variance decomposition on adjusted means using GBLUP models showed large and significant genetic variances for all traits and all designs ***(Table 2, Table S2)***. The decomposition of the genetic variance into GCA and SCA components showed that most of the hybrid variation was due to GCA variance. For the four main traits of interest, the percentage of genetic variance due to SCA ranged from 0% (DMC and MFU) to 11% (DMY) for the F-1H, from 0% (DMC) to 7% (UFL) for the F-4H, from 4% (DMC, DtSilk) to 15% (DMY) for the T-D and from 5% (MFU) to 10% (DMY) for the T-F. These values were lower (except for MFU in F_4H) than the ones estimated based on phenotypic data only. Adding SCA effects in the model induced no or minor changes in the genetic variance components, but reduced slightly the residual variance.

Scenario 1- Using a sparse factorial design (F-1H) to predict new hybrid combinations in new environments

The predictive abilities obtained for new hybrid combinations in new environments (F-4H, T-D and T-F) when calibrating on the F-1H were high for all traits (***Table 3, Table S3***). Considering a prediction model without SCA, they ranged from 0.78 (DMY) to 0.82 (MFU) when predicting all hybrids (Selected + Random) of the F-4H, from 0.74 (DMC) to 0.80 (DMY) for T-D and from 0.69 (DMY) to 0.84 (DMC and DtSilk) for T-F. When predicting the tester designs, the ability to predict the GCA_d was higher than the ability to predict the GCA_f for eight traits out of 11, differences ranged from 0.01 (MFU) to 0.11 (DMY). Considering the SCA in the model did not improve the quality of predictions. Predictive abilities were computed for each hybrid type of the VS (random or selected). They were generally higher for the random hybrids compared to the selected hybrids (hybrids produced from

crossing two selected parental lines) for all VS and all traits but DMC and DMY *(**Table 3**)*. Differences between the predictions of random and selected hybrids were greater for DMY, DtSilk, NDF, CELL and HCELL.

The ability of the F-1H to predict T0, T1 or T2 hybrids was estimated by considering different TRS (***Fig. 3***). The lowest predictive abilities were obtained when predicting T0 hybrids for the four traits presented (and for seven traits out of 11) and ranged from 0.70 (DMC and DtSilk) to 0.79 (MFU). The highest predictive abilities were obtained when predicting T2 hybrids for all presented traits (and for eight traits out of 11), they ranged between 0.77 (DMY and DMC) and 0.82 (MFU). Predictive abilities obtained for T1 hybrids were on average intermediate compared to the ones obtained for T0 and T2 hybrids. For a given trait, differences in predictive abilities between T0, T1 and T2 hybrids were small especially for DMY (ranging from 0.74 to 0.77) and MFU (ranging from 0.79 to 0.82). On average over the 11 traits, the quality of prediction was higher for the T1F hybrids than for the T1D hybrids for all traits (except for DINAG), differences ranged from 0.02 (DINAG) to 0.22 (HCELL) between the highest and the lowest quality of prediction.

Scenario 2- Compare the efficiency of factorial versus tester designs as TRS

We compared the predictive abilities that can be achieved using either the F-4H (363 hybrids) or the tester designs (360 hybrids) as TRS to predict all hybrids of the F-1H *(**Fig. 4**)*. They ranged from 0.55 (DtSilk) to 0.67 (MFU) for the calibration on the factorial design and from 0.54 (DMY) to 0.65 (DMC and MFU) for the calibration on the tester designs. Calibrating on the F-4H gave better predictive abilities than calibrating on the tester designs for eight traits out of 11 and on average over the 11 traits *(**Table S4**)*, but differences were small. Including SCA effects in the GBLUP model did not improve the quality of prediction. The GCA BLUPs predicted using the F-4H as TRS were well correlated to the GCA BLUPs predicted using the tester designs as TRS, correlations ranged from 0.85 (GCA_f for DMY) to 0.95 (GCA_f for DMC) *(**Table 4, Table S5**)*. The coincidence of selection computed for hybrid predictions obtained with the factorial or the tester approaches was 53% for a selection rate of 5% (***Fig.S1***). This illustrated that at this selection rate, the two approaches did not select the same single-cross hybrids. For selection rates below 5%, predictions based on the factorial identified a higher proportion of hybrids that were the top-ranked ones based on their observed performances (ls-means) (***Fig S1***). The predictive abilities were also computed for the different types of hybrids constituting the VS: T0 or T1 hybrids. Overall, predictive abilities of T1 hybrids were higher than the ones of T0 hybrids for all traits and all designs *(**Table 5, Table S6**)*. The differences in predictions between the factorial and tester designs were similar when predicting all, only T1 or T0 hybrids.

At the same number of hybrids (180 hybrids), the impact of the composition of the tester designs used as TRS was investigated. We compared the predictive abilities obtained using either one or two testers or using a factorial design as TRS to predict all 951 hybrids of the F-1H. Across all presented traits and all TRS, the predictive abilities ranged from 0.48 (DtSilk, 2T-180H-90L) to 0.66 (MFU, 2T-180H-180L) *(**Fig. 5**)*. When the TRS was composed of one tester per group, the average predictive ability over the four tester combinations (1T-180H-180L-A, 1T-180H-180L-B, 1T-180H-180L-C, 1T-180H-180L-D) ranged from 0.51 (DMY) to 0.64 (DMC). Predictive abilities varied between the four different combinations of testers (A, B, C and D) and the best combination was different depending on the trait. Differences between the lowest and the highest ranged from 0.005 (DMC) to 0.048

(DMY). Qualities of prediction obtained with the 2T-180H-180L were higher than those obtained with the 2T-180H-90L for all presented traits (and for nine traits out of 11 studied traits). On average, the quality of prediction obtained with 2T-180H-180L as TRS was higher than the ones obtained with the best TRS combination composed of one tester per group (1T-180H-180L-A, B, C, D). At the same number of hybrids, there was always an advantage at using the F-180H-152L instead of the 2T-180H-90L or the best 1T-180H-180L. On average over the 11 traits, predictive abilities were similar between the F-180H-152L and the 2T-180H-180L. It should be noted that predictive abilities obtained when calibrating on 180 hybrids (F-180H-152L) were similar or only slightly lower than the ones obtained when using the whole set of 363 hybrids of the F-4H as TRS. Differences ranged from 0 (DMY) to 0.05 (DtSilk) (***Fig. 4*** and ***Fig. 5***).

