



**HAL**  
open science

# Optimization of genomic selection for hybrids in a reciprocal selection program. Experimental evaluation and simulations on maize

Alizarine Lorenzi

► **To cite this version:**

Alizarine Lorenzi. Optimization of genomic selection for hybrids in a reciprocal selection program. Experimental evaluation and simulations on maize. Plants genetics. Université Paris-Saclay, 2023. English. NNT: 2023UPASB053 . tel-04774825

**HAL Id: tel-04774825**

**<https://pastel.hal.science/tel-04774825v1>**

Submitted on 9 Nov 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Optimization of genomic selection for  
hybrids in a reciprocal selection program.  
Experimental evaluation and simulations  
on maize.

*Optimisation de la mise en œuvre de la sélection génomique hybride  
dans un programme de sélection réciproque. Evaluation expérimentale et  
simulations chez le maïs.*

**Thèse de doctorat de l'université Paris-Saclay**

École doctorale n°581, agriculture, alimentation, biologie, environnement, santé (ABIES)  
Spécialité de doctorat : Sciences végétales  
Graduate School : Biosphera. Référent : AgroParisTech.

Thèse préparée dans l'UMR **Génétique Quantitative et Evolution- Le Moulon**  
(Université Paris-Saclay, INRAE, CNRS, AgroParisTech),  
sous la direction de **Laurence MOREAU**, Directrice de Recherche,  
le co-encadrement de **Alain CHARCOSSET**, Directeur de recherche  
et la co-supervision de **Gaëtan TOUZY**, Cadre scientifique

**Thèse soutenue à Paris-Saclay, le 08 novembre 2023, par**

**Alizarine LORENZI**

**Composition du Jury**

Membres du jury avec voix délibérative

<b>Christine DILLMANN</b> Professeure, Université Paris-Saclay	Présidente
<b>Aaron LORENZ</b> Professeur, Université du Minnesota (USA)	Rapporteur & Examineur
<b>Leopoldo SANCHEZ</b> Directeur de recherche, INRAE (centre Val de Loire)	Rapporteur & Examineur
<b>David CROS</b> Chercheur, Cirad (Montpellier)	Examineur

**Titre :** Optimisation de la mise en œuvre de la sélection génomique hybride dans un programme de sélection réciproque. Evaluation expérimentale et simulations chez le maïs.

**Mots clés :** Génétique quantitative, Sélection génomique, Hybride, Maïs

**Résumé :** La sélection génomique (SG) est désormais couramment utilisée dans les schémas de sélection, ouvrant des perspectives pour réexaminer les schémas en remplaçant une partie du phénotypage par des prédictions basées sur le génotype aux marqueurs. La définition de la population d'entraînement et les étapes où appliquer la SG dans les schémas de sélection sont des questions clé, en particulier pour la sélection hybride. Dans les schémas de sélection hybride du maïs, un défi est d'identifier les meilleurs hybrides parmi tous les croisements possibles entre les lignées candidates produites chaque année dans des groupes hétérotiques. Dans les schémas conventionnels, pour réduire le nombre de combinaisons possibles, les lignées candidates de chaque groupe sont d'abord présélectionnées sur la base de leur valeur en croisement sur testeurs. Ensuite, un sous-ensemble d'hybrides entre ces lignées sélectionnées est évalué afin d'identifier les meilleures variétés hybrides. Récemment, des études ont proposé de remplacer les évaluations sur testeurs par des prédictions génomiques calibrées sur un dispositif factoriel incomplet entre lignées candidates non sélectionnées afin de prédire tous les hybrides possibles à un stade précoce du cycle de sélection. Des simulations et études expérimentales ont montré des qualités prédictives prometteuses. Cette thèse vise à (i) confirmer l'efficacité de factoriels pour prédire les valeurs des lignées et des hybrides par rapport à une approche sur testeurs, (ii) évaluer l'efficacité de factoriels et leur optimisation au cours des cycles et (iii) optimiser l'utilisation de la SG dans les schémas hybrides. Cette thèse s'appuie sur des données expérimentales correspondant à deux cycles de sélection d'un schéma de SG réciproque de maïs impliquant des populations multiparentales connectées des groupes hétérotiques complémentaires flint et dent, sélectionnées pour leurs performances ensilage et obtenues dans le cadre du projet Promaïs "SAMMCR".

Premièrement, les données du premier cycle de sélection ont été utilisées pour évaluer l'efficacité de factoriels incomplets pour prédire les hybrides entre des lignées non sélectionnées de la même génération et la comparer à celle de dispositifs testeur. Les factoriels ont montré des qualités prédictives équivalentes à celles basées sur testeurs et les ont même surpassées pour certains caractères. Les résultats obtenus confirment le potentiel des factoriels incomplets comme alternative fiable aux dispositifs testeurs traditionnels pour prédire les performances hybrides.

Nous avons ensuite évalué la portabilité des prédictions génomiques sur deux cycles de sélection. Les résultats ont permis de valider expérimentalement l'efficacité des dispositifs factoriels incomplets pour prédire les aptitudes générales à la combinaison des lignées et les valeurs hybrides au cours des générations. Ils ont également mis en évidence l'intérêt de recalibrer les populations d'entraînement au cours des cycles et d'optimiser leur composition à l'aide du CDmean.

Pour compléter, des simulations ont permis de comparer à coûts fixes un schéma de sélection phénotypique conventionnel avec cinq schémas de SG différant par la longueur du cycle et l'utilisation de dispositifs testeur et/ou factoriel pour calibrer les prédictions. Les résultats suggèrent que la mise en œuvre de la SG et le raccourcissement du cycle de sélection, en remplaçant des évaluations sur testeurs par un dispositif factoriel incomplet, permettent d'accélérer les cycles et d'augmenter le gain génétique. Ces résultats valident l'efficacité des factoriels incomplets et soulignent le potentiel intérêt de leur intégration dans les programmes de sélection hybride à coûts constants.

Conjointement, nos résultats débouchent sur des recommandations visant à améliorer le schéma de sélection hybride du maïs par l'intégration de la SG et des dispositifs factoriels.

**Title :** Optimization of genomic selection for hybrids in a reciprocal selection program. Experimental evaluation and simulations on maize.

**Keywords :** Quantitative genetics, Genomic selection, Hybrid, Maize

**Abstract :**

Genomic selection (GS) is now used in routine in breeding programs, opening prospects for revisiting breeding schemes by replacing extensive phenotyping with predictions based on marker information. In this context, finding the ideal design for training genomic prediction models and the step(s) at which applying GS are key questions, especially in hybrid breeding. In conventional maize hybrid breeding schemes, one challenge is identifying the best hybrid varieties among all possible hybrid combinations between candidate lines produced each year from each heterotic group. To reduce the number of possible combinations, the candidate inbred lines from each heterotic group are first preselected on their hybrid value estimated in topcross evaluations. Then, a subset of single-cross hybrids between these selected lines is evaluated to identify the best hybrid varieties. Recently, studies proposed to replace topcross evaluations with genomic predictions calibrated on a sparse factorial design between unselected candidate lines to predict all possible single-cross hybrids at an early stage of the breeding cycle. Previous simulations and experimental studies have shown promising predictive abilities using a sparse factorial design instead of tester designs as training sets. This thesis aims to (i) confirm the efficiency of factorial training sets to predict line and hybrid values and compare it to tester training sets, (ii) further investigate the use of factorials and their optimization across two breeding cycles, and (iii) provide recommendations for implementing this approach in maize hybrid breeding schemes. It relies on experimental data corresponding to two breeding cycles of a maize reciprocal genomic selection scheme involving multiparental connected populations from the flint and dent complementary heterotic groups, selected for silage performances, obtained in the framework of the Promaïs "SAMMCR" project. Simulations completed results obtained on experimental data.

First, data from the first breeding cycle were used to evaluate the efficiency of sparse factorial training sets for predicting single-cross hybrids between unselected lines from the same generation and compare it to tester designs. Factorials showed equivalent predictive abilities to tester-based training sets and even outperformed them for some traits. Results obtained with different samplings of our experimental designs support the potential of sparse factorial training sets as a reliable alternative to traditional tester designs for predicting hybrid values.

We then evaluated the portability of genomic predictions across two breeding cycles. The results experimentally validated the efficiency of sparse factorial designs for predicting line general combining abilities and hybrid values across generations. They also highlighted the benefits of updating training sets along breeding cycles and optimizing their composition using the CDmean criterion to ensure high prediction accuracy along breeding cycles.

To complement experimental results, simulations were carried out to compare at a fixed cost a conventional phenotypic breeding scheme with five genomic selection breeding schemes differing by their cycle length and the use of tester and/or factorial designs to calibrate predictions. The simulation results suggested that implementing genomic selection and shortening the breeding cycle by replacing early topcross evaluations by a unique sparse factorial design offers the potential to accelerate the breeding process and increase genetic gain. These findings validate the efficiency of sparse factorial training sets and highlight the potential of integrating them into hybrid breeding programs at fixed costs.

Jointly, our results lead to a set of recommendations to improve the maize hybrid breeding scheme through the integration of genomic selection and factorial designs.

## Additional informations

---

Cette thèse s'inscrit dans le projet « **SAMMCR** » initié par l'association Promais. Elle a eu lieu dans le cadre de la convention CIFRE n° 2020/0032 entre l'INRAE et la société RAGT2n. Elle a été co-financée par les sept partenaires privés du projet SAMMCR : Lidea, Limagrain Europe, Maisadour Semences, Corteva, RAGT 2n, KWS and Syngenta Seeds.

Chloé BOYARD (Cadre scientifique – Limagrain) sera présente à la soutenance de thèse en qualité d'invitée.

## Résumé substantiel en français

---

### Introduction

Les variétés hybrides de maïs sont principalement des croisements simples entre deux lignées issues de groupes hétérotiques différents. En Europe du Nord, les deux principaux groupes hétérotiques utilisés par les sélectionneurs pour produire des hybrides destinés à la production de maïs fourrage sont les cornés et dentés. Dans les programmes de sélection, le nombre de lignées candidates est important et augmente chaque année, ce qui rend impossibles la production et l'évaluation de toutes les combinaisons hybrides possibles. L'approche qui a été utilisée pour contourner cette limitation consiste à présélectionner des lignées dans chaque groupe hétérotique, puis à tester les combinaisons hybrides entre les lignées sélectionnées de chaque groupe.

En 1949, Comstock et al., (1949), ont proposé un schéma de sélection récurrente réciproque, pour améliorer simultanément les deux populations parentales des hybrides en évaluant les lignées candidates d'une population parentale pour leur valeur en croisement avec les lignées du groupe opposé. Dans les programmes de sélection hybride actuels, des variations de ce schéma de sélection récurrente réciproque sont utilisées. Dans la première étape, les lignées candidates d'un groupe hétérotique sont croisées avec quelques lignes du groupe opposé appelées "testeurs". La descendance hybride ainsi produite (topcross) est évaluée phénotypiquement, et les meilleures lignées de chaque groupe hétérotique sont sélectionnées sur la base de leur Aptitude Générale à la Combinaison (AGC). Dans la deuxième étape, les lignées sélectionnées de chaque groupe sont croisées selon un dispositif factoriel incomplet, et les meilleures combinaisons hybrides sont identifiées. La sélection basée sur quelques testeurs facilite la production de semences hybrides lors des premiers tests, mais ne tient pas pleinement compte de la complémentarité entre les groupes (ASC) et peut biaiser l'estimation de l'AGC. De plus, l'utilisation de ce processus de sélection en deux étapes augmente le temps nécessaire au développement d'hybrides commercialisables et nécessite le phénotypage d'un grand nombre de lignées (au moins autant d'hybrides que de lignées dans chaque groupe).

La sélection génomique (SG) (Meuwissen et al. 2001) est désormais couramment utilisée dans les schémas de sélection, ouvrant de nouvelles perspectives pour repenser les schémas de sélection en remplaçant tout ou partie des évaluations phénotypiques par des prédictions génomiques. La définition de la population d'entraînement (ou population de calibration) et les étapes où appliquer la SG dans les schémas de sélection sont des questions clé, particulièrement pour la sélection hybride. Récemment, des études ont proposé de remplacer les évaluations sur testeurs par des prédictions génomiques calibrées sur un dispositif factoriel incomplet entre lignées candidates non sélectionnées afin de prédire tous les hybrides possibles à un stade précoce du cycle de sélection (Giraud, 2016; Kadam et al., 2016). Des simulations (Seye et al., 2020) et des études expérimentales (Burdo et al., 2021; Fristche-Neto et al., 2018; Kadam et al., 2016) ont montré des qualités prédictives prometteuses. Néanmoins, cette approche nécessite le génotypage des lignées candidates et la production d'hybrides par pollinisation manuelle, ce qui est difficile et coûteux. Par conséquent, même si cette approche est attrayante, des études supplémentaires et l'exploration de stratégies d'optimisation sont nécessaires pour évaluer l'intérêt de l'intégrer dans les programmes de sélection. Cette thèse vise à (i) évaluer l'efficacité de dispositifs factoriels pour prédire les valeurs des lignées et des hybrides en comparaison à une approche sur dispositifs testeurs, (ii) évaluer l'efficacité de dispositifs factoriels et leur optimisation au cours des cycles et (iii) optimiser l'utilisation de la SG dans les schémas de sélection hybride du maïs. Cette thèse s'appuie sur des données expérimentales correspondant à deux cycles de sélection d'un schéma de SG réciproque de maïs impliquant des populations multiparentales connectées des groupes hétérotiques complémentaires flint et dent, sélectionnées pour leurs performances ensilage et obtenues dans le cadre du projet Promaïs « SAMMCR » (« Sélection Assistée par Marqueurs Multi-parentale Connectée Réciproque »).

## **Matériel végétal**

Cette thèse s'inscrit dans le cadre du projet SAMMCR qui a débuté en 2010 en collaboration avec huit entreprises privées de sélection, membres de l'association Promaïs : RAGT2n, KWS, Lidea, Corteva, Limagrain, Syngenta and MAS. Le projet SAMMCR repose sur un dispositif expérimental original composé de deux familles multiparental connectées, chacune correspondant à l'un des principaux groupes hétérotiques en Europe du Nord : les cornés et les dentés. Au total, deux générations ont été produites et quatre campagnes expérimentales ont

été menées lors de ce projet et sont détaillées ci-dessous.

Dans chaque groupe hétérotique, quatre lignées fondatrices (F373, F03802, F02803 et F7088 pour le groupe corné, et F98902, F1808, F04401 et F7082 pour le groupe denté) ont été sélectionnées et croisées pour produire six hybrides simples F1. Parmi les quatre lignées fondatrices, trois lignées ont été choisies pour leurs performances agronomiques en matière de production d'ensilage (rendement en biomasse matière sèche, teneur en matière sèche), et la dernière a été choisie pour la qualité de son ensilage (digestibilité). A partir des six hybrides F1, six familles biparentales ont été dérivées par haploïdisation doublée pour les dentés et par cinq à six générations d'autofécondation (Single Seed Descent) pour les cornés. Au total, 821 lignées cornées et 801 lignées dentées ont été produites. L'ensemble des lignées décrites ci-dessus constituent les lignées parentales de la première génération et seront désignées par « **G0** » dans la suite. Les 801 lignées G0 dentées et les 821 lignées G0 cornées ont été croisées selon un plan factoriel très incomplet pour produire 951 hybrides simples corné-denté, formant ainsi le factoriel **G0\_F-1H**. En moyenne, chaque lignée parentale a contribué à la production d'un hybride (20 % des lignées ont contribué à produire deux hybrides). Le dispositif factoriel G0\_F-1H est équilibré avec 22 à 35 hybrides produits à partir de chaque combinaison de familles (pour plus de détails, voir Giraud et al. 2017a).

Puis, dans chaque groupe hétérotique, 60 lignées G0 ont été choisies de manière aléatoire et équilibrée (10 lignées par famille) et 30 ont été sélectionnées sur la base de leurs performances agronomiques et de leur qualité d'ensilage (seules trois familles sur six sont représentées dans les lignées sélectionnées). Les 30 lignées cornées «*sélectionnées*» ont été croisées avec les 30 lignées dentées sélectionnées pour produire 131 hybrides «*sélectionnés*» (G0S), et les 60 lignées dentées «*aléatoires*» ont été croisées avec les 60 lignées cornées «*aléatoires*» pour produire 232 hybrides «*aléatoires*» (G0R). Au total 363 hybrides corné-denté ont été produits et forment le factoriel **G0\_F-4H**. Dans ce dispositif, chaque lignée parentale a contribué à la production de quatre hybrides en moyenne. En parallèle, les mêmes 90 lignées (60 aléatoires + 30 sélectionnées) ont été croisées avec deux testeurs du groupe opposé pour produire un dispositif testeur de 180 hybrides dans chaque groupe hétérotique (G0\_T ou dispositifs testeurs). Les testeurs utilisés ont été choisis parmi les quatre lignées fondatrices du groupe hétérotique complémentaire (F1808 et F98902 pour les testeurs dentés, et F373 et F02803 pour les testeurs cornés) pour être contrastés et avoir un bon potentiel de rendement.



Une nouvelle génération de lignées, appelées « **G1** », a été produite. Dans chaque groupe hétérotique, les 30 lignées G0 sélectionnées précédemment ont été croisées pour produire 40 hybrides simples à partir desquels 351 lignées ont été produites par haploïdisation doublée (HD). Les lignées HD cornées et dentées ainsi produites ont été croisées selon un dispositif factoriel incomplet pour produire 441 hybrides G1 corné-denté. En parallèle, 47 hybrides G0 sélectionnés (G0S) ont été produits en croisant les 30 lignées cornées G0S avec les 30 lignées dentées G0S précédemment décrites. Le nouveau dispositif factoriel est composé de 47 hybrides G0S et de 442 hybrides G1, et il est en moyenne composé d'un hybride par ligne, il sera donc désigné **(G0S+G1)\_F-1H**.

Chaque dispositif expérimental a été évalué dans un total de huit essais en deux ans : le G0\_F-1H a été évalué dans quatre essais en 2013 et quatre en 2014, le G0\_F-4H et le G0\_T ont été évalués conjointement dans trois essais en 2016 et cinq en 2017, et le (G0S+G1)\_F-1H a été évalué dans trois essais en 2019 et cinq en 2020. Tous les hybrides ont été évalués pour 11 caractères, quatre caractères agronomiques (rendement en matière sèche, teneur en matière sèche, date de floraison, hauteur des plantes) et sept caractères de qualité de l'ensilage (MFU, DINAG, DINAGZ, teneur en lignine, cellulose et hémicellulose) mesurés par spectroscopie proche infrarouge (NIRS).

Les lignées fondatrices ont été génotypées sur deux puces différentes : une puce SNP de 50 000 marqueurs (Ganal et al., 2011) et une puce SNP privée de 18 480 marqueurs Affimetrix® (puce Limagrain). Les lignées parentales G0 ont été génotypées sur la puce SNP Affimetrix® de 18 480 marqueurs, et les lignées parentales G1 ont été génotypées sur la puce SNP Illumina® publique de 15 000 marqueurs.

## Résultats

Premièrement, les données du cycle de sélection G0 ont été utilisées pour évaluer l'efficacité de dispositifs factoriels incomplets pour prédire des hybrides entre des lignées non sélectionnées de la même génération et la comparer à celle de dispositifs testeurs. Les dispositifs factoriels ont montré des qualités prédictives équivalentes à celles basées sur testeurs et les ont même surpassées pour certains caractères (PH, DINAG and DINAGZ). Les qualités prédictives élevées obtenues Nous avons montré des qualités prédictives hybrides allant de 0.5 à 0.8 en utilisant un dispositif factoriel très incomplet pour prédire des nouvelles

combinaisons hybrides dans de nouveaux environnements, illustrant l'efficacité d'un tel dispositif. Nous avons évalué l'efficacité de différentes compositions pour le dispositif factoriel en terme de nombre de lignées et de nombre d'hybride par lignée à nombre total d'hybrides identique. Nous avons montré un avantage à augmenter le nombre de lignées évaluées plutôt que le nombre d'hybrides par lignées lorsque l'on prédit des hybrides dont aucun de ses parents n'a contribué à la calibration du modèle de SG. Les résultats obtenus confirment le potentiel des factoriels incomplets comme alternative fiable aux dispositifs testeurs traditionnels pour prédire les performances hybrides.

Nous avons ensuite évalué la portabilité des prédictions génomiques en utilisant les données des deux cycles de sélection (G0 et G1). Les résultats ont permis de valider expérimentalement l'efficacité des dispositifs factoriels incomplets de la première génération pour prédire les aptitudes générales à la combinaison des lignées et les valeurs hybrides au de la seconde génération. Nous avons mis en évidence l'intérêt de recalibrer les populations d'entraînement des modèles des SG au cours des cycles, en ajoutant à la population d'entraînement existante (données historiques de la génération G0) : (i) des hybrides entre les lignées sélectionnées du cycle précédent (G0S) qui ont servi à produire les hybrides de la nouvelle génération (G1) et/ou (ii) des hybrides entre les lignées de la nouvelle génération (G1). Enfin, en utilisant le CDmean, nous avons montré l'efficacité d'optimiser le sous-ensemble d'hybrides de la nouvelle génération (G1) à ajouter à la population d'entraînement pour prédire la nouvelle génération (G1).

Pour compléter, des simulations ont été conduites afin d'évaluer le rapport coût/bénéfice de l'utilisation de dispositifs factoriels couplés à la SG dans les schémas de sélection hybride. Le travail de simulation a permis de comparer à coûts fixes un schéma de sélection phénotypique conventionnel avec cinq schémas basé sur la sélection génomique différant par la longueur du cycle et l'utilisation de dispositifs testeur et/ou factoriel pour calibrer les prédictions. Les résultats suggèrent que la mise en œuvre de la SG et le raccourcissement du cycle de sélection, en remplaçant des évaluations sur testeurs par un dispositif factoriel incomplet, permettent d'accélérer les cycles et d'augmenter le gain génétique. Ces résultats valident l'efficacité des factoriels incomplets et soulignent le potentiel intérêt de leur intégration dans les programmes de sélection hybride à coûts constants.

## Conclusions

Les résultats des chapitres 1 et 2 basés sur des données expérimentales confirment le potentiel des dispositifs factoriels très incomplets en tant qu'alternative efficace aux dispositifs testeurs traditionnels pour prédire les valeurs hybrides en ce qui concerne les qualités prédictives. Nos conclusions restent étroitement liées à la population et aux dispositifs expérimentaux utilisés. De ce fait, il pourrait être intéressant de comparer les dispositifs factoriels et testeurs en utilisant des données expérimentales dans d'autres populations, par exemple avec une base génétique plus large (plus de lignées fondatrices dans la population initiale) et différents niveaux d'ASC, afin d'évaluer l'avantage du dispositif factoriel dans cette situation. Nous avons utilisé des simulations pour mieux comprendre le potentiel de l'approche factorielle dans un programme de sélection plus réaliste. Dans le chapitre 3, nous avons simulé un schéma récurrent réciproque impliquant des populations multi-parentales connectées provenant de deux groupes hétérotiques complémentaires imitant un programme de sélection du maïs. Nous avons montré que l'utilisation de la SG pour réduire de la durée du cycle de sélection en supprimant les étapes d'évaluation phénotypique et l'utilisation de dispositifs factoriels étaient économiquement rentables et permettaient de générer un gain génétique plus rapide.

Un inconvénient majeur de l'incorporation des dispositifs factoriels dans les programmes de sélection était le coût supplémentaire de la production de semences et la main d'œuvre nécessaire pour produire des hybrides factoriels par pollinisation manuelle. Les résultats du chapitre 3 ont montré que le coût supplémentaire de la production de semences hybrides pour un factoriel (deux fois le prix de la production de semences d'hybrides testeurs ou topcross d'après les discussions avec Cyril Bauland et des partenaires privés) pouvait être compensé efficacement en réduisant le nombre de parcelles de phénotypage et en augmentant le nombre de lignées évaluées dans le dispositif factoriel (schéma Y2F1H-GS).

Conjointement, nos résultats débouchent sur des recommandations visant à améliorer le schéma de sélection hybride du maïs par l'intégration de la SG et des dispositifs factoriels.

## Références

- Burdo, B., Leon, N., Kaeppeler, S.M., 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>
- Comstock, R.E., Robinson, H., Harvey, P.H., 1949. A breeding procedure designed to make maximum use of both general and specific combining ability.
- Fristche-Neto, R., Akdemir, D., Jannink, J.-L., 2018. Accuracy of genomic selection to predict maize single-crosses obtained through different mating designs. *Theor. Appl. Genet.* 131, 1153–1162. <https://doi.org/10.1007/s00122-018-3068-8>
- Ganal, M.W., Durstewitz, G., Polley, A., Bérard, A., Buckler, E.S., Charcosset, A., Clarke, J.D., Graner, E.-M., Hansen, M., Joets, J., Paslier, M.-C.L., McMullen, M.D., Montalent, P., Rose, M., Schön, C.-C., Sun, Q., Walter, H., Martin, O.C., Falque, M., 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>
- Giraud, H., 2016. Genetic analysis of hybrid value for silage maize in multiparental designs: QTL detection and genomic selection (thesis). Paris-Saclay.
- Kadam, D.C., Potts, S.M., Bohn, M.O., Lipka, A.E., Lorenz, A.J., 2016. Genomic Prediction of Single Crosses in the Early Stages of a Maize Hybrid Breeding Pipeline. *G3 GenesGenomesGenetics* 6, 3443–3453. <https://doi.org/10.1534/g3.116.031286>
- Meuwissen, T.H.E., Hayes, B.J., Goddard, M.E., 2001. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics* 157, 1819–1829.
- Seye, A.I., 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>



## Aknowledgements

---

Ce manuscrit représente l'accomplissement de plus de trois ans de thèse au sein de l'UMR Génétique Quantitative et Evolution dans l'équipe GQMS. C'est le résultat d'un travail d'équipe et à ce titre, je souhaite adresser ma reconnaissance et mes remerciements à toutes les personnes qui ont permis sa réalisation. Ces quelques mots marquent la fin d'une aventure enrichissante scientifiquement et humainement.

Je remercie en premier lieu la personne qui m'a accordé sa confiance depuis mon stage de master (M2) jusqu'à ma thèse, à savoir ma directrice de thèse, Laurence Moreau. Je tiens à la remercier d'avoir partagé ses connaissances, conseillée et guidée tout au long de cette aventure avec rigueur et passion. Au-delà de son rôle de directrice de thèse, je tiens à la remercier pour son rôle de cheffe d'équipe qui a contribué à la bonne ambiance dans l'équipe. Sa disponibilité, notamment lors des multiples réunions et relectures, ainsi que sa bienveillance, ont été essentielles pour mon apprentissage et la réussite de cette thèse.

Je remercie mes co-encadrants de thèse au sein de l'équipe GQMS, Alain Charcosset et Tristan Mary-Huard, pour leur suivi attentif de mon travail, leurs contributions précieuses lors de nos nombreuses réunions et relectures, ainsi que leurs conseils avisés. En particulier merci à Tristan d'avoir approfondi ma connaissance des modèles statistiques et des codes R et merci à Alain de m'avoir sensibilisée à la gestion de la diversité dans les programmes de sélection. Grâce à leur expertise et à leur soutien, j'ai pu prendre du recul et apporter des améliorations significatives à mon travail.

Je remercie également Christina Lehermeier qui m'a co-encadrée au sein de RAGT2n en début de thèse. Je tiens à lui adresser ma reconnaissance ainsi qu'à Simon Teyssède et au reste de l'équipe de Druelle (RAGT2n) pour leur accueil et le temps qu'ils m'ont accordé lors de mes visites. Celles-ci ont apporté un nouvel éclairage sur mon travail et une meilleure compréhension des programmes de sélection. Je tiens à remercier Gaëtan Touzy qui a assuré le relai ainsi qu'Aurore Béral pour leurs précieux conseils et leur aide dans ma poursuite professionnelle.

Je remercie les deux structures qui ont financé cette thèse : l'ensemble des partenaires du projet SAMMCR (RAGT2n, Limagrain, Lidea, MAS, KWS, Corteva et Syngenta) membres de l'association ProMais et l'ANRT. En particulier, je remercie RAGT2n d'avoir accepté de porter ma thèse. Mes remerciements vont aussi aux expérimentateurs et aux sélectionneurs de ces entreprises qui ont participé à la création et à l'évaluation du matériel SAMMCR. Ce fut un réel plaisir et privilège d'évoluer à l'intersection de la recherche publique et privée. Dans ce contexte, j'ai eu l'occasion de visiter et d'échanger avec différentes entreprises du projet. Je suis reconnaissante à Sofiane Mezouk, Chloé Boyard, Colin Guillaume et Muriel Archipiano d'avoir organisé mes visites et pris le temps de partager leurs connaissances.

J'adresse mes remerciements à Laurence Moreau et Cyril Bauland qui ont initié le projet SAMMCR et sans qui cette thèse n'existerait pas. Je remercie également les coordinateurs privés du projet SAMMCR Christina Lehermeier, Gaëtan Touzy et Colin Guillaume.

Merci à Cyril Bauland pour son rôle dans le projet SAMMCR et pour son suivi des essais aux champs. Je tiens à le remercier pour sa patience et la passion avec laquelle il m'a fait découvrir son métier lors de nos visites. Merci de m'avoir confié une partie de la préparation des essais et de m'avoir permis de comprendre les enjeux de la sélection variétale. Ce fut un vrai plaisir de discuter de notre passion commune pour le vin lors des sessions œnologie au Moulon.

Je remercie l'unité expérimentale de Saint Martin de Hinx qui a assuré le comptage et la préparation des grains, en particulier Carine Palaffre. De même, je remercie Sophie Pin pour le comptage et la

préparation des essais au Moulon. Je remercie également Valérie Combes et Delphine Madur qui ont assuré la préparation des échantillons ADN et l'assemblage des données de géotypage. Le travail conjoint de chacun d'entre eux a rendu cette thèse possible.

Je remercie les membres de mon comité de thèse : David Cros, Zulma Vitezica et Cécile Richard-Molard, pour les échanges stimulants et leurs conseils précieux.

Je remercie Aaron Lorenz et Leopoldo Sanchez d'avoir accepté de rapporter ce document, ainsi que les autres membres de mon jury de thèse, Christine Dillmann, David Cros et Chloé Boyard d'avoir accepté d'évaluer mon travail.

Je remercie les membres de l'équipe GQMS pour leur accueil et les moments de convivialité partagés ensemble. Cette thèse n'aurait pas été réalisable sans le soutien, l'encouragement, et la collaboration de chacun d'entre vous. Je suis profondément reconnaissante d'avoir eu l'opportunité de travailler avec une équipe aussi exceptionnelle.

Je souhaite aussi remercier les membres de l'UMR GQE pour leur sympathie et leur bonne humeur. Merci à Christine Dillmann pour son rôle de directrice d'unité qui se dévoue chaque jour et en particulier qui a su faire face à une situation sanitaire difficile en s'assurant du bien-être de chacun. Je remercie également Christine et Diala Abu-Awad de m'avoir permis de faire l'expérience de l'enseignement.

Je souhaite aussi remercier l'ensemble des thésards que j'ai eu le plaisir de côtoyer. En particulier, merci à Dimitri et Aurélien, avec qui j'ai partagé cette aventure. Merci pour toutes ces discussions qui ont contribué à faire de ma thèse ce qu'elle est. Merci d'avoir été mes compagnons de galère et de réussite et surtout de soirée pancakes ! Je n'oublie pas les anciens thésards : Adama, Antoine, Clément, Pauline et Simon qui m'ont guidée vers le choix de faire une thèse. Le temps est venu de passer le flambeau aux « nouveaux » thésards qui nous ont rejoints : Agustin, Annaïg et Baber, merci pour les bons moments passés ensemble.

*Ces dernières lignes reviennent à ma famille et à mes proches, pour leur soutien sans faille. En particulier, Jocelyn qui a partagé mon quotidien et a fait front avec moi en toutes circonstances.*

## Table of Contents

---

<b>ADDITIONAL INFORMATIONS</b> .....	<b>4</b>
<b>ACKNOWLEDGEMENTS</b> .....	<b>13</b>
<b>TABLE OF CONTENTS</b> .....	<b>15</b>
<b>LIST OF ABBREVIATIONS</b> .....	<b>19</b>
<b>GENERAL INTRODUCTION</b> .....	<b>23</b>
MAIZE BREEDING HISTORY.....	23
<i>From maize domestication to hybrid varieties</i> .....	23
<i>Hybrid breeding schemes</i> .....	26
CONTRIBUTION OF MARKERS TO BREEDING METHODS .....	30
<i>Understanding of the genetic basis of the phenotypic traits, QTL detection</i> .....	30
<i>Application of markers in plant breeding through marker assisted selection</i> .....	30
GENOMIC SELECTION, A PROMISING GENOMEWIDE MARKER-BASED APPROACH FOR SELECTION .....	31
<i>Definition of the concept of genomic selection</i> .....	31
<i>Genomic prediction models for hybrid prediction</i> .....	34
<i>Focus on training set optimization in genomic selection</i> .....	35
IMPLEMENTING GENOMIC SELECTION IN BREEDING PROGRAMS .....	37
REVISITING MAIZE HYBRID BREEDING SCHEME USING GENOMIC SELECTION COMBINED WITH FACTORIAL TRAINING SETS .....	40
OBJECTIVES OF THE THESIS AND ORGANIZATION OF THIS MANUSCRIPT .....	42
<b>CHAPTER 1. 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</b> .....	<b>47</b>
ABSTRACT .....	48
INTRODUCTION .....	49
MATERIALS AND METHODS .....	52
<i>Genetic material</i> .....	52
<i>Field trials</i> .....	53
<i>Variance decomposition on single-plot performances (without marker information)</i> .....	55
<i>Adjusted means</i> .....	57
<i>Genotyping and kinship estimation</i> .....	57
<i>Genomic Best Linear Unbiased Prediction (GBLUP) models</i> .....	58
<i>Predictive ability with different scenarios</i> .....	59
RESULTS.....	62
<i>Variance components and broad-sense heritability at the phenotypic level without marker information</i> .....	62
<i>Variance components obtained using marker information</i> .....	64
<i>Scenario 1-Using a sparse factorial design (F-1H) to predict new hybrid combinations in new environments</i> .....	65
<i>Scenario 2-Compare the efficiency of factorial versus tester designs as TRS</i> .....	67
<i>Scenario 3-Assess the impact of the number of hybrids per line in factorial designs when predicting T0 hybrids</i> .....	70
DISCUSSION .....	71
<i>Importance of SCA and its prediction</i> .....	71



<i>Efficiency of using a factorial design to predict new hybrids in new environments</i> .....	72
<i>Efficiency of the factorial approach compared to the tester approach</i> .....	73
<i>What does it imply for breeding programs?</i> .....	75
REFERENCES .....	76
ACKNOWLEDGMENTS .....	79
FUNDING.....	79
COMPETING INTERESTS.....	79
AUTHOR'S CONTRIBUTIONS .....	79
DATA AVAILABILITY .....	79
<b>CHAPTER 2. PORTABILITY OF GENOMIC PREDICTIONS TRAINED ON SPARSE FACTORIAL DESIGNS ACROSS TWO MAIZE SILAGE BREEDING CYCLES.....</b>	<b>83</b>
ABSTRACT .....	84
INTRODUCTION .....	85
MATERIALS AND METHODS.....	88
<i>Summary of the G0 plant material production and selection of the best candidate lines</i> .....	88
<i>New breeding cycle (G1)</i> .....	90
<i>Genotyping</i> .....	91
<i>Estimation of variance components and trait heritabilities</i> .....	91
<i>Ls-means and genetic gain estimation</i> .....	93
<i>Pedigree based Best Linear Unbiased Prediction (PBLUP) model</i> .....	93
<i>Genomic Best Linear Unbiased Prediction (GBLUP) models</i> .....	94
<i>Prediction scenarios</i> .....	96
<i>Predictive ability and statistical tests</i> .....	99
RESULTS.....	99
<i>Variance components and broad-sense heritability at the phenotypic level without marker     information</i> .....	99
<i>Ls-means and genetic gain</i> .....	102
<i>Scenario 1-Predictive ability within the G1 cycle and GS model comparison</i> .....	102
<i>Scenario 2-Efficiency of a factorial TRS for predictions across breeding cycles and comparison with     tester TRSs</i> .....	102
<i>Scenario 3a-Benefit of updating the factorial TRS across breeding cycles</i> .....	106
<i>Scenario 3b-Optimization of the composition of the factorial TRS for G1 hybrid predictions</i> .....	108
DISCUSSION .....	109
<i>SCA variance and its importance in hybrid breeding</i> .....	109
<i>Genetic gain after selection based on genomic predictions trained on a sparse factorial design</i> .....	110
<i>Predictive ability in the new generation and comparison of different GS models</i> .....	110
<i>Portability of genomic predictions trained on a sparse factorial across breeding cycles</i> .....	111
<i>Efficiency of factorial compared to tester TRSs for predictions across the breeding cycle</i> .....	112
<i>Benefit of updating the factorial TRS along breeding cycles</i> .....	112
<i>Optimization of the G1 hybrid set to phenotype to update the TRS</i> .....	113
CONCLUSIONS.....	115
REFERENCES .....	116
ACKNOWLEDGMENTS .....	120
FUNDING.....	120
COMPETING INTERESTS.....	120
AUTHOR'S CONTRIBUTIONS .....	120
DATA AVAILABILITY .....	120
<b>CHAPTER 3. SHORTENING MAIZE HYBRID BREEDING SCHEMES USING GENOMIC PREDICTIONS CALIBRATED ON FACTORIAL INSTEAD OF TESTER DESIGNS IMPROVES COST EFFICIENCY OVER CYCLES. ....</b>	<b>123</b>

ABSTRACT .....	124
INTRODUCTION .....	125
MATERIALS AND METHODS.....	128
<i>Founder lines</i> .....	128
<i>Simulation of QTL effects and positions</i> .....	129
<i>Initialization</i> .....	130
<i>Phenotypic selection breeding scheme</i> .....	131
<i>GS breeding schemes</i> .....	132
<i>Genomic selection model</i> .....	135
<i>Metrics used to assess the BS efficiency</i> .....	136
RESULTS.....	137
<i>Impact of the different BSs on the genetic gain</i> .....	137
<i>Impact of the different BSs on the genetic variance</i> .....	139
<i>Impact of the different GS breeding schemes on the prediction accuracy</i> .....	140
DISCUSSION .....	142
<i>Impact of implementing GS in a conventional breeding scheme</i> .....	142
<i>Impact of shortening the breeding cycle</i> .....	143
<i>Benefit of factorial designs in GS hybrid breeding schemes</i> .....	145
<i>Prospects for implementing factorial design in hybrid BSs</i> .....	147
CONCLUSION .....	148
REFERENCES .....	150
ACKNOWLEDGMENTS .....	154
FUNDING.....	154
COMPETING INTERESTS.....	154
AUTHOR'S CONTRIBUTIONS .....	154
DATA AVAILABILITY .....	154
<b>GENERAL DISCUSSION AND PERSPECTIVES .....</b>	<b>157</b>
COMPARISON OF PREDICTIVE ABILITIES OBTAINED WITH FACTORIAL AND TESTER TRSs IN EARLY STAGE EVALUATIONS .....	158
SINGLE-CROSS HYBRID PREDICTION USING A FACTORIAL TRS .....	159
<i>Efficiency of genomic predictions calibrated on sparse factorial TRSs</i> .....	159
<i>Genomic predictions calibrated on a sparse factorial TRS across cycles</i> .....	160
<i>SCA and its prediction</i> .....	160
OPTIMIZATION OF THE FACTORIAL TRS.....	161
<i>Composition of the factorial TRS</i> .....	161
<i>Updating factorial TRS across cycles</i> .....	162
IMPLEMENTING FACTORIAL DESIGNS IN GS BREEDING SCHEMES.....	164
<i>Combining GS and factorial TRSs to revisit hybrid breeding schemes using simulations</i> .....	164
<i>Prospects for extending simulations</i> .....	165
PERSPECTIVES TO IMPROVE GS MODELS FOR HYBRID BREEDING .....	167
<i>Accounting for genotype by environment interactions (GxE) in GS models</i> .....	167
<i>Considering multi trait GS models</i> .....	167
<i>Integrating omics data in prediction models, prospects open by phenomic selection</i> .....	168
PERSPECTIVES FOR HYBRID BREEDING SCHEMES INCORPORATING GS.....	169
<i>Sparse tester designs, an interesting compromise between tester and sparse factorials?</i> .....	169
<i>Combining GS with rapid cycling and speed breeding approaches to speed up the breeding process</i> .....	169
<i>Diversity management in hybrid GS breeding schemes</i> .....	170
<b>CITED LITERATURE.....</b>	<b>172</b>
<b>SUPPLEMENTARY MATERIALS. SAMMCR PROJECT AND PLANT MATERIAL .....</b>	<b>185</b>

THE SAMMCR PROJECT.....	185
PLANT MATERIAL .....	186
<i>Production of the G0 parental lines</i> .....	188
<i>The G0_F-1H experimental design</i> .....	189
<i>The G0_F-4H and tester experimental designs</i> .....	190
<i>Production of the G1 parental lines and G1 hybrids</i> .....	191
PHENOTYPING.....	192
GENOTYPING .....	192
REFERENCES .....	194
<b>SUPPLEMENTARY MATERIALS. CHAPTER 1 .....</b>	<b>195</b>
<b>SUPPLEMENTARY MATERIALS. CHAPTER 2 .....</b>	<b>207</b>
<b>SUPPLEMENTARY MATERIALS. CHAPTER 3 .....</b>	<b>217</b>

## List of abbreviations

---

<b>AIC</b>	Akaike Information Criterion
<b>BLUP</b>	Best Linear Unbiased Predictor
<b>CD</b>	Coefficient of Determination
<b>CDmean</b>	Mean of the Coefficient of Determination
<b>CELL</b>	Cellulose content (in %)
<b>cM</b>	CentiMorgan
<b>DH</b>	Doubled Haploid
<b>DMC</b>	Dry Matter Content (in %)
<b>DMY</b>	Dry Matter Yield (in t/ha)
<b>DtSilk</b>	Date of Silking or date of female flowering (in days)
<b>DINAG</b>	Digestibility of the non-starch and non-soluble carbohydrates part of silage (in %)
<b>DINAGZ</b>	Digestibility of the non-starch, non-soluble carbohydrates and non-crude protein part of silage (in %)
<b>GBLUP</b>	Genomic Best Linear Unbiased Prediction
<b>GCA</b>	General Combining Ability
<b>GxE</b>	Genotype by Environment
<b>GQE</b>	Génétique Quantitative et Evolution- Quantitative Genetics and Evolution
<b>GQMS</b>	Génétique Quantitative et Méthodologie pour la Sélection- Quantitative Genetics and Selection Methodology
<b>GS</b>	Genomic Selection
<b>H<sup>2</sup></b>	Broad sense Heritability
<b>HCELL</b>	Hemicellulose content (in %)
<b>HWE</b>	Hardy-Weinberg Equilibrium
<b>iid</b>	independent and identically distributed
<b>ind</b>	independent
<b>INRAE</b>	Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement
<b>LD</b>	Linkage Disequilibrium
<b>LDLA</b>	Linkage Disequilibrium-Linkage Analysis

<b>LIGN</b>	Lignine content (in %)
<b>Ls-means</b>	Least Squares mean
<b>MAF</b>	Minor Allele Frequency
<b>MAS</b>	Marker-Assisted Selection
<b>MFU</b>	Milk Fodder Unit (in MFU/kg)
<b>NDF</b>	Neutral Detergent Fiber (in %)
<b>NIRS</b>	Near Infra-Red Spectroscopy
<b>PEV</b>	Prediction Error Variance
<b>PH</b>	Plant Height (in cm)
<b>PS</b>	Prediction Set
<b>QTL</b>	Quantitative Trait Locus
<b>RR-BULP</b>	Ridge Regression Best Linear Unbiased Prediction
<b>SAMMCR</b>	« Sélection Assistée par Marqueurs Multi-parentale Connectée Réciproque » – Multi-parental Connected Reciprocal Marker Assisted Selection
<b>SCA</b>	Specific Combining Ability
<b>TBV</b>	True Breeding Value
<b>TRS</b>	Training Set
<b>VS</b>	Validation Set

# **GENERAL INTRODUCTION**



## General introduction

---

Maize has been widely used worldwide over centuries for producing grain for food and feed. Its use as forage increased later with the development of silage in North America and Europe. More recently, substantial means have been devoted to energy production, in particular bioethanol in the US (135 million tonnes) (USDA 2021) and biogas in Germany (more than 1.3 million ha in 2021) (German Biogas association). It was grown on 206 million hectares in the world in 2021 and is currently the first cereal in the world in terms of production volume with 1.2 billion tons, followed by rice with 787 million tons and wheat with 770 million tons in 2021 (FAOSTAT 2021). Food security is increasingly a concern in the context of the growing human population size and climate change (Ritchie et al. 2023). Predictions suggest that climate change will significantly diminish crop production and quality in the near future (Merca et al. 2021). In light of these circumstances, scientists and breeders need to speed up the development of new varieties adapted to climate change. Given the growing demand for high-yield, high-quality maize crops and the rapid changes in climatic conditions, it is mandatory to advance the development of modern techniques and tools for maize breeding improvement. In this introduction, we will present maize breeding history and selection schemes before giving an insight into revisiting hybrid breeding schemes using genomic selection (GS) and factorial designs. Finally, we will briefly present the research objectives and organization of this PhD manuscript.

## Maize breeding history

### From maize domestication to hybrid varieties

Maize was domesticated at least 8,700 years ago in the mountainous regions of southeastern Mexico (Tenailon and Charcosset 2011). Genetic analyses using isoenzymes (Doebley et al. 1984) and SSR markers (Matsuoka et al. 2002) support the hypothesis that maize is the result of a single domestication event from teosinte (*Z. mays ssp. parviglumis*). Maize spread across the American continent and then to Europe following two introductions during the 15th century, one in Southern Spain and one in Northern Europe (Tenailon and Charcosset 2011). During its expansion, maize encountered many environmental conditions, resulting in highly differentiated maize populations adapted to local conditions.

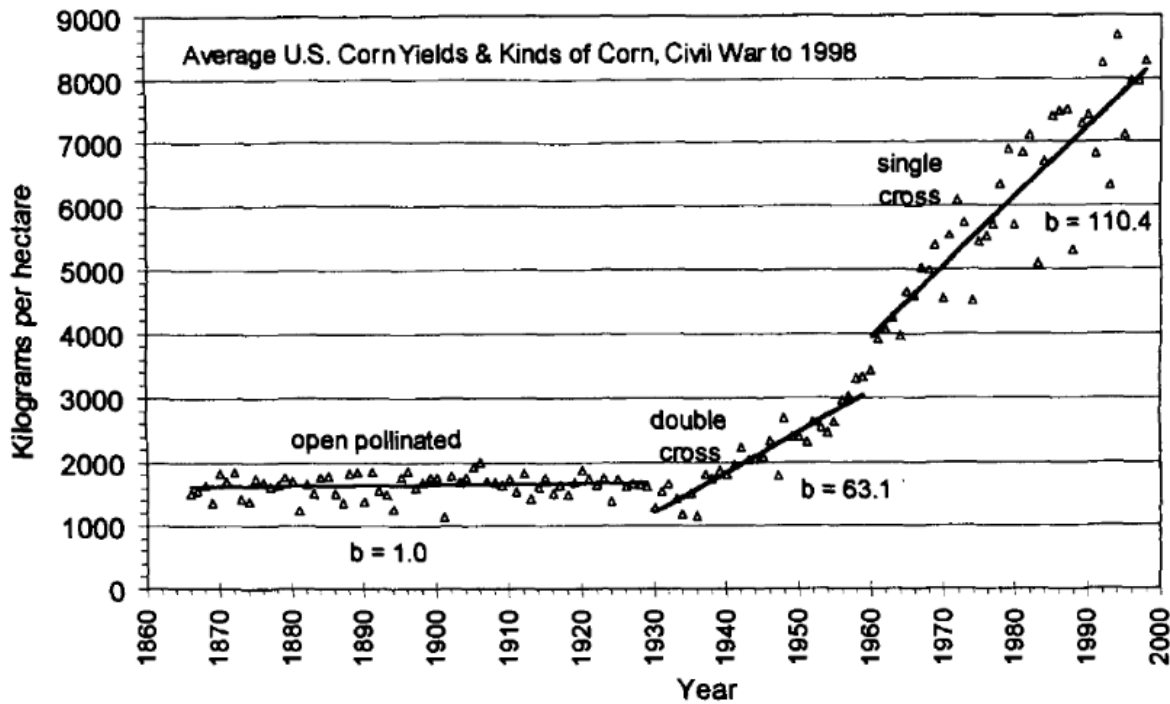


## *General introduction*

Mass selection done by the farmers began with the domestication of maize and continued until the middle of the 19th century. After harvest, farmers selected the ears matching their criteria (size, color, ect...). Then the seeds from the selected individual plants were planted to produce the next generation. In the case of allogamous species such as maize, these selected plants were pollinated at random by other individuals in the production field. The populations produced during this period were very different from each other. In the middle of the 19<sup>th</sup> century, with the work of Vilmorin (1856), the idea of improving the population by increasing their homogeneity emerged. This marked the start of the pedigree selection, in which individuals issued from a given parental plant were evaluated together as a family. The rediscovery of Mendel's laws at the end of the 19th century allowed the identification of the genetic foundations of pedigree selection and its widespread adoption. Following the success of pedigree selection in an allogamous plant (improving the sugar content in sugar beets Hallauer et al. 2010), Hopkins (1899) applied it to corn and referred to it as the "ear-to-row" method. The ear-to-row selection was used until about 1925 with limited success in improving yield but was highly effective in modifying the oil and protein content of the grains (Gallais 2009). This success was nevertheless accompanied by a loss in vigor due to inbreeding depression (see below).

The beginning of the 20th century marked a decisive turning point for corn improvement with the recognition of the benefit of hybridization. The first observation of a greater vigor of hybrids compared to their inbred parents can be traced back to the end of the 19<sup>th</sup> century with Darwin's research work (Darwin 1876). This stimulated further experiments demonstrating the benefits of hybridization between populations for increasing productivity (Beal 1880; Morrow and Gardner 1893). Later, the concomitant public research conducted by Shull (1908) and East (1908) outlined the basis of the breeding methods for developing and producing modern maize hybrids. They reported the deleterious effects of inbreeding when producing maize inbred lines. They showed that single-cross hybrids (F1) between two inbred lines restored the vigor and yield level from the original open-pollinated variety from which the inbred lines were developed, and could exceed it if lines and hybrids were selected. These observations led Shull (1909) to describe a method for producing single-cross maize hybrids: developing lines by self-fertilization before crossing them to produce high-vigor hybrids, and later, to the definition of the concept of "heterosis" (Shull 1914). Heterosis was defined as the greater vigor of the hybrid combination than that of its parents, replacing the numerous expressions describing the "hybrid

vigor". This definition was, however, descriptive and none of the underlying genetic mechanisms governing heterosis were identified at that time.

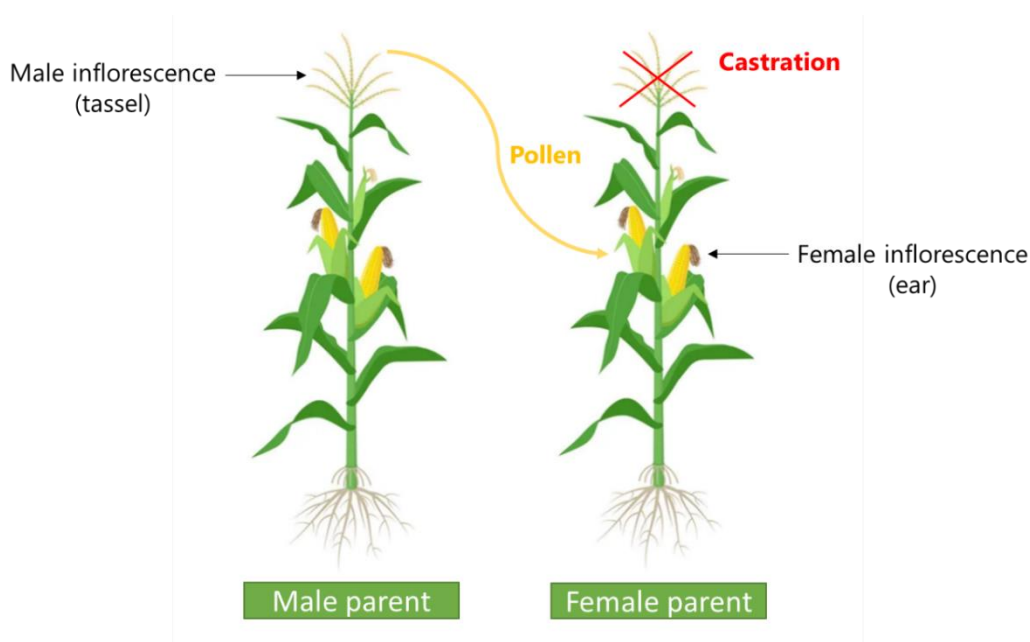


**Fig.1** Evolution of U.S maize yield performances from 1860 to 1998; periods dominated by open-pollinated varieties, double-cross hybrids and single-cross hybrids. Source: Troyer (1999)

Due to the poor vigor of inbred lines that did not allow sufficient seed production, researchers doubted the practical usefulness of this single-cross hybrid selection method. The limitation of poor seed production from inbred lines was overcome by an idea from Jones (1918), suggesting the use of double-cross hybrids resulting from the crossing of two single-cross hybrids. After this discovery, the transition from open-pollinated to double-cross hybrid maize was fast in the US (**Fig.1**) and significantly improved the standability and yield performances of maize. Double-cross hybrids were used because, based on the findings of East (1908) and Shull (1908, 1909), it was the only practical way to produce maize hybrids. In the following years, the continuous improvement in hybrid productivity has been accompanied by an increase in inbred lines per se value (Duvick 2001). The enhancement in the yield of inbred lines permitted the production of single-cross hybrids that were financially viable for farmers and had higher yields than double crosses (Crow 1998). Eventually, from the 1960s, single-cross hybrids started replacing double-cross hybrids (**Fig.1**).

## General introduction

The rapid transition from open pollinated to hybrid varieties was made possible by their easy production in isolation. Hybrid production was facilitated by the monoecy of the maize plant (i.e., a plant with female and male flowers on the same plant), which allowed the castration of male inflorescence without damaging the female inflorescence on the plant used as female, as illustrated in **Fig.2**.



**Fig.2** Hybrid maize seed production, using designated female and male parents and removing the tassels from the female plants before silk emergence, allowing male plants to provide the pollen for fertilizing the silks.

### Hybrid breeding schemes

Maize hybrid selection relies on the method proposed by Shull (1909), which consists of developing inbred lines before evaluating them in crosses to identify the lines that, when crossed, produce the best performing single-cross hybrids (Gallais 2009). This hybrid selection method has allowed breeders to produce genetically identical plants for every hybrid. The straightforward way to identify the best single-cross hybrids would be to produce all possible hybrid combinations between the candidate lines. However, the number of candidate lines can be large and increases every year, making this approach practically undoable. One initial method was to select lines based on their *per se* performance during the selfing generations used to produce inbred lines (Jenkins 1928), but the correlation between the *per se* value of a line and its average value in hybrid crosses is weak for traits exhibiting strong heterosis (Smith 1986). This is the case for yield, which demonstrates a strong heterosis likely due to numerous quantitative trait loci (QTLs) showing dominance effects (Carena et al. 2010). Another method

## General introduction

relied on the evaluation of candidate lines based on the performance of their hybrid progeny. Davis (1927) suggested the use of the testcross methodology (also referred to as topcross), consisting of crossing candidate lines in early-stage selection (not fully homozygous) with one or few “testers” to select and advance only the promising lines to the next selfing generation. This procedure was first tested by Jenkins and Brunson (1932) who showed that the average cross value of a line was better correlated with its cross value with any other “line” (hybrid, inbred line or population) used as “tester” than with its own *per se* value. After Jenkins and Brunson (1932) reported on its successful application, the testcross approach was widely used in breeding programs. Sprague (1939) and Jenkins (1940) suggested that the “combining ability” of the tested lines should be considered at early stages in the inbreeding process. And later, Sprague and Tatum (1942) formally defined combining abilities. They proposed decomposing the hybrid value into the General Combining Abilities (GCA) of each of the hybrid parents and the Specific Combining Ability between the parents (SCA). The GCA of a line  $i$  (respectively  $j$ ), denoted as  $GCA_i$  ( $GCA_j$ ), is defined as the average performance of that line in hybrid combinations. The SCA for the cross between line  $i$  and line  $j$ , denoted as  $SCA_{ij}$ , is defined as the deviation from the expected hybrid performance based on the GCAs of the two lines. After Sprague and Tatum (1942) introduced the notions of combining abilities, new approaches for testcross evaluation were proposed.

The introduction of recurrent selection methods (Hull 1945; Comstock et al. 1949) expanded the possibilities for enhancing maize population improvement. Recurrent selection is generally defined as an iterative process that alternates between phases of (i) evaluation of individuals and (ii) crossing the best individuals to produce the next generation. The purpose of recurrent selection methods is to enhance the frequency of favorable alleles associated with quantitative traits while preserving genetic diversity, thereby enabling continuous genetic enhancement in the long term (Hallauer et al. 2010, Chapter 6). Different recurrent selection methods have been proposed, focusing on the GCA and/or SCA selection (Hull 1945; Comstock et al. 1949). In methods targeting GCA, candidate plants are evaluated based on their hybrid progeny with a tester consisting in a broad base heterogeneous population (i.e., the parental population or any broad genetic base unrelated population). In selection for SCA, a narrow genetic base (inbred line or single-cross) tester is used. It has been shown that the efficiency of these methods depends on the presence of dominance or overdominance in the tested

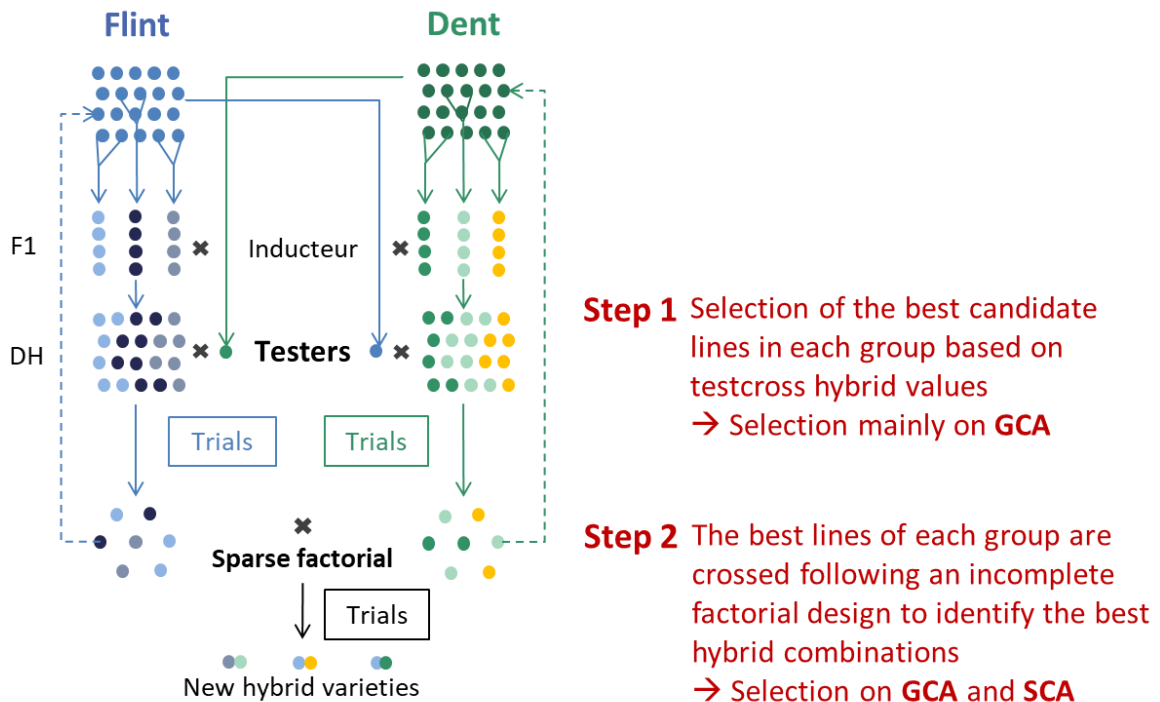
## General introduction

population (Hull 1945; Comstock et al. 1949). In 1949, Comstock proposed a new selection procedure to select for both GCA and SCA, referred to as reciprocal recurrent selection, which consists of simultaneously improving the two parental populations of the hybrid, testing candidate plants from one population by crossing them to candidate plants from the other population. Several examples of reciprocal recurrent selection success in maize have been reported (Coors 1999; Hallauer et al. 2010, chapter 7).

Progressively, the genetic diversity of maize has been structured into heterotic groups that have been defined by identifying different hybrid combinations showing stronger complementarity (Beal 1880; Hallauer et al. 2010). A heterotic group can be defined as “a group of related or unrelated genotypes from the same or different populations, which display similar combining ability and heterotic response when crossed with genotypes from the other genetically distinct germplasm groups” (Melchinger and Gumber 1998). Maize hybrid breeding relies on improving heterotic groups and identifying the inbred parents from distinct heterotic groups that yield superior hybrids when crossed together.

In this context, the challenges of maize hybrid breeders are (i) to select lines that will be used as parents for the next generation and (ii) to identify the best single-cross hybrids among all possible ones to derive new varieties. The most widely used hybrid breeding method today is a two-step process based on a recurrent reciprocal selection procedure. **Fig.3** illustrates this two-step method. Within each heterotic group, inbred lines are derived and improved in a reciprocal recurrent selection scheme designed to enhance their combining ability with the complementary heterotic group so that their crosses improve performance over selection cycles. In the first step, testers from the complementary heterotic group are used to preselect candidate lines based on their GCA. In this scheme, the testers are usually inbred lines or a single-cross hybrids from the complementary heterotic group, which, when crossed with lines to be tested, reveals differences in combining abilities. In the second step, the selected lines from step 1 in each group are crossed according to a factorial design to identify the best hybrid combinations. When lines are developed by self-pollination, this hybrid breeding method remains lengthy. In the last two to three decades, Doubled Haploid (DH) technology has emerged in maize. The technology produces completely homozygous lines in one step from gametes which reduces the length of the breeding scheme and increases its efficiency. DH lines are now used in routine in most commercial maize breeding schemes (see review from Chaikam et al. 2019).

**The conventional reciprocal recurrent selection maize hybrid breeding scheme**



**Fig.3** Conventional reciprocal recurrent selection scheme for maize hybrids between two complementary heterotic groups, using DH to produce inbred lines.

This two-step selection scheme has some limitations. First, the use of tester designs within each heterotic group requires the phenotyping of a large number of hybrids, at least as many hybrids as candidate lines in each group in the first step and the phenotyping of a large number of hybrid combinations in the second step. Secondly, the use of a small number of testers (often one or two) can influence the estimation of GCAs from testcross performances when dominance variance is high and bias selection in favor of lines that combine particularly well with a given tester (Vitezica et al. 2016). Thirdly, given that testcross means reflect the line GCAs, SCA is accounted for only in the incomplete factorial between complementary populations used for commercial hybrid selection in the late stage selection. And last, this method does not allow for the testing of all hybrid combinations between candidate lines living room to miss out on promising superior hybrids.

## **Contribution of markers to breeding methods**

### **Understanding of the genetic basis of the phenotypic traits, QTL detection**

With the development of markers starting in the 1980s, it became possible to study the genes or loci (or QTL, Quantitative Trait Locus) involved in the variation of quantitative traits. QTL detection relies on identifying statistical associations between marker polymorphisms and the considered phenotypic variation. QTL detection can have two objectives: identifying causal genes or polymorphisms and/or using marker-QTL associations in Marker-Assisted Selection (MAS) approaches. A major challenge in quantitative genetics is to map QTL and gain insight into their mode of action to optimize their use in selection. Two main approaches exist for mapping QTL in major crop species. The most common approach, called linkage mapping (or "Linkage Analysis mapping"), involves searching for marker-QTL associations in segregating populations resulting from controlled crosses (often between two inbred lines) (Lander and Botstein 1989; Bernardo 2008). This approach leads to a low resolution of the QTL positions due to the limited number of recombination events within this type of population. Furthermore, each biparental population represents only a small proportion of the genetic variability available for selection. It has therefore appeared interesting to perform QTL detection on a broader genetic basis, using, for example, multiple connected segregating populations (Rebaï et al. 1997; Bardol et al. 2013; Giraud et al. 2014). The second approach, called association mapping (or "linkage disequilibrium mapping"), involves using linkage disequilibrium between markers and QTL in an existing population with a looser pedigree structure, typically a diverse collection. This approach leverages historical recombination events and thus provides better resolution than the linkage analysis mapping approach. However, the history of this population can lead to its structuring and to relatedness among individuals, which can generate linkage disequilibrium between physically unlinked loci (Yu et al. 2006). Meuwissen et al. (2001) proposed the LDLA approach ("Linkage Disequilibrium-Linkage Analysis"), which consists in using segregating populations with known structure and association mapping models and combines the advantages of both QTL identification methods in terms of power and resolution of their positions.

### **Application of markers in plant breeding through marker assisted selection**

Markers linked to QTLs can be used for breeding purposes with the Marker-Assisted Selection (MAS) approach, first discussed by Neimann-Sorensen and Robertson (1961) in the context of

animal breeding. Twenty years later, the advancement of genotyping techniques, opens the way for MAS, and several applications in animal and plant breeding emerged (see reviews by Soller (1994) and Collard and Mackill (2007)). The main application is the selection for traits determined by a major gene that are difficult or costly to phenotype. complex traits influenced by multiple QTL, Lande and Thompson (1990) proposed to estimate the genetic value of individuals by summing the effects at markers significantly associated with QTLs. It has been shown that this approach of MAS provides an advantage over conventional phenotypic selection for traits with relatively low heritability and can accelerate genetic gain by shortening cycles (Hospital et al. 1997; Moreau et al. 1998). However, MAS has several limitations: (1) QTL effects depend on the mapping population used during detection, and there is a possible loss of marker-QTL associations over generations of selection due to recombination events, (2) selection based solely on marker information quickly leads to the fixation of alleles favorable to major QTLs, while unfavorable alleles can become fixed at smaller-effect QTLs that are generally not detected (Hospital et al. 1997), and (3) due to the small QTL detection power, only a subset of QTL are detected as significant, and their effects are overestimated due to a selection bias ("Beavis effect", Beavis 1998), resulting in a reduction in the efficiency of MAS (Moreau et al. 1998, 2004).

## Genomic selection, a promising genomewide marker-based approach for selection

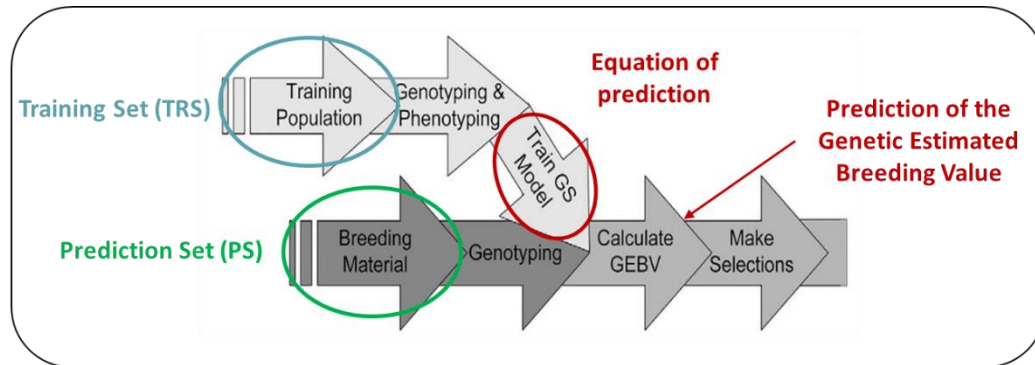
### Definition of the concept of genomic selection

In the 2000s, the emergence of cost-effective high-throughput SNP genotyping and statistical improvements made it possible to utilize genome-wide markers and paved the way for a novel marker-based selection approach: Genomic Selection (GS). While MAS considers only a limited number of markers significantly linked to "major" QTLs, GS uses a large number of markers distributed throughout the genome as predictors of the genetic value of candidate individuals for selection without attempting to locate QTLs precisely or estimate their effects (Meuwissen et al. 2001). GS relies on the assumption that quantitative traits are influenced by a larger number of QTLs with small effects (infinitesimal model by Fisher (1919)). By using genome wide high-density genotyping, it ensures that all QTLs are in Linkage Disequilibrium (LD) with at least one marker.



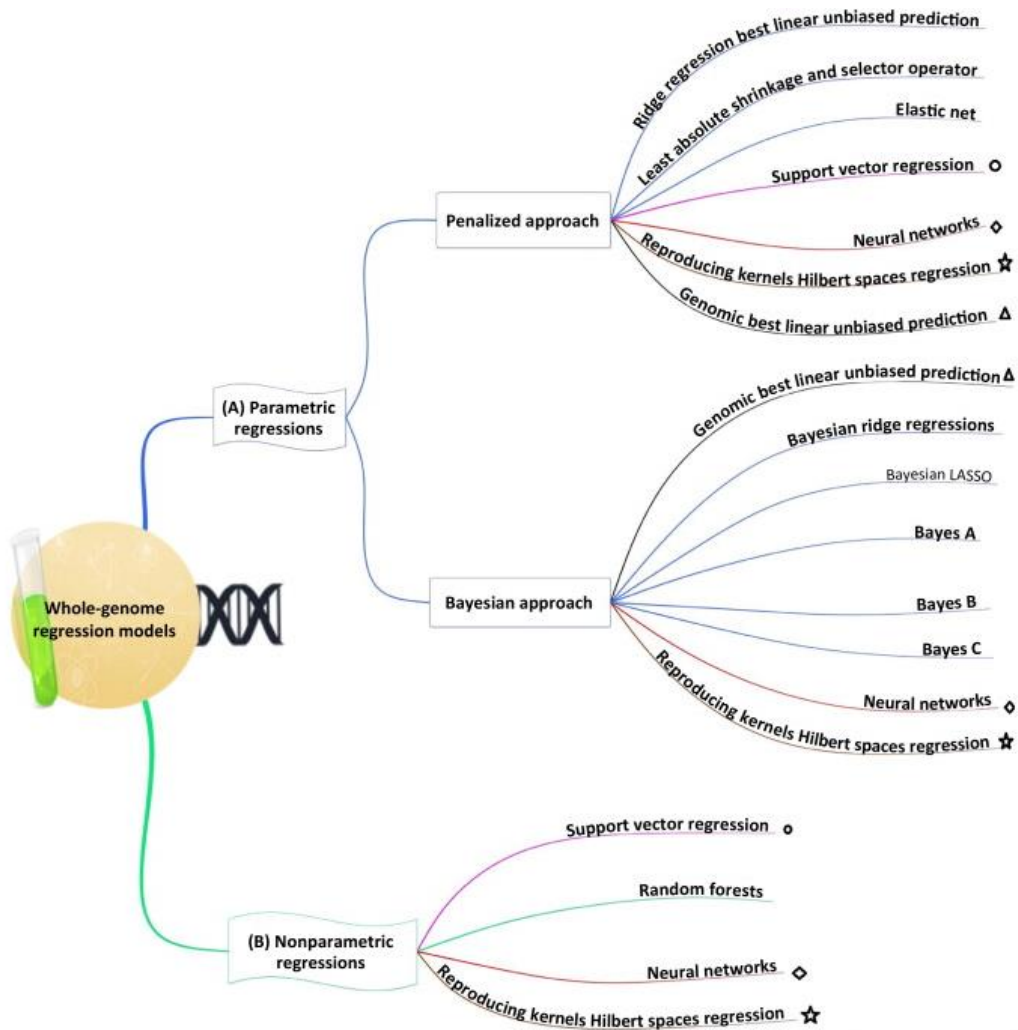
## General introduction

GS relies on a prediction equation trained on a set of genotyped and phenotyped individuals, referred to as a Training Set (TRS), to predict the genetic values of a set of selection candidates based on their marker genotypes, referred to as Prediction Set (PS) (**Fig.4**). In practice, a validation population (Validation Set, VS) that is both genotyped and phenotyped is often used to assess the predictive ability of the model being tested.



**Fig.4** Diagram of genomic selection process, using a training population or training set (TRS) to calibrate an equation of prediction and predict the genomic estimated breeding value (GEBV) of a prediction set (PS) to select individual (Adapted from Heffner et al. 2009)

GS poses several statistical and computational challenges, such as dimensionality, collinearity between markers, and the complexity of quantitative traits. The challenge of dimensionality corresponds to the larger number of effects to be estimated compared to the size of the training population, which makes it impossible to use a classic linear model (fixed effects). Since introducing the first GS models (Whittaker et al. 2000; Meuwissen et al. 2001), several alternative models have been developed to address this problem. Prediction equations are based on various methods: parametric (Bayesian (Bayes A, B, C), GBLUP, RRBLUP) (Meuwissen et al. 2001), or non-parametric (reproducing kernel Hilbert space regression) (see Howard et al. 2022 for a review). A simple scheme representing the main prediction models is presented in **Fig.5**. The simplest, robust and most widely used method, which we will focus on in this work, is the Genomic Best Linear Unbiased Prediction (GBLUP) method. The initial concept of the GBLUP can be traced back to the marker-based prediction proposed by Bernardo (1994). The GBLUP uses markers to estimate the relatedness between individuals in the TRS and PS individuals (VanRaden 2008).



**Fig.5** Classification of the different GS regression models (source Desta and Ortiz 2014)

The classic GBLUP model can be written as:

$$y = X\beta + Za + \varepsilon,$$

where  $y$  is the vector of the  $n$  phenotypic observations,  $\beta$  is the vector of fixed effects (e.g. general mean, trial, year, treatment effect...),  $X$  the incidence matrix of the fixed effects,  $Z$  is the incidence matrix of random effects linking phenotypes to genetic values,  $a$  is column the vector of random effects of the additive genetic values of all individuals (from the TRS and PS) with  $a \sim \mathcal{N}(0, K\sigma_a^2)$  and  $K$  is the genomic relationship matrix (kinship) that models the covariance between individuals of the TRS and PS and  $\sigma_a^2$  the additive genetic variance.  $\varepsilon$  is the error term following  $\varepsilon \sim \mathcal{N}(0, I\sigma_\varepsilon^2)$  where  $\sigma_\varepsilon^2$  is the error variance. The marker effects can be estimated by back-solving of the GBLUP model (Wang et al. 2012) thanks to the statistical equivalence that has been shown with the ridge regression best linear unbiased prediction model (RR-BULP) that estimates marker effects instead of genetic values, assuming that all marker effects are drawn

from a unique normal distribution and considers.

The ultimate assessment of the efficiency of genomic predictions would be to compute the correlation between the True Breeding Value (TBV) and the predicted value, referred to as prediction accuracy. However, in most cases, TBV is not available. Instead predictive abilities are computed as the correlation between the observed value (phenotype) and the predicted value in a suitable Validation Set (VS). Various factors are known to affect genomic prediction accuracies. They can be classified into four categories: trait-specific (heritability, genetic architecture of the trait), population-specific (level of linkage disequilibrium (LD), allelic frequencies), statistical methods used to perform predictions and experiment-specific parameters (marker density, size of the TRS, degree of relatedness between the TRS and PS) (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).

### **Genomic prediction models for hybrid prediction**

Bernardo (1994) was the first to propose a marker-based model for plant hybrid performance prediction. He combined marker-based similarities 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 aimed 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 is now widely used to perform GS in plants and animal (Heslot et al. 2015; Crossa et al. 2017). Since the pioneering work of Bernardo (1994), using a model decomposing genetic effects into General Combining Ability (GCA) and Specific Combining Ability (SCA), different prediction models adapted to hybrid value prediction have been proposed. One method is to decompose genetic effects into additive and dominance effects (Su et al. 2012) based on hybrid genotypes without considering the genetic group of origin of the alleles. This approach has allowed the modeling of dominance effects in animal populations and was further applied to inter-breed hybrids in both animals (Vitezica et al. 2016; Varona et al. 2018) and plants (Piaskowski et al. 2018). Most models used today assume Hardy-Weinberg Equilibrium (HWE), but in many situations, this assumption is not valid (Vitezica et al. 2013). In response to this

## General introduction

observation, (Vitezica et al. 2017) proposed a genomic prediction model that considers additive, dominance, and epistatic effects in a general context (without HWE). In the same context, more recently, González-Diéguez et al. (2021) proposed a new orthogonal model decomposing the hybrid genetic value into GCA and SCA components with a new formula for the computation of the SCA kinship. 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.

### Focus on training set optimization in genomic selection

The effectiveness of genomic selection is closely tied to the prediction accuracy of candidates to selection. The design of the TRS is a key factor affecting prediction accuracies that have attracted interest in the breeding community. GBLUP genomic prediction models assume that TRS and PS individuals are drawn from the same population (Rio et al. 2022b). However, this hypothesis is not met in several situations, and may result in low prediction accuracy. This can be due to differences in LD between molecular markers and QTLs (defined as the non-independence between alleles at different loci on the same gamete), QTL allele frequencies and QTL allele effects between the TRS and PS (Rio et al. 2022b). The first studies addressing the question of the TRS design relied on random sampling without any optimization procedure (Habier et al. 2007; Pszczola et al. 2012; Windhausen et al. 2012; Wientjes et al. 2013). They gave insight into the composition of an ideal TRS and highlighted the importance of the relationship between the TRS and PS. 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; Riedelsheimer et al. 2013; Lorenz and Smith 2015) and minimizing the within TRS relationship to capture a large genetic variance (replicating alleles rather than individuals) (Pszczola et al. 2012).

Random sampling does not always guarantee maximum prediction accuracies. For example, when a strong population structure is involved, random sampling can lead to the under or over-representation of some groups in the TRS, leading to a reduced accuracy (Windhausen et al. 2012; Isidro et al. 2015). This prompted the exploration of optimization procedures for constructing the TRS. Different optimization criteria have been proposed to optimize *a priori* the TRS that can be grouped into "Model-free" and "Model-based" optimization criteria (Rio et al. 2022b). Model-free criteria do not rely on a genomic prediction model, instead, they aim at

## *General introduction*

constructing TRS that resemble the PS. On the contrary, model-based criteria are derived within the mixed model theory of genomic prediction models to maximize the expected prediction accuracy. Rincent et al. (2012) first reported the benefit of using model-based criteria maximizing prediction accuracy. Rincent et al. (2012) derived criteria from the mixed model equations proposed by Laloë (1993) with the objective of either maximizing the Coefficient of Determination (CD), or minimizing the Prediction Error Variance (PEV) of the genetic values of the PS. In particular, he proposed optimizing the TRS to be phenotyped by maximizing the mean of the coefficient of determination (CDmean) of contrasts between each unphenotyped PS individual and the target (PS) population mean. This criterium relies only on molecular information and trait heritabilities of the PS and the individuals that are candidates to be part of the TRS. Numerous studies have shown that building the TRS using the CDmean significantly increases the accuracy of GS models relative to using random sampling (see reviews: Isidro y Sánchez and Akdemir 2021; Rio et al., 2022; Fernández-González et al. 2023). Later, the CD was extended and applied to other situations. Isidro et al. (2015) proposed stratified CDmean to improve the optimization of TRS under population structure effects, Rincent et al. (2017), proposed the CDpop to maximize prediction accuracy in population composed of biparental families by defining contrasts within each family (CDpop). Generalized CD can be derived from multi traits and multi environments genomic prediction models to compute the expected reliability for each individual–trait or individual–environment combination. This was done by Ben-Sadoun et al. (2020) with the CDmutli for the optimization of a TRS under a multi trait context.

Most of the criteria were developed at the line level, and only a few were adapted to the hybrid context. However, model-based criteria rely on genomic prediction models and therefore can be adapted using models specific to hybrid prediction including non-additive effects. This was done for example by Momen and Morota (2018) and Fristche-Neto et al. (2018). They have included non-additive effects in the computation of the CDmean, however, these authors did not find a clear benefit of accounting for dominance in the CDmean computation.

## Implementing genomic selection in breeding programs

While GS methods were first introduced in plant breeding (Bernardo 1994; Whittaker et al. 2000), animal breeders developed GS applications much faster due to the time and cost savings it offered (Hickey et al. 2017a), the rapid development of dense genotyping arrays. Indeed, in animal breeding and more specifically, in dairy cow breeding, time is required for the animal to grow, and often, hundreds of progeny individuals are required to obtain reliable prediction of an adult breeding value, which is cost extensive. This integration of GS made it possible to predict the potential merit of an individual soon after birth in a more precise way than what was possible using only pedigree information, saving time and costs. GBLUP based on the computation of a marker-based kinship (Van Raden 2008) and its derivations, such as single-step BLUP, have been widely used in practical animal breeding (Hayes et al. 2009; Hickey et al. 2017a) replacing for many species evaluation based solely on pedigree-based BLUP. In contrast, the benefit of introducing GS methods into plant breeding programs was initially less clear, given genotyping costs compared to phenotyping and the fact that breeders were not familiar with the use of BLUP models.

A significant advancement enabling the integration of GS in plants was the reduced cost of high-throughput genotyping methods. Numerous empirical studies reported predictive accuracies of different GS models in several plant species (Jannink et al. 2010). Cost-benefit analyses concerning integrating GS into breeding programs indicated that it could be cheaper than phenotypic selection (Crossa et al. 2017). Indeed, estimated genetic values can be used in place of phenotyping to restructure existing breeding programs to: (i) shorten the breeding cycle by replacing early-stage phenotypic evaluations with genomic predictions, (ii) increase selection intensity by increasing the number of candidate lines, and (iii) increase selection accuracy by increasing the precision of estimation of the candidate lines (Heslot et al. 2015; Crossa et al. 2017). The relationship between the additive genetic variance and the expected response to selection is known as the breeder's equation (Lush 1937). Assuming an infinite breeding population and a normally distributed targeted trait, the expected change in mean performance ( $\Delta\mu$ ) per generation can be expressed as:

$$\Delta\mu = \frac{ir\sigma_a}{t},$$

where  $i$  is the selection intensity,  $r$  is the selection accuracy,  $t$  is the generation interval and  $\sigma_a$  is the population additive genetic standard deviation of the targeted trait. By acting upon three parameters of the breeder's equation: reducing the generation interval and increasing the selection intensity and accuracy, GS can increase the genetic gain as reported in several simulation and experimental studies as reviewed for instance by Krishnappa et al. (2021). However, depending on the species, the organization of the breeding programs and the constraints on breeding may be very different. Consequently, most studies are species-specific and can show varying benefit of implementing GS (R2D2 Consortium et al. 2021). GS has now become a routine practice in many breeding programs for major crop species such as maize.

Bernardo and Yu (2007) were the first to report the successful use of GS in a maize hybrid breeding scheme using simulation data in biparental populations. They showed that GS provided 18 to 43% more genetic gain per cycle than marker-assisted recurrent selection based on QTLs. Later, other studies have reported large increases in rates of genetic gain as a result of implementing GS in hybrid breeding programs (Wong and Bernardo 2008; Technow et al. 2012; Massman et al. 2013; Technow et al. 2014; Xu et al. 2014; Beyene et al. 2015; Kadam and Lorenz 2018; Seye et al. 2020; Bernardo 2021a; Beyene et al. 2021). The incorporation of GS in breeding programs can vary in terms of both the stages at which it is applied and the extent to which is it utilized. It is possible to use GS at some steps or at all steps, this was investigated by Bassi et al. (2016) showing a higher genetic gain when applying at all stages. Some experimental and simulation studies went a step further and performed multiple GS cycles without phenotyping to reduce cycle time corresponding to a "rapid cycling" approach (Massman et al. 2013; Beyene et al. 2015; Zhang et al. 2017; Seye et al. 2020; Bernardo 2021a).

The implementation of GS significantly impacts various aspects of the breeding program. This includes allocating resources at a constrained budget, selection pressure at different stages, and the duration of the breeding cycle. There are inherent tradeoffs that emerge, such as deciding how to reallocate resources between phenotyping and genotyping, as well as determining the balance between the number of candidate lines and the accuracy of their evaluation. However, studies rarely address resource allocation at the level of a breeding program and do not compare breeding schemes at fixed costs. In wheat, two simulation studies



## *General introduction*

investigated the benefit of implementing GS in a commercial breeding program at fixed costs and with optimized resource allocation (Longin et al. 2015; Ben-Sadoun et al. 2020). A recent simulation study by Bernardo (2021) investigated the use of different GS strategies at a fixed cost in the early selection stage of a maize breeding program carried out on a biparental population and showed that speeding up selection by using two cycles of GS (without phenotyping) allowed larger genetic gains than phenotypic selection.

When implementing GS in breeding programs, an important factor to consider is the update of the TRS (Isidro y Sánchez and Akdemir 2021; Rio et al. 2022b). In breeding programs, a large dataset from multiple trials and multiple years of phenotyping is often available for model training and can be used to predict new genotypes with high prediction accuracy (Sleper et al. 2020). However, in some cases, it has been shown that including data from previous generations or from genetically distant individuals can deteriorate the prediction accuracy (Lorenz and Smith 2015; Auinger et al. 2016). In this situation, the TRS optimization aims at selecting a subset among the phenotyped individuals to maximize prediction accuracy. Additionally, to maintain high prediction accuracy over generations, the TRS must be regularly updated with newly genotyped and phenotyped individuals (Neyhart et al. 2017; Brandariz and Bernardo 2018; Lopez-Cruz and de los Campos 2021). Often breeders are working with a constrained budget where only a subset of the candidate individuals in the new generation can be phenotyped. Optimizing the TRS represents an opportunity to better allocate resources within breeding programs by reducing the number of individuals that need to be phenotyped to train GS models while maintaining similar accuracy (sparse or selective phenotyping) (Akdemir and Isidro-Sánchez 2019). In this context, 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. This question has rarely been addressed in the studies investigating TRS optimization for hybrid breeding over generations. 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.



## **Revisiting maize hybrid breeding scheme using genomic selection combined with factorial training sets**

The conventional hybrid breeding scheme is divided into early-stage evaluations on testcross hybrids and late-stage evaluations on single-cross hybrids. Several studies evaluated the efficiency of GS to predict hybrids, looking separately at the two steps of hybrid schemes. Some of them have evaluated the interest of GS at the early stage, i.e., the selection of the unselected candidate lines based on testcross hybrids (Lorenzana and Bernardo 2009; Albrecht et al. 2011; Lehermeier et al. 2014; Krchov and Bernardo 2015; Seye et al. 2020; Burdo et al. 2021). Other studies addressed GS for single-cross hybrid value prediction in the second stage of the hybrid breeding scheme, i.e., by using as TRS single-cross hybrids between lines that have already undergone a selection based on their testcross to predict the value of unobserved single-cross hybrids (Maenhout et al. 2010; Technow et al. 2012, 2014; Massman et al. 2013; Kadam et al. 2016). The remarkably high predictive abilities of the single-cross hybrids, ranging from 0.70 to greater than 0.90 has opened prospects for considering the use of single-cross hybrids in early-stage selection and revisiting hybrid breeding schemes (Kadam and Lorenz 2018).

Kadam et al. (2016) and Giraud (2016) proposed replacing testcross testing in the early stage of the process with genomic predictions trained on a sparse factorial design to predict all potential hybrid combinations, as well as GCA in view of parent selection for subsequent generations. A factorial design based on crossing candidate lines from two complementary groups would enable a better utilization of phenotypic information, as each hybrid provides information about the value of candidates from both heterotic groups. In the absence of markers, estimating the GCA of lines and SCA of hybrid combinations requires that each line contributes to several hybrids in the factorial design, which increases the number of hybrids to be evaluated compared to tester-based evaluations. Molecular information could overcome this limitation. Since several hybrids share each allelic combination, it is possible to decompose the total hybrid value into GCA and SCA effects for each marker, even if each line contributes to only one hybrid in the factorial design (Seye et al. 2020). This could reduce the number of hybrids to be phenotyped for the same number of evaluated lines. By eliminating testers, it would be possible to work directly at the hybrid value level to improve both GCA and SCA components from the early stages of selection, avoiding the bias associated with tester choice for GCA estimation. Thus, using factorial design would allow for a better estimation of GCA and

consideration of SCA in early stages of selection. Last, the use of genomic predictions in the early-stage of the process would allow the prediction of all potential hybrids to avoid missing out on the most promising ones.

Several studies investigating the use of factorial designs in early-stage selection found good prediction accuracies for untested hybrids (Giraud 2016; Kadam et al. 2016; Fristche-Neto et al. 2018a; Seye et al. 2020; Burdo et al. 2021). In addition, a study based on simulations by Seye et al. (2020) compared the use of sparse factorial designs with tester designs to train GS models. It revealed equivalent prediction accuracies when there was no SCA variance in the population and an increasing benefit of the factorial design when the SCA variance increased. Moreover, even when no SCA is involved, Seye et al. (2020) illustrated the ability of the GS model trained on very sparse factorials to decouple line GCA components and predict their value accurately. Using factorial design requires genotyping candidate lines and producing hybrids by hand-made pollination, which is challenging and costly. Therefore, even if the use of factorial design in early-stage evaluations is appealing and its efficiency to calibrate GS models compared to testcross evaluations was assessed by simulations (Seye et al. 2020), further investigations and experimental validation are needed to investigate its efficiency along cycles and to optimize the factorial TRS. This also offers the possibility of revisiting the hybrid breeding scheme by collapsing early and late-stage evaluations into one single step using genomic predictions calibrated on a factorial design to predict both line GCAs and unobserved hybrid values. This idea was mentioned by Seye et al. (2020) but has never been tested to our knowledge. It appears necessary to assess the cost-efficiency of GS breeding schemes incorporating factorial designs compared to a conventional hybrid breeding scheme based on testcross evaluations. Studies have yet to be carried out simulating a hybrid breeding scheme including both early and late-stage selection in a reciprocal selection scheme performed in complementary groups at fixed costs.