Scenario 3- Assess the impact of the number of hybrids per line in factorial designs when predicting T0 hybrids

To investigate the impact of the number of lines involved in the calibration set at the same number of hybrids, the F-4H and the F-1H were used in turn as calibration sets to predict T0 hybrids (no common parental lines between the TRS and the VS) *(**Fig. 6, Table S7**)*. The predictive abilities when calibrating on 216 hybrids of the F-1H varied between 0.55 (DMY) and 0.71 (MFU). The predictive abilities when calibrating on 216 hybrids of the F-4H varied between 0.44 (DtSilk) and 0.65 (MFU). For all presented traits (and for nine traits out of 11 and on average), the quality of prediction was higher when calibrating on 216 hybrids of the F-1H (207 dent lines and 209 flint lines) than when calibrating on 216 hybrids of the F-4H (60 dent lines and 60 flint lines) to predict T0 hybrids of the other design. Thus, at a given TRS size of 216, there was an advantage at using more lines and fewer hybrids per line to predict T0 hybrids.

Discussion

Importance of SCA and its prediction

The proportion of SCA variance estimated in the designs using the model without marker information (Model 1.1 and 1.2) was small for all traits, from 0 to 20% for the F-1H and from 0 to 9% for the F-4H. This result is expected in hybrids produced by crossing lines from divergent populations (heterotic groups) (Reif et al. 2007). This relatively small importance of SCA effects compared to GCA effects is consistent with the fact that no SCA QTL could be detected in the F-1H design (Giraud et al. 2017b; Seye et al. 2019). Higher SCA proportions were obtained for the F-1H than for the F-4H, which might be due to differences in environmental conditions or to a different sampling of hybrids from the inbred lines. It should be noted that the estimation of the SCA variance in the F-1H was less accurate (***Table 1***) than in the F-4H design due to the fact that most of the inbred lines contributed to producing only one hybrid. This made it more difficult to separate the GCA from the SCA effects. Therefore, we can assume that the proportion of SCA variance was over-estimated in the F-1H. The proportion of SCA variance in the designs was also estimated through GBLUP models, by including marker information to compute GCA and SCA kinships. The estimated proportion of SCA was lower using the GBLUP model compared to the model without marker information. One possible explanation could be that the GBLUP model was not able to efficiently capture the SCA variance component.

Including SCA effects in the GBLUP models did not improve the predictions. This is consistent with the variance decomposition that showed little SCA variance and was also observed in other studies using data from inter-heterotic single-cross hybrids (Schrag et al. 2006, 2018; Technow et al. 2014; Kadam et al. 2016; Westhues et al. 2017; González-Diéguez et al. 2021). In fact, few studies have shown an increase in single cross prediction accuracy by modeling SCA effects (Technow et al. 2012; Dias et al. 2018). Kadam et al. (2016) reported that including SCA effects in GS models led to higher quality of predictions when predicting hybrids with untested parents (T0) compared to hybrids with one or two tested parents (T1 or T2) but this result was not confirmed in our study. We computed the SCA kinship matrix coefficient as the product, term to term, of the GCA kinship coefficients between the dent and the flint parental lines, similarly to what has been done in previous studies (Bernardo 1994; Technow et al. 2014; Seye et al. 2020; Kadam et al. 2021). González-Diéguez et al., (2021) showed that this computation of the SCA kinship matrix does not capture the whole SCA but only part of it, i.e. the additive-by-additive intergroup epistasis component ($G_{AA}(1,2)$). González-Diéguez et al., (2021) proposed a new SCA kinship formula, but their results as well as ours (results not shown) showed that it did not improve the predictive ability for inter-heterotic group single-cross hybrids.

Efficiency of using a factorial design to predict new hybrids in new environments

We showed in our study that by using a factorial design composed of only one hybrid per line to calibrate GS models achieves good predictive abilities of new hybrids in new environments. We observed good predictive abilities when predicting both factorial and tester designs. The high predictive abilities found when predicting the tester designs clearly illustrate that the genomic prediction model is able to decouple and predict GCAs, even when using a highly sparse factorial design where the inbred lines are parents of only one hybrid. This is in accordance with other studies showing high predictive abilities for yield even with very sparse factorials (for simulations Seye et al. 2020, for experimental data Burdo et al. 2021). When predicting the F-4H, predictive ability was highest for T2 single crosses, followed by T1 and T0 single crosses. This result was also reported in simulations (Technow et al. 2012; Seye et al. 2020) and in maize studies (Technow et al. 2014; Kadam et al. 2016). In the F-4H and the tester designs, two types of hybrids could be predicted: the random hybrids that were produced by crossing parental lines drawn at random from the segregating families, and the hybrids between selected lines that were produced by crossing two lines selected based on their GCA performance. The ability to predict random hybrids was higher than for hybrids between selected lines, which was expected as selection decreased the variance between selected hybrids. Nevertheless, the quality of prediction of the hybrids between selected lines was still high. This shows that GS models calibrated on a sparse factorial can efficiently predict the best hybrid combinations obtained by crossing lines already selected based on their GCA, which is of practical interest in breeding programs.

Efficiency of the factorial approach compared to the tester approach

The main objective of this study was to compare, at the same resource allocation (same number of hybrids), the efficiency of the factorial and tester approaches. To our knowledge our study is the first to compare the use of factorial and tester designs evaluated in the same environment in order to predict a distinct VS composed of new hybrids evaluated in new environments. For the same number of hybrids and lines, our results showed a slight advantage of the factorial design over the tester designs. The simulation study by Seye et al. (2020) showed that

the advantage of the factorial design increases with the proportion of SCA variance. In our study, the proportion of SCA variance was small, which could explain why we observed only a slight advantage of the factorial design. Moreover, the slight advantage of the factorial design over the tester design could also be explained by the fact that we used as testers two of the founder lines of the opposite group. As shown by simulations by Seye et al. (2020), using a founder line as tester reduces the advantage of factorial designs compared to tester designs. A recent study, also compared the potential of using a factorial instead of a tester approach as TRS (Burdo et al. 2021). They considered two multiparental synthetic populations of maize instead of segregating families and used only one tester. As in our study, non-additive effects were small and they used as tester one of the founder lines of the opposite group. They found an advantage of the factorial design over the tester design for some traits (flowering traits), but not for others (grain yield, plant height…). This is globally consistent with our results even if differences in their designs, traits, calibration set sizes and number of lines considered prevent a direct comparison with our results.