## **Objectives of the thesis and organization of this manuscript**

The PhD work presented in the following manuscript is part of the **SAMMCR** project ("Marker-Assisted Multi-Cross Reciprocal Selection") initiated by INRAE and seven breeding companies within the framework of the **Promais Association** (<http://pro-maize-corn.com/>): KWS, Limagrain, Lidea, Syngenta, MAS, Corteva and RAGT2n. The project aims at studying and implementing alternative hybrid breeding methods, and validate their interest in a pilot breeding program and using simulations. More specifically, the objective was to develop marker-assisted prediction approaches based on QTLs or genomic predictions in a connected reciprocal multi-parental factorial design to improve hybrid value for silage maize. Several research questions were addressed in previous PhD thesis regarding (i) the effectiveness of sparse factorial designs for detecting QTLs involved in GCA and SCA (Giraud 2016; Giraud et al. 2017a, b; Seye 2019; Seye et al. 2019), (ii) the quality of GCA and SCA predictions using marker-assisted selection or genomic selection approaches (Giraud 2016; Seye 2019), (iii) the optimization of the experimental factorial design, and its comparison with a traditional tester-based design by simulation. This thesis aims at complementing results from these previous theses by including new experimental data with a focus on investigating the efficiency of replacing testcross evaluations by genomic predictions calibrated on a factorial designs to improve maize hybrid breeding schemes.

The main objectives of this thesis were to (i) validate on experimental data the efficiency of factorial TRSs for the prediction of untested single-cross hybrids and compare it with tester-based TRSs, (ii) further investigate the use of factorials and their optimization to predict line GCAs and hybrid values across two breeding cycles, and (iii) provide recommendations for implementing genomic predictions calibrated on factorial designs into maize hybrid breeding schemes. This thesis work relies on experimental data corresponding to two breeding cycles of a maize reciprocal genomic selection scheme involving multiparental connected populations from the flint and dent complementary heterotic groups, selected for silage performances. Experimental data were completed by simulations. The results are presented in three main chapters, each written in a scientific paper format, followed by a general discussion and perspectives. An overview of the SAMMCR project, the plant material and the different experimental designs developed during the project and used in this thesis are presented in the Supplementary materials section at the end of the document. Figures and tables are numeroted

within each section of the manuscript.

In **Chapter 1**, data from the first breeding cycle was used to evaluate the efficiency of sparse factorial TRSs for predicting single-cross hybrids between unselected lines from the same generation and compare it to tester designs. At a same number of hybrids and lines, the factorial design was as efficient as the tester designs, and, for some traits, outperformed them. We showed an advantage of increasing the number of lines involved in the TRS, by (1) allocating each of them to a different tester for the tester design, or (2) reducing the number of hybrids per line for the factorial design. This chapter corresponds to a paper published in *Theor. Appl. Genet.* (<https://doi.org/10.1007/s00122-022-04176-y>)

In **Chapter 2**, data from the two breeding cycles was used to evaluate the portability of genomic predictions calibrated on a factorial TRS across breeding cycles and investigate strategies to update and optimize the factorial TRS along cycles. The results validated experimentally the efficiency of sparse factorial designs for predicting line GCAs and hybrid values across two generations. They also highlighted the benefits of updating TRSs along breeding cycles and optimizing their composition using the CDmean criterion to ensure high prediction accuracy. This chapter was submitted in *Theor. Appl. Genet.*

In **Chapter 3**, to complement experimental results, simulations were carried out to investigate the use of factorial designs into GS hybrid breeding schemes. We compared at a fixed cost a conventional phenotypic breeding scheme with five genomic selection breeding schemes differing by their cycle length and the use of tester and/or factorial designs to calibrate genomic predictions. The simulation results suggested that implementing genomic selection and shortening the breeding cycle by replacing early testcross evaluations by a unique sparse factorial design offers the potential to accelerate the breeding process and increase genetic gain. These findings not only validate the efficiency of sparse factorial TRSs but also highlight the potential of integrating them into hybrid breeding programs at fixed costs. This chapter corresponds to a paper draft that we plan to submit before the PhD defense.

This thesis led to a set of recommendations to improve maize hybrid breeding scheme through the integration of genomic selection and factorial designs.



# { CHAPTER 1 }



## Chapter 1. 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<sup>1</sup>; Cyril Bauland<sup>1</sup>; Tristan Mary-Huard<sup>1,2</sup>; Sophie Pin<sup>1</sup>; Carine Palafre<sup>3</sup>; Colin Guillaume<sup>4</sup>; Christina Lehermeier<sup>5</sup>; Alain Charcosset<sup>1</sup>; Laurence Moreau<sup>1</sup>

Communicated by Matthias Frisch.

<sup>1</sup> Génétique Quantitative et Evolution - Le Moulon, INRAE, CNRS, AgroParisTech, Université Paris-Saclay, 91190 Gif-sur-Yvette, France

<sup>2</sup> MIA, INRAE, AgroParisTech, Université Paris-Saclay, 75005 Paris, France

<sup>3</sup> UE 0394 SMH, INRAE, 2297 Route de l'INRA, 40390 Saint-Martin-de-Hinx, France

<sup>4</sup> Maïsadour Semences SA, 40001 Mont de Marsan Cedex, France

<sup>5</sup> RAGT2n, Genetics and Analytics Unit, 12510 Druelle, France

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

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

This study has been published in the peer-reviewed journal TAG: Theoretical and Applied Genetics. The publisher version of this article is available at:

<https://doi.org/10.1007/s00122-022-04176-y>

Received: 14 January 2022 / Accepted: 6 July 2022



## 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 (1) allocating each of them to a different tester for the tester design, or (2) reducing the number of hybrids per line for the factorial design. Our results confirm the potential of sparse factorial designs for genomic hybrid breeding.

## 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 train 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 training population size (Jannink et al. 2010; Technow et al. 2014; Seye et al. 2020), the relationship between the 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. 2018b; 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. 2018b) 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.

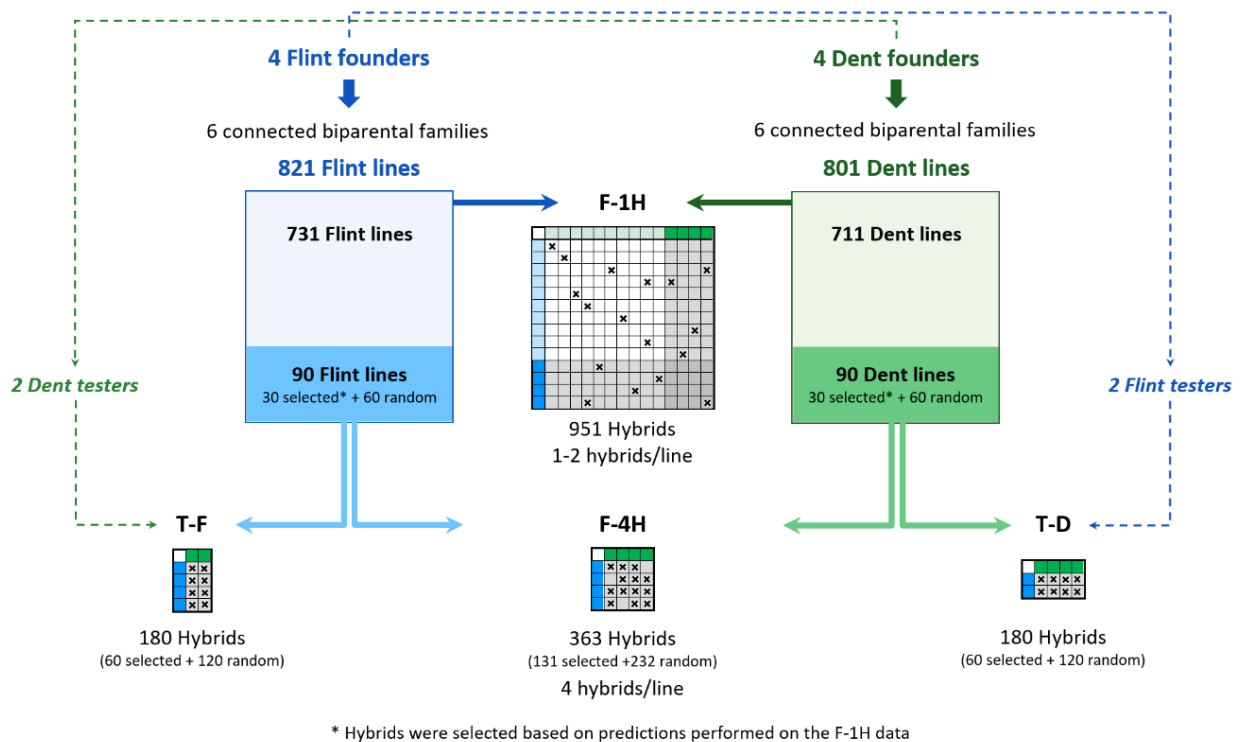
In the present study, our objective is to further evaluate the interest of using a factorial design instead of a tester design as TRS in early stages of the breeding pipeline. Original factorial and tester experimental designs were produced. In the flint and the dent heterotic group, biparental populations were derived from intercrossing four founder lines. From these segregating 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. All these designs were used as TRS or validation set (VS) to evaluate their potential for training GS models for different prediction objectives. The aim of this study was to (1) assess the efficiency of a sparse factorial TRS to predict new hybrid combinations obtained with a tester or another factorial design, (2)

compare the efficiency of factorial versus tester designs of equal size to train GS models, 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 (**Fig. 1**).



**Fig. 1** 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. 2011a). 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  $l$  (each environment corresponding to the combination of a location and a year of experiment), located at row  $x$ , column  $y$  and in block  $z$ .  $\mu$  is the intercept,  $\lambda_l$  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),  $\tau_h$  is the fixed factor with 18 levels corresponding to the control hybrids,  $\rho_{lh}$  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:  $U_k \sim \mathcal{N}(0, \sigma_{GCA_f}^2)$  and respectively  $U'_{k'} \sim \mathcal{N}(0, \sigma_{GCA_d}^2)$ .  $\sigma_{GCA_f}^2$  and  $\sigma_{GCA_d}^2$  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  $S_{kk'} \sim \mathcal{N}(0, \sigma_{SCA}^2)$  iid with  $\sigma_{SCA}^2$  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  $(U\lambda)_{kl} \sim \mathcal{N}(0, \sigma_{GCA_f \times E}^2)$  iid,  $(U'\lambda)_{k'l} \sim \mathcal{N}(0, \sigma_{GCA_d \times E}^2)$  iid and  $(S\lambda)_{kk'l} \sim \mathcal{N}(0, \sigma_{SCA \times E}^2)$  iid.  $\sigma_{GCA_f \times E}^2$ ,  $\sigma_{GCA_d \times E}^2$  and  $\sigma_{SCA \times E}^2$  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:  $E_{hkk'lxyz} \sim \mathcal{N}(0, \sigma_{E_l}^2)$  and are independent and identically distributed within trial and independent between trials,  $\sigma_{E_l}^2$  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:

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

where  $\lambda_l$ ,  $t_h$ ,  $\tau_h$  and  $\rho_{lh}$  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  $x$ , column  $y$  and in block  $z$ .  $\gamma_m$  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  $U_k \sim \mathcal{N}(0, \sigma_{GCA_f}^2)$  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  $S_{km} \sim \mathcal{N}(0, \sigma_{SCA_t}^2)$  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  $(U\lambda)_{kl} \sim \mathcal{N}(0, \sigma_{GCA_f \times E}^2)$  iid and  $(S\lambda)_{kml} \sim \mathcal{N}(0, \sigma_{SCA_t \times E}^2)$  iid.  $E_{hkmxyz}$  is the error term. We assume that the errors follow  $E_{hkmxyz} \sim \mathcal{N}(0, \sigma_{E_l}^2)$  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:

$$H^2 = \frac{\sigma_H^2}{\sigma_H^2 + \frac{\sigma_{H \times E}^2}{n_{site}} + \frac{\sigma_{E_{moy}}^2}{n_{rep} \times n_{site}}},$$

where  $\sigma_H^2$  is the hybrid genetic variance. For the factorial designs it is computed as:  $\sigma_H^2 = \sigma_{GCA_f}^2 + \sigma_{GCA_d}^2 + \sigma_{SCA}^2$  and for the T-F (respectively T-D) it corresponds to the  $\sigma_{GCA_f}^2$  (respectively  $\sigma_{GCA_d}^2$ ).  $\sigma_{H \times E}^2$  is the total genotype by trial variance decomposed as:  $\sigma_{H \times E}^2 = \sigma_{GCA_f \times E}^2 + \sigma_{GCA_d \times E}^2 + \sigma_{SCA \times E}^2$  for the factorial designs and as:  $\sigma_{H \times E}^2 = \sigma_{GCA_f \times E}^2$  for the T-F (respectively  $\sigma_{H \times E}^2 = \sigma_{GCA_d \times E}^2$  for the T-D).  $\sigma_{E_{moy}}^2$  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:

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

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  $l$ .  $\mu$  is the intercept,  $\lambda_l$  is the fixed effect of environment  $l$ ,  $\gamma_h$  is the fixed genetic effect of hybrid  $h$ .  $E_{hrl}$  is the error term of environment  $l$ , with  $E_{hrl} \sim \mathcal{N}(0, \sigma_{El}^2)$  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. Markers with more than 20% of missing values within the dent and flint parental lines, markers with more than 5% (10%) of heterozygosity among the dent (flint) parental lines and markers with Minor Allele Frequency (MAF) inferior to 5% were discarded. After quality control, 9,548 SNP polymorphic markers (in at least the flint or dent

population) were conserved and mapped on a consensus map (Giraud et al. 2017a).

Kinship matrices for the flint and dent GCA ( $\mathbf{K}_{GCA_f}$  and  $\mathbf{K}_{GCA_d}$ ) 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  $\mathbf{K}_{GCA_d}$  was computed similarly. The coefficient of the SCA kinship matrix ( $\mathbf{K}_{SCA}$ ) 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:

$$\mathbf{y} = \mathbf{1}_n \cdot \mu + \mathbf{Z}_d \mathbf{g}_{GCA_d} + \mathbf{Z}_f \mathbf{g}_{GCA_f} + \mathbf{Z}_{df} \mathbf{g}_{SCA_{df}} + \mathbf{E}, \quad (5.1)$$

where  $\mathbf{y}$  is the vector of ls-means of the  $n$  phenotyped hybrids,  $\mathbf{1}_n$  is a vector of  $n$  ones and  $\mu$  is the intercept.  $\mathbf{g}_{GCA_f}$  (respectively  $\mathbf{g}_{GCA_d}$ ) is the vector of random GCA effects of the  $n_f$  flint parental lines (respectively  $n_d$  dent lines), with  $\mathbf{g}_{GCA_f} \sim \mathcal{N}(0, \mathbf{K}_{GCA_f} \sigma_{GCA_f}^2)$  (respectively  $\mathbf{g}_{GCA_d} \sim \mathcal{N}(0, \mathbf{K}_{GCA_d} \sigma_{GCA_d}^2)$ ) where  $\mathbf{K}_{GCA_f}$  (respectively  $\mathbf{K}_{GCA_d}$ ) is the genomic relatedness matrix between the flint lines (respectively dent lines).  $\sigma_{GCA_f}^2$  and  $\sigma_{GCA_d}^2$  are the flint and dent GCA variances.  $\mathbf{g}_{SCA_{df}}$  is the vector of SCA random effects of the  $n$  hybrids, accounting for the interactions between the flint and dent parental lines, with  $\mathbf{g}_{SCA_{df}} \sim \mathcal{N}(0, \mathbf{K}_{SCA_{df}} \sigma_{SCA_{df}}^2)$  where  $\mathbf{K}_{SCA_{df}}$  is the SCA kinship matrix of the hybrids (phenotyped or not) and  $\sigma_{SCA_{df}}^2$  the SCA variance.  $\mathbf{Z}_d$ ,  $\mathbf{Z}_f$  and  $\mathbf{Z}_{df}$  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.  $\mathbf{E}$  is the vector of error terms, with  $\mathbf{E} \sim \mathcal{N}(0, \mathbf{I}\sigma_E^2)$ . 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:

$$\mathbf{y} = \mathbf{1}_n \cdot \mu + \mathbf{X}\mathbf{v} + \mathbf{Z}_f \mathbf{g}_{GCA_f} + \mathbf{Z} \mathbf{g}_{SCAt} + \mathbf{E}, \quad (5.2)$$

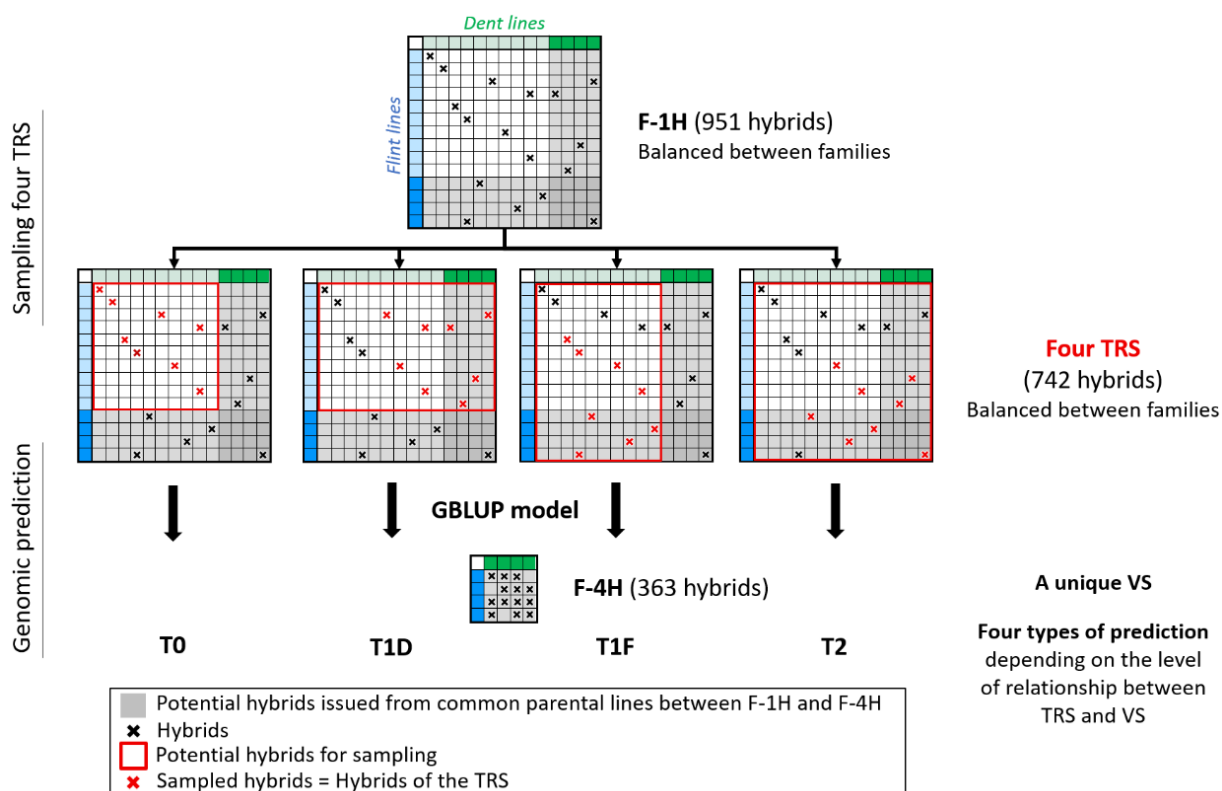
where  $\mathbf{y}$  is the vector of ls-means of the  $n$  phenotyped hybrids,  $\mathbf{1}_n$  is a vector of  $n$  ones and  $\mu$  is the intercept.  $\mathbf{v}$  is the vector of fixed effects of the  $n_t$  testers.  $\mathbf{g}_{GCA_f}$  is the vector of random GCA effects of the  $n_f$  flint parental lines, with  $\mathbf{g}_{GCA_f} \sim \mathcal{N}(0, \mathbf{K}_{GCA_f} \sigma_{GCA_f}^2)$  where  $\mathbf{K}_{GCA_f}$  is the genomic relatedness matrix between the flint lines and  $\sigma_{GCA_f}^2$  is the flint GCA variance.  $\mathbf{g}_{SCAt}$  is the vector of random effects of the interaction between the flint line and the dent tester, with  $\mathbf{g}_{SCAt} \sim \mathcal{N}(0, \mathbf{I}_2 \otimes \mathbf{K}_{GCA_f} \sigma_{SCAt}^2)$  where  $\sigma_{SCAt}^2$  is the SCA variance.  $\mathbf{X}$ ,  $\mathbf{Z}_f$  and  $\mathbf{Z}$  are the corresponding incidence matrices of dimensions  $[n \times n_t]$ ,  $[n \times n_f]$ , and  $[n \times n_t n_f]$  respectively.  $\mathbf{E}$  is the vector of error terms, with  $\mathbf{E} \sim \mathcal{N}(0, \mathbf{I}\sigma_E^2)$ . The different random effects are assumed to be independent.

### Predictive ability with different scenarios

Three objectives were investigated using three different scenarios. In all scenarios 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 SCA effect was included in the model).

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. In scenario 1a, all 951 hybrids of the F-1H, were used as TRS to predict the F-4H, T-F and T-D. 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 scenario 1b, 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. The TRS size of 742 was chosen since it corresponded to the maximum number of hybrids that could be sampled for the T0 prediction.



**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 composed of 742 hybrids were sampled from the F-1H 10 times. 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. The constraint was to sample hybrids within the red box. For the T2 hybrids, an additional constraint was to include at least one hybrid per line that contributed to the VS (F-4H)

In scenario 2, we compared the efficiency of training a GBLUP model with a factorial or a tester design. In scenario 2a, the F-4H (composed of 363 hybrids) or the tester designs (360 hybrids) were used to train respectively models (5.1) or (5.2) to predict the F-1H design (951 hybrids). For each TRS, the predictive ability was computed. In addition, the GCA BLUPs predicted when training 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 each trait. 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. In scenario 2b we 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 2b, TRS were sampled only once.

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 in scenarios 2a and 2b, Williams tests (Williams 1959) were performed (with a risk level  $\alpha=0.05$ ) using the "r.test" function of the *psych* R-package (Revelle 2021) for dependent correlations with a common variable. In scenario 2b, since 21 pairwise tests were performed for each trait, a Bonferroni correction (multiple comparison correction) was applied per trait. 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 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.

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

<sup>a</sup> Variance component defined as in Model (1.1) and (1.2). For the tester designs,  $\sigma_{SCA}^2$  corresponds to  $\sigma_{SCAt}^2$

<sup>b</sup> Percentage of SCA variance computed as  $\frac{\sigma_{SCA}^2}{\sigma_{GCA_d}^2 + \sigma_{GCA_f}^2 + \sigma_{SCA}^2} \times 100$

<sup>c</sup> Minimum residual variance and maximum residual variance across all environments

<sup>d</sup> Broad-sense heritability

<sup>e</sup> Standard error in brackets

<sup>f</sup> Significance of variance components assessed by likelihood ratio test with  $\chi^2$  mixed distributions ( $\alpha=0.05$ ). Non-significant variance indicated by **ns**



### 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.

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

<sup>a</sup> For the tester designs,  $\sigma_{SCA}^2$  corresponds to  $\sigma_{SCAt}^2$

<sup>b</sup> Percentage of SCA variance computed for the factorial designs as  $\frac{\sigma_{SCA}^2}{\sigma_{GCA_d}^2 + \sigma_{GCA_f}^2 + \sigma_{SCA}^2} \times 100$

<sup>c</sup> Standard error in brackets

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

In scenario 1a, 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 (**Table 3**). 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 for F-4H and T-D and DMY for T-F (**Table 3**). Differences between the predictions of random and selected hybrids were greater for DMY, DtSilk, NDF, CELL and HCELL.

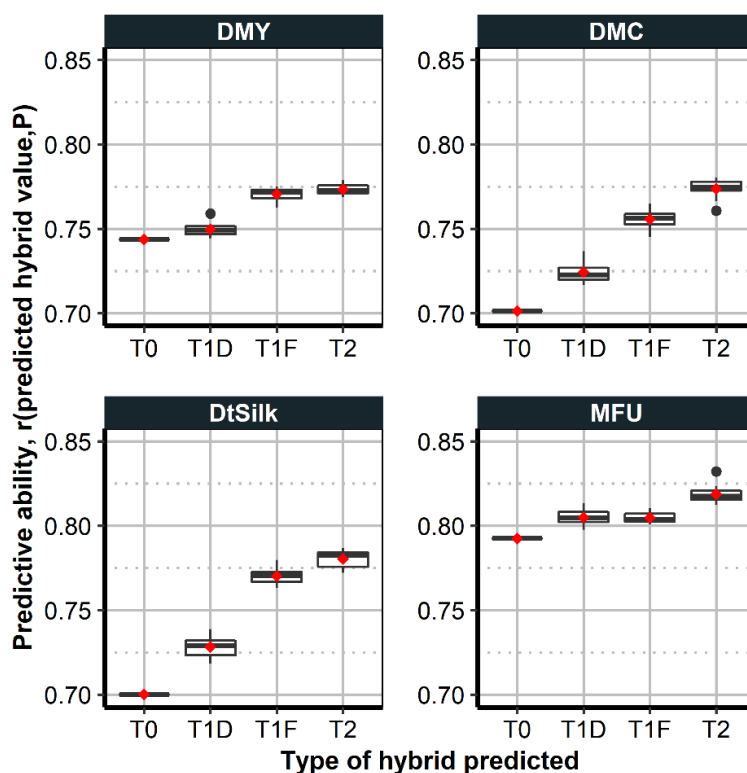
**Table 3** Predictive abilities obtained for the F-4H and for each of the tester designs by training the GBLUP model on the F-1H in scenario 1a.

Validation Set	Predicted hybrid value component(s)	Hybrid type in the VS <sup>a</sup>	DMY	DMC	DtSilk	MFU
<b>F-4H</b>	$GCA_f + GCA_d$	<b>Selected+Random (363)<sup>b</sup></b>	<b>0.78</b>	<b>0.78</b>	<b>0.79</b>	<b>0.82</b>
		Selected (127)	0.48	0.80	0.69	0.73
		Random (236)	0.67	0.78	0.76	0.80
	$GCA_f + GCA_d + SCA$	<b>Selected+Random (363)</b>	<b>0.77</b>	<b>0.78</b>	<b>0.79</b>	<b>0.82</b>
		Selected (127)	0.46	0.80	0.71	0.73
		Random (236)	0.67	0.78	0.77	0.80
<b>T-D</b>	$GCA_d$	<b>Selected+Random (180)</b>	<b>0.80</b>	<b>0.74</b>	<b>0.76</b>	<b>0.79</b>
		Selected (60)	0.66	0.76	0.69	0.73
		Random (120)	0.71	0.73	0.71	0.76
<b>T-F</b>	$GCA_f$	<b>Selected+Random (180)</b>	<b>0.69</b>	<b>0.84</b>	<b>0.84</b>	<b>0.79</b>
		Selected (60)	0.68	0.84	0.78	0.71
		Random (120)	0.63	0.84	0.86	0.79

<sup>a</sup> Predictive ability was computed for all hybrids (Selected+Random) and also for each hybrid type (Selected and Random) of the VS

<sup>b</sup> Number of hybrids in brackets

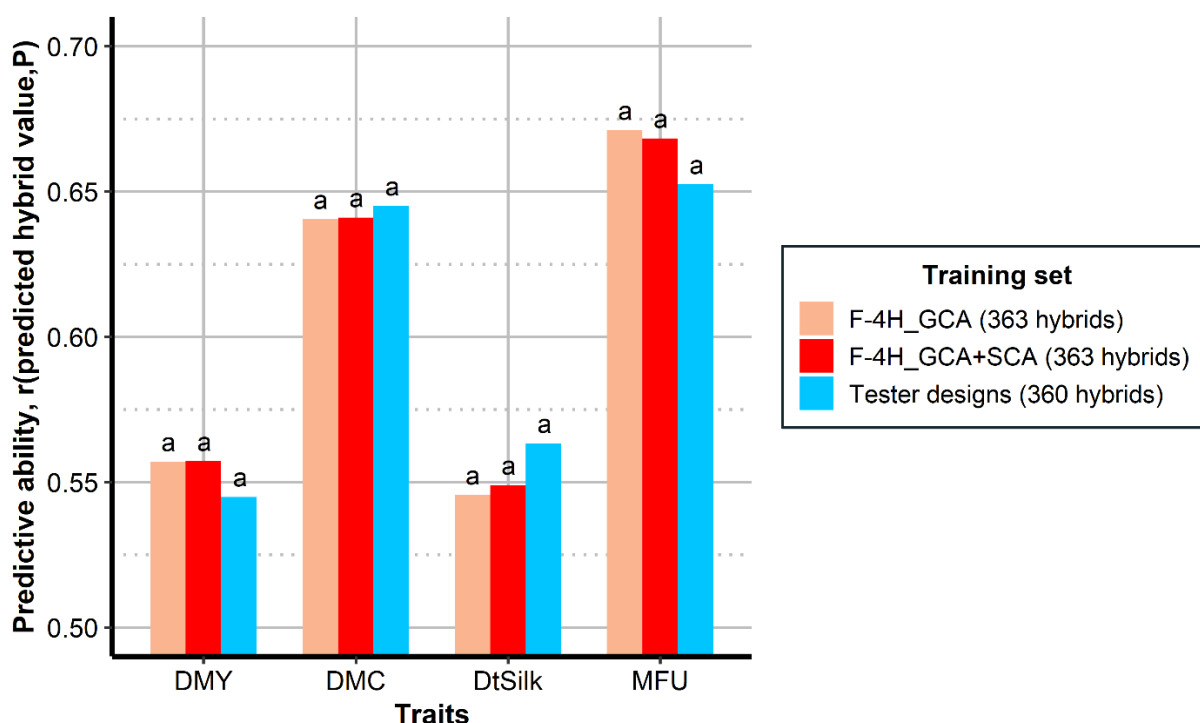
In scenario 1b, the ability of the F-1H to predict T0, T1 or T2 hybrids was estimated by considering different TRS (**Fig.3**). The lowest predictive abilities were obtained when predicting T0 hybrids for the four traits presented (and for seven traits out of 11) and ranged from 0.70 (DMC and DtSilk) 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. It should be noted that for the T0 predictions, sampling was only performed once which explains the zero-variance.



**Fig.3** Predictive abilities obtained for the F-4H (363 hybrids) by training the model on four different TRS composed of 742 hybrids sampled from the F-1H in scenario 1b. 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). For the T1D, T1F and T2 TRS 10 samplings were performed. The red diamond shaped point represents the mean of the predictive ability of the 10 samplings. For the T0 TRS, sampling was performed once (all 742 available hybrids with no common parental lines between the TRS and VS were sampled), this explains the zero variance observed on the figure.

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

In scenario 2a, 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 tended to give 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 not significant according to Williams tests ( $\alpha=0.05$ ). Including SCA effects in the GBLUP model did not improve the quality of prediction.



**Fig.4** Predictive abilities obtained for all hybrids of the F-1H (951 hybrids) by training the model on the F-4H (363 hybrids) (including or not the SCA effect in the model) or on the tester designs (360 hybrids) in scenario 2a. Williams tests were performed ( $\alpha=0.05$ ) and significant differences were indicated with letters: two different letters indicate a significant difference and at least one common letters indicate no significant difference.

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 for a selected rate of 5% was equal to 53% for DMY, 75% for DMC, 56% for DtSilk and 68% for MFU (**Fig.S1**). This illustrated that at this selection rate, the two approaches did not select the same single-cross hybrids.

Interestingly, for DMY, which is the major trait of interest in breeding, for low selection rates (<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.

**Table 4** Correlations between the GCA BLUPs predicted for the F-1H (951 hybrids) using either the F-4H (363 hybrids) or the tester designs (360 hybrids) as TRS in scenario 2a.

BLUPs correlated	DMY	DMC	DtSilk	MFU
GCA <sub>f</sub>	0.85	0.95	0.94	0.93
GCA <sub>d</sub>	0.90	0.88	0.86	0.92

**Table 5** Predictive abilities obtained by training the GS model on the F-4H (363 hybrids) or on the tester designs (360 hybrids) to predict different hybrid types of the F-1H (951 hybrids) in scenario 2a.

TRS	Validation Set		DMY	DMC	DtSilk	MFU
	Hybrid type <sup>a</sup>	Number of hybrids				
<b>F-4H<sup>b</sup></b> (363) <sup>c</sup>	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
	<b>All</b>	<b>951</b>	<b>0.56</b>	<b>0.64</b>	<b>0.55</b>	<b>0.67</b>
<b>Tester designs</b> (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
	<b>All</b>	<b>951</b>	<b>0.54</b>	<b>0.64</b>	<b>0.56</b>	<b>0.65</b>

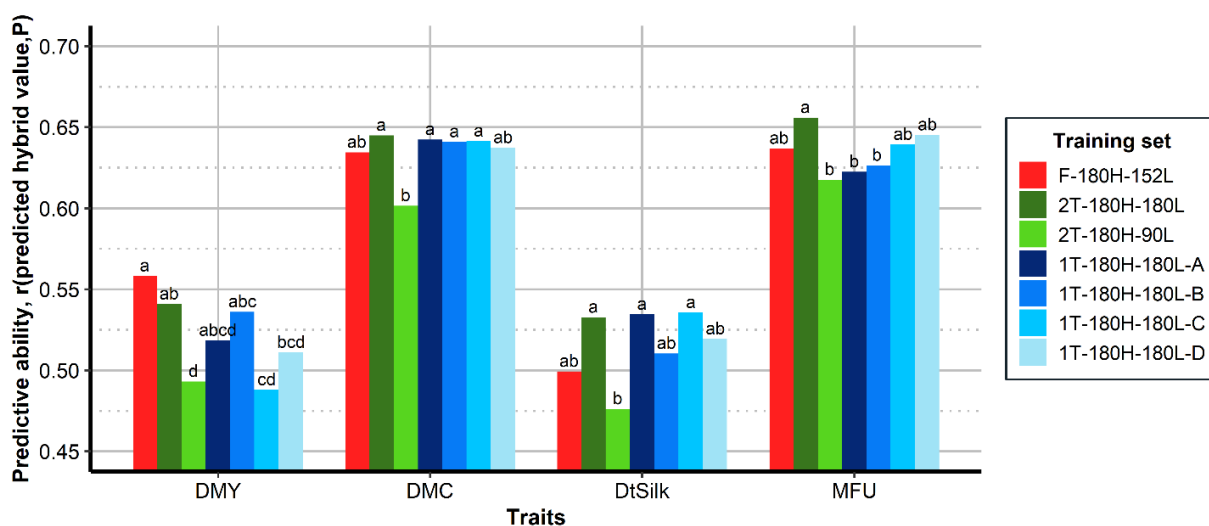
<sup>a</sup> VS hybrids had none of their parental lines in the TRS (T0), only their flint parental line in the TRS (T1F), only their dent parental line in the TRS (T1D)

<sup>b</sup> Model (5.1) including SCA effects

<sup>c</sup> Number of hybrids in the TRS

At the same number of hybrids (180 hybrids), the impact of the composition of the tester designs used as TRS was investigated in scenario 2b. 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-

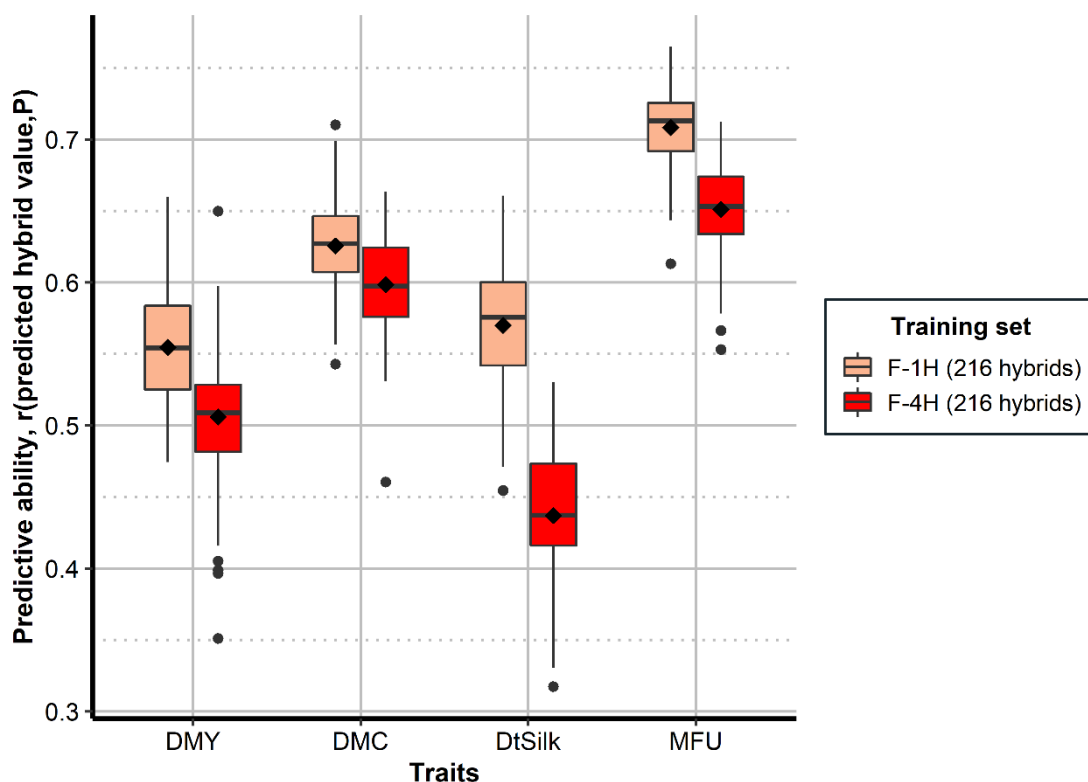
180H-180L-A, 1T-180H-180L-B, 1T-180H-180L-C, 1T-180H-180L-D) ranged from 0.51 (DMY) to 0.64 (DMC). Predictive abilities varied between the four different combinations of testers (A, B, C and D) and the best combination was different depending on the trait. Differences between the lowest and the highest ranged from 0.005 (DMC) to 0.048 (DMY). Qualities of prediction obtained with the 2T-180H-180L were significantly 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 equivalent or significantly higher to the ones obtained with TRS composed of one tester per group (1T-180H-180L-A, B, C, D). At the same number of hybrids, using the F-180H-152L instead of the 2T-180H-90L or the 1T-180H-180L tended to give higher predictive abilities on average over the 11 traits. This advantage was significant for DMY and DMC (between the F-180H-152L and the 2T-180H-90L). Differences were not significant between the F-180H-152L and the 2T-180H-180L (for the 11 traits studied). 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**).



**Fig.5** Predictive abilities obtained when training the model on different tester design compositions or on a factorial design in scenario 2b. Each tester TRS was 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). The factorial design was 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 ( $\alpha=0.05$ ) and was indicated with letters: two different letters indicate a significant difference and at least one common letters indicate no significant difference.

### 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, using more lines and fewer hybrids per line to predict T0 hybrids tended to give better qualities of prediction in our scenario.



**Fig.6** Predictive abilities obtained when training the GS model on 216 hybrids of the F-1H design to predict 216 hybrids of the F-4H design in light orange and when training the model on 216 hybrids of the F-4H design to predict the F-1H in red in scenario 3. For each approach, 100 samplings were performed. The black diamond-shaped point represents the mean of the predictive abilities over the 100 repetitions.

## 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**

One of our objectives was to evaluate the interest of using a sparse factorial design to train GS models in early breeding stages. To this end, we compared different TRS designs derived from unselected segregating biparental families that correspond to the population structure of candidate lines generated by breeders in breeding programs. We showed in scenario 1a that by using a factorial design composed of only one hybrid per line to train GS models achieves good predictive abilities of new hybrids in new environments (new years and locations). We observed good predictive abilities when predicting both factorial and tester designs. There were two main explanations for these high predictive abilities. First, all parental lines of the VS hybrids were also evaluated in the TRS and for the testers designs two founder lines from the complementary group were used as testers. Therefore, the predicted hybrids in the F-4H (VS) were only T2 hybrids. The impact of predicting T2 hybrids compared to T1 or T0 hybrids was investigated in scenario 1b. Results showed that predictive abilities were highest for T2 single-crosses, followed by T1 and T0 single-crosses. This 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). Another explanation is the limited number of founders (four) in each group and the fact that they were chosen to have contrasted performances. This created a strong population structure that was accounted for implicitly in the GS models. A benchmark prediction model considering only the fixed effects of the founder line origin of the hybrids was implemented (**File S2**). Its good predictive ability confirmed that population structure alone could predict part of the hybrid performances (**Fig.S2**). 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). We noted differences in predictive abilities depending on the tester design (dent of flint) used as VS. For most of the traits, a higher predictive ability was observed for the group where the larger GCA

variance component was estimated (**Table 1**).

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 (scenario 1a). 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 (scenario 2a). 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 2b). 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. Our scenario 3 aimed at addressing this question by comparing the efficiency of the F-1H and the F<sub>4</sub>H as TRS. It suggests 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 T<sub>0</sub> hybrids. Since we used reciprocally the F-1H and the F-4H as VS, this might have biased results. Nevertheless, a simulation study by Seye et al. (2020) comparing factorial designs composed of different numbers of hybrid per line when predicting a third independent design is in accordance with our results from scenario 3. Therefore, we hypothesize that a factorial composed of 180 hybrids and 360 lines could outperform all the tester designs in scenario 2b.

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 T<sub>0</sub> hybrids, as also observed by Seye et al. (2020) with simulations. Yet, within that same study they also showed that when predicting T<sub>2</sub> 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 a factorial design composed of more hybrids per line in a second screening would be a promising lead. 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) Prév́ision 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](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, Kaepler 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>
- Fritsche-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-Dié́guez 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/bfpg/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>

- 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>
- Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* 124:743
- Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics* 157:1819–1829
- Molenberghs G, Verbeke G (2007) Likelihood Ratio, Score, and Wald Tests in a Constrained Parameter Space. *The American Statistician* 61:22–27. <https://doi.org/10.1198/000313007X171322>
- Peyrat J, Nozière P, Féraud A, et al (2016) «Prévoir la digestibilité et la valeur énergétique du maïs fourrage: Guide des nouvelles références». ARVALIS - Institut du végétal - INRA
- Powell O, Gaynor RC, Gorjanc G, et al (2020) A Two-Part Strategy using Genomic Selection in Hybrid Crop Breeding Programs. *bioRxiv* 2020.05.24.113258. <https://doi.org/10.1101/2020.05.24.113258>
- Pszczola M, Strabel T, Mulder HA, Calus MPL (2012) Reliability of direct genomic values for animals with different relationships within and to the reference population. *Journal of Dairy Science* 95:389–400. <https://doi.org/10.3168/jds.2011-4338>
- R Core Team (2020) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria
- Reif JC, Gumpert F-M, Fischer S, Melchinger AE (2007) Impact of Interpopulation Divergence on Additive and Dominance Variance in Hybrid Populations. *Genetics* 176:1931. <https://doi.org/10.1534/genetics.107.074146>
- Revelle W (2021) psych: Procedures for Psychological, Psychometric, and Personality Research. Northwestern University, Evanston, Illinois
- Saatchi M, Miraei-Ashtiani SR, Javaremi AN, Moradi-Shahrehabak M (2010) The impact of information quantity and strength of relationship between training set and validation set on accuracy of genomic estimated breeding values. *African Journal of Biotechnology* 9:438–442. <https://doi.org/10.5897/AJB09.1024>
- 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
- 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>

## Chapter 1

- 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>

## Acknowledgments

We thank Lidea, Limagrain Europe, Maïsadour Semences, Corteva, RAGT 2n, KWS 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).

## 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.





## { CHAPTER 2 }



## Chapter 2. Portability of genomic predictions trained on sparse factorial designs across two maize silage breeding cycles

---

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

<sup>1</sup> Université Paris-Saclay, INRAE, CNRS, AgroParisTech, Génétique Quantitative et Evolution - Le Moulon, 91190 Gif-sur-Yvette, France

<sup>2</sup> UE 0394 SMH, INRAE, 2297 Route de l'INRA, 40390, Saint-Martin-de-Hinx, France

<sup>3</sup> Maïsadour Semences SA, F-40001 Mont-de-Marsan Cedex, France

<sup>4</sup> RAGT2n, Genetics and Analytics Unit, 12510 Druelle, France

<sup>5</sup> MIA, INRAE, AgroParisTech, Université Paris-Saclay, 75005, Paris, France

**Key message.** 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.

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

This paper has been submitted in the peer-reviewed journal TAG: Theoretical and Applied Genetics. The preprint version of this article is available at:

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

Review, August 2023.

## 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.

## 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). Rincet 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**.

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

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

<sup>a</sup> G0R hybrids were produced by crossing two random lines from the G0 cycle

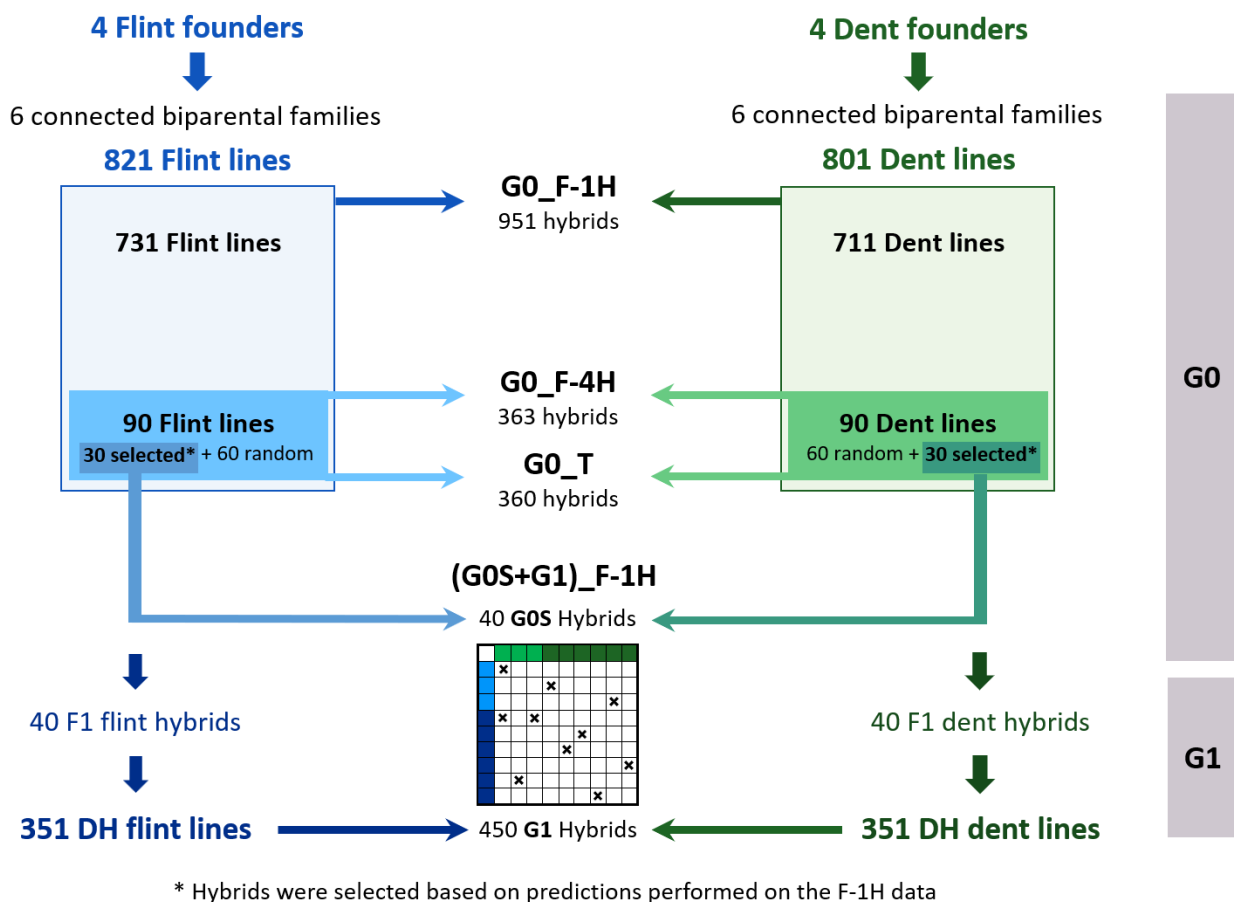
<sup>b</sup> G0S hybrids were produced by crossing two selected lines from the G0 cycle

<sup>c</sup> A reference was indicated for data that was already analyzed in previous studies

### 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 2017 (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.



**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).

### **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. 2011a). 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:

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

where  $Y_{hii'jl}$  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  $l$ .  $\mu$  is the intercept,  $\lambda_l$  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),  $\tau_h$  is the fixed effect of control hybrids with 19 levels (2 for commercial hybrids + 16 for founder hybrids + one for non-control hybrids),  $\rho_{lh}$  is the effect of the interaction between trial  $l$  and control hybrid  $h$ ,  $\pi_j$  is the fixed effect of the generation with two levels (G0S or G1 hybrids).  $H_{h(ii')j}$  is the random genetic effect

## Chapter 2

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:

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

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

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

where  $(U\lambda)_{ilj}$  and  $(U'\lambda)_{i'lj}$  are the random effects of the flint GCA effect by trial interaction, respectively dent GCA by trial interaction of generation  $j$  and  $(S\lambda)_{ii'lj}$  is the random effect of the SCA by trial interaction of generation  $j$ . With  $(U\lambda)_{ilj} \sim \mathcal{N}(0, \sigma_{GCA \times E_f^j}^2)$ ,  $(U'\lambda)_{i'lj} \sim \mathcal{N}(0, \sigma_{GCA \times E_d^j}^2)$  and  $(S\lambda)_{ii'lj} \sim \mathcal{N}(0, \sigma_{SCA \times E_j}^2)$ .  $\sigma_{GCA \times E_f^j}^2$ ,  $\sigma_{GCA \times E_d^j}^2$  and  $\sigma_{SCA \times E_j}^2$  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:  $E_{hii'jl} \sim \mathcal{N}(0, \sigma_{E_l}^2)$  and are iid within trial and independent between trials,  $\sigma_{E_l}^2$  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:

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

where  $\sigma_{H_j}^2$  is the hybrid genetic variance of generation  $j$  computed as  $\sigma_{H_j}^2 = \sigma_{GCA_f^j}^2 + \sigma_{GCA_d^j}^2 + \sigma_{SCA_j}^2$ ,  $\sigma_{H \times E_j}^2$  is the total genotype by trial variance of generation  $j$  decomposed as:

$\sigma_{H \times E_j}^2 = \sigma_{GCA \times E_f^j}^2 + \sigma_{GCA \times E_d^j}^2 + \sigma_{SCA \times E^j}^2$ , and  $\sigma_{E_{moy}}^2$  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.

### Least square-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  $l$ .  $\mu$  is the intercept,  $\lambda_l$  is the fixed effect of trial  $l$ ,  $\gamma_h$  is the fixed genetic effect of hybrid  $h$ .  $E_{hrl}$  is the error term of environment  $l$ , with  $E_{hrl} \sim \mathcal{N}(0, \sigma_{E_l}^2)$  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:

$$\mathbf{y} = \mathbf{1}_n \cdot \mu + \mathbf{Z}\mathbf{g} + \mathbf{E}, \quad (3)$$

where  $\mathbf{y}$  is the vector of ls-means of the  $n$  phenotyped hybrids,  $\mathbf{1}_n$  is a vector of  $n$  ones and  $\mu$  is the intercept.  $\mathbf{g}$  is the vector of random hybrid effects, with  $\mathbf{g} \sim \mathcal{N}(0, \mathbf{K}\sigma_h^2)$  where  $\mathbf{K}$  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).  $\sigma_h^2$  is the hybrid variance.  $\mathbf{Z}$  is the corresponding incidence matrix.  $\mathbf{E}$  is the vector of error

terms, with  $\mathbf{E} \sim \mathcal{N}(0, \mathbf{I}\sigma_E^2)$ . 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:

$$\mathbf{y} = \mathbf{1}_n \cdot \mu + \mathbf{Z}_d \mathbf{g}_{GCA_d} + \mathbf{Z}_f \mathbf{g}_{GCA_f} + \mathbf{Z} \mathbf{g}_{SCA_{df}} + \mathbf{E}, \quad (4.1)$$

where  $\mathbf{y}$  is the vector of ls-means of the  $n$  phenotyped hybrids,  $\mathbf{1}_n$  is a vector of  $n$  ones and  $\mu$  is the intercept.  $\mathbf{g}_{GCA_f}$  (respectively  $\mathbf{g}_{GCA_d}$ ) is the vector of random GCA effects of the  $n_f$  flint parental lines (respectively  $n_d$  dent lines), with  $\mathbf{g}_{GCA_f} \sim \mathcal{N}(0, \mathbf{K}_{GCA_f} \sigma_{GCA_f}^2)$  (respectively  $\mathbf{g}_{GCA_d} \sim \mathcal{N}(0, \mathbf{K}_{GCA_d} \sigma_{GCA_d}^2)$ ) where  $\mathbf{K}_{GCA_f}$  (respectively  $\mathbf{K}_{GCA_d}$ ) 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).  $\sigma_{GCA_f}^2$  and  $\sigma_{GCA_d}^2$  are the flint and dent GCA variances.  $\mathbf{g}_{SCA_{df}}$  is the vector of SCA random effects of the  $n$  hybrids, accounting for the interactions between the flint and dent parental lines, with  $\mathbf{g}_{SCA_{df}} \sim \mathcal{N}(0, \mathbf{K}_{SCA_{df}} \sigma_{SCA_{df}}^2)$  where  $\mathbf{K}_{SCA_{df}}$  is the SCA kinship matrix of the hybrids (phenotyped or not) and  $\sigma_{SCA_{df}}^2$  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).  $\mathbf{Z}_d$ ,  $\mathbf{Z}_f$  and  $\mathbf{Z}$  are the corresponding incidence matrices.  $\mathbf{E}$  is the vector of error terms, with  $\mathbf{E} \sim \mathcal{N}(0, \mathbf{I}\sigma_E^2)$ . The different random effects are assumed to be independent.

The model implemented on the G0\_T-F was:

$$\mathbf{y} = \mathbf{1}_n \cdot \mu + \mathbf{X}\mathbf{v} + \mathbf{Z}_f \mathbf{g}_{GCA_f} + \mathbf{Z} \mathbf{g}_{SCAt} + \mathbf{E}, \quad (4.2)$$

where  $\mathbf{y}$  is the vector of ls-means of the  $n$  phenotyped hybrids,  $\mathbf{1}_n$  is a vector of  $n$  ones and  $\mu$  is the intercept.  $\mathbf{v}$  is the vector of fixed effects of the two dent testers.  $\mathbf{g}_{GCA_f}$  is the vector of random GCA effects of the  $n_f$  flint parental lines, with  $\mathbf{g}_{GCA_f} \sim \mathcal{N}(0, \mathbf{K}_{GCA_f} \sigma_{GCA_f}^2)$  where  $\mathbf{K}_{GCA_f}$  is the genomic relatedness matrix between the flint lines and  $\sigma_{GCA_f}^2$  is the flint GCA variance.  $\mathbf{g}_{SCAt}$  is the vector of random effects of the interaction between the flint line and the dent testers, with  $\mathbf{g}_{SCAt} \sim \mathcal{N}(0, \mathbf{I}_2 \otimes \mathbf{K}_{GCA_f} \sigma_{SCAt}^2)$  where  $\sigma_{SCAt}^2$  is the SCA variance. The kinship matrix was computed for all the flint parental lines following method 1 from VanRaden (2008).  $\mathbf{X}$ ,  $\mathbf{Z}_f$  and  $\mathbf{Z}$  are the corresponding incidence matrices.  $\mathbf{E}$  is the vector of error terms, with  $\mathbf{E} \sim \mathcal{N}(0, \mathbf{I} \sigma_E^2)$ . 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-Diéguéz et al. (2021), where the genetic effect is defined according to gamete origin. The fullest model for the factorial TRS was:

$$\mathbf{y} = \mathbf{1}_n \cdot \mu + \mathbf{Z}_d \mathbf{g}_{A_d} + \mathbf{Z}_f \mathbf{g}_{A_f} + \mathbf{Z} \mathbf{g}_D + \mathbf{Z}_d \mathbf{g}_{AA_d} + \mathbf{Z}_f \mathbf{g}_{AA_f} + \mathbf{Z} \mathbf{g}_{AA_{df}} + \mathbf{E}, \quad (6)$$

where  $\mathbf{y}$  is the vector of ls-means of the  $n$  phenotyped hybrids,  $\mathbf{1}_n$  is a vector of  $n$  ones and  $\mu$  is the intercept.  $\mathbf{g}_{A_f}$  and  $\mathbf{g}_{A_d}$  are the vectors of the random additive effect from the flint and dent parental lines with  $\mathbf{g}_{A_f} \sim \mathcal{N}(0, \mathbf{K}_{A_f} \sigma_{A_f}^2)$  and  $\mathbf{g}_{A_d} \sim \mathcal{N}(0, \mathbf{K}_{A_d} \sigma_{A_d}^2)$ , respectively.  $\mathbf{g}_D$  is the vector of random dominance effect with  $\mathbf{g}_D \sim \mathcal{N}(0, \mathbf{K}_D \sigma_D^2)$ ,  $\mathbf{g}_{AA_f}$  is the vector of random additive-by-additive epistatic effect within the flint (resp. dent) population with  $\mathbf{g}_{AA_f} \sim \mathcal{N}(0, \mathbf{K}_{AA_f} \sigma_{AA_f}^2)$  (resp.  $\mathbf{g}_{AA_d} \sim \mathcal{N}(0, \mathbf{K}_{AA_d} \sigma_{AA_d}^2)$ ) and  $\mathbf{g}_{AA_{df}}$  is the vector of random additive-by-additive epistatic effect across the flint and dent populations  $\mathbf{g}_{AA_{df}} \sim \mathcal{N}(0, \mathbf{K}_{AA_{df}} \sigma_{AA_{df}}^2)$ .  $\mathbf{K}_{A_f}$ ,  $\mathbf{K}_{A_d}$ ,  $\mathbf{K}_D$ ,  $\mathbf{K}_{AA_f}$ ,  $\mathbf{K}_{AA_d}$  and  $\mathbf{K}_{AA_{df}}$  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-Diéguéz et al. 2021.  $\sigma_{A_f}^2$ ,  $\sigma_{A_d}^2$ ,  $\sigma_D^2$ ,  $\sigma_{AA_f}^2$ ,  $\sigma_{AA_d}^2$  and  $\sigma_{AA_{df}}^2$  are the corresponding



variances.  $\mathbf{Z}_f$ ,  $\mathbf{Z}_d$  and  $\mathbf{Z}$  are the incidence matrices.  $\mathbf{E}$  is the vector of error terms, with  $\mathbf{E} \sim \mathcal{N}(0, \mathbf{I}\sigma_E^2)$ . 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:

$$\mathbf{y} = \mathbf{1}_n \cdot \mu + \mathbf{Z}\mathbf{g}_A + \mathbf{Z}\mathbf{g}_D + \mathbf{Z}\mathbf{g}_{AA} + \mathbf{E}, \quad (5)$$

where  $\mathbf{y}$  is the vector of ls-means of the  $n$  phenotyped hybrids,  $\mathbf{1}_n$  is a vector of  $n$  ones and  $\mu$  is the intercept.  $\mathbf{g}_A$  is the vector of the random additive effect with  $\mathbf{g}_A \sim \mathcal{N}(0, \mathbf{K}_A \sigma_A^2)$ ,  $\mathbf{g}_D$  is the vector of random dominance effect with  $\mathbf{g}_D \sim \mathcal{N}(0, \mathbf{K}_D \sigma_D^2)$  and  $\mathbf{g}_{AA}$  is the vector of random additive-by-additive epistasis effect with  $\mathbf{g}_{AA} \sim \mathcal{N}(0, \mathbf{K}_{AA} \sigma_{AA}^2)$ .  $\mathbf{K}_A$ ,  $\mathbf{K}_D$  and  $\mathbf{K}_{AA}$  are respectively the additive, dominance and additive-by-additive epistasis genomic relatedness matrices computed following Vitezica et al. 2017.  $\sigma_A^2$ ,  $\sigma_D^2$  and  $\sigma_{AA}^2$  are the corresponding variances and  $\mathbf{Z}$  is the incidence matrix.  $\mathbf{E}$  is the vector of error terms, with  $\mathbf{E} \sim \mathcal{N}(0, \mathbf{I}\sigma_E^2)$ . 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 (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.

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

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

<sup>a</sup> 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.

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  $\lambda$  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  $\alpha=0.05$ , and a Bonferroni correction (multiple comparison correction) was applied per trait. In Scenario 2b, Williams tests (Williams 1959) were performed with a risk level  $\alpha=0.05$  using the "r.test" function of the psych R-package (Revelle 2021). In Scenario 2b', t-tests with a risk level  $\alpha=0.05$  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 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.  $\sigma_{GCA_f}^2$  was always larger than  $\sigma_{GCA_d}^2$  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,  $\sigma_{SCA}^2$  was larger than  $\sigma_{SCA \times E}^2$  for all traits. For G1 hybrids,  $\sigma_{SCA \times E}^2$  was larger than  $\sigma_{SCA}^2$ , except for DtSilk.

**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 (MFUx10 <sup>2</sup> /kg)	
	G0S	G1	G0S	G1	G0S	G1	G0S	G1
$\sigma_{GCA_f}^2$	0.50(0.24) <sup>d</sup>	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)
$\sigma_{GCA_d}^2$	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)
$\sigma_{SCA}^2$	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)
$\sigma_{GCA_f \times E}^2$	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)
$\sigma_{GCA_d \times E}^2$	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)
$\sigma_{SCA \times E}^2$	0.00	0.23(0.07)	0.00	0.00	0.00	0.24(0.11)	0.00	0.11(0.09)
$\sigma_E^2$ <sup>a</sup>	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 <sup>b</sup>	19	0	0	10	1	1	30	0
$H^2$ <sup>c</sup>	0.87	0.80	0.90	0.92	0.93	0.94	0.56	0.62

<sup>a</sup> Minimum and maximum residual variance across all environments

<sup>b</sup> Percentage of SCA variance computed as  $\frac{\sigma_{SCA}^2}{\sigma_{GCA_d}^2 + \sigma_{GCA_f}^2 + \sigma_{SCA}^2} \times 100$

<sup>c</sup> Broad-sense heritability

<sup>d</sup> 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.

	<b>Hybrid type</b>	<b>Component</b>	<b>DMY</b> (t/ha)	<b>DMC</b> (%)	<b>DtSilk</b> (days)	<b>MFU</b> (MFUx10 <sup>2</sup> /kg)
<b>Ls-means</b>	<b>Commercial</b>	<b>Mean</b>	<b>17.96</b> (17.18-18.74) <sup>b</sup>	<b>34.69</b> (34.37-35.00)	<b>201.89</b> (200.27-203.50)	<b>95.54</b> (95.34-95.74)
		<b>Sd<sup>c</sup></b>	1.10	0.45	2.28	0.28
	<b>Founder</b>	<b>Mean</b>	<b>15.80</b> (14.27-17.47)	<b>34.10</b> (30.53-37.48)	<b>203.14</b> (199.80-206.16)	<b>95.28</b> (91.13-98.14)
		<b>Sd</b>	0.84	1.94	1.50	2.05
	<b>G0S</b>	<b>Mean</b>	<b>17.35</b> (14.82-18.89)	<b>33.33</b> (30.54-37.13)	<b>204.97</b> (201.80-208.79)	<b>93.17</b> (89.04-97.06)
		<b>Sd</b>	0.92	1.40	1.82	1.56
	<b>G1</b>	<b>Mean</b>	<b>17.33</b> (14.33-19.72)	<b>33.43</b> (29.39-39.18)	<b>205.04</b> (197.98-211.40)	<b>93.13</b> (87.59-101.36)
		<b>Sd</b>	0.85	1.63	1.97	1.71
<b>Genetic gain<sup>a</sup></b>	<b>G0S</b>		1.55	-0.77	1.83	-2.11
	<b>G1</b>		1.52	-0.67	1.90	-2.15
<b>Predicted genetic gain<sup>d</sup></b>	<b>G0S</b>		1.45	-0.33	1.18	-1.35
	<b>G1</b>		1.41	-0.28	0.91	-1.14

<sup>a</sup> Genetic gain computed as the difference between the mean performance of the experimental hybrids and the founder hybrids

<sup>b</sup> Minimum and maximum mean performance in brackets

<sup>c</sup> Standard deviation of the ls-means of the experimental hybrid performances

<sup>d</sup> Predicted genetic gain based on genomic predictions trained on the G0\_F-1H

### **Ls-means and genetic gain**

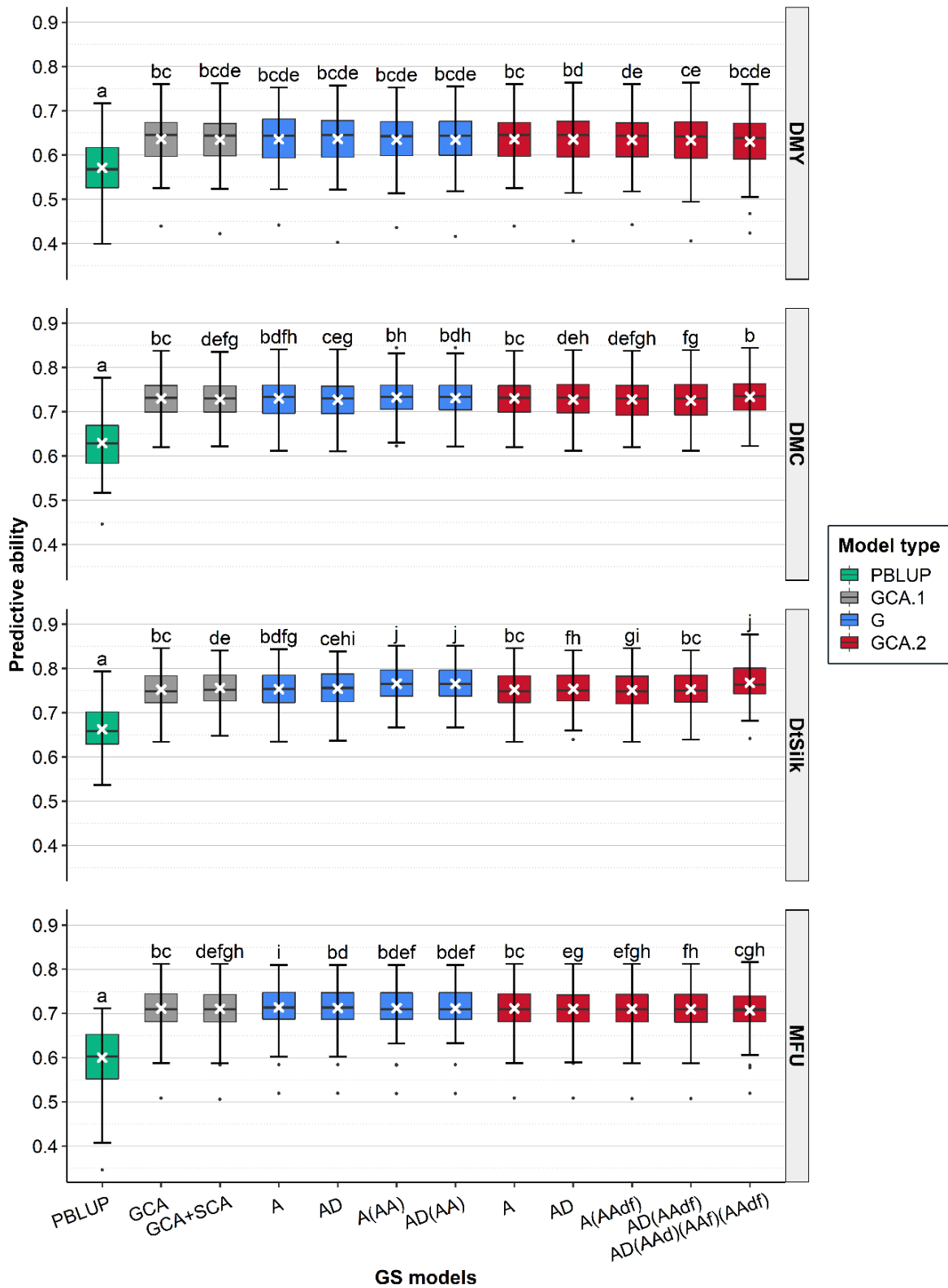
On average, G0S and G1 hybrids performed similarly (**Table 4**). Compared to the 16 founder hybrids, which are representative of the performance of the unselected G0 hybrids (G0R), G0S and G1 hybrids showed a gain in performance for DMY (+1.55 t/ha for G0S and +1.52 t/ha for G1). This gain was associated 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 MFUx10<sup>2</sup>/kg for G0S and -2.15 MFUx10<sup>2</sup>/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**).

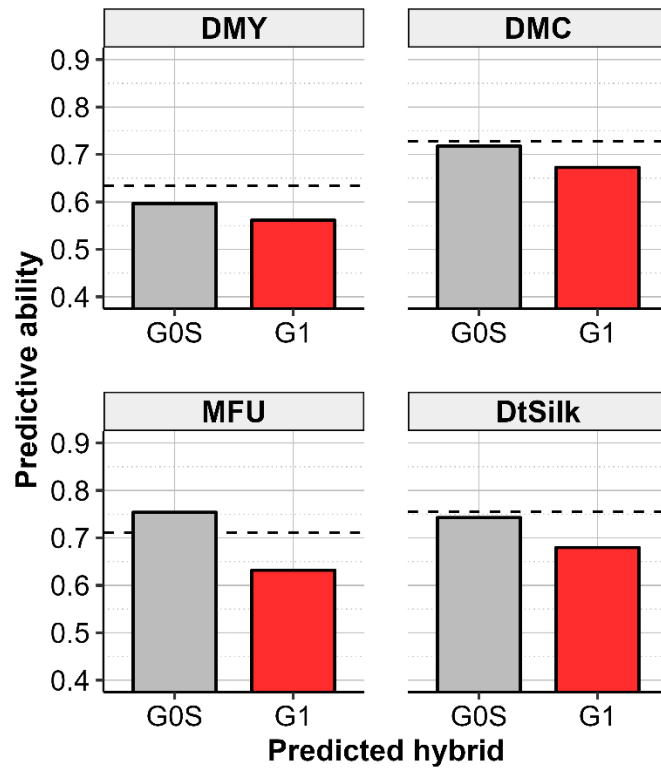


**Fig.2** Predictive abilities obtained by cross-validations within the 442 G1 hybrids using different prediction models (PBLUP, GCA.1, GCA.2 or G models) in Scenario 1. The mean predictive ability over the 100 replicates is represented by a white cross. Significant differences (as obtained by paired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

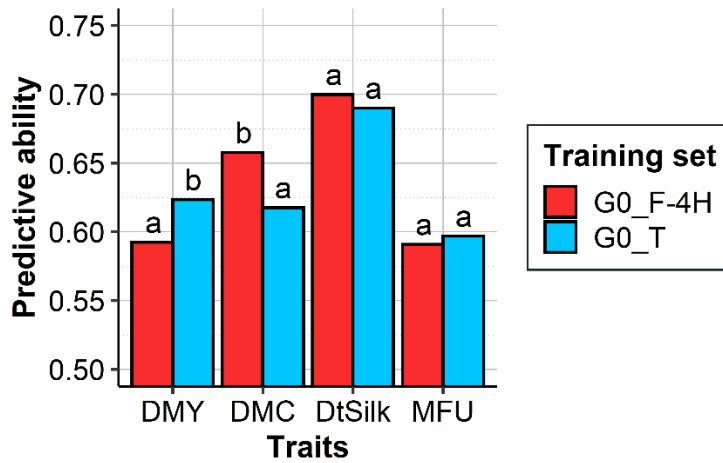


In Scenario 2b, we compared predictive abilities obtained using either the G0\_F-4H (363 hybrids) or the G0 tester designs (360 hybrids) as TRS to predict all G1 hybrids (442 hybrids) (**Fig.4**). They ranged from 0.59 (DMY and MFU) to 0.70 (DtSilk) when training on the G0\_F-4H and from 0.60 (MFU) to 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.

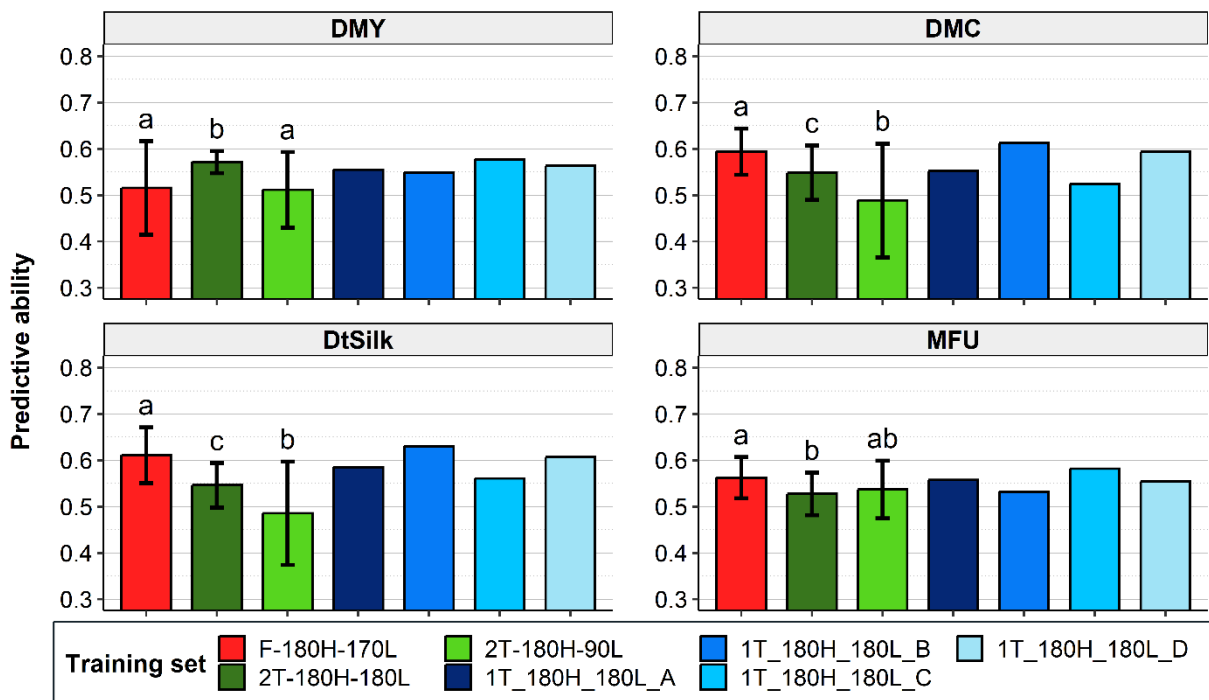
In Scenario 2b', we investigated the efficiency of different G0 tester design compositions to predict G1 hybrids (442 hybrids) at the same number of hybrids (180) and compared them with a factorial design of same size (**Fig.5**). Predictive abilities varied between the four one-tester TRS ranging from 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**).



**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.



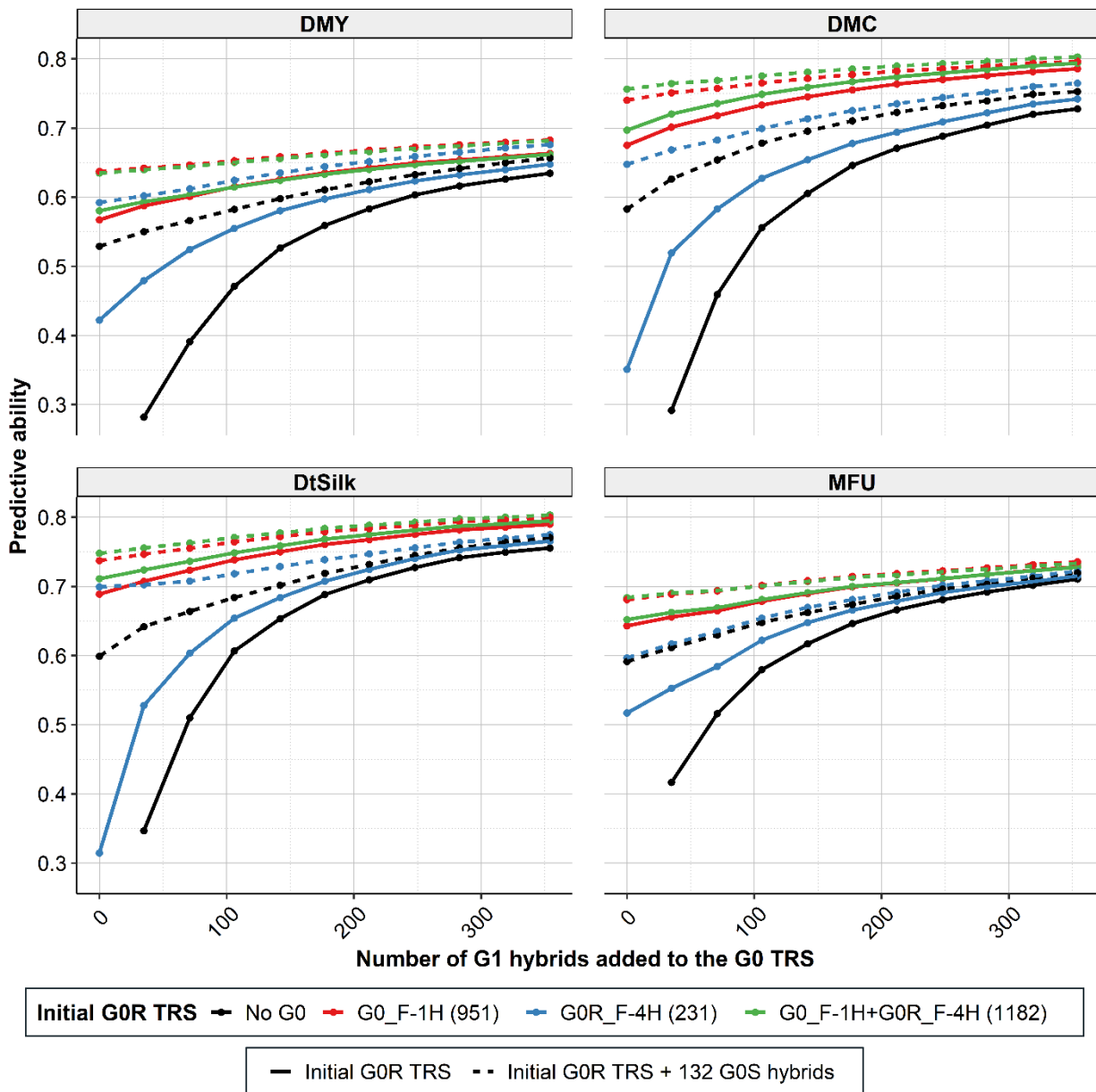
**Fig.4** Predictive abilities obtained for the G1 hybrids (442) by training the GS model on the G0\_F-4H (363) or the G0\_T (360) TRSs in Scenario 2b. Williams tests were performed ( $\alpha=0.05$ ) and significant differences were indicated with letters: two different letters indicate a significant 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 ( $\alpha=0.05$ ) 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.

### 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.



**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 dotted 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.

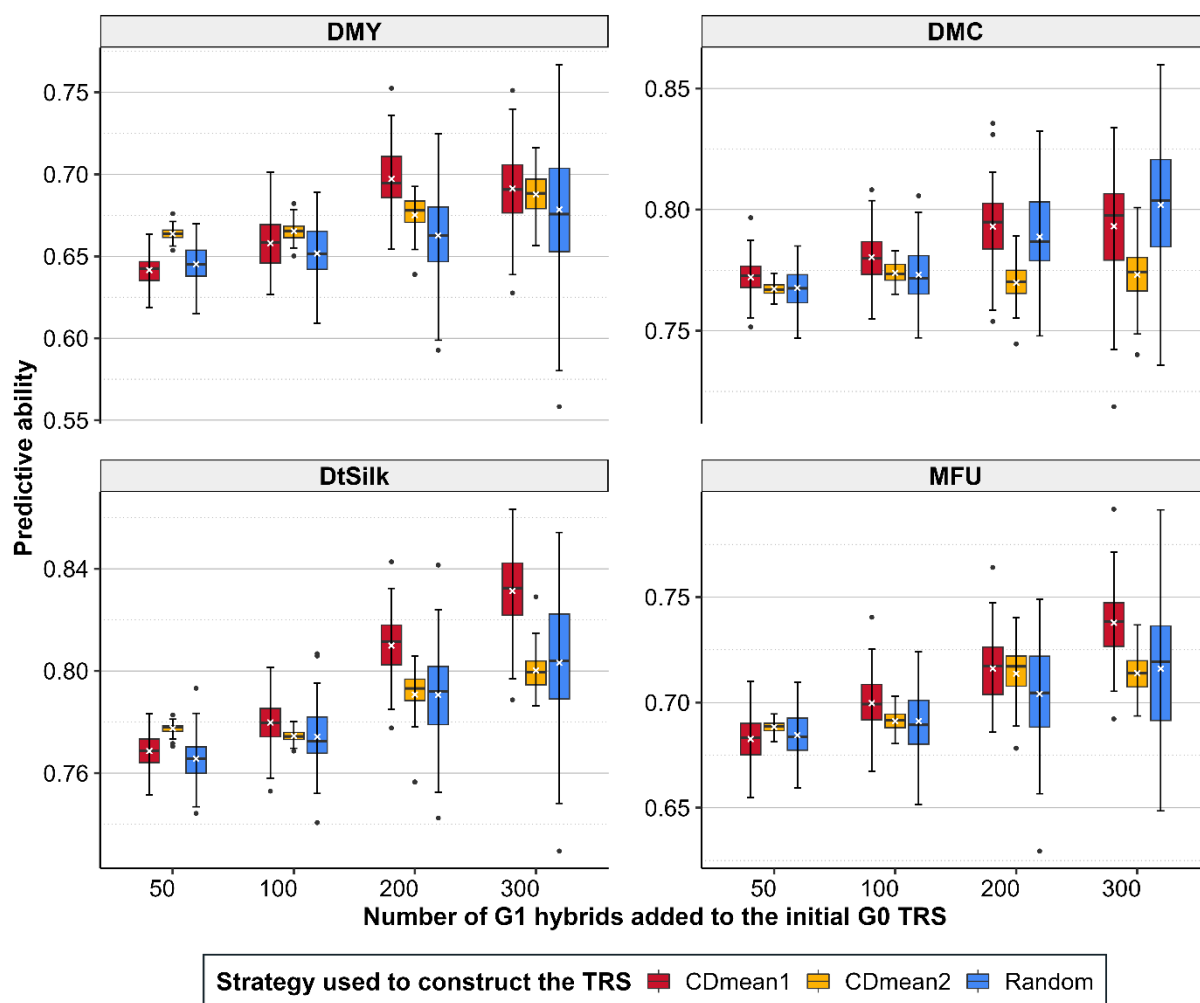
Adding 132 G0S hybrids to the initial G0R TRSs (G0\_F-1H, G0R\_F-4H, or G0\_F-1H+G0R\_F-4H) increased predictive abilities (with a gain on average of 0.10 for DMY, 0.14 for DMC, 0.16 for DtSilk and 0.05 for MFU). The largest gain in predictive ability was observed for the G0R\_F-4H TRS, which was also 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).



**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.

## 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-Diéguéz et al. 2021; Lorenzi et al. 2022). Note that the new SCA kinship formula proposed by González-Diéguéz 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-Diéguéz 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 (Rincen 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 and potentially hybrids from the new generation. When incorporating hybrids from the new generation, choosing them based on a criterion such as the CDmean was beneficial.

## References

- Akdemir D, Sanchez JI, Jannink J-L (2015) Optimization of genomic selection training populations with a genetic algorithm. *Genet Sel Evol* 47:38. <https://doi.org/10.1186/s12711-015-0116-6>
- Alves FC, Granato ÍSC, Galli G, et al (2019) Bayesian analysis and prediction of hybrid performance. *Plant Methods* 15:14. <https://doi.org/10.1186/s13007-019-0388-x>
- Amadeu RR, Cellon C, Olmstead JW, et al (2016) AGHmatrix: R Package to Construct Relationship Matrices for Autotetraploid and Diploid Species: A Blueberry Example. *The Plant Genome* 9:plant-genome2016.01.0009. <https://doi.org/10.3835/plantgenome2016.01.0009>
- Andrieu J (1995) Préviation 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>
- Auinger H-J, Schönleben M, Lehermeier C, et al (2016) Model training across multiple breeding cycles significantly improves genomic prediction accuracy in rye (*Secale cereale* L.). *Theor Appl Genet* 129:2043–2053. <https://doi.org/10.1007/s00122-016-2756-5>
- Barrière Y, Emile JC (2000) Le maïs fourrage. III - Evaluation et perspectives de progrès génétiques sur les caractères de valeur alimentaire
- Ben-Sadoun S, Rincenc R, Auzanneau J, et al (2020) Economical optimization of a breeding scheme by selective phenotyping of the calibration set in a multi-trait context: application to bread making quality. *Theor Appl Genet* 133:2197–2212. <https://doi.org/10.1007/s00122-020-03590-4>
- 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>
- Brandariz SP, Bernardo R (2018) Maintaining the Accuracy of Genomewide Predictions when Selection Has Occurred in the Training Population. *Crop Science* 58:1226–1231. <https://doi.org/10.2135/cropsci2017.11.0682>
- 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>
- Coombes NE (2009) DiGGER, a spatial design program. *Biometric Bulletin* NSW Department of Primary Industries, Orange, NSW
- Denis M, Bouvet J-M (2013) Efficiency of genomic selection with models including dominance effect in the context of Eucalyptus breeding. *Tree Genetics & Genomes* 9:37–51. <https://doi.org/10.1007/s11295-012-0528-1>
- Fernández-González J, Akdemir D, Isidro y Sánchez J (2023) A comparison of methods for training population optimization in genomic selection. *Theor Appl Genet* 136:30. <https://doi.org/10.1007/s00122-023-04265-6>
- Fristche-Neto R, Akdemir D, Jannink J-L (2018) Accuracy of genomic selection to predict maize single-crosses obtained through different mating designs. *Theor Appl Genet* 131:1153–1162. <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-Diéguéz 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 Prediction Genomic 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>
- Laporte F, Charcosset A, Mary-Huard T (2022) Efficient ReML inference in variance component mixed models using a Min-Max algorithm. *PLOS Computational Biology* 18:e1009659. <https://doi.org/10.1371/journal.pcbi.1009659>
- Laporte F, Mary-Huard T (2020) MM4LMM: Inference of Linear Mixed Models Through MM Algorithm
- Legarra A, Gonzalez-Dieguéz DO, Charcosset A, Vitezica ZG (2023) Impact of interpopulation distance on dominance variance and average heterosis in hybrid populations within species. *Genetics* 224:iyad059. <https://doi.org/10.1093/genetics/iyad059>
- Lorenzi A, Bauland C, Mary-Huard T, et al (2022) 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. *Theor Appl Genet*. <https://doi.org/10.1007/s00122-022-04176-y>

## Chapter 2

- Maenhout S, De Baets B, Haesaert G (2010) Prediction of maize single-cross hybrid performance: support vector machine regression versus best linear prediction. *Theoretical and applied genetics* 120:415–427
- Mangin B, Rincint R, Rabier C-E, et al (2019) Training set optimization of genomic prediction by means of EthAcc. *PLoS ONE* 14:e0205629. <https://doi.org/10.1371/journal.pone.0205629>
- Merrick LF, Carter AH (2021) Comparison of genomic selection models for exploring predictive ability of complex traits in breeding programs. *The Plant Genome* 14:e20158. <https://doi.org/10.1002/tpg2.20158>
- Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics* 157:1819–1829
- Michel S, Ametz C, Gungor H, et al (2016) Genomic selection across multiple breeding cycles in applied bread wheat breeding. *Theor Appl Genet* 129:1179–1189. <https://doi.org/10.1007/s00122-016-2694-2>
- Momen M, Morota G (2018) Quantifying genomic connectedness and prediction accuracy from additive and non-additive gene actions. *Genetics Selection Evolution* 50:45. <https://doi.org/10.1186/s12711-018-0415-9>
- Mrode RA, Thompson R (2005) *Linear models for the prediction of animal breeding values*, 2nd ed. CABI Pub, Wallingford, UK; Cambridge, MA
- Neyhart JL, Tiede T, Lorenz AJ, Smith KP (2017) Evaluating Methods of Updating Training Data in Long-Term Genomewide Selection. *G3 Genes|Genomes|Genetics* 7:1499–1510. <https://doi.org/10.1534/g3.117.040550>
- Pszczola M, Calus MPL (2016) Updating the reference population to achieve constant genomic prediction reliability across generations. *animal* 10:1018–1024. <https://doi.org/10.1017/S1751731115002785>
- Pszczola M, Strabel T, Mulder HA, Calus MPL (2012) Reliability of direct genomic values for animals with different relationships within and to the reference population. *Journal of Dairy Science* 95:389–400. <https://doi.org/10.3168/jds.2011-4338>
- R Core Team (2020) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria
- Reif JC, Gumpert F-M, Fischer S, Melchinger AE (2007) Impact of Interpopulation Divergence on Additive and Dominance Variance in Hybrid Populations. *Genetics* 176:1931. <https://doi.org/10.1534/genetics.107.074146>
- Revelle W (2021) *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University, Evanston, Illinois
- Rincint R, Charcosset A, Moreau L (2017) Predicting genomic selection efficiency to optimize calibration set and to assess prediction accuracy in highly structured populations. *Theor Appl Genet* 130:2231–2247. <https://doi.org/10.1007/s00122-017-2956-7>
- Rincint 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>



## 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).

## 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.

## { CHAPTER 3 }



## Chapter 3. Shortening maize hybrid breeding schemes using genomic predictions calibrated on factorial instead of tester designs improves cost efficiency over cycles.