At the same number of hybrids, the impact of the composition of the tester designs was investigated and compared to a factorial design (Scenario 2). The use of only one tester revealed that the quality of prediction varied according to the tester used and that the best tester differed from one trait to another. Depending on the alleles carried by the tester, each tester is expected to mask part of the genetic variation for traits showing dominance. At the same number of hybrids, using more testers while maximizing the number of lines evaluated was always beneficial (the 2T-180H-180L TRS outperformed the 2T-180H-90L for nine traits out of 11 and all the 1T-180H-180L combinations). This strategy which maximizes the number of lines evaluated by crossing a given line to only one tester and using several testers in the tester designs is close to the one applied when considering a very sparse factorial design. Our experimental designs did not allow the direct comparison of a factorial design composed of only one hybrid per line to tester designs at the same number of hybrids. Nevertheless, results in Scenario 3 suggested that increasing the number of lines instead of increasing the number of hybrids per line was more efficient at the same number of hybrids when predicting T0 hybrids; this was also reported by Seye et al. (2020). Therefore, we hypothesize that a factorial composed of 180 hybrids and 360 lines would outperform all the tester designs in Scenario 2.

A major issue in breeding programs is resource allocation. It is therefore important to optimize the factorial design at a given number of hybrids. Our results showed the advantage of using more lines instead of more hybrids per line when predicting T0 hybrids, as also observed by Seye et al. (2020) with simulations. Yet, within that same study they also showed that when predicting T2 hybrids it was more efficient to use a factorial with four hybrids per line instead of one hybrid per line. In light of these preliminary results, we can argue that different objectives of selection could lead to considering different compositions for the factorial TRS to maximize the quality of prediction. Given our results, we hypothesize that using a factorial design composed of only one hybrid per line for preliminary screening and using a factorial design composed of four hybrids per line in a second screening would be optimal. This could be integrated into the two-part strategy proposed in the hybrid context by Powell et al. (2020). We propose using a factorial design composed of one hybrid per line in the population improvement part of the program and a factorial design composed of more than one hybrid per line in the product development part when the objective is to identify commercial hybrids.

What does it imply for breeding programs?

Our study revealed a significant advantage (Williams tests) of the factorial approach compared to the tester approach for some traits (DINAG, DINAGZ, PH) and at least an equivalence for the rest of the traits studied. According to our prediction results, topcross evaluations could be replaced by evaluations on a sparse factorial design and lead to similar or higher predictive abilities. At the same number of lines, a factorial design composed of one hybrid per line requires half as much phenotyping effort as the tester design. However, creating single-cross hybrids is more challenging than test-cross hybrids since hand-made pollination is necessary. Therefore, the factorial design could decrease the number of plots needed for phenotypic evaluation but its production could be more costly. A preliminary study conducted by Seye et al. (2020) showed that the increase in production costs would be compensated by the diminution in field plots needed.

This study relies on original experimental designs derived from segregating families, which allow the testing of several hypotheses. It has given some insights into the potential of replacing topcross evaluations by genomic predictions calibrated on a factorial design in breeding programs. Since our conclusions are closely related to the experimental designs and populations we considered (genetic variability available in the founder lines, number of founder lines…), future studies could take into consideration other populations. Another line of investigation would be the optimization of the TRS and the portability along breeding cycles.

References

Albrecht T, Wimmer V, Auinger H-J, et al (2011) Genome-based prediction of testcross values in maize. Theor Appl Genet 123:339–350. https://doi.org/10.1007/s00122-011-1587-7

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

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

Baker CW, Givens DI, Deaville ER (1994) Prediction of organic matter digestibility in vivo of grass silage by near infrared reflectance spectroscopy: effect of calibration method, residual moisture and particle size. Animal Feed Science and Technology 50:17–26. https://doi.org/10.1016/0377-8401(94)90006-X

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

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

Butler D (2019) asreml: Fits the Linear Mixed Model

Dias KODG, Gezan SA, Guimarães CT, et al (2018) Improving accuracies of genomic predictions for drought tolerance in maize by joint modeling of additive and dominance effects in multi-environment trials. Heredity 121:24–37. https://doi.org/10.1038/s41437-018-0053-6

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

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

Heslot N, Yang H-P, Sorrells ME, Jannink J-L (2012) Genomic Selection in Plant Breeding: A Comparison of Models. Crop Science 52:146–160. https://doi.org/10.2135/cropsci2011.06.0297

Jannink J-L, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. Briefings in Functional Genomics 9:166–177. https://doi.org/10.1093/bfgp/elq001

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

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

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

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

594 Molenberghs G, Verbeke G (2007) Likelihood Ratio, Score, and Wald Tests in a Constrained Parameter Space.
595 The American Statistician 61:22–27. https://doi.org/10.1198/000313007X171322

596 Peyrat J, Nozière P, Férard A, et al (2016) «Prévoir la digestibilité et la valeur énergétique du maïs fourrage :
597 Guide des nouvelles références ». ARVALIS - Institut du végétal - INRA

598 Powell O, Gaynor RC, Gorjanc G, et al (2020) A Two-Part Strategy using Genomic Selection in Hybrid Crop
599 Breeding Programs. bioRxiv 2020.05.24.113258. https://doi.org/10.1101/2020.05.24.113258

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

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

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

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

610 Saatchi M, Miraei-Ashtiani SR, Javaremi AN, Moradi-Shahrebabak M (2010) The impact of information quantity
611 and strength of relationship between training set and validation set on accuracy of genomic estimated
612 breeding values. African Journal of Biotechnolog 9:438–442. https://doi.org/10.5897/AJB09.1024

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

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

Self SG, Liang K-Y (1987) Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests under Nonstandard Conditions. Journal of the American Statistical Association 82:605–610. https://doi.org/10.1080/01621459.1987.10478472

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

Sprague GF, Tatum LA (1942) General vs. Specific Combining Ability in Single Crosses of Corn1. Agronomy Journal 34:923–932. https://doi.org/10.2134/agronj1942.00021962003400100008x

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

Technow F, Riedelsheimer C, Schrag TA, Melchinger AE (2012) Genomic prediction of hybrid performance in maize with models incorporating dominance and population specific marker effects. Theoretical and Applied Genetics 125:1181–1194. https://doi.org/10.1007/s00122-012-1905-8

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

Westhues M, Schrag TA, Heuer C, et al (2017) Omics-based hybrid prediction in maize. Theoretical and Applied Genetics 130:1927–1939. https://doi.org/10.1007/s00122-017-2934-0

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

Statements and Declarations

Funding

Lidea, Limagrain Europe, Maïsadour Semences, Corteva, RAGT 2n, and Syngenta Seeds grouped in the frame of the ProMais funded the SAM-MCR project. A.L. phD contract was funded by RAGT 2n and ANRT contract n° 2020/0032 and receive support from the "Investissement d'Avenir" project "Amaizing" (Amaizing, ANR-10-BTBR-0001). 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 CL and CG. AC, LM and CL supervised this work. SP and CP contributed to the development of the genetic material. 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 Experimental designs.