---

Alizarine Lorenzi<sup>1,2</sup>, Gaëtan Touzy<sup>2</sup>, Cyril Bauland<sup>1</sup>, Colin Guillaume<sup>3</sup>, Tristan Mary-Huard<sup>1,4</sup>, Alain Charcosset<sup>1</sup>, Laurence Moreau<sup>1</sup>

<sup>1</sup> Université Paris-Saclay, INRAE, CNRS, AgroParisTech, Génétique Quantitative et Evolution - Le Moulon, 91190 Gif-sur-Yvette, France

<sup>2</sup> RAGT2n, Genetics and Analytics Unit, 12510 Druelle, France

<sup>3</sup> Maïsadour Semences SA, F-40001 Mont-de-Marsan Cedex, France

<sup>4</sup> MIA, INRAE, AgroParisTech, Université Paris-Saclay, 75005, Paris, France

**Key message** Genomic predictions calibrated on a factorial design hold great potential for revisiting maize hybrid breeding schemes.

**Key words** Hybrid breeding scheme, Genomic selection, Factorial design, Simulations

## Abstract

Genomic selection has the potential to increase the rate of genetic gain by acting on three parameters of the breeder's equation: reducing the generation interval, increasing the selection intensity and the selection accuracy. In conventional maize hybrid breeding schemes, candidate inbred lines are selected 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. Recent studies proposed to replace tester-based evaluations with genomic predictions calibrated on a factorial design to predict single-cross hybrids at an early stage and have shown promising predictive abilities to improve hybrid breeding schemes. This study aims to further investigate the use of genomic selection trained on factorial designs in hybrid breeding programs. We simulated a reciprocal recurrent scheme involving multiparental connected populations from two complementary heterotic groups mimicking a maize breeding program. We compared a conventional phenotypic selection breeding scheme at fixed costs against five other schemes based on genomic predictions combined or not with (i) shortening the breeding cycle time by removing phenotypic evaluation steps and (ii) using factorial instead of tester designs to train genomic selection models. The simulations were carried out considering two scenarios, with or without specific combining ability (SCA) effects. Results suggest that implementing GS and shortening the breeding cycle by replacing the two early-stage topcross evaluations with a unique sparse factorial design hold great potential to fasten the breeding process while increasing the genetic gain under both SCA scenarios.

## Introduction

Modern maize breeding emerged in the 20<sup>th</sup> century, replacing open-pollinated populations with hybrids. Progressively, maize genetic diversity has been structured into complementary groups, and current varieties are mainly single-cross hybrids between two inbred lines issued from different heterotic groups, which maximizes the expression of heterosis at the hybrid level (Shull 1908). Hybrid selection involves developing inbred lines by self-fertilization or doubled haploidization and then crossing them to produce and identify the best-performing single-cross hybrids. In conventional breeding programs, hundreds of candidate lines can be produced yearly in each group, making it impossible to evaluate all possible single-cross hybrid combinations. To circumvent this issue, the selection process is divided into two steps. In the first step, candidate inbred lines from each group are crossed with one or a few “testers”, usually inbred lines, from the complementary group to produce topcross hybrids. Based on these topcross hybrids, an early-stage selection is done to reduce the number of candidate lines. In the second step, selected candidate lines from the first step are crossed between the two groups, and the resulting single-cross hybrid combinations are evaluated to select the future hybrid varieties and parental lines for initiating the next cycle. The early stage selection based on a few testers can bias the estimation of the capacity of a line to generate good hybrids, referred to as its General Combining Ability (GCA), by confounding the line GCA with its Specific Combining Ability (SCA) with the testers, masking part of the GCA variance among candidate lines (Hallauer et al. 2010). Therefore, this two-stage selection process does not guarantee the identification of inbred lines that will produce the best hybrid combinations. This breeding process is also time-consuming and requires extensive phenotyping in both groups since each topcross hybrid only provides information on one candidate line from one group.

Genomic selection (GS) is now a routine practice in many breeding programs for major crop species such as maize. It uses markers across the whole genome to predict the genetic value of complex traits (Meuwissen et al. 2001). It relies on a prediction model trained on a set of genotyped and phenotyped individuals (training set, TRS) that is then used to predict the genetic values of a set of selection candidates (prediction set, PS) based on their marker genotypes. Several prediction methods have been proposed, but the most robust and common is the genomic best linear unbiased prediction (GBLUP). Genomic predictions can be used to restructure existing breeding programs to: (i) shorten the breeding cycle by suppressing

phenotypic evaluation steps, (ii) increase selection intensity by increasing the number of candidate lines and (iii) increase selection accuracy of the candidate lines for traits difficult or costly to phenotype (Heslot et al. 2015; Crossa et al. 2017). By acting upon three parameters of the breeder's equation: reducing the generation interval and increasing the selection intensity and accuracy, GS can increase the genetic gain, as reported in several simulation and experimental studies reviewed for instance by Krishnappa et al. (2021). However, depending on the species, the organization of the breeding programs and the constraints on breeding may be very different. As a consequence, the benefit of using GS is species-specific (R2D2 Consortium et al. 2021).

Bernardo and Yu (2007) were the first to report the successful use of GS in a maize breeding scheme using simulation results. They focused on using GS in the early selection stage, considering only one heterotic group and topcross evaluations. Following their work, numerous simulation and field studies have shown the great potential of GS in hybrid breeding (Technow et al. 2012, 2014; Massman et al. 2013; Xu et al. 2014; Kadam et al. 2016; Kadam and Lorenz 2018; Seye 2019; Lorenzi et al. 2022). Some of them have evaluated the interest of GS at the early stage, i.e., the selection of the candidate lines based on topcross hybrids (Lorenzana and Bernardo 2009; Albrecht et al. 2011; Lehermeier et al. 2014; Krchov and Bernardo 2015; Seye et al. 2020; Burdo et al. 2021; Lorenzi et al. 2022). Other studies addressed GS for hybrid value prediction in the second stage of the hybrid breeding scheme, i.e., by using as TRS single-cross hybrids between lines that have already undergone a selection based on their testcross to predict the value of unobserved single-cross hybrids (Maenhout et al. 2010; Technow et al. 2012, 2014; Massman et al. 2013).

Whatever the species considered, implementing GS may impact various aspects of a breeding program. This includes reallocating phenotyping resources, changing selection pressure at different stages, and modifying the duration of the breeding cycle. With a constrained budget, inherent tradeoffs emerge, such as deciding how to reallocate resources between phenotyping and genotyping, as well as determining the balance between the number of candidate lines and the accuracy of their evaluation. Simulation studies published so far for evaluating the relative interest of GS rarely addressed resource allocation at fixed costs. In wheat, two simulation studies investigated the benefit of implementing GS in a commercial breeding program at fixed costs and with optimized resource allocation (Longin et al. 2015; Ben-Sadoun

et al. 2020). A recent simulation study by Bernardo (2021) investigated the use of different GS strategies at a fixed cost in the early selection stage of a maize breeding program carried out on a biparental population and showed that speeding up selection by using two cycles of GS (without phenotyping) allowed larger genetic gains than phenotypic selection. To the best of our knowledge, no study has been carried out simulating a more complex hybrid breeding scheme (BS) at a fixed cost including both early and late-stage selection in a reciprocal selection scheme performed in complementary groups.

To further improve hybrid breeding programs taking advantage of GS, Kadam et al. (2016) and Giraud (2016) proposed replacing topcross testing in the early stage of the process with genomic predictions trained on a sparse factorial design to predict all potential hybrid combinations. Both studies found good prediction accuracies for untested hybrids using sparse factorial design to train GS models. In addition, studies based on simulations by Seye et al. (2020) and experimental data by Lorenzi et al. (2022) and (Lorenzi et al. 2023) compared the use of sparse factorial designs with tester designs to train GS models. They revealed equivalent prediction accuracies when there is no SCA variance in the population and an increasing benefit of the factorial design when the SCA variance increases. Moreover, even when no SCA is involved, these studies illustrated the ability of the GS model trained on very sparse factorials to decouple line GCA components and predict their value accurately. The potential benefits of using a sparse factorial design instead of tester designs are (i) to evaluate more lines with the same number of plots and (ii) to select jointly on the GCA and SCA components in the early stage of the selection process. This also offers the possibility of revisiting the hybrid breeding scheme by collapsing early and late-stage evaluations into one single step using genomic predictions calibrated on a factorial design to predict both line GCAs and unobserved hybrid values. This idea was mentioned by Seye et al. (2020) but has never been tested to our knowledge. The main drawback of implementing factorial designs into maize breeding programs is the production of single-cross hybrids by handmade pollination instead of isolation plots for testcross evaluation, which can be challenging and costly. Therefore, even if the approach is appealing, it is necessary to assess the cost-efficiency of breeding schemes incorporating factorial designs compared to a conventional hybrid breeding scheme based on topcross evaluation.



The objective of our study is to further investigate the use of genomic selection trained on factorial designs in hybrid breeding programs. We used simulations to evaluate how the global efficiency of a conventional hybrid breeding program is affected by the implementation of GS and investigated several breeding scheme strategies combining (i) shortening the breeding cycle time by omitting one or several phenotypic evaluation steps, and (ii) using factorial designs instead of tester designs to train GS models to improve hybrid breeding schemes. Since breeders operate with a fixed budget, we compared all breeding scheme strategies at fixed annual costs. We used the generic public simulator MoBPS (Pook et al. 2020) to simulate a conventional phenotypic selection breeding scheme, considered as a benchmark, and five GS breeding schemes. The simulations were carried out considering a trait showing SCA or not. In both cases, our results illustrated the benefit of replacing test-cross evaluations by GS calibrated on a sparse factorial design to fasten the breeding process and revisit reciprocal recurrent selection schemes.

### Materials and Methods

We aimed at comparing, at the same budget, the efficiency of different breeding schemes (BS) in a recurrent reciprocal breeding context. To do so, we simulated two segregating populations from two complementary heterotic groups. We implemented five BSs using GS differing in their duration and the use of sparse factorials vs. tester evaluation. We compared them to a benchmark BS relying on phenotypic selection only. The simulations were performed using R (R Core Team 2020) and the R package MoBPS (Pook et al. 2020). Genomic predictions were performed using homemade scripts and the R package MM4LMM (Laporte et al. 2022).

#### Founder lines

We used the real genotypic data of 15 dent and 15 flint lines to simulate realistic LD patterns and allelic frequencies, as used in previous simulations (Seye et al. 2020). The genotypic data originated from two projects: 11 flint and 11 dent lines from the Dent and Flint NAM design of the European CornFed project (Bauer et al. 2013; Giraud et al. 2014), and four flint and four dent lines from the French SAMMCR (“Sélection Assistée par Marqueurs Multi-parental Connectée Réciproque”) project (Giraud et al. 2017b; Seye et al. 2020; Lorenzi et al. 2022). Lines were genotyped with single-nucleotide polymorphism (SNP) from the Illumina MaizeSNP50 BeadChip (Ganal et al. 2011).

Marker quality control and imputation were done following Seye et al. 2020, considering jointly genotypic data from the European CornFed and SAMMCR projects (308 dent and 397 flint lines). We applied filters to keep PANZEA markers (Gore et al. 2009) that were mapped on the CornFed NAM consensus genetic map (Bauer et al. 2013; Giraud et al. 2014) and were polymorphic in the set of 30 founder lines considered for the simulations. In total, 20,766 SNPs were used in our simulations. Among them, 88% were polymorphic in both groups, 7% were monomorphic in the flint founder group, and 6% were monomorphic in the dent founder group. The genetic positions of these markers were retrieved from the CornFed NAM consensus genetic map (Bauer et al. 2013; Giraud et al. 2014).

### **Simulation of QTL effects and positions**

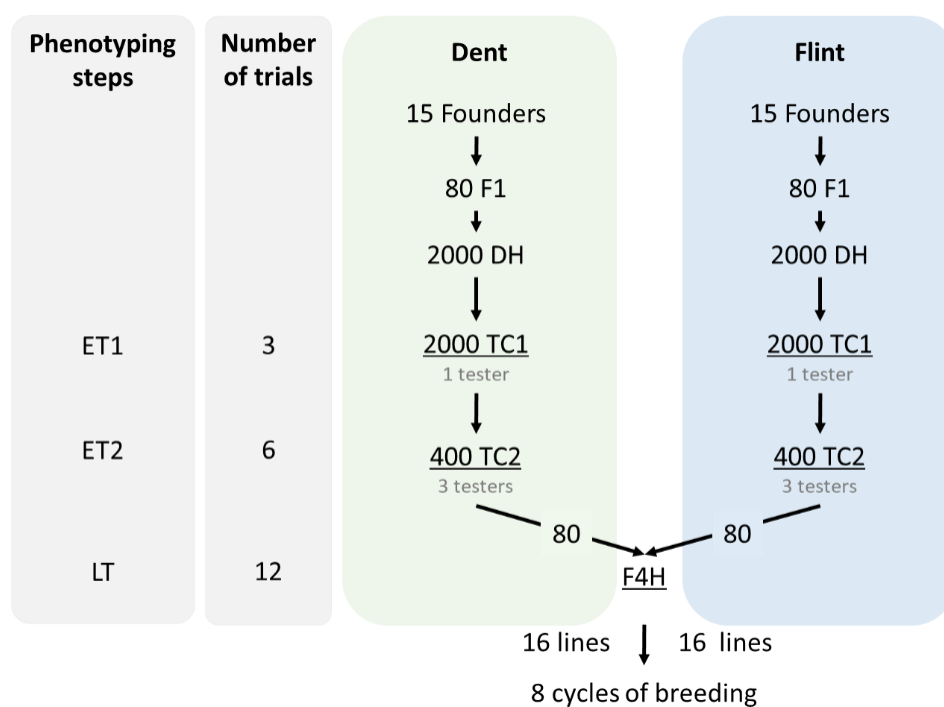
Each simulation replication was initiated by simulating one quantitative trait controlled by 300 biallelic QTLs. We sampled 300 QTLs among the polymorphic markers in both groups with a minimum interval of 0.2 cM between consecutive QTLs to avoid linkage between too close QTLs. Then, the 300 markers selected as QTLs were excluded from the list of markers used in the simulations to calibrate genomic predictions.

To simulate QTLs, we considered two scenarios with different proportions of SCA variance: one with no SCA and the other with a proportion of SCA variance similar to what could be observed in a maize breeding program between the dent and flint groups. In the absence of epistasis effects, SCA is generated by dominance effects at QTLs, with a magnitude depending on the divergence in allele frequencies between the two populations (Technow et al. 2012). The QTLs contributing to SCA variance are the ones that are polymorphic in both heterotic groups and showing dominance effects (Legarra et al. 2023). To simulate two levels of SCA variance, we considered different levels of additive and dominance QTL effects. We considered the same QTL effects for both flint and dent groups. In the purely additive scenario (no SCA variance), additive effects were drawn from a Gamma distribution with parameters scale=1.66 and shape=0.4 (Meuwissen et al. 2001; Technow et al. 2012), and for each locus, the favorable allele was randomly chosen (with probability  $P = 0.5$ ). In the scenario with SCA variance, additive effects were simulated as described previously, and the dominance effects were obtained as the product between the absolute value of the additive effect and the degree of dominance at each QTL. The degree of dominance was drawn from a Normal distribution of mean 1.0 and variance 0.75, corresponding to experimental estimates for grain yield in maize (Technow et al. 2012).

This simulates QTLs with mostly positive dominance effects for the favorable effect and the possibility of overdominance effects. This scenario generated on average 29% of genetic variance due to SCA in the initial cycle.

### Initialization

Before starting any of the six BSs, one initialization cycle was performed to produce populations that mimic those of breeding programs and phenotypic data to train the GS model in the first cycle of the GS BSs (see **Fig.1**). The 15 initial founder lines in each group were crossed following a half diallel to produce 80 single-cross F1 hybrids. 25 DH lines were generated for each single-cross, reaching a total of 2000 DH lines per group. The breeding scheme encompassed three different evaluation steps: two early-stage topcross hybrid testings, further called early testing 1 and 2 (ET1 and ET2), and a late testing stage (LT) for the final selection of candidates. In the ET1 step, DH lines were crossed to one tester from the opposite group and evaluated in three trials for yield performance. In the ET2 step, candidate lines issued from the ET1 step were crossed to three testers from the opposite group and evaluated in six trials. And last, in the LT step, the dent and flint candidate lines issued from the ET2 step were crossed following a factorial design, with each line contributing to four hybrids. These hybrids of the LT step were evaluated in 12 trials. Testers used in ET1 and ET2 steps were chosen randomly at each breeding cycle and phenotyping step among the founder lines of the population (15 flint founder and 15 dent founder lines). In this initialization cycle, 20% of the candidate lines were randomly sampled after each phenotypic evaluation step. This random sampling aimed at avoiding reducing genetic variance before starting BSs. At the end of the initialization cycle, 16 lines were used to be parents of the subsequent breeding cycles. After the initialization, six different BSs were conducted for eight non-overlapping breeding cycles (i.e., the initialization is common to all breeding schemes).



**Fig. 1** Initialization breeding scheme comprising three different evaluation steps, two early-stage topcross hybrid testings further called, early testing 1 and 2 (ET1 and ET2) and a late testing stage (LT)

### Phenotypic selection breeding scheme

A convention phenotypic selection BS, further called “Y3-Pheno”, was implemented to be used as a benchmark. Candidate individuals were assumed to be doubled-haploid (DH) lines at each cycle, as is currently the case in most maize breeding schemes. The Y3-Pheno scheme was designed considering realistic costs for a maize breeding scheme. Costs per unit of each operation within a breeding scheme (production, multiplication, phenotyping, and genotyping) are shown in **Table 1**. We considered the cost of production of the crosses between selected individuals to generate segregating DH families at each cycle as being negligible. The costs were assumed to be equal for genotyping, testcross seed production, and phenotyping at one location. The seed production cost of a single-cross hybrid (factorial) was considered twice the cost of a testcross hybrid. These costs were defined based on the expertise of breeders involved in the SAMMCR project.

The Y3-Pheno scheme is described in **Fig. 2**. After the initialization cycle, at each cycle, the 16 selected lines in each group from the previous breeding cycle were crossed following a half diallel to produce 80 single-crosses. For each single-cross, 25 DH lines were generated, reaching a total of 2000 DH lines per group with equal contributions from all families. The breeding scheme encompassed the same three evaluation steps (ET1, ET2, LT) already described

in the initialization cycle. After each phenotyping step (ET1, ET2, and LT), the best 20% candidate lines were selected based on the average phenotypic performance of their progeny and passed on to the next selection step. At the end of the cycle, the 16 best dent lines and the 16 best flint lines from the LT step were used as parents for the subsequent breeding cycle.

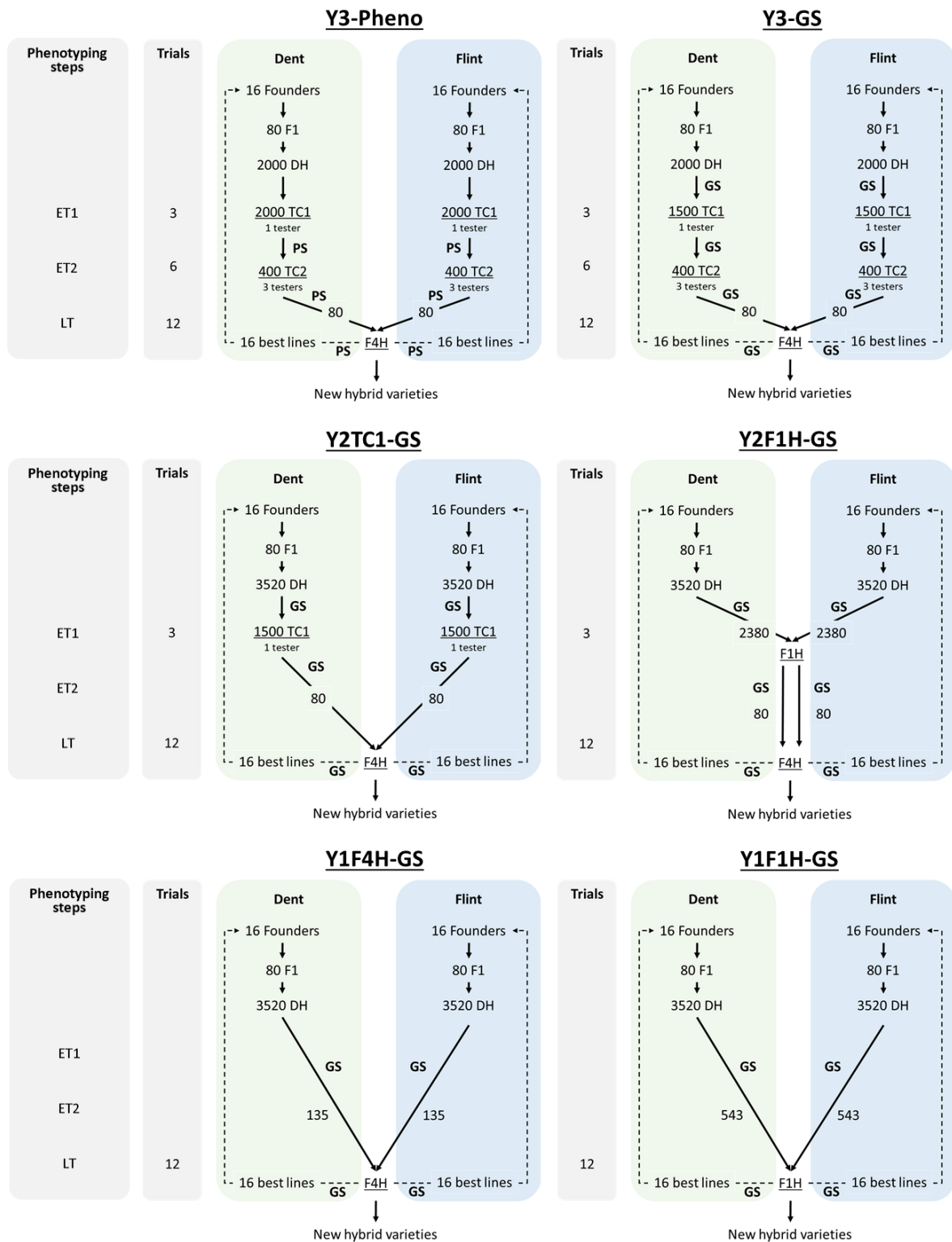
Genetic values were generated from the genotype at QTLs of the hybrids. Then, environmental errors were added to generate phenotypic values. Environmental errors were sampled from a normal distribution with a zero mean and an error variance  $\sigma_E^2$  defined by the initial repeatability in the founder population:  $r = \frac{\sigma_G^2}{\sigma_G^2 + \sigma_E^2} = 0.25$ . This led to an initial heritability in the founder population of  $h^2 = 0.50$  for a multi-trial phenotypic evaluation in three environments (ET1),  $h^2 = 0.67$  in six environments (ET2) and  $h^2 = 0.80$  in 12 environments (LT). Note that the heritability varied along breeding cycles with the evolution of the genetic variance  $\sigma_G^2$ .

**Table 1** Cost per unit in euros of each operation in the breeding scheme

Operation	Unit	Cost per unit
Production + Multiplication	DH line	25
Hybrid seed production	Testcross hybrid	15
	Single-cross hybrid	30
Genotyping	DH line	15
Phenotyping	Field plot	15

### GS breeding schemes

Five different BSs using GS were considered and compared to the Y3-Pheno scheme described above (**Fig.2**). We compared all BSs at a constant averaged annual cost, using the Y3-Pheno annual cost as a reference. Compared to the Y3-Pheno scheme, GS BSs included the additional genotyping cost of candidate lines. Resources were reallocated in the BSs (see below). We did not modify the number of trials for a given phenotyping evaluation step (three in ET1, six in ET2 and 12 in LT) and the same number of selected lines to be used as parents for the subsequent cycle (16 selected lines in each group). Details on the sizes and costs of each BS are given in **Fig.2** and **Table 2**.



**Fig.2** Definition of the benchmark phenotypic selection BS and the five GS BSs. For each BS and at each step, the number of lines is indicated along with the method - phenotypic selection (PS) or genomic selection (GS) - used to select the lines. The phenotyping steps are underlined and the corresponding number of trials is indicated (Trials). The phenotyping steps ET1 and ET2 correspond to the early testing of candidate lines based on topcross hybrids and the LT step to the late-stage evaluation of the hybrid combinations obtained by crossing the selected candidate lines following a factorial design with four hybrids per line (F4H).

**Table 2** Summary of hybrid BSs sizes (number of parents, F1 hybrids, DH lines and for each phenotypic evaluation step, the number of lines  $n_l$  and hybrids  $n_h$ ), number of years per cycle and costs (in euros).

Breeding scheme	Parents	F1	DH	ET1		ET2		LT		$n_{\text{trial}}$	Years per cycle	Cost per year	Cost per cycle
				$n_l$	$n_h$	$n_l$	$n_h$	$n_l$	$n_h$				
<b>Y3-Pheno</b>	32	160	4000	4000	4000	800	2400	160	320	21	5	131840	659200
<b>Y3-GS</b>	32	160	4000	3000	3000	800	2400	160	320	21	5	131840	659200
<b>Y2TC1-GS</b>	32	160	7040	3000	3000	-	-	160	320	15	4	132200	528800
<b>Y2F1H-GS</b>	32	160	7040	4760	2380	-	-	160	320	15	4	131825	527300
<b>Y1F4H-GS</b>	32	160	7040	-	-	-	-	270	540	12	3	131667	395000
<b>Y1F1H-GS</b>	32	160	7040	-	-	-	-	1086	543	12	3	131877	395630

We implemented the Y3-GS scheme to evaluate the impact of introducing GS while keeping the same steps as the Y3-Pheno scheme at fixed costs (**Fig.2**). In this GS BS and all the following ones, the candidate lines were selected based on genomic predictions of the candidate lines GCAs (GCA BLUPs). The GS model was trained on two generations for all GS BSs. The TRS comprised all phenotypic data available from the previous generation and the phenotypic data from the current generation was added to the TRS along the different steps of the breeding cycle. Note that in all GS BSs, the candidate DH lines were first selected based on predictions calibrated on the phenotypic data from the previous cycle. In contrast in the ET1, ET2 and LT selection stages, the selection included hybrid performances of the candidate lines. It should also be noted that depending on the GS BS, the size of the TRS can differ significantly. The description of the GS model used to calibrate the prediction model is given in the “Genomic selection” section. To compensate for the additional costs of genotyping in the Y3-GS scheme, we reduced the number of evaluated lines in the ET1 phenotyping step from 2000 in each group in the Y3-Pheno to 1500 lines in the Y3-GS scheme.

Then, taking advantage of GS, we investigated the impact of shortening the breeding cycle by suppressing one or two phenotypic evaluation steps. The Y2TC1-GS scheme omitted the ET2 phenotyping step, reducing the length of the scheme by one year compared to the Y3-GS (**Fig.2**). The budget saved by suppressing the ET2 step was reallocated to produce more DH lines (3520 DH lines in each group in the Y2TC1-GS instead of 2000 in the Y3-GS scheme) while keeping the other parameters identical to the Y3-GS scheme. The Y1F4H-GS scheme omitted both ET1 and ET2 phenotyping steps, reducing the length of the scheme by two years compared

to the Y3-GS (**Fig.2**). The budget saved by suppressing the ET1 and ET2 steps was reallocated to produce more DH lines (3520 DH lines in each group) and evaluate more lines in the LT phenotyping step (135 in each group in the Y1F4H-GS instead of 80 in the Y3-GS scheme).

Combined with shortening the breeding cycle, we also investigated the impact of replacing tester designs with a factorial design. We implemented the Y2F1H-GS scheme to evaluate the efficiency of suppressing the ET2 phenotyping step and using a factorial design instead of tester designs in the ET1 step (**Fig.2**). We used the Y2TC1-GS scheme as a framework and replaced the tester designs from the ET1 step with a factorial design composed of one hybrid per line. To make the Y2F1H-GS and Y2TC1-GS schemes comparable, we adjusted the number of evaluated lines in the ET1 phase (2380 lines in each group in the Y2F1H-GS instead of 1500 in the Y2TC1-GS) while keeping all other parameters identical.

Last, we investigated the impact of changing the composition of the factorial design in the Y1F4H-GS scheme. We implemented the Y1F1H-GS scheme to assess the impact of increasing the number of evaluated lines at the same number of plots by replacing the factorial composed of four hybrids per line with one composed of one hybrid per line in the LT step (**Fig.2**). To make the Y1F4H-GS and Y1F1H-GS schemes comparable, we adjusted the number of evaluated lines in the LT step without changing any other parameters of the BS.

### Genomic selection model

To perform genomic predictions, we used a unique GBLUP GS model considering phenotypic and genotypic data from both groups (flint and dent) jointly. For BSs comprising tester designs, tester lines were considered as any other line. The GS model was defined as:

$$\mathbf{y} = \mathbf{1}_n \cdot \mu + \mathbf{Z}_d \mathbf{g}_{GCA_d} + \mathbf{Z}_f \mathbf{g}_{GCA_f} + \mathbf{E},$$

where  $\mathbf{y}$  is the vector of ls-means of the  $n$  phenotyped hybrids,  $\mathbf{1}_n$  is a vector of  $n$  ones and  $\mu$  is the intercept.  $\mathbf{g}_{GCA_f}$  (respectively  $\mathbf{g}_{GCA_d}$ ) is the vector of random GCA effects of the  $n_f$  flint parental lines of the hybrids (respectively  $n_d$  dent lines), with  $\mathbf{g}_{GCA_f} \sim \mathcal{N}(0, \mathbf{K}_{GCA_f} \sigma_{GCA_f}^2)$  (respectively  $\mathbf{g}_{GCA_d} \sim \mathcal{N}(0, \mathbf{K}_{GCA_d} \sigma_{GCA_d}^2)$ ) where  $\mathbf{K}_{GCA_f}$  (respectively  $\mathbf{K}_{GCA_d}$ ) is the genomic relatedness matrix between the flint lines (respectively dent lines) computed following method 1 from VanRaden (2008).  $\sigma_{GCA_f}^2$  and  $\sigma_{GCA_d}^2$  are the flint and dent GCA variances.  $\mathbf{Z}_d$  and  $\mathbf{Z}_f$  are



the corresponding incidence matrices of dimensions  $[n \times n_d]$  and  $[n \times n_f]$ , respectively, that relate the observations to the GCA effects of lines considered in the model.  $\mathbf{E}$  is the vector of error terms, with  $\mathbf{E} \sim \mathcal{N}(0, \mathbf{I}\sigma_E^2)$ . The different random effects are assumed to be independent. The GS model was fitted using the package MM4LMM for R (Laporte and Mary-Huard 2020).

### **Metrics used to assess the BS efficiency**

To assess the efficiency of a BS we estimated the genetic gain and the genetic variance after each breeding cycle. And for GS BSs, we computed the prediction accuracy at two selection steps where genomic predictions were used: the initial selection of the DH lines before any early-stage phenotyping, and the LT selection step to select the final individuals.

To estimate the genetic gain, we estimated the average true genetic values of all possible single-cross hybrids between the 16 selected flint candidate lines and the 16 selected dent candidate lines ( $\mu$ ). We also computed the average of the ten superior true genetic values to mimic the mean value of the best varieties that could be released from the breeding program ( $\mu_{10}$ ). The genetic gain was computed following:  $G = \frac{\mu - \mu_0}{\sigma_0}$ , where  $\mu_0$  and  $\sigma_0$  were the mean and standard deviation of the initial population (hybrids produced by crossing the 2000 DH lines between the dent and flint groups in the initialization cycle).

To estimate the GCA and SCA genetic components of the population of a given breeding cycle, we produced a factorial design between all DH candidate lines of the flint and dent groups, with each DH line contributing to 50 single-cross flint-dent hybrids. The GCA of a DH line was estimated as the mean true genetic value of its progeny, and the SCA of a hybrid was estimated as the difference between the expected value based on the mean of the parental GCAs and the observed genetic value of the hybrid.

The prediction accuracy of the GS BSs was computed at two selection steps: at the selection stage of the candidate DH lines (prior to their evaluation) and at the LT selection stage of the final 16 best lines in each group. For the DH selection step, the prediction accuracy of the line GCAs was computed based on the hybrids of the incomplete factorial design between all the flint and the dent DH lines used to obtain variance components. The prediction accuracy was computed as the correlation between the GCA BLUP and the true line genetic value, where the true genetic value of the line was estimated as the average true genetic value of the hybrid

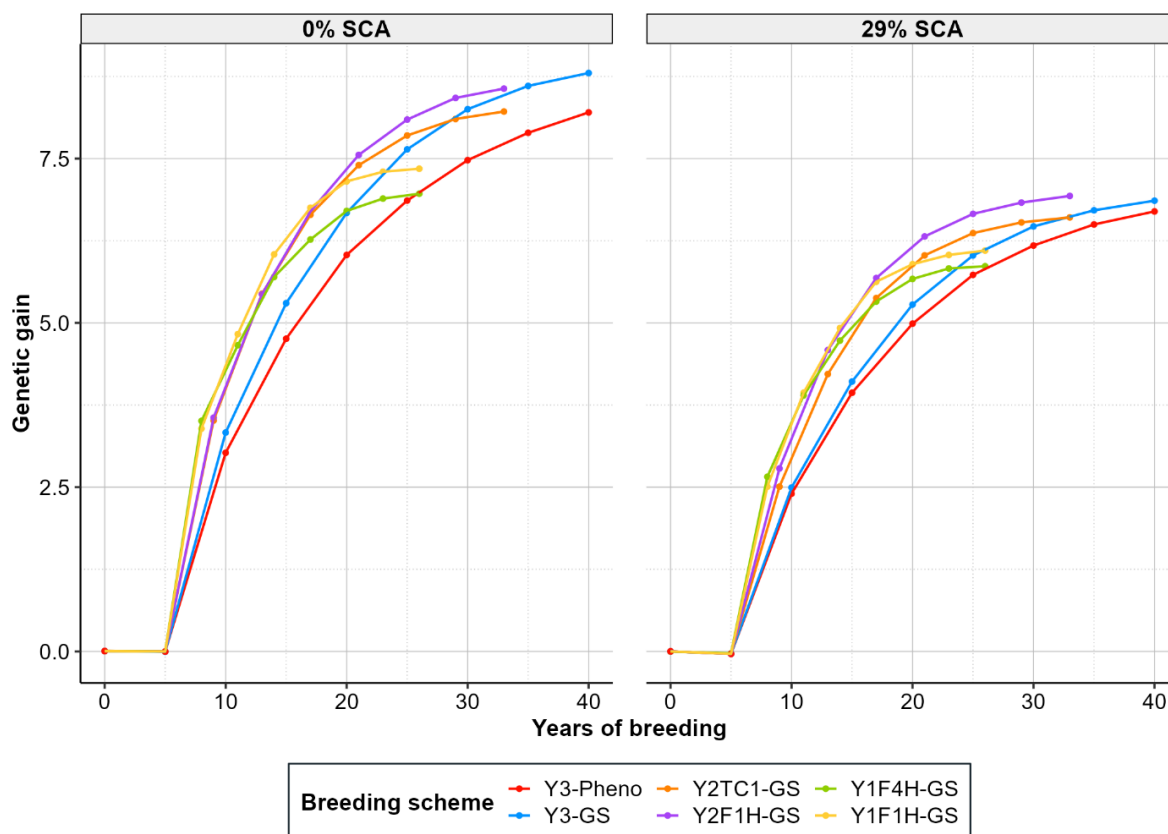
progeny of the line. For the LT selection step, the prediction accuracies of the line GCAs and the hybrids of the LT factorial design were computed using all performances available in the TRS (including the LT evaluation phenotypes). The prediction accuracy of the line GCAs was computed as described previously, using the true line genetic values estimated in the DH selection step. The hybrid prediction accuracy was computed as the correlation between the predicted hybrid value (sum of the GCA BLUPs) and the true hybrid genetic value.

## Results

### Impact of the different BSs on the genetic gain

We evaluated first how genetic gain was affected by the implementation of GS in a conventional BS, including early testing of DH candidates based on testers. At fixed costs, the genetic gain along years (**Fig.3**) and breeding cycles (**Fig.S1**) was higher for the Y3-GS than the Y3-Pheno scheme. The advantage of the Y3-GS scheme was higher for the purely additive scenario (0% SCA) compared to the SCA variance scenario (29% SCA) (**Fig.S1**). The genetic gain computed on the top ten hybrids was higher than the one computed on the total population for both BSs under both SCA scenarios. Differences between the Y3-Pheno and Y3-GS schemes were also lower in the top ten hybrids than in the total population, however, the Y3-GS scheme still outperformed the Y3-Pheno scheme (**Fig. S2** and **Fig. S3**).

Then, we investigated the impact of shortening the breeding cycle by suppressing one (Y2TC1-GS) or two phenotypic topcross evaluation steps (Y1F4H-GS) and compared it to the Y3-GS scheme. Shortening the breeding cycle in the Y2TC1-GS and Y1F4H-GS schemes resulted in a faster and higher short-term genetic gain per year than in the Y3-GS scheme. However, the cumulated genetic gain after eight cycles was generally lower (**Fig.3**, **Fig.S1**). The Y2TC1-GS scheme, which omitted the ET2 phenotyping step, combined a faster short-term genetic gain with a high long-term genetic gain, slightly below and above the one of the Y3-GS at 0% and 29% SCA, respectively. The Y1F4H-GS scheme, which omitted ET1 and ET2 phenotyping steps, generated the highest short-term genetic gain (first 15 years) but the lowest long-term genetic gain. Overall, the genetic gain in the scenario with 29% SCA was lower than in the purely additive scenario, but the global ranking between BSs was unchanged. The genetic gain computed on the top ten hybrids showed the same ranking and similar differences between BSs to what was observed on the total population (**Fig.S2** and **Fig.S3**).



**Fig.3** Evolution of the genetic gain over time of all BSs in two SCA scenarios with 0% or 29% SCA in the initial population. Genetic gain is expressed as the difference between the mean genetic value of the hybrids and the mean hybrid genetic value in the initial population divided by the standard deviation in the initial population. Each point represents the overall mean of 10 replicates. Year 0 corresponds to the initial population, year 5 corresponds to the initialization and the breeding starts after year 5.

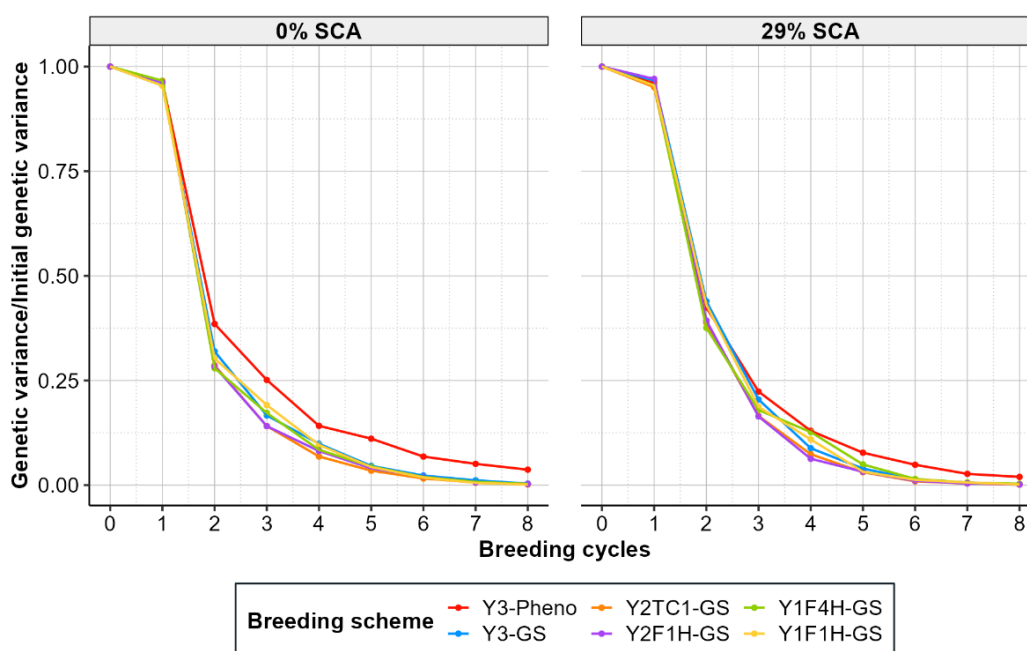
Last, we investigated the impact of combining shorter breeding cycles with factorial designs and compared different factorial design compositions. First, we compared the Y2TC1-GS and Y2F1H-GS schemes to investigate the impact of replacing tester designs with a sparse factorial design (F1H) in the ET1 step. The Y2F1H-GS scheme, using a factorial design in the ET1 step, enabled the evaluation of more lines than Y2TC1-GS and generated a higher genetic gain per year (**Fig.3**) and per breeding cycle (**Fig.S1**) regardless of the SCA scenario. The advantage of the Y2F1H-GS scheme was stronger when 29% SCA variance was considered in the initial population where it outperformed all the other BSs (**Fig.3**). The Y1F4H-GS and Y1F1H-GS schemes, which only involve one evaluation step (LT) were slightly more efficient per year in the short-term and less efficient in the long-term compared to all the other BSs (**Fig.3**). The comparison between the Y1F1H-GS and the Y1F4H-GS schemes, allowed us to assess the efficiency of evaluating more lines (F1H) instead of more hybrids per line (F4H) at the same number of plots. The Y1F1H-GS scheme generated a higher genetic gain per year (**Fig.3**) and per breeding cycle (**Fig.S1**) than the Y1F4H-GS. The results were similar between the two SCA

scenarios, however, differences were smaller in the scenario with 29% SCA. The genetic gain computed on the top ten hybrids showed the same ranking and similar differences between BSs to what was observed on the total population (**Fig.S2** and **Fig.S3**).

### Impact of the different BSs on the genetic variance

The total genetic variance decreased rapidly along breeding cycles. This decrease was faster in the scenario with 0% SCA variance than in the scenario with 29% SCA (**Fig.4**). After eight cycles, almost all initial variance was exhausted, especially for GS BSs. The Y3-Pheno scheme was associated with a slower decrease in the genetic variance per breeding cycle compared to all the other BSs. The implementation of GS resulted in a faster decrease in the genetic variance per cycle. However, the shortening of the breeding cycle and the use of factorial designs had no strong impact on the genetic variance per cycle.

We decomposed the genetic variance into its GCA and SCA components (results presented in supplementary materials in **Fig.S4**). We observed the same trends and ranking between the BSs as described previously on the total genetic variance. When SCA was involved (29% SCA scenario), the SCA variance decreased faster than the GCA variances for all BS, as illustrated by the decreasing percentage of SCA variance along breeding cycles (**Fig.S4**).

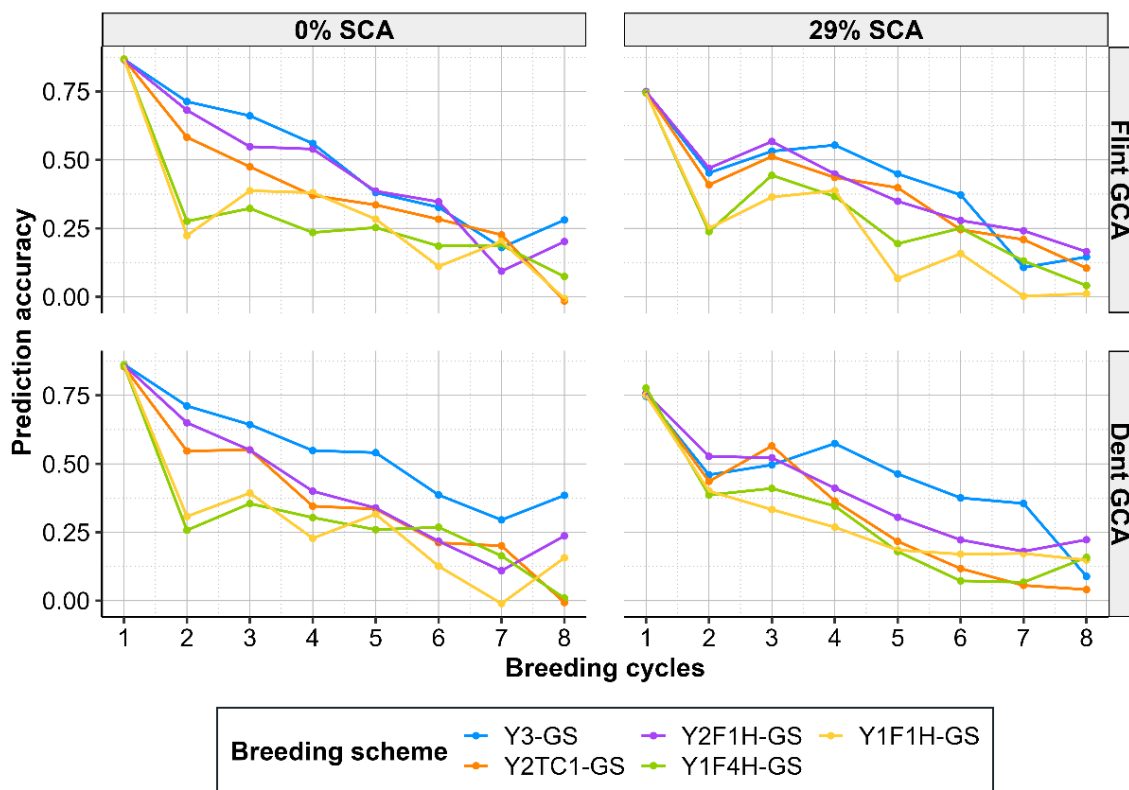


**Fig.4** Evolution of the genetic variance along breeding cycles for all BSs for two SCA scenarios, purely additive or with 29% SCA. Genetic variance is expressed as the ratio between the genetic variance in the current population and the genetic variance in the initial population. Each point represents the overall mean of 10 replicates. Breeding cycle 0 corresponds to the initialization and the breeding starts at 1.