Fig. 2 Definition of the four different TRS and a unique VS used in Scenario 1 to assess the impact of the level of relationship between the TRS and the VS.The F-4H composed of 363 hybrids was used as the unique VS and four TRS were sampled form the F-1H, each composed of 742 hybrids. The four TRS led to consider four types of prediction: T0 hybrids where none of their parental lines contributed to the TRS, T1C (and T1D) hybrids where only their flint (dent) parental lines contributed to the TRS, and T2 hybrids where both of their parental lines contributed to the TRS.

Fig. 3 Predictive abilities obtained for the F-4H by calibration on four different TRS sampled from the F-1H. Depending on the TRS, VS hybrids had none of their parental lines in the TRS (T0), only their flint parental line in the TRS (T1F), only their dent parental line in the TRS (T1D) or their two parental lines in the TRS (T2). The red diamond shaped point represents the mean of the predictive ability of the 10 samplings.

Fig. 4 Predictive abilities obtained for all hybrids of the F-1H by calibration on the F-4H (including or not the SCA effect in the model) or on the tester designs. Williams tests were performed (α=0.05), no significant difference was obeserved for the presented traits.

Fig. 5 Predictive abilities obtained when calibrating on different tester designs composed of 180 hybrids: 180 hybrids produced by crossing 90 lines to one tester in each group (1T-180H-180L-A, 1T-180H-180L-B, 1T-180H-180L-C, 1T-180H-180L-D), 180 hybrids produced by crossing 90 lines to two testers in each group (2T-180H-180L), 180 hybrids produced by crossing 45 lines to two testers in each group (2T-180H-90L) or when calibrating on a factorial design composed of 180 hybrids produced by crossing 76 flint lines to 76 dent lines (F-180H-152L) to predict all 951 hybrids of the F-1H. The significance of the difference in predictive abilities was assessed by William tests (α=0.05) and was indicated with letters: two different letters indicate a significant difference and at least one common letters indicate no significant difference.

Fig. 6 Predictive abilitiies obtained when calibrating on 216 hyrbids of the F-1H design to predict 216 hybrids of the F-4H design in light orange and when calibrating on 216 hyrbids of the F-4H design to predict the F-1H in red. For each approach, 100 samplings were performed. The black diamond-shaped point represents the mean of the predictive abilities over the 100 repetitions.

Table 1 Broad-sense heritability (H^2), percentage of genetic variance asigned to SCA variance (%SCA) and variance components estimated on phenotypic data corrected for the spatial effects for all the designs (F-1H, F-4H, T-F and T-D) without using marker information.

Trait	Design	$\sigma^2_{GCA_d}$ [a]	$\sigma^2_{GCA_f}$ [a]	σ^2_{SCA} [a]	$\sigma^2_{GCA_d \times E}$ [a]	$\sigma^2_{GCA_f \times E}$ [a]	$\sigma^2_{SCA \times E}$ [a]	%SCA [b]	σ^2_E [a] (min-max) [c]	H^2 [d]
DMY	**F-1H**	0.72 (0.10)[e]	0.25 (0.10)	0.25 (0.11)	0.03 (0.06)	0.19 (0.06)	0.12 (0.09) **ns**	**20**	0.70(0.06)-3.09(0.17)	**0.84**
	F-4H	0.74 (0.14)	0.54 (0.11)	0.13 (0.03)	0.18 (0.03)	0.17 (0.03)	0.11 (0.05)	**9**	0.44(0.06)-3.15(0.26)	**0.89**
	T-D	0.70 (0.13)	-	0.17 (0.04)	0.03 (0.04) **ns**[f]	-	0.21 (0.06)	**21**	0.26(0.06)-2.22(0.27)	**0.88**
	T-F	-	0.52 (0.10)	0.14 (0.03)	-	0.09 (0.03)	0.09 (0.05)	**20**	0.34(0.07)-1.42(0.18)	**0.87**
DMC	**F-1H**	0.84 (0.27)	2.25 (0.26)	0.55 (0.27)	0.33 (0.10)	0.30 (0.10)	0.20 (0.15) **ns**	**15**	0.71(0.07)-4.96(0.28)	**0.91**
	F-4H	1.32 (0.24)	2.78 (0.46)	0.28 (0.05)	0.40 (0.06)	0.45 (0.06)	0.25 (0.09)	**6**	0.68(0.10)-4.15(0.37)	**0.93**
	T-D	1.89 (0.34)	-	0.32 (0.08)	0.56 (0.10)	-	0.35 (0.12)	**14**	0.59(013)-2.63(0.37)	**0.89**
	T-F	-	3.00 (0.51)	0.44 (0.09)	-	0.30 (0.07)	0.12 (0.10)	**13**	0.50(0.10)-2.35(0.31)	**0.94**
DtSilk	**F-1H**	1.26 (0.18)	0.73 (0.19)	0.41 (0.18)	0.12 (0.05)	0.16 (0.06)	0 (0.08) **ns**	**17**	0.66(0.06)-5.57(0.28)	**0.91**
	F-4H	0.76 (0.14)	1.55 (0.26)	0.14 (0.03)	0.15 (0.03)	0.20 (0.03)	0.09 (0.06) **ns**	**6**	0.69(0.08)-1.42(0.14)	**0.93**
	T-D	0.63 (0.11)	-	0.08 (0.03)	0.06 (0.03)	-	0.09 (0.06)	**11**	0.37(0.07)-2.06(0.25)	**0.86**
	T-F	-	1.07 (0.19)	0.22 (0.05)	-	0.14 (0.04)	0.01 (0.08)	**17**	0.51(0.09)-3.34(0.39)	**0.90**
MFU (x10²)	**F-1H**	2.06 (0.30)	1.03 (0.33)	0.33 (0.34) **ns**	0.13 (0.23)	0.01 (0.24)	0.86 (0.39)	**10**	3.39(0.28)-12.36(0.68)	**0.79**
	F-4H	1.74 (0.31)	1.67 (0.30)	0.10 (0.06) **ns**	0.37 (0.08)	0.29 (0.07)	0.00 **ns**	**3**	1.29(0.12)-7.61(0.60)	**0.86**
	T-D	2.23 (0.40)	-	0.28 (0.10)	0.46 (0.12)	-	0.00	**11**	1.04(0.16)-6.31(0.74)	**0.85**
	T-F	-	1.46 (0.27)	0.23 (0.09)	-	0.21 (0.11)	0.23 (0.16)	**14**	0.46(0.12)-6.53(0.77)	**0.78**

[a] Variance component defined as in Model (1.1) and (1.2). For the tester designs, σ^2_{SCA} corresponds to σ^2_{SCAt}

[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] Minimum residual variance and maximum residual variance across all environments

[d] Broad-sense heritability

[e] Standard error in brackets

[f] Significance of the variance components assessed by likelihood ratio test with χ^2 mixed distributions (α=0.05). Non-significant variance component is indicated by **ns**

Table 2 Variance components and percentage of genetic variance asigned to SCA variance (%SCA) estimated with marker information (GBLUP model) for all the designs (F-1H, F-4H, T-D and T-F).

Trait	Design	$\sigma^2_{GCA_d}$	$\sigma^2_{GCA_f}$	σ^2_{SCA} [a]	%SCA[b]	σ^2_E
DMY	**F-1H**	0.62 (0.12)[c]	0.48 (0.09)	0.13 (0.06)	**11**	0.52 (0.06)
	F-4H	0.68 (0.15)	1.04 (0.21)	0.03 (0.05)	**2**	0.31 (0.05)
	T-D	0.47 (0.13)	-	0.08 (0.06)	**15**	0.25 (0.05)
	T-F	-	0.86 (0.17)	0.10 (0.05)	**10**	0.11 (0.03)
DMC	**F-1H**	1.41 (0.25)	1.79 (0.28)	0.00	**0**	1.07 (0.08)
	F-4H	2.79 (0.53)	4.45 (0.78)	0.01 (0.08)	**0**	0.62 (0.10)
	T-D	4.27 (0.79)	-	0.19 (0.12)	**4**	0.41 (0.09)
	T-F	-	3.95 (0.76)	0.28 (0.18)	**7**	0.51 (0.12)
DtSilk	**F-1H**	1.21 (0.21)	2.03 (0.29)	0.09 (0.06)	**3**	0.73 (0.08)
	F-4H	1.67 (0.32)	2.65 (0.46)	0.11 (0.06)	**2**	0.24 (0.05)
	T-D	0.99 (0.23)	-	0.04 (0.04)	**4**	0.21 (0.05)
	T-F	-	2.29 (0.42)	0.17 (0.10)	**7**	0.24 (0.06)
MFU ($x10^2$)	**F-1H**	1.14 (0.22)	1.31 (0.24)	0.00	**0**	1.50 (0.1)
	F-4H	1.19 (0.27)	2.17 (0.42)	0.26 (0.14)	**7**	0.48 (0.11)
	T-D	1.38 (0.38)	-	0.07 (0.10)	**5**	0.81 (0.14)
	T-F	-	2.11 (0.44)	0.12 (0.13)	**5**	0.55 (0.11)

[a] For the tester designs, σ^2_{SCA} corresponds to σ^2_{SCAt}

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

[c] Standard error in brackets

Table 3 Predictive abilities obtained for the F-4H and for each of the tester designs by calibrating the GBLUP model on the F-1H.

Validation Set	Predicted hybrid value component(s)	Hybrid type in the VS[a]	DMY	DMC	DtSilk	MFU
F-4H	$GCA_f + GCA_d$	**Selected+Random (363)**[b]	**0.78**	**0.78**	**0.79**	**0.82**
		Selected (127)	0.48	0.80	0.69	0.73
		Random (236)	0.67	0.78	0.76	0.80
	$GCA_f + GCA_d + SCA$	**Selected+Random (363)**	**0.77**	**0.78**	**0.79**	**0.82**
		Selected (127)	0.46	0.80	0.71	0.73
		Random (236)	0.67	0.78	0.77	0.80
T-D	GCA_d	**Selected+Random (363)**	**0.80**	**0.74**	**0.76**	**0.79**
		Selected (60)	0.66	0.76	0.69	0.73
		Random (120)	0.71	0.73	0.71	0.76
T-F	GCA_f	**Selected+Random (363)**	**0.69**	**0.84**	**0.84**	**0.79**
		Selected (60)	0.68	0.84	0.78	0.71
		Random (120)	0.63	0.84	0.86	0.79

[a] Predictive ability was computed for all hybrids (Selected+Random) and also for each hybrid type (Selected and Random) of the VS

[b] Number of hybrids in brackets

Table 4 Correlations between the GCA BLUPs predicted for the F-1H using etiher the F-4H or the tester designs as TRS.

BLUPs correlated	DMY	DMC	DtSilk	MFU
$\mathbf{GCA_f}$	0.85	0.95	0.94	0.93
$\mathbf{GCA_d}$	0.90	0.88	0.86	0.92

Table 5 Predictive abilities obtained by calibrating a GS model on the F-4H or on the tester designs to predict different hybrid types of the F-1H (VS).

TRS	Validation Set		DMY	DMC	DtSilk	MFU
	Hybrid type[a]	Number of hybrids				
F-4H[b] (363)[c]	T1D	106	0.61	0.72	0.61	0.70
	T1F	95	0.74	0.81	0.72	0.78
	T0	742	0.49	0.61	0.50	0.64
	All	**951**	**0.56**	**0.64**	**0.55**	**0.67**
Tester designs (360)	T1D	107	0.60	0.73	0.63	0.68
	T1F	95	0.75	0.78	0.73	0.75
	T0	741	0.49	0.62	0.52	0.63
	All	**951**	**0.54**	**0.64**	**0.56**	**0.65**

[a] VS hybrids had none of their parental lines in the TRS (T0), only their flint parental line in the TRS (T1F), only their dent parental line in the TRS (T1D)