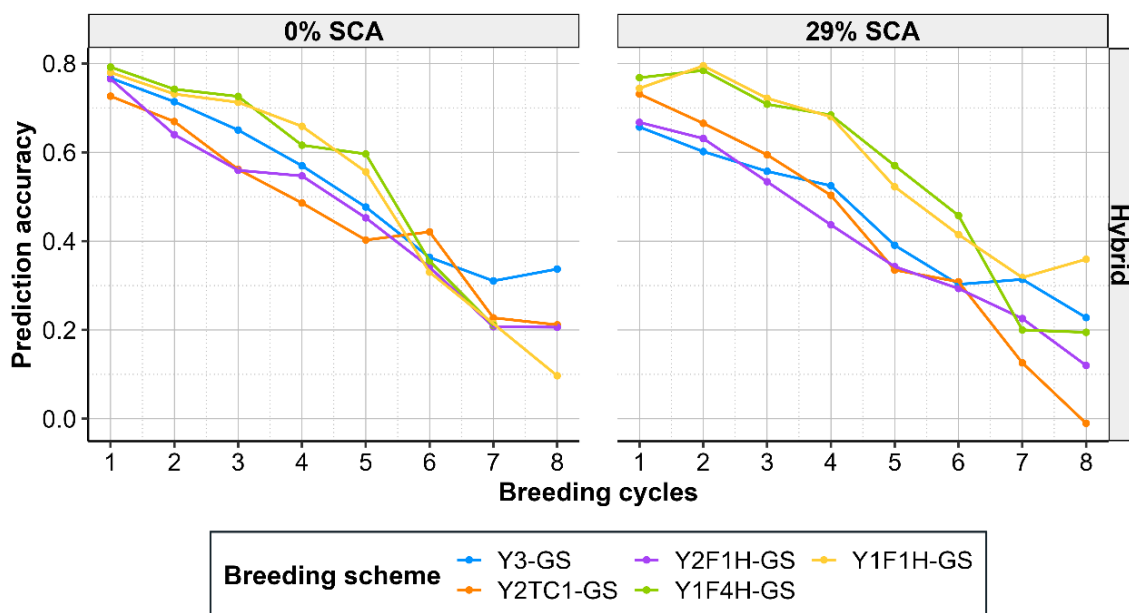
### Impact of the different GS breeding schemes on the prediction accuracy

We computed the prediction accuracy for two selection steps: the initial selection of the DH lines (**Fig.5**) and the late-stage selection of the lines (LT selection step) (**Fig.6**). At the DH selection step, since the objective was to select the candidate lines to be phenotyped, we computed the prediction accuracy of the DH line GCAs for each breeding cycle (**Fig.5**). Genomic predictions at this step were performed using only the phenotyping data from the previous breeding cycle to train the model. Overall, the prediction accuracy decreased along breeding cycles regardless of the SCA scenario. There was a strong decrease in prediction accuracy between the first and second breeding cycle, particularly for the Y1F4H-GS and Y1F1H-GS schemes. This drop is associated with moving from predictions calibrated on data from the initialization cycle, which includes three phenotyping steps, to predictions calibrated on the first selection cycle, which for these BS includes only one phenotyping step. The ranking based on the prediction accuracy was: Y3-GS, Y2F1H-GS, Y2TC1-GS, Y1F1H-GS and Y1F4H-GS. The lowest prediction accuracies were observed for the Y1F1H-GS and Y1F4H-GS, which were also the schemes with the smallest TRSs. The prediction accuracies were lower in the scenario with SCA (29% SCA) compared to the purely additive scenario (0% SCA).

Overall, in the LT selection step, the hybrid prediction accuracy decreased along breeding cycles regardless of the SCA scenario (**Fig.6**). The highest prediction accuracies were observed for the Y1F1H-GS and Y1F4H-GS with no clear difference between them. The prediction accuracies of the other BSs were similar but lower than the Y1F4H-GS and the Y1F1H-GS schemes. Similar results were observed between the two SCA scenarios with a similar prediction accuracy level.



**Fig.5** Evolution of the prediction accuracy of the line GCAs in the DH initial selection step along breeding cycles for two SCA scenarios, purely additive or with 29% SCA. Prediction accuracy is expressed as the correlation between the line GCA true genetic values and the GCA BLUPs. Each point represents the overall mean of 10 replicates.



**Fig.6** Evolution of the hybrid prediction accuracy in the LT step along breeding cycles for two SCA scenarios, purely additive or with 29% SCA. Prediction accuracy is expressed as the correlation between the hybrid true genetic values and the predicted hybrid value. Each point represents the overall mean of 10 replicates.

## Discussion

### Impact of implementing GS in a conventional breeding scheme

The first objective of this study was to assess the efficiency of implementing GS in a conventional recurrent reciprocal BS by comparing the Y3-GS and Y3-Pheno schemes. The Y3-GS scheme showed a higher genetic gain at fixed costs regardless of the SCA scenario (**Fig.3, Fig.S1**). Numerous studies reported an increased genetic gain with GS compared to phenotypic selection in hybrid BSs (Lin et al. 2016; Gaynor et al. 2017; Gorjanc et al. 2018; Voss-Fels et al. 2019a, b). However, these studies were not conducted at fixed costs and often took advantage of GS to do more cycles per year. Comparison of Y3-GS and Y3-Pheno schemes at fixed costs with the same cycle length still showed an advantage of Y3-GS, even if fewer candidate lines were evaluated in the ET1 phenotyping step to offset the genotyping costs in the Y3-GS scheme. In the Y3-GS, these candidate lines were preselected based on genomic predictions before being evaluated in the ET1 step, which added one selection step compared to Y3-Pheno.

The Y3-Pheno scheme only relied on phenotypic performances to select candidate individuals. It could have been possible to use pedigree BLUP (PBLUP) predictions in the Y3-Pheno scheme to improve prediction accuracy. However, all DH candidate lines were phenotyped in the Y3-Pheno scheme, and the benefit of using information from relatives to increase the GCA prediction accuracy within segregating families is expected to be low in this case. Several plant studies compared GS with PBLUP selection and generally reported an advantage of GS (Beyene et al. 2015; Vivek et al. 2017; Zhang et al. 2017; Viana et al. 2019).

Implementing GS improved genetic gains but resulted in a faster decrease in the genetic variance compared to the Y3-Pheno scheme. This accelerated loss of genetic diversity under GS has already been reported in previous studies, especially when coupled with high selection intensity (Gaynor et al. 2017; Tessema et al. 2020; Li et al. 2022). It is due to the rapid fixation of favorable but also unfavorable alleles at QTL (Falconer 1996; Li et al. 2008, 2022; Jannink 2010). Maintaining genetic diversity is therefore essential for the sustainability of GS breeding programs (Goddard 2009). Optimum contribution selection (OCS) methods have been proposed to maintain diversity in breeding programs, showing promising results (Gorjanc et al. 2018; Allier et al. 2019a). It would be interesting to apply such methods to the hybrid context to maximize genetic gain while controlling the loss of genetic diversity in GS breeding programs. We

simulated closed breeding programs, i.e. no new material was introduced along cycles, which explains the decrease in genetic variance that we observed for all BSs after eight generations. In real breeding programs, external diversity is regularly introduced and GS was helpful in optimizing introductions (Allier et al. 2020; Sanchez et al. 2023). Introduction of genetic resources raises specific questions when applied to reciprocal BS, notably the maintenance of heterotic group differentiations. These questions were beyond the scope of this paper but would deserve further attention.

#### **Impact of shortening the breeding cycle**

This section discusses the impact of shortening the breeding cycle without changing the phenotypic evaluation designs. In the Y2TC1-GS and Y1F4H-GS schemes, we investigated the benefit of reducing the cycle length by suppressing one or two early-stage phenotyping steps compared to a conventional GS BS (Y3-GS). The shorter BSs allowed for a faster short-term genetic gain (particularly the Y1F4H-GS) but led to a lower genetic gain in the long term. Shortening the breeding cycle made it possible to do more breeding cycles per time unit compared to the Y3-GS scheme, which increased the genetic gain per time unit in the first years (**Fig.3**). However, shortening breeding cycles led to a lower cumulated genetic gain after eight breeding cycles compared to the Y3-GS as shown in **Fig.S1**. To explain this result, we looked at different parameters that impact genetic gain: the selection intensity, the genetic variance and the prediction accuracy. When reallocating resources in the Y2TC1-GS and Y1F4H-GS schemes to fit the constrained budget, we increased the number of candidate DH lines without changing the number of selected lines to be used as parents for the subsequent cycle, so we increased the selection intensity compared to the Y3-GS. This increased selection intensity could result in a faster decrease in genetic variance. However, genetic variances along breeding cycles were similar between GS BSs (**Fig.4**), indicating that it did not explain the lower long-term genetic gain. So, to explain this result we looked at prediction accuracies. Prediction accuracies were computed at two genomic selection steps: the initial DH lines selection stage (before field evaluation) and the late-stage selection (LT step) (**Fig.5, Fig.6**). Lower prediction accuracies were observed at the initial DH selection step for GS BSs with shorter cycle lengths than for the Y3-GS scheme, whereas in the LT selection step, the prediction accuracies were higher for the Y1F4H-GS scheme than for the Y3-GS and Y2TC1-GS schemes. As most selection is done at the DH phase, the prediction accuracies at this step have a stronger impact on the genetic gain than



in the LT step. Thus, even if the prediction accuracies for the Y1F4H-GS scheme were higher in the LT step, it did not compensate for the decrease in prediction accuracy from the DH phase. So, the lower long-term genetic gain observed for the shorter GS BSs (Y2TC1-GS, Y1F4H-GS) compared to the Y3-GS scheme is mostly due to lower prediction accuracy in the DH selection step, which was not compensated by the increased selection intensity.

The size of the TRS is a major factor affecting prediction accuracies (cf review by Isidro y Sánchez and Akdemir 2021). Since incorporating phenotypic data from previous years in the TRS can increase prediction accuracies (Lorenzi et al. 2023 in review, Jannink 2010; Denis and Bouvet 2013; Neyhart et al. 2017; Sleper et al. 2020), we chose to include the phenotyping data of two generations in the TRSs regardless of the BS. As a consequence, the TRS sizes were smaller for the BSs with a reduced number of phenotyping years, impacting the prediction accuracy. Indeed, in the DH selection step in the Y3-GS scheme, genomic predictions were calibrated on the performances of 5720 hybrids, whereas it was calibrated on 3320 in the Y2TC1-GS and 540 in the Y1F4H-GS. The effect of the TRS size on the prediction accuracy was particularly visible between the first and second breeding cycles, where there was a strong decrease of prediction accuracy for the Y1F4H-GS BS. This decrease in prediction accuracy is associated with moving from genomic predictions calibrated on data from the initialization cycle comprising three phenotyping steps (5720 hybrids) to predictions calibrated on the first breeding cycle comprising for the Y1F4H-GS BSs only one phenotyping step. To avoid the reduction in the size of TRS for the shortest BSs, it might have been interesting to include more than two generations in the TRS (particularly for the Y1F4H-GS). Additionally, optimization strategies to construct the TRS could have been investigated (see Rio et al. 2022 for a review). Note that the same trend in terms of prediction accuracies was observed for the Y1F1H-GS, which is also based only one evaluation step.

We could have tested other schemes with different allocations for the phenotyping, but we had to limit the number of schemes to facilitate their exploitation. In the Y2TC1-GS scheme, we chose to suppress the ET2 phenotyping step instead of the ET1 for two reasons. The ET1 is the step where the greatest number of candidate lines for selection are phenotyped, which enables eliminating the worst candidate lines. It is also the cheapest phenotyping step due to the limited number of trials. So, suppressing ET2 instead of ET1 allowed for more efficient resource reallocation. We considered shortening the BS by suppressing some phenotyping steps

(one or two) but we did not consider the possibility of further shortening BS by performing selection cycles without any phenotyping, as proposed for instance by Bernardo (2021).

The phenotyping stages and the number of phenotyping cycles to be conserved or not for calibrating GS predictions is an important question. More generally, the choice of the individuals to phenotype along each stage of a BS is crucial. In conventional phenotypic breeding programs, only the most promising individuals are phenotyped in the late stage, but keeping only these individuals in the TRS might not lead to the highest prediction accuracy in GS. Results found by Neyhart et al. (2017) and Brandariz and Bernardo (2018), showed that when constructing a TRS, one should consider keeping hybrids produced between unselected lines to maintain high prediction accuracy. It is still unclear whether it would be advantageous at a fixed budget to evaluate only the best lines based on their predicted value or also some of the worst ones to construct an efficient TRS.

#### **Benefit of factorial designs in GS hybrid breeding schemes**

Our simulations align with a previous simulation study conducted by Seye et al. (2020), which showed promising results for replacing testcross evaluations with genomic predictions calibrated on a factorial design in early-stage maize hybrid BSs. We aimed at extending this previous simulation study to consider more realistic BSs addressing both the line evaluation and the identification of superior hybrids at fixed costs. We investigated the benefit of replacing the tester designs with a factorial design in a two-step phenotyping BSs (Y2F1H-GS) and also collapsing early- and late-stage phenotyping into a single phenotyping stage relying on a factorial design (Y1F4H-GS and Y1F1H-GS). This last option was mentioned by Seye et al. (2020) but has never been tested.

We showed that reducing the cycle length by one year and replacing tester designs with a sparse factorial design in the Y2F1H-GS scheme improved the genetic gain compared with the Y2TC1-GS and the Y3-GS schemes (**Fig.3, Fig.4**). The use of a factorial design in the Y2F1H-GS instead of a tester designs slightly increased the prediction accuracies and allowed evaluating 1760 more lines in the ET1 step than in the Y2TC1-GS scheme, which increased the genetic gain per cycle. Interestingly, the Y2F1H-GS scheme was the only scheme with a reduced cycle length showing an advantage both in the short- and long-term compared to the Y3-GS scheme. The Y2F1H-GS even outperformed the Y3-GS scheme in the long term when SCA was involved.

Prediction accuracies in the DH selection step were slightly higher for the Y2F1H-GS than the Y3-GS in the first three breeding cycles, which could explain the increased genetic gain. This result agrees with the simulations by Seye et al. (2020) reporting an increasing advantage of the factorial compared to the tester for the prediction of line and hybrid values when the SCA variance increases. While SCA levels simulated in both studies were comparable (23%, for Seye et al. 2020 and 29% in our study), the increased advantage of the factorial was not as strong in our study as the one observed by Seye et al (2020). It may be due to the fact that we considered a more complex and realistic breeding scheme involving more founder lines and more families at each cycle, which is less favorable to predict GCAs in a sparse factorial design. We also did not include SCA effects in the GS models because the computation time was too long to be compatible with the simulation of different BSs. However, studies have shown that including SCA effects in GS models often does not increase or only slightly improve the prediction accuracy of single-cross hybrids (Vitezica et al. 2017; Seye et al. 2020; González-Diéguéz et al. 2021; Lorenzi et al. 2022). In this context, we think the benefit of using factorial designs could have been similar or slightly improved if SCA effects were included in the GS prediction models, but it would not have changed our main conclusions.

We also investigated BSs with only one phenotyping step and compared two BSs with different resource allocations (Y1F4H-GS and Y1F1H-GS). The observations mentioned for the Y1F4H-GS scheme in the previous section also apply to the Y1F1H-GS, both BSs were less efficient in the long term than the Y2F1H-GS. The Y1F1H-GS outperformed the Y1F4H-GS in terms of genetic gain per year and per cycle. Yet, the number of produced candidate DH lines, the genetic variances and prediction accuracies along cycles were similar between the two schemes. The benefit of the Y1F1H-GS can be explained by the increased number of lines that were evaluated in the factorial design F1H compared to the F4H. Increasing the number of evaluated lines in the factorial design decreases the selection intensity at the DH selection step and increases it in the LT step when predictions are based on the performances of the lines still in the selection process and are more accurate than in the LT step. However, it was surprising that there was no clear difference in prediction accuracy in the DH selection step between the Y2F4H-GS and Y1F1H-GS schemes. Results from previous studies showed that it was beneficial to use more lines instead of more hybrids per line when predicting untested hybrids (Seye et al. 2020; Lorenzi et al. 2022), thus we were expecting an increased prediction accuracy in the Y1F1H-

GS scheme. One possible explanation is that in these studies, the TRS and PS were composed of the same generation, whereas in our simulations, we used performances from the previous generation to predict the next one. Compared to other GS BSs, the strategy using only one year of phenotyping (Y1F4H-GS and Y1F1H-GS), even when considering the use of a factorial with one hybrid per line (Y1F1H-GS), showed the lowest genetic gains. Some optimization of the resource allocation could be investigated to improve Y1 BSs, however, in light of our results, we do not recommend using this BSs.

#### **Prospects for implementing factorial design in hybrid BSs**

We intended simulating a realistic breeding program. To do so, we used real genotypic data to create our initial population, considered two SCA level scenarios, compared BSs at fixed costs, and resource allocation was performed based on discussions with the private companies and on results from previous studies (Seye et al. 2020; Lorenzi et al. 2022; Lorenzi et al. 2023). Still, for the sake of simplicity, some hypotheses were made. We simulated a closed breeding program and no overlapping generations. We could have tested other schemes, but we had to limit the number of schemes to facilitate the interpretation of our results. Among the BSs tested, the best one is the Y2F1H-GS scheme, in which the two conventional testcross evaluations (ET1 and ET2) are replaced with a very sparse factorial (F1H), followed by a LT evaluation based on factorial with more hybrids per line to better estimate their GCAs. This strategy was more efficient in the long term than the strategies (Y1F1H-GS and Y1F4H-GS) that collapse the two evaluation stages of a conventional hybrid BS by using only one factorial design per cycle. As already discussed, this is due in our simulations to the strong reduction of the number of lines evaluated for their hybrid value in these BS compared to the Y2F1H-GS, which reduced prediction accuracies for the early DH selection of the next cycle. We chose to keep the same number of trials in this LT stage, which limited the number of lines involved in this factorial. The presented BSs might not be optimal and it could be interesting to further optimize the resource allocation to assess the optimum genetic gain that could be achieved.

While our results confirm the interest in using factorial in the early selection stage, a major drawback is the cost and the workforce necessary for producing single-cross hybrids. Handmade pollination increases the costs per unit of single-cross hybrids seed production compared to topcross seed production (**Table 2**). Since one benefit of the factorial is to be able to evaluate more lines than a tester design with fewer plots, the increased cost of production of

single-crosses can be compensated by a reduced cost of phenotyping for the same number of tested lines (Seye et al. 2020). This strategy was used to implement factorial designs in our BSs at fixed costs. We showed that it is cost-efficient to suppress one early-stage phenotyping evaluation (ET2) and use a factorial design that maximizes the number of evaluated lines in early stages. Besides financial aspects, producing single-cross hybrids requires restructuring the breeding programs to manage the challenges of single-cross production. Indeed, handmade pollination requires larger workforces but can also result in a lower rate of success of seed production compared to topcross hybrids and requires synchronizing the flowering period of the two new candidate DH lines, which may not be known with precision. Sparse tester designs are an alternative to factorial designs showing equivalent predictive abilities for predicting single-cross hybrids (Lorenzi et al. 2022). It corresponds to a tester-based design, which maximizes the number of evaluated lines by crossing more lines, each with a different tester. Using such a design in the ET1 step in a Y2 BS could have been interesting since it would benefit from the factorial strategy (maximizing the number of evaluated lines) without requiring major restructuration of the breeding program.

The integration of GS into breeding programs can vary in terms of the stages at which it is applied and the extent to which it is utilized, depending on the constraints encountered. GS can be coupled with other approaches to even speed up the process in several species. Recently, methods aiming at accelerating plant development (greenhouses or controlled environment facilities) have been proposed to reduce the generation time and accelerate the breeding process. This opens the way to the so-called “speed breeding” approach (Hickey et al. 2017b; Wanga et al. 2021). Simulation studies in several species have investigated combining GS and speed breeding to increase the breeding cycles and generate high genetic gains per unit time (Watson et al. 2018; Jighly et al. 2019; Pandey et al. 2022). A possible improvement for the BSs we proposed could be to combine GS factorial BSs with speed breeding.

## Conclusion

This study aimed at investigating the use of genomic selection trained on factorial designs in maize hybrid BSs and give some insight regarding its practical implementation in commercial breeding programs at fixed costs. Using factorial design is challenging because handmade pollination is necessary for producing single-cross hybrids and increases the cost per unit of

single-cross hybrid seed production. Our results clearly showed that even when considering additional seed production costs, replacing tester designs with a sparse factorial design is beneficial. Combining genomic selection with shortening the breeding cycle by replacing the two early-stage topcross evaluations with a unique sparse factorial design (Y2F1H-GS scheme) was the most promising scheme, holding great potential to fasten maize hybrid breeding while increasing the genetic gain. Implementing genomic predictions calibrated on factorial designs will require reorganizing the logistics of BSs but is potentially worth the effort in light of our results.

## 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>
- Allier A, Lehermeier C, Charcosset A, et al (2019) Improving Short- and Long-Term Genetic Gain by Accounting for Within-Family Variance in Optimal Cross-Selection. *Front Genet* 10. <https://doi.org/10.3389/fgene.2019.01006>
- Allier A, Teyssède S, Lehermeier C, et al (2020) Genomic prediction with a maize collaborative panel: identification of genetic resources to enrich elite breeding programs. *Theor Appl Genet* 133:201–215. <https://doi.org/10.1007/s00122-019-03451-9>
- Bauer E, Falque M, Walter H, et al (2013) Intraspecific variation of recombination rate in maize. *Genome Biol* 14:R103. <https://doi.org/10.1186/gb-2013-14-9-r103>
- Ben-Sadoun S, Rincenc R, Auzanneau J, et al (2020) Economical optimization of a breeding scheme by selective phenotyping of the calibration set in a multi-trait context: application to bread making quality. *Theor Appl Genet* 133:2197–2212. <https://doi.org/10.1007/s00122-020-03590-4>
- Bernardo R (2021) Upgrading a maize breeding program via two-cycle genomewide selection: Same cost, same or less time, and larger gains. *Crop Science* 61:2444–2455. <https://doi.org/10.1002/csc2.20516>
- Bernardo R, Yu J (2007) Prospects for Genomewide Selection for Quantitative Traits in Maize. *Crop Science* 47:1082–1090. <https://doi.org/10.2135/cropsci2006.11.0690>
- Beyene Y, Semagn K, Mugo S, et al (2015) Genetic Gains in Grain Yield Through Genomic Selection in Eight Bi-parental Maize Populations under Drought Stress. *Crop Science* 55:154–163. <https://doi.org/10.2135/cropsci2014.07.0460>
- Brandariz SP, Bernardo R (2018) Maintaining the Accuracy of Genomewide Predictions when Selection Has Occurred in the Training Population. *Crop Science* 58:1226–1231. <https://doi.org/10.2135/cropsci2017.11.0682>
- 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>
- Crossa J, Pérez-Rodríguez P, Cuevas J, et al (2017) Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. *Trends in Plant Science* 22:961–975. <https://doi.org/10.1016/j.tplants.2017.08.011>
- Denis M, Bouvet J-M (2013) Efficiency of genomic selection with models including dominance effect in the context of Eucalyptus breeding. *Tree Genetics & Genomes* 9:37–51. <https://doi.org/10.1007/s11295-012-0528-1>
- Falconer DS (1996) Introduction to quantitative genetics. Pearson Education India
- 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>
- Gaynor RC, Gorjanc G, Bentley AR, et al (2017) A Two-Part Strategy for Using Genomic Selection to Develop Inbred Lines. *Crop Science* 57:2372–2386. <https://doi.org/10.2135/cropsci2016.09.0742>
- 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 (2017) 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>
- Giraud H, Lehermeier C, Bauer E, et al (2014) Linkage Disequilibrium with Linkage Analysis of Multiline Crosses Reveals Different Multiallelic QTL for Hybrid Performance in the Flint and Dent Heterotic Groups of Maize. *Genetics* 198:1717–1734. <https://doi.org/10.1534/genetics.114.169367>



- Goddard M (2009) Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica* 136:245–257. <https://doi.org/10.1007/s10709-008-9308-0>
- González-Diéguéz 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>
- Gore MA, Chia J-M, Elshire RJ, et al (2009) A First-Generation Haplotype Map of Maize. *Science* 326:1115–1117. <https://doi.org/10.1126/science.1177837>
- Gorjanc G, Gaynor RC, Hickey JM (2018) Optimal cross selection for long-term genetic gain in two-part programs with rapid recurrent genomic selection. *Theor Appl Genet* 131:1953–1966. <https://doi.org/10.1007/s00122-018-3125-3>
- Hallauer AR, Carena MJ, Filho JBM (2010) *Quantitative Genetics in Maize Breeding*. Springer Science & Business Media
- Heslot N, Jannink J-L, Sorrells ME (2015) Perspectives for Genomic Selection Applications and Research in Plants. *Crop Science* 55:1–12. <https://doi.org/10.2135/cropsci2014.03.0249>
- Hickey LT, Germán SE, Pereyra SA, et al (2017) Speed breeding for multiple disease resistance in barley. *Euphytica* 213:64. <https://doi.org/10.1007/s10681-016-1803-2>
- 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>
- Jighly A, Lin Z, Pembleton LW, et al (2019) Boosting Genetic Gain in Allogamous Crops via Speed Breeding and Genomic Selection. *Frontiers in Plant Science* 10:
- 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>
- Krchov L-M, Bernardo R (2015) Relative Efficiency of Genomewide Selection for Testcross Performance of Doubled Haploid Lines in a Maize Breeding Program. *Crop Science* 55:2091–2099. <https://doi.org/10.2135/cropsci2015.01.0064>
- Krishnappa G, Savadi S, Tyagi BS, et al (2021) Integrated genomic selection for rapid improvement of crops. *Genomics* 113:1070–1086. <https://doi.org/10.1016/j.ygeno.2021.02.007>
- Laporte F, Charcosset A, Mary-Huard T (2022) Efficient ReML inference in variance component mixed models using a Min-Max algorithm. *PLOS Computational Biology* 18:e1009659. <https://doi.org/10.1371/journal.pcbi.1009659>
- Laporte F, Mary-Huard T (2020) MM4LMM: Inference of Linear Mixed Models Through MM Algorithm
- Legarra A, Gonzalez-Dieguéz DO, Charcosset A, Vitezica ZG (2023) Impact of interpopulation distance on dominance variance and average heterosis in hybrid populations within species. *Genetics* 224:iyad059. <https://doi.org/10.1093/genetics/iyad059>
- Lehermeier C, Krämer N, Bauer E, et al (2014) Usefulness of Multiparental Populations of Maize (*Zea mays* L.) for Genome-Based Prediction. *Genetics* 198:3–16. <https://doi.org/10.1534/genetics.114.161943>
- Li Y, Kadarmideen HN, Dekkers JCM (2008) Selection on multiple QTL with control of gene diversity and inbreeding for long-term benefit. *Journal of Animal Breeding and Genetics* 125:320–329. <https://doi.org/10.1111/j.1439-0388.2007.00717.x>
- Li Y, Kaur S, Pembleton LW, et al (2022) Strategies of preserving genetic diversity while maximizing genetic response from implementing genomic selection in pulse breeding programs. *Theor Appl Genet* 135:1813–1828. <https://doi.org/10.1007/s00122-022-04071-6>



- Lin Z, Cogan NOI, Pembleton LW, et al (2016) Genetic Gain and Inbreeding from Genomic Selection in a Simulated Commercial Breeding Program for Perennial Ryegrass. *The Plant Genome* 9:plant-genome2015.06.0046. <https://doi.org/10.3835/plantgenome2015.06.0046>
- Longin CFH, Mi X, Würschum T (2015) Genomic selection in wheat: optimum allocation of test resources and comparison of breeding strategies for line and hybrid breeding. *Theor Appl Genet* 128:1297–1306. <https://doi.org/10.1007/s00122-015-2505-1>
- Lorenzana RE, Bernardo R (2009) Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. *Theor Appl Genet* 120:151–161. <https://doi.org/10.1007/s00122-009-1166-3>
- Lorenzi A, Bauland C, Mary-Huard T, et al (2022) 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. *Theor Appl Genet*. <https://doi.org/10.1007/s00122-022-04176-y>
- Lorenzi A, Bauland C, Pin S, et al (2023) Portability of genomic predictions trained on sparse factorial designs across two maize silage breeding cycles. In Review
- Maenhout S, De Baets B, Haesaert G (2010) Prediction of maize single-cross hybrid performance: support vector machine regression versus best linear prediction. *Theoretical and applied genetics* 120:415–427
- Massman JM, Gordillo A, Lorenzana RE, Bernardo R (2013) Genomewide predictions from maize single-cross data. *Theor Appl Genet* 126:13–22. <https://doi.org/10.1007/s00122-012-1955-y>
- Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics* 157:1819–1829
- Neyhart JL, Tiede T, Lorenz AJ, Smith KP (2017) Evaluating Methods of Updating Training Data in Long-Term Genomewide Selection. *G3 Genes|Genomes|Genetics* 7:1499–1510. <https://doi.org/10.1534/g3.117.040550>
- Pandey S, Singh A, Parida SK, Prasad M (2022) Combining speed breeding with traditional and genomics-assisted breeding for crop improvement. *Plant Breeding* 141:301–313. <https://doi.org/10.1111/pbr.13012>
- Pook T, Schlather M, Simianer H (2020) MoBPS - Modular Breeding Program Simulator. *G3 Genes|Genomes|Genetics* 10:1915–1918. <https://doi.org/10.1534/g3.120.401193>
- R Core Team (2020) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria
- R2D2 Consortium RC, Fugerey-Scarbel A, Bastien C, et al (2021) Why and How to Switch to Genomic Selection: Lessons From Plant and Animal Breeding Experience. *Frontiers in Genetics* 12:
- Rio S, Charcosset A, Mary-Huard T, et al (2022) 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
- Sanchez D, Sadoun SB, Mary-Huard T, et al (2023) Improving the use of plant genetic resources to sustain breeding programs' efficiency. *Proceedings of the National Academy of Sciences* 120:e2205780119. <https://doi.org/10.1073/pnas.2205780119>
- 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>
- Shull GH (1908) The Composition of a Field of Maize. *Journal of Heredity* os-4:296–301. <https://doi.org/10.1093/jhered/os-4.1.296>

## Chapter 3

- Sleper JA, Sweet PK, Mukherjee S, et al (2020) Genomewide selection utilizing historic datasets improves early stage selection accuracy and selection stability. *Crop Science* 60:772–778. <https://doi.org/10.1002/csc2.20017>
- 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>
- Tessema BB, Liu H, Sørensen AC, et al (2020) Strategies Using Genomic Selection to Increase Genetic Gain in Breeding Programs for Wheat. *Frontiers in Genetics* 11:
- VanRaden PM (2008) Efficient Methods to Compute Genomic Predictions. *Journal of Dairy Science* 91:4414–4423. <https://doi.org/10.3168/jds.2007-0980>
- Viana JMS, Pereira HD, Piepho H-P, e Silva FF (2019) Efficiency of Genomic Prediction of Nonassessed Testcrosses. *Crop Science* 59:2020–2027. <https://doi.org/10.2135/cropsci2019.02.0118>
- 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>
- Vivek B s., Krishna GK, Vengadessan V, et al (2017) Use of Genomic Estimated Breeding Values Results in Rapid Genetic Gains for Drought Tolerance in Maize. *The Plant Genome* 10:plantgenome2016.07.0070. <https://doi.org/10.3835/plantgenome2016.07.0070>
- Voss-Fels KP, Cooper M, Hayes BJ (2019a) Accelerating crop genetic gains with genomic selection. *Theor Appl Genet* 132:669–686. <https://doi.org/10.1007/s00122-018-3270-8>
- Voss-Fels KP, Herzog E, Dreisigacker S, et al (2019b) Chapter 14 - “SpeedGS” to Accelerate Genetic Gain in Spring Wheat. In: Miedaner T, Korzun V (eds) *Applications of Genetic and Genomic Research in Cereals*. Woodhead Publishing, pp 303–327
- Wanga MA, Shimelis H, Mashilo J, Laing MD (2021) Opportunities and challenges of speed breeding: A review. *Plant Breeding* 140:185–194. <https://doi.org/10.1111/pbr.12909>
- Watson A, Ghosh S, Williams MJ, et al (2018) Speed breeding is a powerful tool to accelerate crop research and breeding. *Nature Plants* 4:23–29. <https://doi.org/10.1038/s41477-017-0083-8>
- Xu S, Zhu D, Zhang Q (2014) Predicting hybrid performance in rice using genomic best linear unbiased prediction. *Proceedings of the National Academy of Sciences* 111:12456–12461. <https://doi.org/10.1073/pnas.1413750111>
- Zhang X, Pérez-Rodríguez P, Burgueño J, et al (2017) Rapid Cycling Genomic Selection in a Multiparental Tropical Maize Population. *G3 Genes|Genomes|Genetics* 7:2315–2326. <https://doi.org/10.1534/g3.117.043141>

## Acknowledgments

We thank Lidea, Limagrain Europe, Maisadour Semences, Corteva, RAGT 2n, KWS and Syngenta Seeds grouped in the frame of the ProMais SAM-MCR project for their funding. We are grateful to scientists from these companies and scientists of INRAE for their input regarding the organization of maize breeding programs and their help for defining the economical parameters used in our simulations. We thank the members of the INRAE-CIRAD "R2D2" network for helpful discussions on the implementation of genomic selection in breeding programs. 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).

## Funding

Lidea, Limagrain Europe, Maisadour 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. 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.

# GENERAL DISCUSSION



## **General discussion and Perspectives**

---

Typical maize hybrid breeding programs are based on reciprocal selection in complementary heterotic groups. These schemes involve (i) the creation of biparental families in each group, (ii) the evaluation of candidate lines based on topcross hybrids, and (iii) the identification of the best hybrid combinations between candidate lines selected in each group. This method is convenient, but leads to a lower gain than an approach that would involve testing all possible hybrid combinations between candidate lines right from the start in the hybrid breeding process (Fehr 1987). While it is practically unfeasible to field test all potential hybrids, Genomic Selection (GS) enables predicting their value based on the performances of a set of hybrids used as training set (TRS), holding great promise for revisiting hybrid breeding schemes.

Recent studies proposed to revisit the maize hybrid breeding scheme by replacing tester-based evaluations with genomic predictions calibrated on a factorial design to predict all possible single-cross hybrids at an early stage (Giraud 2016; Kadam et al. 2016). This could enable the identification of the most promising single-cross hybrids early in the breeding process, ultimately reducing the time needed to bring new hybrid varieties to the market. Using a factorial design to calibrate GS models showed promising predictive abilities of line GCAs and hybrid values (Kadam et al. 2016; Seye et al. 2020; Burdo et al. 2021). This PhD aimed to further investigate the use of GS trained on factorial designs and discuss the cost-efficiency of its implementation in maize breeding programs compared to a conventional breeding scheme.

This main objective was addressed in three chapters each focused on a specific question: (i) validate on experimental data the efficiency of factorial TRSs for the prediction of untested single-cross hybrids and compare it with tester-based TRSs, (ii) assess the portability of genomic predictions calibrated on a factorial design across two generations and investigate the optimization of factorial TRS across cycles, and (iii) investigate by simulations the benefit of breeding schemes using genomic predictions calibrated on a factorial design to improve maize hybrid breeding schemes. In the following, the three chapters are discussed and put into perspective regarding the benefits and limitations of implementing GS in hybrid breeding programs.

## **Comparison of predictive abilities obtained with factorial and tester TRSs in early stage evaluations**

In **Chapter 1**, we compared predictive abilities obtained by calibrating GS models on a factorial and two tester designs of the same size (360 hybrids), derived from the same candidate inbred lines and evaluated in the same environments (G0-F4H and G0-T designs). Predictive abilities obtained with the factorial TRS were higher for some traits (plant height, DINAG, DINAGZ) and equivalent for others (dry matter yield, date of female flowering ...), and on average, there was a slight advantage of the factorial over the tester designs. We also compared the efficiency of the factorial and tester TRSs of the first generation (G0) for predicting the new generation (G1) of hybrids (**Chapter 2**). We found the advantage of using factorial compared to tester TRSs was lower when predicting the next generation (G1) than the G0 generation. Recently, another experimental study reported similar results in a multiparent synthetic population of maize (Burdo et al. 2021), showing an advantage of the factorial design for some traits (flowering traits) but not for others (grain yield, plant height...).

Simulations by Seye et al. (2020) showed that the advantage of factorial compared to tester TRSs increased with the SCA variance proportion. The small SCA variances estimated in our experimental factorial and tester designs certainly explain why we only observed a slight advantage of the factorial over the tester designs. Additionally, as in Burdo et al. (2021), we used founder lines of the complementary population as testers. As shown by simulations (Seye et al. 2020), using as tester one of the founder line of the complementary group instead of a non-founder line reduces the advantage of factorial compared to tester designs. Yet, even if the predictive abilities were similar between factorial and tester TRSs, predictions based on the two TRSs did not select the same subset of individuals. Interestingly, for yield (DMY), the predictions calibrated on the factorial design identified a higher proportion of the best-performing hybrids. This result might be specific to our data set, but the factorial TRS was more efficient to identify the best-performing hybrids in advance.

In tester-based designs, the choice of the testers is crucial, as illustrated in **Chapter 1** and **Chapter 2** comparing several tester design compositions to factorial designs. We showed that predictive abilities of tester TRSs varied according to the tester used in each group and that the one-tester TRS that best predicted single-cross hybrids depended on the trait. For a trait showing dominance, the dominant QTL alleles carried by the tester mask part of the genetic

variability available in the line population (Hallauer et al. 2010). This limits the ability of models trained on one tester to predict hybrids obtained with another tester or in a factorial. As QTLs and QTL effects differ between traits, the tester that provided the best GS predictions might vary depending on the trait. One benefit of using a factorial design is getting rid of that variability linked to the choice of the tester.

## **Single-cross hybrid prediction using a factorial TRS**

### **Efficiency of genomic predictions calibrated on sparse factorial TRSs**

In **Chapter 1 and Chapter 2**, we evaluated experimentally the prediction accuracy of factorial TRSs derived from unselected segregating biparental families. We confirmed the efficiency of factorial TRSs to predict single-cross hybrids, as also reported in previous studies (Kadam et al. 2016; Fristche-Neto et al. 2018; Seye et al. 2020, Burdo et al., 2021). We focused our analyses on a sparse factorial design in which one line contributed to one hybrid (G0-F-1H). In the absence of markers, estimating the GCA of lines and SCA of hybrid combinations requires that each line contributes to several hybrids in the factorial design. With molecular markers, since each allelic combination is shared by several hybrids, it is possible to decompose the total hybrid value into GCA and SCA effects for each marker. In **Chapter 1**, we showed that models calibrated with a sparse factorial TRS (G0\_F-1H) could efficiently predict performances in tester designs, which illustrates the ability of the GS models to decouple the GCA effects from each parental line. This aligns with results found by Seye et al. (2020) and Burdo et al. (2021) using sparse factorial TRSs. Compared to the other experimental studies, our predictive abilities were particularly high. The small number of founder lines at the origin of the population could explain this result. We confirmed that population structure, specifically the founder line origin alone, could predict part of the hybrid values.

We confirmed that predictive abilities calibrated on a factorial design (G0\_F-1H) were higher when parental lines of the PS (G0\_F-4H) hybrids contributed to the TRS (T2 and T1 hybrids) than when none of them were represented in the TRS (T0 hybrids) (**Chapter 1**), as also reported in simulations (Technow et al. 2012; Seye et al. 2020) and maize studies (Technow et al. 2014; Kadam et al. 2016). Interestingly, predictive abilities of T0 hybrids in our study were still high, illustrating that a sparse factorial design can efficiently predict hybrids between lines that have never been evaluated. The small differences in predictive abilities observed for T0, T1 or T2



hybrids might be the consequence of our population design. As each line belongs to a large biparental family, even when the parental line of a hybrid does not contribute to the TRS, several of its full-sib lines contribute to the TRS, ensuring a good prediction of the hybrid value. Last, we investigated the ability of GS models calibrated on a sparse factorial TRS (G0\_F-1H) to efficiently predict new hybrid combinations obtained by crossing lines already selected based on their GCAs and showed high predictive abilities (G0\_F-4H) (**Chapter 1**). GS is able to identify in advance the best potential single-crosses between selected inbred lines. This supports the idea of using factorial designs in early stage of the process (as simulated in the **Chapter 3**).

### **Genomic predictions calibrated on a sparse factorial TRS across cycles**

An important objective was to evaluate the portability of genomic predictions calibrated on factorial designs across two breeding cycles (**Chapter 2**). The first factorial design (G0\_F-1H) was used to select the best candidate lines from which the new-generation hybrids (G1) were produced. We confirmed the efficiency of sparse factorial designs (G0-F1H) for predicting line general combining abilities (GCAs) and hybrid values in the new generation (G1). We found that predictions across two breeding cycles without updating the TRS led to a decrease in prediction accuracy, as reported in several 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 decreased prediction accuracy is likely due to changes in allelic frequencies and a decrease in linkage disequilibrium (LD) between markers and QTL due to recombination events (Auinger et al. (2016). To maintain high prediction accuracy over generations, we investigated strategies to update the factorial TRS and optimize its composition, which are discussed below.

### **SCA and its prediction**

The relative proportion of GCA and SCA components is an important factor for organizing hybrid selection schemes (Seye et al. 2020). In principle, factorial designs allow for considering the SCA component (or dominance) of the hybrid value and, compared to tester design, eliminate a potential bias from using a limited number of testers from the complementary group. Across **Chapter 1** and **Chapter 2**, we highlighted a good ability to predict GCA but showed limits with respect to SCA prediction, which is consistent with the small proportion of SCA variance estimated across experimental designs (ranging from 0 to 20% for dry matter yield depending

on the experimental design). This relatively small importance of SCA effects compared to GCA effects is consistent with the fact that no SCA QTL was detected in the sparse factorial design (G0\_F-1H) (Giraud et al. 2017b; Seye et al. 2019). Along breeding cycles, a decrease in the SCA variance is expected as recurrent reciprocal selection increases divergence between groups (Reif et al. 2007; Gerke et al. 2015; Legarra et al. 2023), which reinforces the difficulty to predict SCA for hybrids obtained by crossing lines issued from heterotic groups derived from a reciprocal selection process as it is the case for the flint and dent group that we used in this project.

To improve prediction accuracies in our designs, we compared different prediction models, decomposing the hybrid genetic effect into GCA and SCA components or additive, dominance, and epistasis. We found that including non-additive genetic effects had little to no impact 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). The small SCA variances estimated in our experimental designs were consistent with this result. The high multicollinearity between GCA and SCA kinship matrices makes it challenging to separate their variances (Sweet and Bernardo 2023). Using orthogonal models for GCA and SCA (Vitezica et al. 2017; González-Diéguez et al. 2021) could have reduced the confusion of effects. However, using such models did not increase prediction accuracies, as reported in other studies (González-Diéguez et al. 2021). The estimation of SCA (dominance) variance components that we obtained using GBLUP models were much lower than estimates given directly by field plot data analysis, suggesting SCA (dominance) was maybe not well captured by markers. Including non-additive effects in GS models must be reconsidered in other hybrid contexts when heterotic groups are not well established. The interest of factorial TRSs and GS models including non-additive effects are expected to be higher in this context.

## **Optimization of the factorial TRS**

### **Composition of the factorial TRS**

The composition of the TRS is an important factor to consider in the implementation of GS in breeding programs (Isidro y Sánchez and Akdemir 2021; Rio et al. 2022b). In **Chapter 1 and 2**, we took advantage of the two factorial designs (G0-F1H and G0-F4H) to investigate different factorial design compositions regarding the number of inbred lines and hybrids per line. Results

showed that at a fixed number of hybrids, *there is a benefit to increasing the number of evaluated lines instead of the number of hybrids per line* when predicting T0 hybrids (**Chapter 1**, Seye et al. 2020). The two factorials were being used reciprocally as TRS and PS. Simulation results (Seye et al. 2020) comparing the same number of hybrids factorials composed of one hybrid per line or four hybrids per line and predicting a distinct PS are in accordance with our findings. This was further investigated using simulations (**Chapter 3**) and is discussed below.

### **Updating factorial TRS across cycles**

Large datasets from multiple trials and multiple years of phenotyping are often available for model training in GS and can be used to predict new genotypes with high prediction accuracy (Sleper et al. 2020). We showed the benefit of keeping data from previous generations in the TRS (**Chapter 2**). More interestingly, we observed the benefit of including hybrids between unselected lines from the previous generations evaluated in different environments. This observation aligns with results found by Neyhart et al. (2017) and Brandariz and Bernardo (2018).