[b] Model (5.1) including SCA effects

[c] Number of hybrids in the TRS

Figures

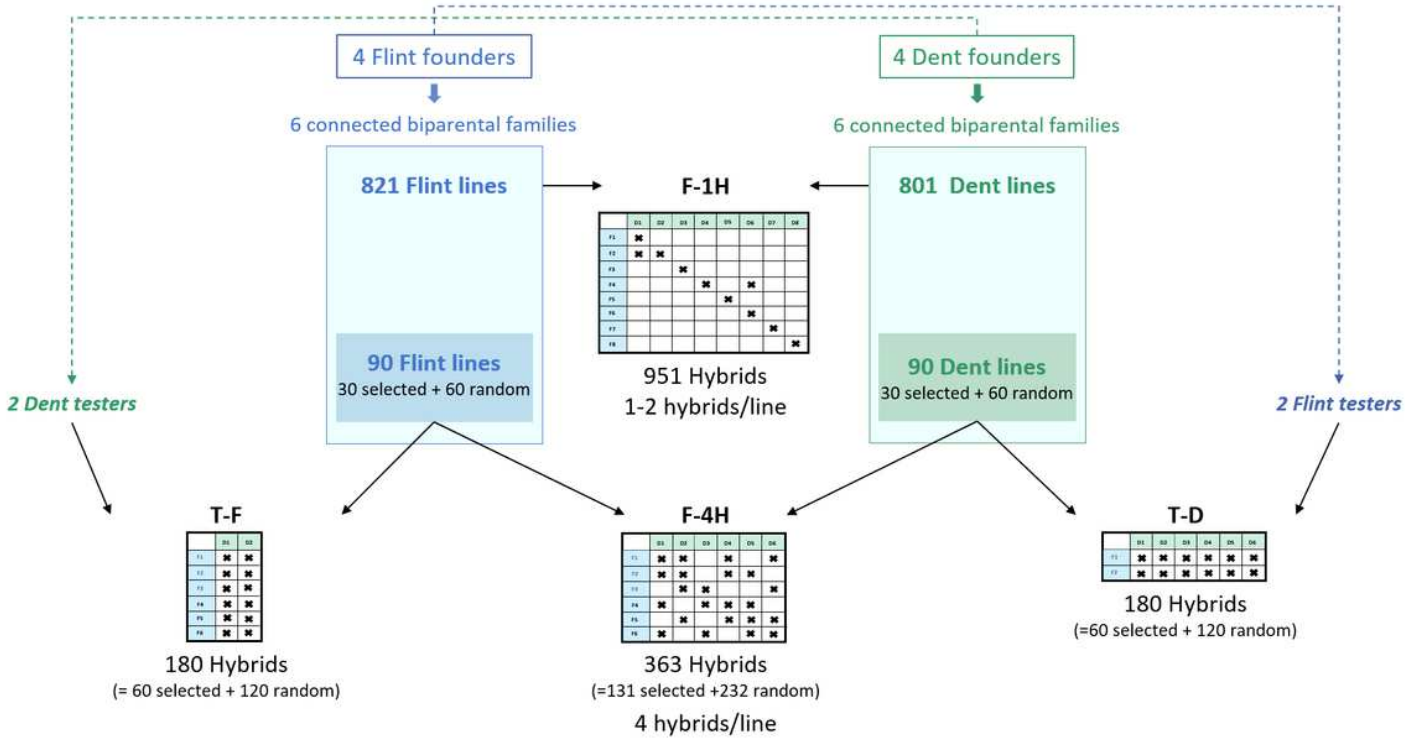

Figure 1

Experimental designs.

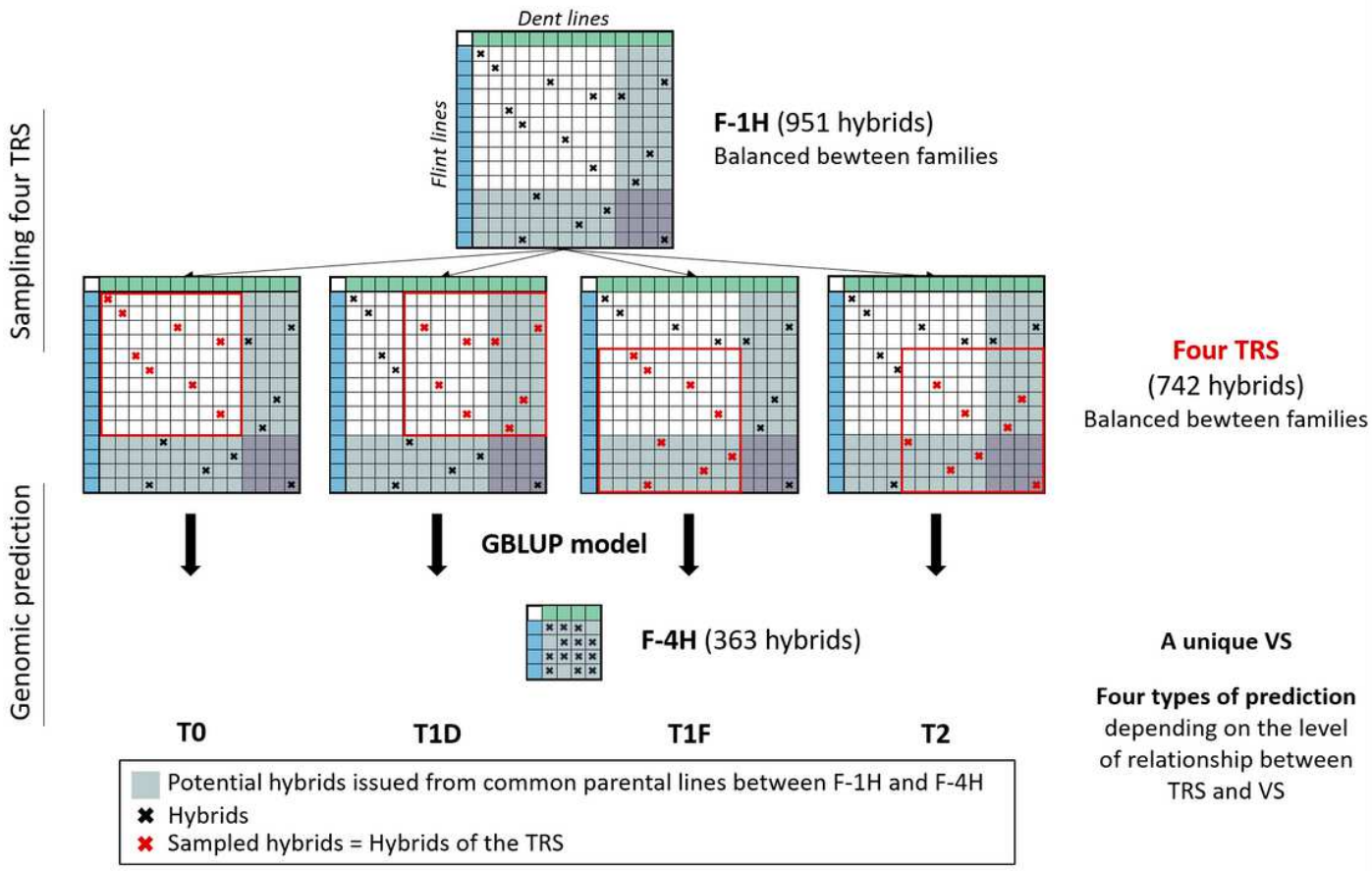

Figure 2

Definition of the four different TRS and a unique VS used in Scenario 1 to assess the impact of the level of relationship between the TRS and the VS.The F-4H composed of 363 hybrids was used as the unique VS and four TRS were sampled form the F-1H, each composed of 742 hybrids. The four TRS led to consider four types of prediction: T0 hybrids where none of their parental lines contributed to the TRS, T1C (and T1D) hybrids where only their flint (dent) parental lines contributed to the TRS, and T2 hybrids where both of their parental lines contributed to the TRS.

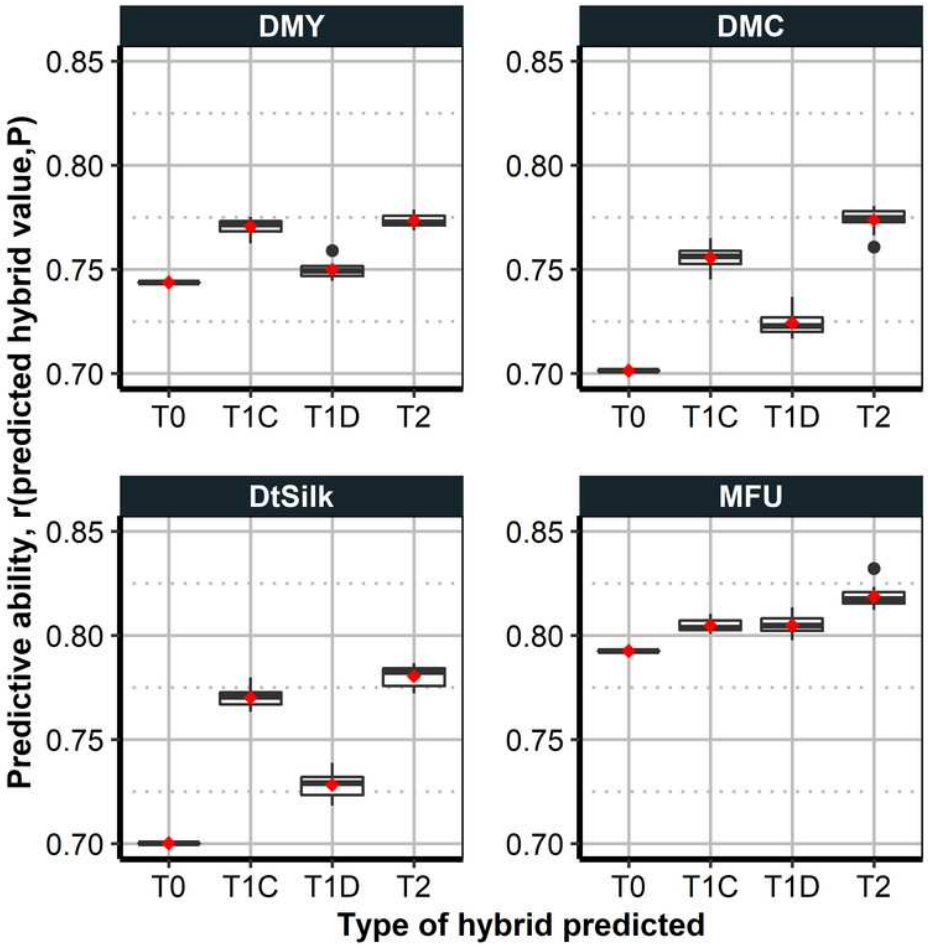

Figure 3

Predictive abilities obtained for the F-4H by calibration on four different TRS sampled from the F-1H. Depending on the TRS, VS hybrids had none of their parental lines in the TRS (T0), only their flint parental line in the TRS (T1F), only their dent parental line in the TRS (T1D) or their two parental lines in the TRS (T2). The red diamond shaped point represents the mean of the predictive ability of the 10 samplings.