The TRS must be regularly updated with newly genotyped and phenotyped individuals to maintain high prediction accuracy over generations (Neyhart et al. 2017; Brandariz and Bernardo 2018; Lopez-Cruz and de los Campos 2021). We considered two options to update the TRS, evaluating hybrids between (i) selected lines of the previous generation in new environments (G0S) (ii) and/or new-generation hybrids (G1). We showed the benefits of including hybrids between selected lines from the previous generation to predict hybrids from the new generation. And even when new-generation hybrids were already in the TRS, including selected hybrids from previous generations was still beneficial. These results suggest that it might be interesting to reevaluate some hybrids between the best-performing lines from the previous generations while evaluating the new-generation hybrids. This provides a better estimation of the breeding value of the parental lines of the new generation and information on their values in additional environmental conditions.

Breeders work with a constrained budget so that, often, only a subset of the candidate individuals of the new generation can be phenotyped. Optimizing the TRS represents an opportunity to better allocate resources within breeding programs by reducing the number of individuals that need to be phenotyped to train GS models while maintaining a similar accuracy (sparse or selective phenotyping) (Akdemir and Isidro-Sánchez 2019). In this context, 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. This question has rarely been addressed in studies investigating TRS optimization for hybrid breeding. In **Chapter 2**, we used the CDmean (Rincent et al. 2012) and extended its application 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. Two different computations of the CDmean were compared, considering or not the marker information from the historical data in the optimization process. Both ways of computing the CDmean generated a gain in predictive ability compared to a randomly sampled TRS. Interestingly, it was the strategy that did not account for the historical information (CDmean1) when optimizing the new-generation hybrids that performed best. The strategy that did not consider the information from the historical data (CDmean1) likely selected representative hybrids of the whole range of new generation hybrids. In contrast, the strategy considering the genotypic information from the historical data (CDmean2) likely maximized the diversity by selecting new-generation hybrids genetically distant from the historical data. The new- and the previous-generation hybrids were not evaluated in the same environment. This could not be accounted for by the CDmean2, which relies only on genotyping information without accounting for phenotyping information and for possible genotype-by-environment interactions. To circumvent this problem, we could have considered each environment as a different trait and used the correlation value between the two environments to compute the CDmean, as Ben-Sadoun et al. (2020) suggested in the context of multi trait predictions. 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 be 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.

To optimize the TRS, we decided to only present the results of the CDmean (**Chapter 2**), but other criteria have been considered. The CDmean presents several limits: (i) requires *a priori* determining the size of the TRS, and (ii) when phenotyping is available, it does not consider

phenotypic observations to optimize the TRS. For example, we considered another approach called “sparse selection index” (SSI) (Lopez-Cruz and de los Campos 2021). The SSI is a model aiming to optimize the TRS for each individual of the PS without requiring the size of the TRS to be known *a priori*. Results showed no increase in prediction accuracy and required long computation time. However, we tested the SSI on one generation only. It could be interesting to use it across generations when using historical data and predicting the next generation.

## Implementing factorial designs in GS breeding schemes

### Combining GS and factorial TRSs to revisit hybrid breeding schemes using simulations

Results from **Chapter 1** and **Chapter 2** based on the experimental designs support the potential of sparse factorial TRSs as a reliable alternative to traditional tester TRSs for predicting hybrid values regarding predictive abilities. Our conclusions are closely related to the population and composition of the experimental designs. It could be interesting to compare factorial and tester TRSs using experimental data in other populations, for example, with a larger genetic base (more founders lines in the initial population) and different SCA levels, to evaluate the advantage of the factorial TRS in this situation. We used simulations to get further insight into the potential of such an approach in a more realistic breeding program. In **Chapter 3**, we simulated a reciprocal recurrent scheme involving multiparental connected populations from two complementary heterotic groups mimicking a maize breeding program. We compared a conventional phenotypic selection BS at fixed costs against GS BSs combined or not with (i) shortening the breeding cycle length by removing phenotypic evaluation steps and (ii) using factorial instead of tester designs to train genomic selection models.

As in numerous studies, we reported an increased genetic gain with GS compared to phenotypic selection in hybrid BSs (Lin et al. 2016; Gaynor et al. 2017; Gorjanc et al. 2018; Voss-Fels et al. 2019a, b), even at fixed costs. Then, we investigated the impact of shortening the BS by suppressing one or two phenotyping steps in “conventional” programs involving testers. We kept remaining trials with a constant size and reallocated spared resources by producing and genotyping more DH lines, thus increasing the selection intensity thanks to predictions. The shorter BSs allowed for a faster short-term genetic gain (more cycles per time unit) but led to a lower genetic gain in the long term (lower genetic gain per cycle). This lower genetic gain was due to lower prediction accuracies, particularly in the first selection step of the candidate DH

lines, that was not compensated by the increased selection intensity. Last, we investigated the benefit of replacing tester designs with a sparse factorial design in a two-step phenotyping BSs (Y2F1H-GS) and also collapsing early- and late-stage phenotyping into a single phenotyping stage relying on a factorial design (Y1F4H-GS and Y1F1H-GS). The one-year BSs (Y1F4H-GS and Y1F1H-GS) generated the highest short-term genetic gain (increased number of cycles per year) and the lowest long-term genetic gain. Thus, considering the actual parameters and resource allocation, we do not recommend using this one-year BSs. On the opposite, using a two-year BS and replacing tester designs with a sparse factorial design (Y2F1H-GS) improved the genetic gain compared to the same BS using tester designs (Y2TC1-GS) and outperformed the conventional GS BS (Y3-GS) in the short term and was equivalent in the long term.

A major drawback of incorporating factorial designs in breeding programs was the additional seed production costs and workforce necessary for producing single-cross hybrids by handmade pollination. Results from **Chapter 3** showed that the additional cost of single-cross seed production (twice the price of topcross seed production based on discussion with Cyril Bauland and private partners) could be efficiently compensated by reducing the number of phenotyping plots and increasing the number of evaluated lines in the factorial TRS (Y2F1H-GS scheme).

The Y2F1H-GS scheme is in line with the initial proposition of the SAMMCR project: using a very sparse factorial design in the first step with a large number of candidate lines and using a factorial between the most promising lines with more hybrid per lines in the second step to better estimate their GCAs. In light of all our results (**Chapter 1**, **Chapter 2** and **Chapter 3**), *it appears that implementing GS and shortening the breeding cycle by replacing the two early-stage topcross evaluations with a unique sparse factorial design (Y2F1H-GS) hold great potential to fasten the breeding process while increasing the genetic gain.*

### **Prospects for extending simulations**

The simulation program developed in this thesis could be used to test other BSs in connected reciprocal multi-parental breeding programs. However, optimizing the algorithm for computer resource (RAM) consumption and computation time is essential, as these two factors constrained the number of BSs we were able to test in **Chapter 3**. In addition to considering other populations and BSs, several improvements could be considered.

One improvement concerns the inclusion of SCA effects in the GS model. Because the computation time was too long, we did not include SCA effects in the GS model. Studies relying on experimental data have shown that including SCA effects in GS models often does not increase or only slightly improve the prediction accuracy of single-cross hybrids (Vitezica et al. 2017; Seye et al. 2020; González-Diéguéz et al. 2021; Lorenzi et al. 2022). In addition, a simulation study considering a scenario with a proportion of SCA over the total genetic variance of 24%, reported that including SCA effects in the model had little effect on prediction accuracies (Seye et al. 2020). In this context, the benefit of using factorial designs could have been similar or slightly improved if SCA effects were included in the GS prediction models.

Secondly, we compared GS BSs at a fixed number of generations in the TRS, resulting in a smaller TRS size for shorter BSs (including one or two phenotyping steps). The small size of the TRS considered for these shorter BSs certainly limited the prediction accuracies (Akdemir and Isidro-Sánchez 2019). It would have been interesting to include data from more than two generations to increase the size of the TRSs and evaluate its impact on prediction accuracies. Additionally, optimization strategies to construct the TRS could be investigated (see Rio et al. 2022 for a review) to maintain high prediction accuracies even with numerous generations in the TRS, such as the ones investigated in **Chapter 2**.

Another possible improvement is the optimization of the resource allocation in the different BSs. To find the optimum resource allocation for a given BS, testing all combinations of parameters (number of lines, number of hybrids, number of trials...) is impossible. To facilitate the optimization of the resource allocation, it would be ideal to move from stochastic simulations to more deterministic simulation approaches as done, for instance, by Lorenz (2013), Riedelsheimer and Melchinger (2013), or Endelman et al. (2014) for the optimization of phenotypic resource allocation in GS programs at a given generation. For GS, a key point is to anticipate the impact of different resource allocations on the prediction accuracy. Deterministic formulae of GS accuracy have been proposed (Daetwyler et al. 2008), but their application depends on parameters that are difficult to know in practice, especially if one wants to anticipate the effect of selection in complex designs along generations.

The balance between the costs and benefits of replacing topcross evaluations by genomic predictions calibrated on a sparse factorial TRS can vary depending on the costs and resources



available in a given breeding program. Additionally, as illustrated in the previous sections, several factors can affect the benefit of factorial compared to tester TRSs, such as the level of SCA variance or the factorial TRS composition. It appears necessary to investigate further the impact of the different parameters of the BS on the benefit of factorial TRS. Using simulations to test several BSs and resource allocations corresponding to the real breeding program of the company could help guide the decision of including factorial designs or not.

## Perspectives to improve GS models for hybrid breeding

### Accounting for genotype by environment interactions (GxE) in GS models

In breeding programs, TRSs often assemble data from multi environmental trials, requiring properly modelling GxE interactions to maximize prediction accuracy (Jighly et al. 2021). GS models that account for GxE interactions have been widely investigated in plant breeding and improved prediction accuracies in most cases (Crossa et al. 2017). Current GS models exploit the genetic correlation among different environments to model GxE effects (Burgueño et al. 2012; Lopez-Cruz et al. 2015; Cuevas et al. 2016, 2017); other consider environmental covariates to improve the prediction accuracy for multi environmental trials (Jarquín et al. 2014; Heslot et al. 2015; He et al. 2019). In the SAMMCR project, each hybrid panel was evaluated in eight environments. Significant GxE variance estimations confirmed the importance of GxE interactions in the phenotypic data of the project. We detected more SCAXE variance than SCA variance for most traits, suggesting that non-additive effects contributed to hybrid adaptation to environmental conditions. It would be interesting to incorporate GxE effects into GS models to improve prediction accuracies.

### Considering multi trait GS models

Although yield is usually the primary trait of interest in most crops, the economic value of a variety always relies on several other traits (Bernardo 2021b). For example, in silage maize, agronomic (DMY, DMC...) and digestibility quality (MFU, DINAG...) traits are improved simultaneously. The correlation between traits can make multi-trait selection challenging. One objective of the SAMMCR project was to carry out a breeding program as a proof of concept of the interest of our approach. Hybrids were selected to combine a good yield potential with a good digestibility. The candidate lines were selected for an index combining yield performance (DMY), dry matter content (DMC), and digestibility (MFU) based on genomic predictions



calibrated on the G0\_F-1H sparse factorial design. The observed genetic gain for DMY was similar to the predicted one, but responses to selection obtained on silage quality traits (MFU) were worse than anticipated. This illustrates the importance of considering multiple trait aspects in GS, which was not done in our case.

An approach to account for multi-trait in GS is to estimate a genomic matrix of covariance between different traits (Schulthess et al. 2016). The application of multi-trait models in GS has been the subject of several studies in maize (dos Santos et al. 2016; Lyra et al. 2017), showing a small gain in prediction accuracy. The challenge of such an approach is to optimize multiple functions simultaneously to find a compromise. Akdemir et al. (2019) proposed a multi-optimized framework for breeders to optimize breeding schemes that improve multiple traits. They showed with simulations that this approach led to higher genetic gain for the different traits than the selection index.

### **Integrating omics data in prediction models, prospects open by phenomic selection**

More recently, studies showed a gain for genomic prediction by including various omics data in the models (Rice and Lipka 2021). Schrag et al. (2018) showed that predicting hybrid performance in maize using models combining genotyping data and transcriptomic data outperformed models only considering one of the two data types. However, studying gene expression generates additional costs that need to be evaluated to assess the economic feasibility of these models and also raises the question of the stability of these expressions in different environments. The choice of organs, stages and conditions for sampling can complicate this approach. Another promising approach in silage maize is the use of NIRS data that are already available in most silage breeding programs and can be seen as a cheap way to access metabolomic information. Rinent et al. (2018) have proposed using a kinship matrix based on NIRS spectra to calibrate prediction equations. This “phenomic” selection approach can be viewed as an alternative or complementary approach to GS and has shown promising predictive abilities in wheat (Robert et al. 2022). This approach will be tested in a hybrid context using populations and data produced in the SAMMCR project” (PhenoMaize project 2022-2023 led by R. Rinent ).

## **Perspectives for hybrid breeding schemes incorporating GS**

### **Sparse tester designs, an interesting compromise between tester and sparse factorials?**

A remaining question not directly addressed in this thesis is the challenges of producing single-cross hybrids that may require restructuring part of breeding programs. Handmade pollination requires larger workforces and can result in a lower success rate than topcross hybrid production. Additionally, single-crosses require more precise management of the precocity of the inbred lines to cross. Producing single-cross hybrids represents new constraints to account for and developing new protocols to optimize the production of single-crosses may be necessary.

Sparse tester designs are an alternative to factorial designs, showing promising predictive abilities for predicting single-cross hybrids and relaxing the constraint of their production (**Chapter 1** and **2**). It corresponds to a tester-based design, which maximizes the number of evaluated lines by crossing more lines, each with a different tester. The sparse tester designs were compared to classic tester designs (where all lines are crossed to all testers) and factorial designs on experimental data in **Chapter 1** and **Chapter 2**. Using sparse tester designs in early testing would benefit from the factorial strategy (maximizing the number of evaluated lines) without requiring major restructuring of the breeding program. We could have investigated a BS using this approach in the simulation work (**Chapter 3**) to assess its cost efficiency compared to the Y2F1H-GS BS using a factorial design in early-stage evaluation.

### **Combining GS with rapid cycling and speed breeding approaches to speed up the breeding process**

The incorporation of GS in breeding programs can vary in terms of both the stages at which it is applied and the extent to which it is utilized. In **Chapter 3**, we chose to have at least part of the candidates phenotyped at each cycle. Some simulation studies went a step further and evaluated the interest of “rapid cycling” approaches in which GS cycles without phenotyping are performed to accelerate breeding (Massman et al. 2013; Beyene et al. 2015; Zhang et al. 2017; Bernardo 2021a). Recently, Bernardo (2021b) showed that a BS using two cycles of GS without updating the TRS (no phenotyping) outperformed a conventional BS at the same costs and cycle time regarding genetic gains.

Combining GS with other modern breeding strategies can enhance its efficiency. Recently, “speed breeding” has been promising for accelerating breeding programs by reducing the generation time (Hickey et al. 2017b; Wanga et al. 2021). Speed breeding corresponds to the methods and protocols aiming at accelerating plant development using rapid generation advance systems such as greenhouses or controlled environment facilities where plants are grown at high density under conditions that promote flowering and production of viable seed (e.g., controlled temperature and constant light exposure). Simulation studies in several species have investigated combining GS and speed breeding to increase the breeding cycles and generated high genetic gains per unit time (Watson et al. 2018; Jighly et al. 2019; Pandey et al. 2022).

### **Diversity management in hybrid GS breeding schemes**

The implementation of GS can accelerate the loss of genetic diversity per time unit, especially when coupled with high selection intensity, as reported in several studies (Gaynor et al. 2017; Tessema et al. 2020; Li et al. 2022) and as reported in **Chapter 3**. This genetic variance loss is due to the rapid fixation of favorable and unfavorable alleles at QTLs (Falconer 1996; Li et al. 2008, 2022; Jannink 2010). To maximize genetic gains in the long term, monitoring genetic diversity is essential (Goddard 2009). In addition to the prediction of the breeding values of candidates, GS can also be used to predict the variance in the next generation and identify the most promising crosses between candidate lines. Several studies have shown that GS could be used to predict the usefulness criterion (UC) of a cross (i.e. the expected performance of the best progeny; Mohammadi et al. 2015; Tiede et al. 2015; Lehermeier et al. 2017) and the contribution of each candidate lines to the next generation after selection (using the UCPC proposed by Allier et al. (2019b) and therefore the level of diversity in the next generation. Optimum contribution selection (OCS) methods can be used to optimize the crossing list of individuals to maximize the genetic gain with a constraint on diversity. These methods have proved efficient in increasing long-term genetic gain, with minor short-term penalties (Clark et al. 2013; Woolliams et al. 2015; Gorjanc et al. 2018 ). Recently, the UCPC was incorporated into the OCS and led to higher long-term genetic gain than OCS alone (Allier et al. 2019a). In real breeding programs, external diversity is regularly introduced, and GS can help optimize introductions by identifying promising donors and the best crosses between these donors and the elite lines of the populations (Allier et al. 2020; Sanchez et al. 2023). It would be interesting to apply such

## *General discussion and Perspectives*

methods to the hybrid context. Incorporating genetic resources into reciprocal breeding schemes raises specific questions, particularly the maintenance of heterotic group differentiations that deserve further research.

## Cited literature

---

- Akdemir D, Beavis W, Fritsche-Neto R, et al (2019) Multi-objective optimized genomic breeding strategies for sustainable food improvement. *Heredity* 122:672–683. <https://doi.org/10.1038/s41437-018-0147-1>
- Akdemir D, Isidro-Sánchez J (2019) Design of training populations for selective phenotyping in genomic prediction. *Sci Rep* 9:1446. <https://doi.org/10.1038/s41598-018-38081-6>
- 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>
- Allier A, Lehermeier C, Charcosset A, et al (2019a) Improving Short- and Long-Term Genetic Gain by Accounting for Within-Family Variance in Optimal Cross-Selection. *Front Genet* 10. <https://doi.org/10.3389/fgene.2019.01006>
- Allier A, Moreau L, Charcosset A, et al (2019b) Usefulness Criterion and Post-selection Parental Contributions in Multi-parental Crosses: Application to Polygenic Trait Introgression. *G3 Genes|Genomes|Genetics* 9:1469–1479. <https://doi.org/10.1534/g3.119.400129>
- Allier A, Teyssède S, Lehermeier C, et al (2020) Genomic prediction with a maize collaborative panel: identification of genetic resources to enrich elite breeding programs. *Theor Appl Genet* 133:201–215. <https://doi.org/10.1007/s00122-019-03451-9>
- Auinger H-J, Schönleben M, Lehermeier C, et al (2016) Model training across multiple breeding cycles significantly improves genomic prediction accuracy in rye (*Secale cereale* L.). *Theor Appl Genet* 129:2043–2053. <https://doi.org/10.1007/s00122-016-2756-5>
- Bardol N, Ventelon M, Mangin B, et al (2013) Combined linkage and linkage disequilibrium QTL mapping in multiple families of maize (*Zea mays* L.) line crosses highlights complementarities between models based on parental haplotype and single locus polymorphism. *Theor Appl Genet* 126:2717–2736. <https://doi.org/10.1007/s00122-013-2167-9>
- Bassi FM, Bentley AR, Charmet G, et al (2016) Breeding schemes for the implementation of genomic selection in wheat (*Triticum* spp.). *Plant Science* 242:23–36. <https://doi.org/10.1016/j.plantsci.2015.08.021>
- Beal W (1880) Report of the Michigan Board of Agriculture
- Beavis WD (1998) QTL analyses: power, precision, and accuracy. *Molecular dissection of complex traits* 1998:145–162
- Ben-Sadoun S, Rincet R, Auzanneau J, et al (2020) Economical optimization of a breeding scheme by selective phenotyping of the calibration set in a multi-trait context: application to bread making quality. *Theor Appl Genet* 133:2197–2212. <https://doi.org/10.1007/s00122-020-03590-4>
- 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>
- Bernardo R (2008) Molecular Markers and Selection for Complex Traits in Plants: Learning from the Last 20 Years. *Crop Science* 48:1649–1664. <https://doi.org/10.2135/cropsci2008.03.0131>
- Bernardo R (2021a) Upgrading a maize breeding program via two-cycle genomewide selection: Same cost, same or less time, and larger gains. *Crop Science* 61:2444–2455. <https://doi.org/10.1002/csc2.20516>
- Bernardo R (2021b) Predictive breeding in maize during the last 90 years. *Crop Science* 61:2872–2881. <https://doi.org/10.1002/csc2.20529>
- Bernardo R, Yu J (2007) Prospects for Genomewide Selection for Quantitative Traits in Maize. *Crop Science* 47:1082–1090. <https://doi.org/10.2135/cropsci2006.11.0690>
- Beyene Y, Gowda M, Pérez-Rodríguez P, et al (2021) Application of Genomic Selection at the Early Stage of Breeding Pipeline in Tropical Maize. *Frontiers in Plant Science* 12:

## Cited literature

- Beyene Y, Semagn K, Mugo S, et al (2015) Genetic Gains in Grain Yield Through Genomic Selection in Eight Bi-parental Maize Populations under Drought Stress. *Crop Science* 55:154–163. <https://doi.org/10.2135/cropsci2014.07.0460>
- Brandariz SP, Bernardo R (2018) Maintaining the Accuracy of Genomewide Predictions when Selection Has Occurred in the Training Population. *Crop Science* 58:1226–1231. <https://doi.org/10.2135/cropsci2017.11.0682>
- 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>
- Burgueño J, de los Campos G, Weigel K, Crossa J (2012) Genomic prediction of breeding values when modeling genotypex environment interaction using pedigree and dense molecular markers. *Crop Science* 52:707–719
- Carena MJ, Hallauer AR, Miranda Filho JB (2010) *Quantitative Genetics in Maize Breeding*. Springer New York, New York, NY
- Chaikam V, Molenaar W, Melchinger AE, Boddupalli PM (2019) Doubled haploid technology for line development in maize: technical advances and prospects. *Theor Appl Genet* 132:3227–3243. <https://doi.org/10.1007/s00122-019-03433-x>
- Clark SA, Kinghorn BP, Hickey JM, van der Werf JH (2013) The effect of genomic information on optimal contribution selection in livestock breeding programs. *Genet Sel Evol* 45:1–8. <https://doi.org/10.1186/1297-9686-45-44>
- Collard BCY, Mackill DJ (2007) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philosophical Transactions of the Royal Society B: Biological Sciences* 363:557–572. <https://doi.org/10.1098/rstb.2007.2170>
- Comstock RE, Robinson H, Harvey PH (1949) A breeding procedure designed to make maximum use of both general and specific combining ability
- Coors JG (1999) Selection Methodology and Heterosis. In: *Genetics and Exploitation of Heterosis in Crops*. John Wiley & Sons, Ltd, pp 225–245
- Crossa J, Pérez-Rodríguez P, Cuevas J, et al (2017) Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. *Trends in Plant Science* 22:961–975. <https://doi.org/10.1016/j.tplants.2017.08.011>
- Crow JF (1998) 90 Years Ago: The Beginning of Hybrid Maize. *Genetics* 148:923–928. <https://doi.org/10.1093/genetics/148.3.923>
- Cuevas J, Crossa J, Montesinos-Lopez OA, et al (2017) Bayesian Genomic Prediction with Genotype x Environment Interaction Kernel Models. *G3-Genes Genomes Genet* 7:41–53. <https://doi.org/10.1534/g3.116.035584>
- Cuevas J, Crossa J, Soberanis V, et al (2016) Genomic Prediction of Genotype x Environment Interaction Kernel Regression Models. *Plant Genome* 9:. <https://doi.org/10.3835/plantgenome2016.03.0024>
- Daetwyler HD, Villanueva B, Woolliams JA (2008) Accuracy of Predicting the Genetic Risk of Disease Using a Genome-Wide Approach. *PLOS ONE* 3:e3395. <https://doi.org/10.1371/journal.pone.0003395>
- Darwin C (1876) *The effects of cross and self fertilisation in the vegetable kingdom*, 2d. ed. 5th impression. J. Murray, London:
- Davis RL (1927) Report of the Plant Breeder. Rep Puerto Rico Agric Exp Stn 14–15
- Desta ZA, Ortiz R (2014) Genomic selection: genome-wide prediction in plant improvement. *Trends in Plant Science* 19:592–601. <https://doi.org/10.1016/j.tplants.2014.05.006>
- Doebley JF, Goodman MM, Stuber CW (1984) Isoenzymatic Variation in Zea (Gramineae). *Systematic Botany* 9:203. <https://doi.org/10.2307/2418824>
- dos Santos JPR, Vasconcellos RC de C, Pires LPM, et al (2016) Inclusion of Dominance Effects in the Multivariate GBLUP Model. *PLoS ONE* 11:e0152045. <https://doi.org/10.1371/journal.pone.0152045>

## Cited literature

- Duvick DN (2001) Biotechnology in the 1930s: the development of hybrid maize. *Nat Rev Genet* 2:69–74. <https://doi.org/10.1038/35047587>
- East EM (1908) Inbreeding in corn. *Rep Conn Agric Exp Stn* 1907:419–428
- Endelman JB, Atlin GN, Beyene Y, et al (2014) Optimal Design of Preliminary Yield Trials with Genome-Wide Markers. *Crop Science* 54:48–59. <https://doi.org/10.2135/cropsci2013.03.0154>
- Falconer DS (1996) Introduction to quantitative genetics. Pearson Education India
- FAOSTAT (2021) FAOSTAT. <https://www.fao.org/faostat/en/#data/QCL/visualize>. Accessed 11 Sep 2023
- Fehr WR (1987) Principles of cultivar development. Volume 1. Theory and technique. Principles of cultivar development Volume 1 Theory and technique
- Fernández-González J, Akdemir D, Isidro y Sánchez J (2023) A comparison of methods for training population optimization in genomic selection. *Theor Appl Genet* 136:30. <https://doi.org/10.1007/s00122-023-04265-6>
- Fisher RA (1919) XV.—The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Earth and Environmental Science Transactions of The Royal Society of Edinburgh* 52:399–433. <https://doi.org/10.1017/S0080456800012163>
- Fristche-Neto R, Akdemir D, Jannink J-L (2018) Accuracy of genomic selection to predict maize single-crosses obtained through different mating designs. *Theor Appl Genet* 131:1153–1162. <https://doi.org/10.1007/s00122-018-3068-8>
- Gallais A (2009) Hétérosis et variétés hybrides en amélioration des plantes. Editions Quae
- Gaynor RC, Gorjanc G, Bentley AR, et al (2017) A Two-Part Strategy for Using Genomic Selection to Develop Inbred Lines. *Crop Science* 57:2372–2386. <https://doi.org/10.2135/cropsci2016.09.0742>
- 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>
- Giraud H, Lehermeier C, Bauer E, et al (2014) Linkage Disequilibrium with Linkage Analysis of Multiline Crosses Reveals Different Multiallelic QTL for Hybrid Performance in the Flint and Dent Heterotic Groups of Maize. *Genetics* 198:1717–1734. <https://doi.org/10.1534/genetics.114.169367>
- Goddard M (2009) Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica* 136:245–257. <https://doi.org/10.1007/s10709-008-9308-0>
- 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>
- Gorjanc G, Gaynor RC, Hickey JM (2018) Optimal cross selection for long-term genetic gain in two-part programs with rapid recurrent genomic selection. *Theor Appl Genet* 131:1953–1966. <https://doi.org/10.1007/s00122-018-3125-3>
- Habier D, Fernando RL, Dekkers JCM (2007) The Impact of Genetic Relationship Information on Genome-Assisted Breeding Values. *Genetics* 177:2389–2397. <https://doi.org/10.1534/genetics.107.081190>
- Hallauer AR, Carena MJ, Filho JBM (2010) Quantitative Genetics in Maize Breeding. Springer Science & Business Media
- Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME (2009) Invited review: Genomic selection in dairy cattle: Progress and challenges. *Journal of Dairy Science* 92:433–443. <https://doi.org/10.3168/jds.2008-1646>



## Cited literature

- He S, Thistlethwaite R, Forrest K, et al (2019) Extension of a haplotype-based genomic prediction model to manage multi-environment wheat data using environmental covariates. *Theor Appl Genet* 132:3143–3154. <https://doi.org/10.1007/s00122-019-03413-1>
- Heffner EL, Sorrells ME, Jannink J-L (2009) Genomic Selection for Crop Improvement. *Crop Science* 49:1–12. <https://doi.org/10.2135/cropsci2008.08.0512>
- Heslot N, Jannink J-L, Sorrells ME (2015) Perspectives for Genomic Selection Applications and Research in Plants. *Crop Science* 55:1–12. <https://doi.org/10.2135/cropsci2014.03.0249>
- Hickey JM, Chiurugwi T, Mackay I, Powell W (2017a) Genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery. *Nat Genet* 49:1297–1303. <https://doi.org/10.1038/ng.3920>
- Hickey LT, Germán SE, Pereyra SA, et al (2017b) Speed breeding for multiple disease resistance in barley. *Euphytica* 213:64. <https://doi.org/10.1007/s10681-016-1803-2>
- 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>
- Hopkins CG (1899) Improvement in the chemical composition of the corn kernel. *Journal of the American Chemical Society* 21:1039–1057
- Hospital F, Moreau L, Lacoudre F, et al (1997) More on the efficiency of marker-assisted selection: *Theor Appl Genet* 95:1181–1189. <https://doi.org/10.1007/s001220050679>
- Howard R, Jarquin D, Crossa J (2022) Overview of Genomic Prediction Genomic 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
- Hull FH (1945) Recurrent Selection for Specific Combining Ability in Corn. *Agronomy Journal* 37:134–145. <https://doi.org/10.2134/agronj1945.00021962003700020006x>
- 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>
- 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>
- Jarquín D, Crossa J, Lacaze X, et al (2014) A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theoretical and applied genetics* 127:595–607
- Jenkins MT (1940) The segregation of genes affecting yield of grain in maize. *Journal of the American Society of Agronomy* 32:55–63
- Jenkins MT (1928) Correlation studies with inbred and crossbred strains of maize. 266
- Jenkins MT, Brunson AM (1932) Methods of Testing Inbred Lines of Maize in Crossbred Combinations. *Agronomy Journal* 24:523–530. <https://doi.org/10.2134/agronj1932.00021962002400070004x>
- Jighly A, Hayden M, Daetwyler H (2021) Integrating genomic selection with a genotype plus genotype x environment (GGE) model improves prediction accuracy and computational efficiency. *Plant, Cell & Environment* 44:3459–3470. <https://doi.org/10.1111/pce.14145>
- Jighly A, Lin Z, Pembleton LW, et al (2019) Boosting Genetic Gain in Allogamous Crops via Speed Breeding and Genomic Selection. *Frontiers in Plant Science* 10:
- Jones DF (1918) The Effect of Inbreeding and Crossbreeding Upon Development. *Proceedings of the National Academy of Sciences* 4:246–250. <https://doi.org/10.1073/pnas.4.8.246>



## Cited literature

- 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>
- Krchov L-M, Bernardo R (2015) Relative Efficiency of Genomewide Selection for Testcross Performance of Doubled Haploid Lines in a Maize Breeding Program. *Crop Science* 55:2091–2099. <https://doi.org/10.2135/cropsci2015.01.0064>
- Krishnappa G, Savadi S, Tyagi BS, et al (2021) Integrated genomic selection for rapid improvement of crops. *Genomics* 113:1070–1086. <https://doi.org/10.1016/j.ygeno.2021.02.007>
- Laloë D (1993) Precision and information in linear models of genetic evaluation. *Genetics Selection Evolution* 25:557. <https://doi.org/10.1186/1297-9686-25-6-557>
- Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* 124:743
- Lander ES, Botstein D (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121:185
- Legarra A, Gonzalez-Dieguez DO, Charcosset A, Vitezica ZG (2023) Impact of interpopulation distance on dominance variance and average heterosis in hybrid populations within species. *Genetics* 224:iyad059. <https://doi.org/10.1093/genetics/iyad059>
- Lehermeier C, Krämer N, Bauer E, et al (2014) Usefulness of Multiparental Populations of Maize (*Zea mays* L.) for Genome-Based Prediction. *Genetics* 198:3–16. <https://doi.org/10.1534/genetics.114.161943>
- Lehermeier C, Teyssède S, Schön C-C (2017) Genetic Gain Increases by Applying the Usefulness Criterion with Improved Variance Prediction in Selection of Crosses. *Genetics* 207:1651–1661. <https://doi.org/10.1534/genetics.117.300403>
- Li Y, Kadarmideen HN, Dekkers JCM (2008) Selection on multiple QTL with control of gene diversity and inbreeding for long-term benefit. *Journal of Animal Breeding and Genetics* 125:320–329. <https://doi.org/10.1111/j.1439-0388.2007.00717.x>
- Li Y, Kaur S, Pembleton LW, et al (2022) Strategies of preserving genetic diversity while maximizing genetic response from implementing genomic selection in pulse breeding programs. *Theor Appl Genet* 135:1813–1828. <https://doi.org/10.1007/s00122-022-04071-6>
- Lin Z, Cogan NOI, Pembleton LW, et al (2016) Genetic Gain and Inbreeding from Genomic Selection in a Simulated Commercial Breeding Program for Perennial Ryegrass. *The Plant Genome* 9:plant-genome2015.06.0046. <https://doi.org/10.3835/plantgenome2015.06.0046>
- Longin CFH, Mi X, Würschum T (2015) Genomic selection in wheat: optimum allocation of test resources and comparison of breeding strategies for line and hybrid breeding. *Theor Appl Genet* 128:1297–1306. <https://doi.org/10.1007/s00122-015-2505-1>
- Lopez-Cruz M, Crossa J, Bonnett D, et al (2015) Increased prediction accuracy in wheat breeding trials using a marker × environment interaction genomic selection model. *G3: Genes, Genomes, Genetics* 5:569–582
- Lopez-Cruz M, de los Campos G (2021) Optimal breeding-value prediction using a sparse selection index. *Genetics*. <https://doi.org/10.1093/genetics/iyab030>
- Lorenz AJ (2013) Resource Allocation for Maximizing Prediction Accuracy and Genetic Gain of Genomic Selection in Plant Breeding: A Simulation Experiment. *G3 Genes|Genomes|Genetics* 3:481–491. <https://doi.org/10.1534/g3.112.004911>

## Cited literature

- Lorenz AJ, Smith KP (2015) Adding Genetically Distant Individuals to Training Populations Reduces Genomic Prediction Accuracy in Barley. *Crop Science* 55:2657–2667. <https://doi.org/10.2135/cropsci2014.12.0827>
- Lorenzana RE, Bernardo R (2009) Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. *Theor Appl Genet* 120:151–161. <https://doi.org/10.1007/s00122-009-1166-3>
- Lorenzi A, Bauland C, Mary-Huard T, et al (2022) 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. *Theor Appl Genet*. <https://doi.org/10.1007/s00122-022-04176-y>
- Lush J (1937) *Animal Breeding Plans* (Ames, Iowa, USA: Iowa State College Press)
- Lyra DH, de Freitas Mendonça L, Galli G, et al (2017) Multi-trait genomic prediction for nitrogen response indices in tropical maize hybrids. *Molecular breeding* 37:80
- Maenhout S, De Baets B, Haesaert G (2010) Prediction of maize single-cross hybrid performance: support vector machine regression versus best linear prediction. *Theoretical and applied genetics* 120:415–427
- Massman JM, Gordillo A, Lorenzana RE, Bernardo R (2013) Genomewide predictions from maize single-cross data. *Theor Appl Genet* 126:13–22. <https://doi.org/10.1007/s00122-012-1955-y>
- Matsuoka Y, Vigouroux Y, Goodman MM, et al (2002) A single domestication for maize shown by multi-locus microsatellite genotyping. *Proceedings of the National Academy of Sciences* 99:6080–6084. <https://doi.org/10.1073/pnas.052125199>
- Melchinger AE, Gumber RK (1998) Overview of Heterosis and Heterotic Groups in Agronomic Crops. In: *Concepts and Breeding of Heterosis in Crop Plants*. John Wiley & Sons, Ltd, pp 29–44
- Merca NC, Rusu T, Merca I, Ona AD (2021) AGROECOLOGY: A REAL OPPORTUNITY TO FIGHT AGAINST THE CLIMATE CHALLENGES. 21:
- Merrick LF, Carter AH (2021) Comparison of genomic selection models for exploring predictive ability of complex traits in breeding programs. *The Plant Genome* 14:e20158. <https://doi.org/10.1002/tpg2.20158>
- Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics* 157:1819–1829
- Michel S, Ametz C, Gungor H, et al (2016) Genomic selection across multiple breeding cycles in applied bread wheat breeding. *Theor Appl Genet* 129:1179–1189. <https://doi.org/10.1007/s00122-016-2694-2>
- Mohammadi M, Tiede T, Smith KP (2015) PopVar: A Genome-Wide Procedure for Predicting Genetic Variance and Correlated Response in Biparental Breeding Populations. *Crop Science* 55:2068–2077. <https://doi.org/10.2135/cropsci2015.01.0030>
- Momen M, Morota G (2018) Quantifying genomic connectedness and prediction accuracy from additive and non-additive gene actions. *Genetics Selection Evolution* 50:45. <https://doi.org/10.1186/s12711-018-0415-9>
- Moreau L, Charcosset A, Gallais A (2004) Experimental evaluation of several cycles of marker-assisted selection in maize. *Euphytica* 137:111–118. <https://doi.org/10.1023/B:EUPH.0000040508.01402.21>
- Moreau L, Charcosset A, Hospital F, Gallais A (1998) Marker-Assisted Selection Efficiency in Populations of Finite Size. *Genetics* 148:1353
- Morrow GE, Gardner FD (1893) Field experiments with corn, 1892. *Bulletin (University of Illinois (Urbana-Champaign campus) Agricultural Experiment Station); no 25*
- Neimann-Sorensen A, Robertson A (1961) The Association between Blood Groups and Several Production Characteristics in Three Danish Cattle Breeds. *Acta Agriculturae Scandinavica* 11:163–196. <https://doi.org/10.1080/00015126109433054>

## Cited literature

- Neyhart JL, Tiede T, Lorenz AJ, Smith KP (2017) Evaluating Methods of Updating Training Data in Long-Term Genomewide Selection. *G3 Genes|Genomes|Genetics* 7:1499–1510. <https://doi.org/10.1534/g3.117.040550>
- Pandey S, Singh A, Parida SK, Prasad M (2022) Combining speed breeding with traditional and genomics-assisted breeding for crop improvement. *Plant Breeding* 141:301–313. <https://doi.org/10.1111/pbr.13012>
- Piaskowski J, Hardner C, Cai L, et al (2018) Genomic heritability estimates in sweet cherry reveal non-additive genetic variance is relevant for industry-prioritized traits. *BMC Genet* 19:23. <https://doi.org/10.1186/s12863-018-0609-8>
- Pszczola M, Calus MPL (2016) Updating the reference population to achieve constant genomic prediction reliability across generations. *animal* 10:1018–1024. <https://doi.org/10.1017/S1751731115002785>
- Pszczola M, Strabel T, Mulder HA, Calus MPL (2012) Reliability of direct genomic values for animals with different relationships within and to the reference population. *Journal of Dairy Science* 95:389–400. <https://doi.org/10.3168/jds.2011-4338>
- R2D2 Consortium RC, Fugerey-Scarbel A, Bastien C, et al (2021) Why and How to Switch to Genomic Selection: Lessons From Plant and Animal Breeding Experience. *Frontiers in Genetics* 12:
- Rebai A, Blanchard P, Perret D, Vincourt P (1997) Mapping quantitative trait loci controlling silking date in a diallel cross among four lines of maize: *Theor Appl Genet* 95:451–459. <https://doi.org/10.1007/s001220050582>
- Reif JC, Gumpert F-M, Fischer S, Melchinger AE (2007) Impact of Interpopulation Divergence on Additive and Dominance Variance in Hybrid Populations. *Genetics* 176:1931. <https://doi.org/10.1534/genetics.107.074146>
- Rice BR, Lipka AE (2021) Diversifying maize genomic selection models. *Mol Breeding* 41:33. <https://doi.org/10.1007/s11032-021-01221-4>
- Riedelsheimer C, Endelman JB, Stange M, et al (2013) Genomic Predictability of Interconnected Biparental Maize Populations. *Genetics* 194:493–503. <https://doi.org/10.1534/genetics.113.150227>
- Riedelsheimer C, Melchinger AE (2013) Optimizing the allocation of resources for genomic selection in one breeding cycle. *Theor Appl Genet* 126:2835–2848. <https://doi.org/10.1007/s00122-013-2175-9>
- Rincent R, Charcosset A, Moreau L (2017) Predicting genomic selection efficiency to optimize calibration set and to assess prediction accuracy in highly structured populations. *Theor Appl Genet* 130:2231–2247. <https://doi.org/10.1007/s00122-017-2956-7>
- Rincent R, Charpentier J-P, Faivre-Rampant P, et al (2018) Phenomic Selection Is a Low-Cost and High-Throughput Method Based on Indirect Predictions: Proof of Concept on Wheat and Poplar. *G3 Genes|Genomes|Genetics* 8:3961–3972. <https://doi.org/10.1534/g3.118.200760>
- 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
- Ritchie H, Rodés-Guirao L, Mathieu E, et al (2023) Population Growth. *Our World in Data*
- Robert P, Brault C, Rincent R, Segura V (2022) Phenomic Selection: A New and Efficient Alternative to Genomic Selection. *Genomic selection (GS)*. In: Ahmadi N, Bartholomé J (eds) *Genomic Prediction of Complex Traits: Methods and Protocols*. Springer US, New York, NY, pp 397–420

## Cited literature

- 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>
- Sanchez D, Sadoun SB, Mary-Huard T, et al (2023) Improving the use of plant genetic resources to sustain breeding programs' efficiency. *Proceedings of the National Academy of Sciences* 120:e2205780119. <https://doi.org/10.1073/pnas.2205780119>
- 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
- Schulthess AW, Wang Y, Miedaner T, et al (2016) Multiple-trait-and selection indices-genomic predictions for grain yield and protein content in rye for feeding purposes. *Theoretical and Applied Genetics* 129:273–287
- 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>
- Shull GH (1908) The Composition of a Field of Maize. *Journal of Heredity* os-4:296–301. <https://doi.org/10.1093/jhered/os-4.1.296>
- Shull GH (1914) Duplicate genes for capsule-form in *Bursa bursa-pastoris*. 53
- Shull GH (1909) A pure-line method in corn breeding. *Journal of Heredity* 51–58
- Sleper JA, Sweet PK, Mukherjee S, et al (2020) Genomewide selection utilizing historic datasets improves early stage selection accuracy and selection stability. *Crop Science* 60:772–778. <https://doi.org/10.1002/csc2.20017>
- Smith OS (1986) Covariance between Line per se and Testcross Performance1. *Crop Science* 26:540. <https://doi.org/10.2135/cropsci1986.0011183X002600030023x>
- Soller M (1994) Marker assisted selection - an overview. *Animal Biotechnology* 5:193–207. <https://doi.org/10.1080/10495399409525821>
- Sprague G (1939) An Estimation of the Number of Top-Crossed Plants Required for Adequate Representation of a Corn Variety 1. *Agronomy Journal* 31:11–16
- 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>
- Su G, Christensen OF, Ostersen T, et al (2012) Estimating Additive and Non-Additive Genetic Variances and Predicting Genetic Merits Using Genome-Wide Dense Single Nucleotide Polymorphism Markers. *PLoS ONE* 7:e45293. <https://doi.org/10.1371/journal.pone.0045293>
- Sweet PK, Bernardo R (2023) Reciprocal testcross design for genome-wide prediction of maize single-cross performance. *Theor Appl Genet* 136:184. <https://doi.org/10.1007/s00122-023-04435-6>
- 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>

## Cited literature

- 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>
- Tenaillon MI, Charcosset A (2011) A European perspective on maize history. *Comptes Rendus Biologies* 334:221–228. <https://doi.org/10.1016/j.crvi.2010.12.015>
- Tessema BB, Liu H, Sørensen AC, et al (2020) Strategies Using Genomic Selection to Increase Genetic Gain in Breeding Programs for Wheat. *Frontiers in Genetics* 11:
- Tiede T, Kumar L, Mohammadi M, Smith KP (2015) Predicting genetic variance in bi-parental breeding populations is more accurate when explicitly modeling the segregation of informative genomewide markers. *Mol Breeding* 35:199. <https://doi.org/10.1007/s11032-015-0390-6>
- Troyer AF (1999) Background of U.S. Hybrid Corn. *Crop Science* 39:601–626. <https://doi.org/10.2135/cropsci1999.0011183X003900020001x>
- USDA (2021) United States Department of Agriculture, Economic Research Service, Feed Grains Yearbook. <https://www.ers.usda.gov/topics/crops/corn-and-other-feed-grains/feed-grains-sector-at-a-glance/>. Accessed 12 Sep 2023
- 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, Elsen J-M, et al (2016) Genomic BLUP including additive and dominant variation in purebreds and F1 crossbreds, with an application in pigs. *Genet Sel Evol* 48:6. <https://doi.org/10.1186/s12711-016-0185-1>
- 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>
- Voss-Fels KP, Cooper M, Hayes BJ (2019a) Accelerating crop genetic gains with genomic selection. *Theor Appl Genet* 132:669–686. <https://doi.org/10.1007/s00122-018-3270-8>
- Voss-Fels KP, Herzog E, Dreisigacker S, et al (2019b) Chapter 14 - “SpeedGS” to Accelerate Genetic Gain in Spring Wheat. In: Miedaner T, Korzun V (eds) *Applications of Genetic and Genomic Research in Cereals*. Woodhead Publishing, pp 303–327
- Wang H, Misztal I, Aguilar I, et al (2012) Genome-wide association mapping including phenotypes from relatives without genotypes. *Genetics Research* 94:73–83. <https://doi.org/10.1017/S0016672312000274>
- 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>
- Wanga MA, Shimelis H, Mashilo J, Laing MD (2021) Opportunities and challenges of speed breeding: A review. *Plant Breeding* 140:185–194. <https://doi.org/10.1111/pbr.12909>
- Watson A, Ghosh S, Williams MJ, et al (2018) Speed breeding is a powerful tool to accelerate crop research and breeding. *Nature Plants* 4:23–29. <https://doi.org/10.1038/s41477-017-0083-8>
- Whittaker JC, Thompson R, Denham MC (2000) Marker-assisted selection using ridge regression. *Genetics Research* 75:249–252. <https://doi.org/10.1017/S0016672399004462>

## Cited literature

- Wientjes YCJ, Veerkamp RF, Calus MPL (2013) The Effect of Linkage Disequilibrium and Family Relationships on the Reliability of Genomic Prediction. *Genetics* 193:621–631. <https://doi.org/10.1534/genetics.112.146290>
- Windhausen VS, Atlin GN, Hickey JM, et al (2012) Effectiveness of Genomic Prediction of Maize Hybrid Performance in Different Breeding Populations and Environments. *G3 Genes|Genomes|Genetics* 2:1427–1436. <https://doi.org/10.1534/g3.112.003699>
- Wong CK, Bernardo R (2008) Genomewide selection in oil palm: increasing selection gain per unit time and cost with small populations. *Theor Appl Genet* 116:815–824. <https://doi.org/10.1007/s00122-008-0715-5>
- Woolliams J a., Berg P, Dagnachew B s., Meuwissen T h. e. (2015) Genetic contributions and their optimization. *Journal of Animal Breeding and Genetics* 132:89–99. <https://doi.org/10.1111/jbg.12148>
- Xu S, Zhu D, Zhang Q (2014) Predicting hybrid performance in rice using genomic best linear unbiased prediction. *Proceedings of the National Academy of Sciences* 111:12456–12461. <https://doi.org/10.1073/pnas.1413750111>
- Yu J, Pressoir G, Briggs WH, et al (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet* 38:203–208. <https://doi.org/10.1038/ng1702>
- Zhang X, Pérez-Rodríguez P, Burgueño J, et al (2017) Rapid Cycling Genomic Selection in a Multiparental Tropical Maize Population. *G3 Genes|Genomes|Genetics* 7:2315–2326. <https://doi.org/10.1534/g3.117.043141>
- 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>





# **~ SUPPLEMENTARY MATERIALS ~**





## Supplementary materials. SAMMCR project and Plant Material

---

In this chapter, we introduce the “**SAMMCR**” project and present the original plant material produced within this framework. This chapter gives the reader an overview of the plant material used in this Ph.D.