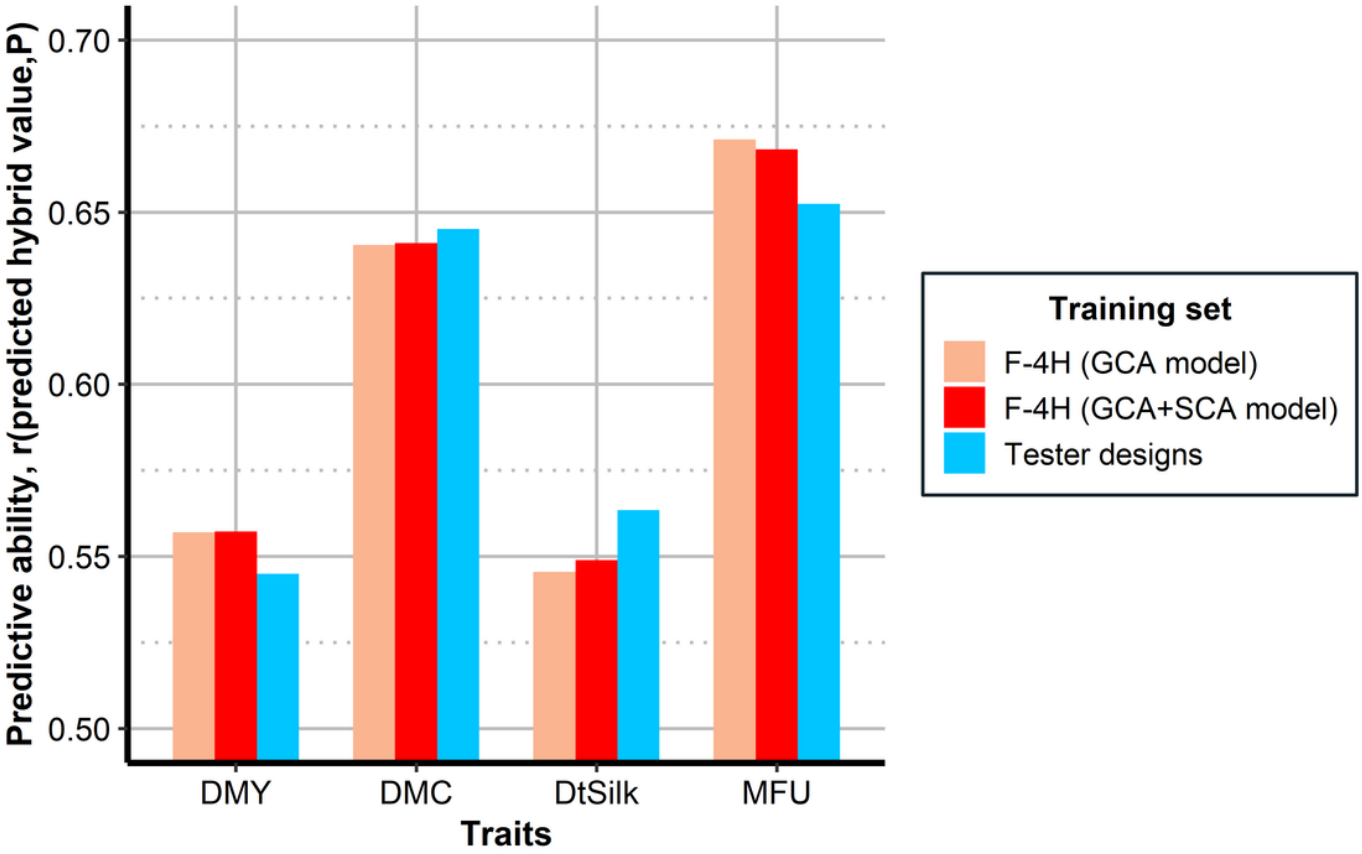

Figure 4

Predictive abilities obtained for all hybrids of the F-1H by calibration on the F-4H (including or not the SCA effect in the model) or on the tester designs. Williams tests were performed (α=0.05), no significant difference was obeserved for the presented traits.

Figure 5

Predictive abilities obtained when calibrating on different tester designs composed of 180 hybrids: 180 hybrids produced by crossing 90 lines to one tester in each group (1T-180H-180L-A, 1T-180H-180L-B, 1T-180H-180L-C, 1T-180H-180L-D), 180 hybrids produced by crossing 90 lines to two testers in each group (2T-180H-180L), 180 hybrids produced by crossing 45 lines to two testers in each group (2T-180H-90L) or when calibrating on a factorial design composed of 180 hybrids produced by crossing 76 flint lines to 76 dent lines (F-180H-152L) to predict all 951 hybrids of the F-1H. The significance of the difference in predictive abilities was assessed by William tests (α=0.05) and was indicated with letters: two different letters indicate a significant difference and at least one common letters indicate no significant difference.

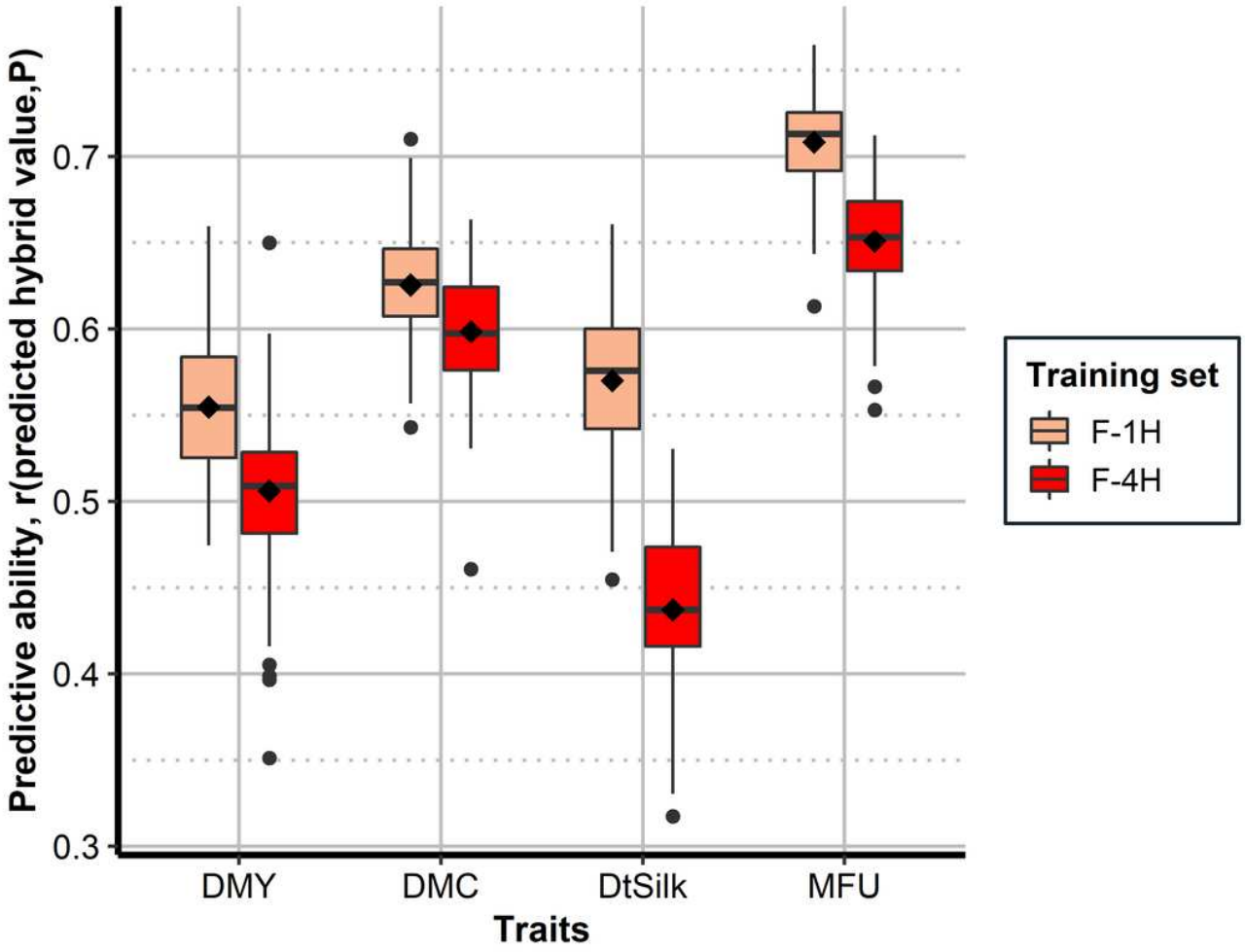

Figure 6

Predictive abilitiies obtained when calibrating on 216 hyrbids of the F-1H design to predict 216 hybrids of the F-4H design in light orange and when calibrating on 216 hyrbids of the F-4H design to predict the F-1H in red. For each approach, 100 samplings were performed. The black diamond-shaped point represents the mean of the predictive abilities over the 100 repetitions.

Supplementary Files

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

- 20220113SupplementaryMaterial.pdf