### The SAMMCR project

This thesis relies on an original plant material that was produced in the framework of the **SAMMCR** project (“*Sélection Assistée par Marqueur Multi-parentale Connectée Réciproque*”). The project started in 2010 and has been supported by 7 private partners members of the ProMaïs association: Syngenta, RAGT 2n, Maisadour, Limagrain, Lidea, KWS, and Corteva. The project aims to study and implement alternative hybrid breeding methods using molecular information in a reciprocal multiparental design, validate their interest in a pilot breeding program on silage maize, and use simulations to complement experimental data. The SAMMCR project encompasses three successive phases, each with different objectives.

**SAMMCR1** aimed to develop segregating populations in two complementary heterotic groups and evaluate a factorial design between them with the double objective of (i) acquiring knowledge on the genetic determinism of GCA and SCA components for silage and agronomic traits and (ii) evaluating the efficiency of marker-assisted selection and specifically genomic selection calibrated on this sparse factorial design. The segregating populations of this phase are referred to as the G0 generation. The analysis of the experimental data of the SAMMCR1 campaign was done in the framework of two previous PhD thesis (Giraud 2016; Seye 2019) who performed QTL detection and performed the first genomic predictions. From these analyses, the best individuals in each group were selected.

**SAMMCR2** objectives were to (i) assess the efficiency of factorial designs to train GS models to predict line GCAs and hybrids compared to tester-based training sets and (ii) to create a new generation of inbred lines (called G1) from the lines selected based on genomic predictions calibrated on a factorial design and test their hybrid performances. Some preliminary analyses of the data obtained in the SAMMCR2 phase by comparing the tester and factorial designs were performed during Adama Seye’s PhD thesis (Seye 2019) and continued during the

present study (Lorenzi et al. 2022, 2023).

**SAMMCR3** objectives were to evaluate the portability of genomic predictions trained on a factorial design across two breeding cycles (G0 and G1) using the hybrid performances of the new generation evaluated in the SAMMCR2 phase and used simulations to propose an alternative breeding scheme taking advantage of GS and factorial designs (Lorenzi et al. 2023).

The main objective of this thesis was to evaluate the benefit of replacing tester-based evaluations with genomic predictions trained on a factorial design to predict line GCAs of all candidate lines and hybrid values of all single-crosses in the early stages of the selection process. To achieve this objective, the data from the three experimental campaigns was analyzed jointly.

## Plant material

The plant material produced in the project is a reciprocal recurrent breeding experiment aiming at improving the silage performance of maize single cross hybrids produced between the dent and flint heterotic groups (used for silage maize in Northern Europe). Two breeding cycles were produced, **G0** and **G1**, comprising several experimental designs described in detail in the following and in **Fig. 1**, **Fig. 2** and **Table 1**.

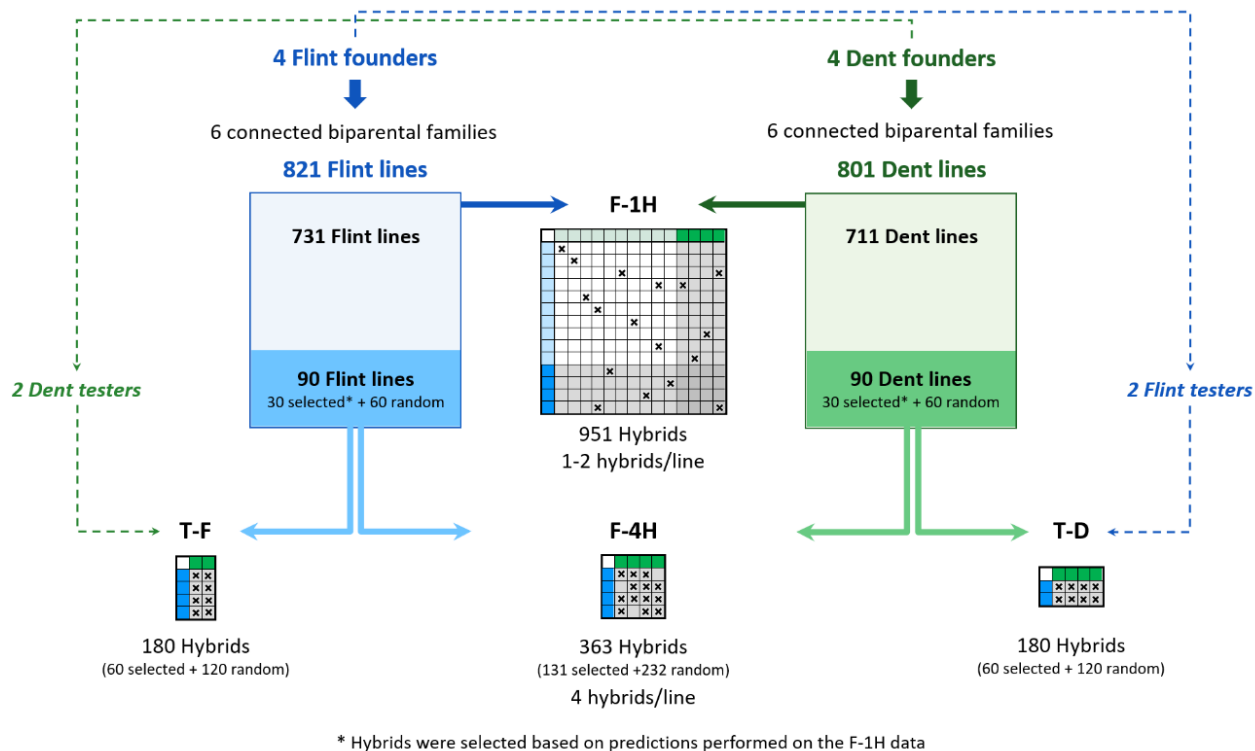
**Table 1** Description of all experimental designs used in this study.

Years of phenotyping	Breeding cycle	Design	Name	Hybrids within the design <sup>a</sup>	References <sup>c</sup>
2013, 2014	G0	Factorial	G0_F-1H	G0R <sup>a</sup>	Giraud et al. 2017a, b; Seye et al. 2019; Lorenzi et al. 2022
2015, 2016	G0	Factorial	G0_F-4H	G0R + G0S <sup>b</sup>	(Seye 2019; Lorenzi et al. 2022, 2023)Seye 2019; Lorenzi et al. 2022, 2023
		Tester	G0_T		
2019, 2020	G0+G1	Factorial	(G0S+G1)_F-1H	G0S + G1	Lorenzi et al. 2023

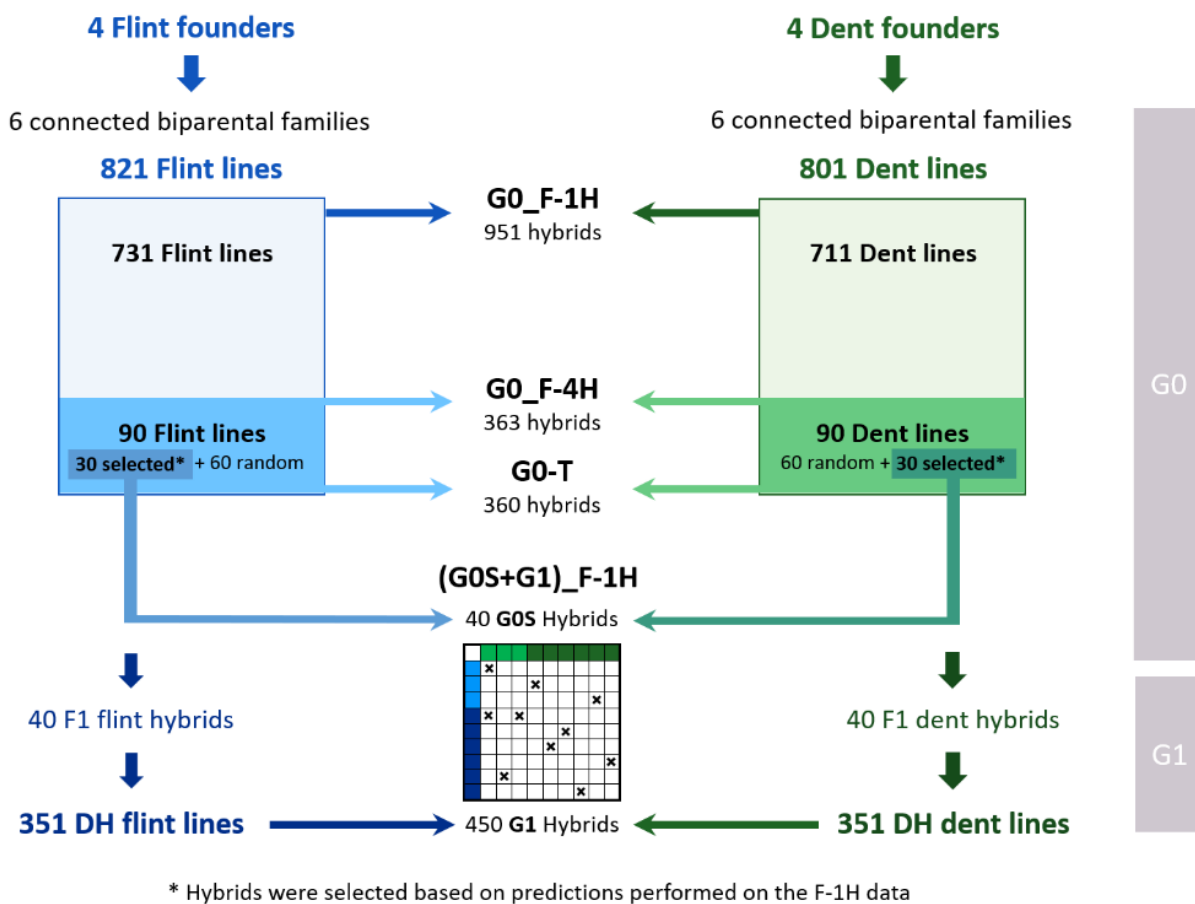
<sup>a</sup> G0R hybrids were produced by crossing two random lines from the G0 cycle

<sup>b</sup> G0S hybrids were produced by crossing two selected lines from the G0 cycle

<sup>c</sup> Reference of the publication corresponding to the analysis of the corresponding design.



**Fig.1** Experimental designs of the G0 generation



**Fig.2** Experimental designs of the G1 generation

### Production of the G0 parental lines

Two multiparental connected families were produced, each corresponding to one of the major heterotic groups used for maize silage in Northern Europe: the flint and the dent. In each heterotic group, four founder lines were chosen: three (F373, F03802 and F02803 for the flint group and F98902, F1808 and F04401 for the dent group) for their agronomical performances for silage production (dry matter biomass yield and dry matter content) and one (F7088 for the flint group and F7082 for the dent group) for its silage (digestibility) quality. The four founder lines were intercrossed to produce six single-cross F1 hybrids from which six biparental connected families were generated. In total, 801 G0 dent lines were obtained by double haploidization and 822 G0 flint lines were produced by five to six generations of selfing using a single-seed descend (SSD) process. The crosses at the origin of the six biparental families and the number of G0 lines in each family is given in **Table 2** for the flints and **Table 3** for the dents.

**Table 2** Half-diallel design between the four flint founder lines at the origin of the six flint biparental families (F1, F2, F3, F4, F5 and F6) and, between brackets, the corresponding number of flint G0 SSD lines produced in each family.

	<b>F373</b>	<b>F03802</b>	<b>F02803</b>	<b>F7088</b>
<b>F373</b>		F2 (139)	F1 (162)	F4 (154)
<b>F03802</b>			F3 (111)	F6 (138)
<b>F02803</b>				F5 (118)
<b>F7088</b>				

The green founder line corresponds to the one chosen for its digestibility quality

**Table 3** Half-diallel design between the dent founder lines at the origin of the six dent biparental families (D1, D2, D3, D4, D5 and D6) and, between brackets, the corresponding number of dent G0 DH lines produced in each family.

	<b>F1808</b>	<b>F98902</b>	<b>F04401</b>	<b>F7082</b>
<b>F1808</b>		D2 (144)	D1 (139)	D5 (138)
<b>F98902</b>			D3 (140)	D6 (99)
<b>F04401</b>				D4 (142)
<b>F7082</b>				

The green founder line corresponds to the one chosen for its digestibility quality

### The G0\_F-1H experimental design

The 822 flint and 801 dent parental lines (**G0**) thus obtained were crossed following a very incomplete factorial to produce 951 hybrids, further referred to as **G0\_F-1H** (also referred to as F-1H in Chapter 2, Lorenzi et al. 2022). The originality of this factorial design lies in its composition, where most parental lines contributed to producing only one hybrid (on average 1.2 hybrids per line). This factorial design is balanced between families where 22 to 35 hybrids were produced from each biparental family combination (**Table 4**). The hybrids of this design were evaluated in eight trials in 2013 and 2014 for 11 traits (four agronomic traits and seven silage quality traits). The data from the G0\_F-1H was first analyzed during Heloise Giraud’s thesis (Giraud 2016) for agronomical traits and during Adama Seye’s thesis for the silage quality traits (Seye 2019). QTL(s) involved in GCA and SCA were detected for agronomic traits (Giraud et al. 2017c, d) and silage quality traits (Seye et al. 2019). The first genomic predictions were obtained with a “simple” genomic prediction model (Giraud, 2016; Seye, 2019) and were validated by cross-validations. Good qualities of prediction were obtained for GCA components but not for SCA components. In this thesis, the G0\_F-1H allowed us assessing using cross-validations the efficiency of using a sparse factorial composed of one hybrid per line to train for genomic predictions. Questions remained about the efficiency of such a design compared to tester designs along breeding cycles and the optimization of the composition of the factorial design (number of lines and number of hybrids per line). The advantage of using a factorial design instead of a tester design as a training set for genomic predictions was assessed by simulations (Seye et al. 2020). However, this advantage had to be confirmed experimentally.

**Table 4** Number of hybrids per biparental family in the G0\_F-1H design

	<b>F1</b>	<b>F2</b>	<b>F3</b>	<b>F4</b>	<b>F5</b>	<b>F6</b>
<b>D1</b>	31	27	22	30	25	27
<b>D2</b>	34	27	22	30	23	32
<b>D3</b>	28	29	23	30	26	27
<b>D4</b>	28	28	25	30	27	27
<b>D5</b>	33	29	23	28	24	30
<b>D6</b>	24	23	15	25	21	18

### The G0\_F-4H and tester experimental designs

In each heterotic group, 30 lines were selected among the G0 lines (G0S lines) based on genomic prediction obtained in the G0\_F-1H for an index combining silage yield (DMY), moisture content at harvest (DMC) and silage quality (MFU). This index corresponds to the one used for silage hybrid registration in France (Seye 2019). Note that in each heterotic group, only three families out of six were represented in the selected lines (**Table 5**). Additionally, in each heterotic group, 60 lines were chosen at random in a balanced manner, with ten lines per family (G0R lines) (**Table 5**). The **G0\_F-4H** factorial design (also referred to as F-4H in Chapter 2, Lorenzi et al. 2022) comprises 363 new single-cross hybrids produced by crossing randomly (i) the 60 dent G0R lines to the 60 flint G0R lines to produce 236 random hybrids and, (ii) the 30 dent G0S lines to the 30 flint G0S lines to produce 127 selected hybrids (**Table 6**). On average, each parental line contributed to produce four hybrids (F4H). Note that the G0\_F-1H and the G0\_F-4H designs were issued from the same inbred line populations (G0) 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 G0\_F-4H than in the G0\_F-1H. One reason for increasing the number of hybrids per line, compared to the G0\_F-1H, was to better estimate the GCA and SCA components of each hybrid performance. In parallel, two tester designs were produced by crossing the same 90 G0 lines to two founder lines from the complementary heterotic group. In each group, the 30 G0S and the 60 G0R lines were crossed to two testers from the complementary heterotic group that were also the founders of the population (F1808 and

F98902 for the dent testers and F373 and F02803 for the flint testers). The testers were chosen to be genetically distant and with good yield performances. The G0\_F-4H and the G0\_T design are both composed of two types of hybrids: (i) hybrids obtained by crossing lines with the best predicted values GCA (G0S) and (ii) hybrids that correspond to a random sample of potential hybrids that can be obtained from the G0 generation. This makes it possible to distinguish the two types of hybrids in the analyses. The G0S hybrids correspond to hybrids produced after a first selection stage of a breeding program.

**Table 5** Distribution of the G0S or G0R lines per family in the G0\_F-4H design

	Dent						Flint					
	D1	D2	D3	D4	D5	D6	F1	F2	F3	F4	F5	F6
<b>G0S lines</b>	0	21	0	0	2	7	20	0	5	0	5	0
<b>G0R lines</b>	10	10	10	10	10	10	10	10	10	10	10	10

**Table 6** Number of hybrids per biparental family in the G0\_F-4H design

	F1	F2	F3	F4	F5	F6
<b>D1</b>	6	7	7	6	7	7
<b>D2</b>	60	7	22	7	23	6
<b>D3</b>	7	7	6	7	7	6
<b>D4</b>	7	6	7	7	6	7
<b>D5</b>	15	6	8	7	9	6
<b>D6</b>	30	6	8	6	10	7

### Production of the G1 parental lines and G1 hybrids

Last, a new generation of inbred lines was produced to evaluate the portability of the genomic predictions along breeding cycles (called **G1**). In each group, 40 intragroup single cross hybrids were produced by crossing the 30 G0S lines previously described. 351 dent and 351 flint DH lines (**G1**) were derived by haplo-diploidization from the 40 F1 single-cross hybrids in the dent and flint group, respectively. The G1 parental lines thus obtained were crossed following a sparse factorial design (**G1\_F-1H**) to produce 442 G1 hybrids. In parallel, the 30 G0S lines were also



crossed following a factorial design to produce 47 G0S hybrids. Together, they form the new factorial design referred to as **(G0S+G1)\_F-1H** comprising 489 hybrids (47 G0S + 442 G1). Adding some G0S hybrids in this design makes it possible to compare the hybrid value of the G0S parents and their G1 progeny.

## **Phenotyping**

Each experimental design was evaluated in eight trials in the North of France and Germany. Across the three experimental campaigns, we kept the trial network in the same geographical area. Hybrids from the G0\_F-1H design were evaluated in four trials in 2013 and four in 2014, hybrids from the G0\_F-4H and the tester designs were evaluated in the same field in three trials in 2016 and five in 2017 and hybrids from the (G0S+G1)\_F-1H design were evaluated in three trials in 2019 and five in 2020. 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. 2011b). 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.

## **Genotyping**

The founder lines were genotyped with a 50 K SNP array (Ganal et al. 2011). The founder lines and the G0 lines were genotyped for 18,480 SNPs using an Affymetrix® array provided by Limagrain. Markers with more than 20% of missing values within the dent and flint parental lines, markers with more than 5% (10%) of heterozygosity among the dent (flint) parental lines and markers with Minor Allele Frequency (MAF) inferior to 5% were discarded. After quality control, 9,548 SNP polymorphic markers (in at least the flint or dent population) were conserved and mapped on a consensus map (Giraud et al. 2017a). For the analyses, we considered the Affymetrix® genotyping data for the founder lines and when needed, replaced missing data with the genotypes obtained with the 50 K chip when it was available. Missing genotypes of the G0 lines were imputed with Beagle v3.0. (Browning and Browning 2007) by family, using genetic maps, and putting the founder lines in the dataset. Phasing of the flint lines and of the founder lines that presented residual heterozygosity was done simultaneously with missing genotypes imputation.

The G1 lines were genotyped using a custom-made chip comprising a subset of 15,000 SNPs of the Illumina® MaizeSNP50 BeadChip (Ganal et al. 2011). The same quality control filters as for the G0 genotypes were applied. 4,812 SNP polymorphic markers (in at least the flint or dent population) were common with the 9,548 SNP considered for the G0 lines.

## References

- Browning SR, Browning BL (2007) Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies By Use of Localized Haplotype Clustering. *The American Journal of Human Genetics* 81:1084–1097. <https://doi.org/10.1086/521987>
- 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>
- 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>
- Giraud H, Bauland C, Falque M, et al (2017c) 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&#58; Genes|Genomes|Genetics* 7:3649–3657. <https://doi.org/10.1534/g3.117.300121>
- Giraud H, Bauland C, Falque M, et al (2017d) 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>
- Lorenzi A, Bauland C, Mary-Huard T, et al (2022) 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. *Theor Appl Genet*. <https://doi.org/10.1007/s00122-022-04176-y>
- Lorenzi A, Bauland C, Pin S, et al (2023) Portability of genomic predictions trained on sparse factorial designs across two maize silage breeding cycles. In Review
- 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>
- Williams E, Piepho H-P, Whitaker D (2011) Augmented p-rep designs. *Biometrical Journal* 53:19–27. <https://doi.org/10.1002/bimj.201000102>

## Supplementary materials. Chapter 1

---

### Chapter 1. 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<sup>1</sup>; Cyril Bauland<sup>1</sup>; Tristan Mary-Huard<sup>1,2</sup>; Sophie Pin<sup>1</sup>; Carine Palafre<sup>3</sup>; Colin Guillaume<sup>4</sup>; Christina Lehermeier<sup>5</sup>; Alain Charcosset<sup>1</sup>; Laurence Moreau<sup>1</sup>

<sup>1</sup> Génétique Quantitative et Evolution - Le Moulon, INRAE, CNRS, AgroParisTech, Université Paris-Saclay, 91190 Gif-sur-Yvette, France

<sup>2</sup> MIA, INRAE, AgroParisTech, Université Paris-Saclay, 75005 Paris, France

<sup>3</sup> UE 0394 SMH, INRAE, 2297 Route de l'INRA, 40390 Saint-Martin-de-Hinx, France

<sup>4</sup> Maisadour Semences SA, 40001 Mont de Marsan Cedex, France

<sup>5</sup> RAGT2n, Genetics and Analytics Unit, 12510 Druelle, France

#### File S1 Spatial corrections

The BLUPs of spatial effects were estimated on raw performance using model (S.1) on the factorial designs and model (S.2) on the tester designs.

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

$$Y_{hkk'lxyz} = \mu + \lambda_l + \tau_h \times t_h + H_{h(kk')} \times (1 - t_h) + (R_{x(l)} + C_{y(l)}) \times (1 - d_l) + B_{z(l)} \times d_l + E_{hkk'lxyz}, \text{ (S.1)}$$

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  $z$ .  $\mu$  is the intercept,  $\lambda_l$  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 controls,  $\tau$  is the vector of fixed effect of the control with 18 levels (two for the commercial hybrids and 16 for the founder controls). With  $H_{h(kk')}$  the random genetic effect of hybrid  $h$  produced by crossing the flint line  $k$  and the dent line  $k'$ . The genetic value of the hybrid  $H_{h(km)}$ , is defined 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  $U_k$  ( $U'_{k'}$ ) are independent and identically distributed (iid) and follow a normal distribution:  $U_k \sim \mathcal{N}(0, \sigma_{GCA_f}^2)$  (and  $U'_{k'} \sim \mathcal{N}(0, \sigma_{GCA_d}^2)$ ).  $S_{kk'}$  is the random SCA effect of the

interaction between the parental lines  $k$  and  $k'$ , with  $S_{kk'} \sim \mathcal{N}(0, \sigma_{SCA}^2)$  iid.  $R_{x(l)}$  is the random effect of row  $x$  in environment  $l$ ,  $C_{y(l)}$  is the random effect of column  $y$  in environment  $l$  and  $B_{z(l)}$  is the random effect of block  $z$  in environment  $l$ , with  $R_{x(l)} \sim \mathcal{N}(0, \sigma_{R_l}^2)$ ,  $C_{y(l)} \sim \mathcal{N}(0, \sigma_{C_l}^2)$  and  $B_{z(l)} \sim \mathcal{N}(0, \sigma_{B_l}^2)$  which are assumed to be independent and identically distributed (iid).  $d_l$  indicates the spatial effect to consider in environment  $l$ , it is set to 1 if the spatial effect is modeled by a block effect and by 0 if modeled by row and column effects. To construct the vector  $d$ , for each trait and each trial, two types of spatial effects were tested the line and column effects or the block effect. The best spatial effect was chosen based on the AIC criterion.  $E_{hkk'lxyz}$  is the error term of the model, with  $E_{hkk'lxyz} \sim \mathcal{N}(0, \sigma_{E_l}^2)$  iid within trial and independent between trials. The different random effects of the model are assumed to be independent.

The model implemented on the T-F was:

$$Y_{hkmlyz} = \mu + \lambda_l + \tau_h \times t_h + \gamma_m + H_{h(km)} \times (1 - t_h) + (R_{x(l)} + C_{y(l)}) \times (1 - d_l) + B_{z(l)} \times d_l + E_{hkmlyz}, \quad (S.2)$$

where  $\lambda_l$ ,  $t_h$ ,  $\tau_h$ ,  $\rho_{lh}$ ,  $R_{x(l)}$ ,  $C_{y(l)}$ ,  $B_{z(l)}$  and  $d_l$  are defined as in model (S.1).  $Y_{hlmxyz}$  is the phenotypic value of hybrid  $h$  produced by crossing the dent parental line  $m$  used as tester and the flint parental line  $k$ , evaluated in environment  $l$  located at row  $x$ , column  $y$  and in block  $z$ .  $\gamma_m$  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 parental line  $m$  used as tester and the flint parental line  $k$  evaluated for its GCA. The genetic value of the hybrid  $H_{h(km)}$  is defined as follows:

$$H_{h(km)} = U_k + S_{km},$$

where  $U_k$  is the random GCA effect of the flint line  $k$ , with  $U_k \sim \mathcal{N}(0, \sigma_{GCA_f}^2)$  iid and  $S_{km}$  is the random effect of the interaction between the flint line  $k$  and the dent line  $m$  used as tester, with  $S_{km} \sim \mathcal{N}(0, \sigma_S^2)$  iid.  $E_{hkmlyz}$  is the error term of the model, we assume that it follows:  $E_{hkk'lxyz} \sim \mathcal{N}(0, \sigma_{E_l}^2)$  iid. The same model was adapted and implemented on T-D. The different random effects of the model are assumed to be independent.

**Table S1** Broad-sense heritability ( $H^2$ ), percentage of genetic variance assigned 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-D and T-F) without using marker information.

Trait	Design	$\sigma_{GCA_d}^2$	$\sigma_{GCA_f}^2$	$\sigma_{SCA}^2$	$\sigma_{GCA_d \times E}^2$	$\sigma_{GCA_f \times E}^2$	$\sigma_{SCA \times E}^2$	%SCA <sup>a</sup>	$\sigma_E^2$ min-max <sup>b</sup>	$H^2$ <sup>c</sup>
<b>DMY</b>	<b>F-1H</b>	0.72(0.10)	0.25(0.10)	0.25(0.11)	0.03(0.06)	0.19(0.06)	0.12(0.09) ns	<b>20</b>	0.70(0.06)-3.09(0.17)	<b>0.84</b>
	<b>F-4H</b>	0.74(0.14)	0.54(0.11)	0.13(0.03)	0.18(0.03)	0.17(0.03)	0.11(0.05)	<b>9</b>	0.44(0.06)-3.15(0.26)	<b>0.89</b>
	<b>T-D</b>	0.70(0.13)	-	0.17(0.04)	0.03(0.04) ns	-	0.21(0.06)	<b>20</b>	0.26(0.06)-2.22(0.27)	<b>0.88</b>
	<b>T-F</b>	-	0.52(0.10)	0.14(0.03)	-	0.09(0.03)	0.09(0.05)	<b>21</b>	0.34(0.07)-1.42(0.18)	<b>0.87</b>
<b>DMC</b>	<b>F-1H</b>	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	<b>15</b>	0.71(0.07)-4.96(0.28)	<b>0.91</b>
	<b>F-4H</b>	1.32(0.24)	2.78(0.46)	0.28(0.05)	0.40(0.06)	0.45(0.06)	0.25(0.09)	<b>6</b>	0.68(0.10)-4.15(0.37)	<b>0.93</b>
	<b>T-D</b>	1.89(0.34)	-	0.32(0.08)	0.56(0.10)	-	0.35(0.12)	<b>14</b>	0.59(0.13)-2.63(0.37)	<b>0.89</b>
	<b>T-F</b>	-	3.00(0.51)	0.44(0.09)	-	0.30(0.07)	0.12(0.10)	<b>13</b>	0.50(0.10)-2.26(0.31)	<b>0.94</b>
<b>DtSilk</b>	<b>F-1H</b>	1.26(0.18)	0.73(0.19)	0.41(0.18)	0.12(0.05)	0.16(0.06)	0.00(0.08) ns	<b>17</b>	0.66(0.06)-5.57(0.28)	<b>0.91</b>
	<b>F-4H</b>	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	<b>6</b>	9.69(0.08)-2.34(0.22)	<b>0.93</b>
	<b>T-D</b>	0.63(0.11)	-	0.08(0.03)	0.06(0.03)	-	0.09(0.06)	<b>11</b>	0.37(0.07)-2.06(0.25)	<b>0.86</b>
	<b>T-F</b>	-	1.07(0.19)	0.22(0.05)	-	0.14(0.04)	0.01(0.08)	<b>17</b>	0.51(0.09)-3.34(0.39)	<b>0.90</b>
<b>PH</b>	<b>F-1H</b>	89.26(10.94)	41.69(11.68)	23.40(11.90)	8.36(3.50)	1.95(3.76)	3.03(5.78) ns	<b>15</b>	33.87(3.39)-215.58(11.31)	<b>0.90</b>
	<b>F-4H</b>	120.75(20.40)	57.37(10.36)	5.40(1.96)	11.68(2.38)	6.99(2.16)	0.00 ns	<b>3</b>	55.38(6.60)-169.29(14.29)	<b>0.92</b>
	<b>T-D</b>	115.58(18.90)	-	5.80(2.49)	10.66(3.68)	-	0.00	<b>5</b>	46.16(7.77)-128.96(16.56)	<b>0.92</b>
	<b>T-F</b>	-	56.86(11.64)	24.98(5.22)	-	10.12(3.28) ns	0.00	<b>31</b>	38.26(7.06)-91.75(12.70)	<b>0.89</b>
<b>DINAG</b>	<b>F-1H</b>	1.55(0.18)	0.90(0.18)	0.00 ns	0.09(0.11)	0.11(0.12)	0.23(0.18) ns	<b>0</b>	1.45(0.13)-6.47(0.34)	<b>0.85</b>
	<b>F-4H</b>	2.11(0.36)	1.17(0.21)	0.08(0.05)	0.36(0.07)	0.23(0.06)	0.00 ns	<b>2</b>	1.12(0.11)-5.36(0.43)	<b>0.87</b>
	<b>T-D</b>	2.68(0.45)	-	0.15(0.08) ns	0.38(0.12)	-	0.24(0.20)	<b>5</b>	1.28(0.24)-3.27(0.46)	<b>0.88</b>

Supplementary materials. Chapter 1

	<b>T-F</b>	-	1.25(0.23)	0.21(0.08)	-	0.18(0.10) <b>ns</b>	0.04(0.17)	<b>14</b>	0.99(0.21)-3.60(0.44)	<b>0.81</b>
<b>DINAGZ</b>	<b>F-1H</b>	1.47(0.17)	0.86(0.15)	0.00 <b>ns</b>	0.21(0.12)	0.25(0.13)	0.00 <b>ns</b>	<b>0</b>	2.10(0.16)-5.88(0.33)	<b>0.79</b>
	<b>F-4H</b>	1.88(0.34)	0.97(0.20)	0.11(0.08) <b>ns</b>	0.43(0.10)	0.28(0.09)	0.00 <b>ns</b>	<b>4</b>	2.13(0.25)-8.84(0.71)	<b>0.78</b>
	<b>T-D</b>	2.53(0.44)	-	0.08(0.10) <b>ns</b>	0.21(0.14)	-	0.36(0.25)	<b>3</b>	1.16(0.24)-5.67(0.71)	<b>0.82</b>
	<b>T-F</b>	-	1.10(0.23)	0.25(0.12)	-	0.10(0.14) <b>ns</b>	0.00 <b>ns</b>	<b>19</b>	1.39(0.25)-5.78(0.68)	<b>0.71</b>
<b>MFU (x10<sup>2</sup>)</b>	<b>F-1H</b>	2.06(0.30)	1.03(0.33)	0.33(0.34) <b>ns</b>	0.13(0.23)	0.01(0.24)	0.86(0.39)	<b>10</b>	3.39(0.28)-12.83(0.69)	<b>0.79</b>
	<b>F-4H</b>	1.74(0.31)	1.67(0.30)	0.10(0.06) <b>ns</b>	0.37(0.08)	0.29(0.07)	0.00 <b>ns</b>	<b>3</b>	1.29(0.12)-7.61(0.60)	<b>0.86</b>
	<b>T-D</b>	2.23(0.40)	-	0.28(0.10)	0.46(0.12)	-	0.00	<b>11</b>	1.04(0.16)-6.31(0.74)	<b>0.85</b>
	<b>T-F</b>	-	1.46(0.27)	0.23(0.09)	-	0.21(0.11)	0.23(0.16)	<b>14</b>	0.46(0.12)-6.53(0.77)	<b>0.78</b>
<b>NDF</b>	<b>F-1H</b>	0.80(0.21)	0.89(0.22)	0.25(0.24) <b>ns</b>	0.06(0.19)	0.13(0.19)	0.71(0.31)	<b>13</b>	2.51(0.22)-10.23(0.55)	<b>0.72</b>
	<b>F-4H</b>	0.71(0.15)	1.31(0.24)	0.12(0.07) <b>ns</b>	0.17(0.07)	0.23(0.08)	0.23(0.19) <b>ns</b>	<b>6</b>	2.07(0.24)-5.32(0.47)	<b>0.80</b>
	<b>T-D</b>	0.79(0.19)	-	0.27(0.11)	0.48(0.14)	-	0.00	<b>25</b>	1.42(0.27)-4.03(0.53)	<b>0.71</b>
	<b>T-F</b>	-	1.46(0.29)	0.31(0.12)	-	0.08(0.14) <b>ns</b>	0.62(0.24)	<b>18</b>	1.15(0.23)-4.30(0.59)	<b>0.79</b>
<b>LIGN (x10<sup>2</sup>)</b>	<b>F-1H</b>	0.08(0.01)	0.02(0.01)	0.00(0.01) <b>ns</b>	0.01(0.00)	0.00	0.01(0.01) <b>ns</b>	<b>0</b>	0.05(0.00)-1.03(0.06)	<b>0.89</b>
	<b>F-4H</b>	0.13(0.02)	0.03(0.01)	0.00 <b>ns</b>	0.02(0.00)	0.01(0.00)	0.01(0.01) <b>ns</b>	<b>0</b>	0.05(0.01)-0.11(0.01)	<b>0.86</b>
	<b>T-D</b>	0.12(0.02)	-	0.01(0.00) <b>ns</b>	0.01(0.00)	-	0.00	<b>8</b>	0.04(0.01)-0.09(0.01)	<b>0.88</b>
	<b>T-F</b>	-	0.05(0.01)	0.00 <b>ns</b>	-	0.01(0.00)	0.01(0.01)	<b>0</b>	0.03(0.01)-0.12(0.02)	<b>0.70</b>
<b>CELL (x10<sup>2</sup>)</b>	<b>F-1H</b>	0.21(0.03)	0.10(0.03)	0.01(0.03) <b>ns</b>	0.05(0.02)	0.03(0.02)	0.00 <b>ns</b>	<b>3</b>	0.34(0.03)-1.03(0.06)	<b>0.78</b>
	<b>F-4H</b>	0.40(0.08)	0.25(0.05)	0.02(0.02) <b>ns</b>	0.10(0.03)	0.07(0.03)	0.03(0.04) <b>ns</b>	<b>3</b>	0.26(0.04)-12.98(0.96)	<b>0.57</b>
	<b>T-D</b>	0.40(0.08)	-	0.07(0.03)	0.12(0.04)	-	0.00(0.06)	<b>15</b>	0.29(0.06)-7.65(0.83)	<b>0.58</b>
	<b>T-F</b>	-	0.20(0.05)	0.03(0.03) <b>ns</b>	-	0.09(0.04)	0.01(0.06)	<b>13</b>	0.28(0.06)-7.03(0.82)	<b>0.39</b>
<b>HCELL (x10<sup>2</sup>)</b>	<b>F-1H</b>	0.55(0.06)	0.21(0.06)	0.00(0.05) <b>ns</b>	0.08(0.03)	0.04(0.03)	0.00 <b>ns</b>	<b>0</b>	0.54(0.04)-2.01(0.11)	<b>0.84</b>
	<b>F-4H</b>	0.94(0.17)	0.36(0.08)	0.04(0.03) <b>ns</b>	0.20(0.04)	0.17(0.04)	0.05(0.06) <b>ns</b>	<b>3</b>	0.38(0.05)-9.81(0.76)	<b>0.73</b>

Supplementary materials. Chapter 1

<b>T-D</b>	0.89(0.16)	-	0.12(0.05)	0.19(0.06)	-	0.00	<b>12</b>	0.49(0.10)-6.22(0.70)	<b>0.76</b>
<b>T-F</b>	-	0.32(0.07)	0.04(0.04) <b>ns</b>	-	0.17(0.06)	0.05(0.08)	<b>11</b>	0.39(0.08)-7.41(0.82)	<b>0.48</b>

<sup>a</sup> Percentage of SCA variance computed as  $\frac{\sigma_{SCA}^2}{\sigma_{GCA_d}^2 + \sigma_{GCA_f}^2 + \sigma_{SCA}^2} \times 100$

<sup>b</sup> Minimum residual variance and maximum residual variance across all environments

<sup>c</sup> Broad-sense heritability

<sup>d</sup> Standard error in brackets

<sup>e</sup> Significance of the variance components assessed by likelihood ratio test with  $\chi^2$  mixed distributions ( $\alpha=0.05$ ). Non-significant variance component is indicated by **ns**



**Table S2** Variance components and percentage of genetic variance assigned 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	$\sigma_{GCA_d}^2$	$\sigma_{GCA_f}^2$	$\sigma_{SCA}^2$	%SCA <sup>a</sup>	$\sigma_E^2$
<b>DMY</b>	<b>F-1H</b>	0.62(0.12) <sup>b</sup>	0.48(0.09)	0.13(0.06)	<b>11</b>	0.52(0.06)
	<b>F-4H</b>	0.68(0.15)	1.04(0.21)	0.03(0.05)	<b>2</b>	0.31(0.05)
	<b>T-D</b>	0.47(0.13)	-	0.08(0.06)	<b>15</b>	0.25(0.05)
	<b>T-F</b>	-	0.86(0.17)	0.10(0.05)	<b>10</b>	0.11(0.03)
<b>DMC</b>	<b>F-1H</b>	1.41(0.25)	1.79(0.28)	0.00	<b>0</b>	1.07(0.08)
	<b>F-4H</b>	2.79(0.53)	4.45(0.78)	0.01(0.08)	<b>0</b>	0.62(0.10)
	<b>T-D</b>	4.27(0.79)	-	0.19(0.12)	<b>4</b>	0.41(0.09)
	<b>T-F</b>	-	3.95(0.76)	0.28(0.18)	<b>7</b>	0.51(0.12)
<b>DtSilk</b>	<b>F-1H</b>	1.21(0.21)	2.03(0.29)	0.09(0.06)	<b>3</b>	0.73(0.08)
	<b>F-4H</b>	1.67(0.32)	2.65(0.46)	0.11(0.06)	<b>2</b>	0.24(0.05)
	<b>T-D</b>	0.99(0.23)	-	0.04(0.04)	<b>4</b>	0.21(0.05)
	<b>T-F</b>	-	2.29(0.42)	0.17(0.11)	<b>7</b>	0.24(0.06)
<b>PH</b>	<b>F-1H</b>	70.90(11.74)	68.85(11.28)	6.59(4.31)	<b>5</b>	33.18(4.62)
	<b>F-4H</b>	76.78(14.91)	103.47(18.80)	0.61(2.81)	<b>0</b>	20.42(3.16)
	<b>T-D</b>	75.91(15.80)	-	1.12(2.38)	<b>1</b>	16.75(3.11)
	<b>T-F</b>	-	100.13(21.72)	17.53(10.18)	<b>15</b>	20.85(5.92)
<b>DINAG</b>	<b>F-1H</b>	0.78(0.13)	0.60(0.11)	0.00	<b>0</b>	0.69(0.05)
	<b>F-4H</b>	1.03(0.22)	1.10(0.23)	0.08(0.09)	<b>4</b>	0.47(0.09)
	<b>T-D</b>	1.67(0.39)	-	0.00	<b>0</b>	0.63(0.09)
	<b>T-F</b>	-	1.13(0.28)	0.47(0.24)	<b>29</b>	0.32(0.11)
<b>DINAGZ</b>	<b>F-1H</b>	0.84(0.15)	0.58(0.12)	0.00	<b>0</b>	0.90(0.06)
	<b>F-4H</b>	1.11(0.26)	0.79(0.21)	0.00	<b>0</b>	0.92(0.09)
	<b>T-D</b>	1.97(0.49)	-	0.00	<b>0</b>	0.85(0.12)
	<b>T-F</b>	-	0.90(0.28)	0.24(0.22)	<b>21</b>	0.72(0.15)
<b>MFU (x10<sup>2</sup>)</b>	<b>F-1H</b>	1.14(0.22)	1.31(0.24)	0.00	<b>0</b>	1.50(0.10)
	<b>F-4H</b>	1.19(0.27)	2.17(0.42)	0.26(0.14)	<b>7</b>	0.48(0.11)
	<b>T-D</b>	1.38(0.38)	-	0.07(0.10)	<b>5</b>	0.81(0.14)
	<b>T-F</b>	-	2.11(0.44)	0.12(0.13)	<b>5</b>	0.55(0.11)
<b>NDF</b>	<b>F-1H</b>	0.77(0.16)	0.99(0.19)	0.02(0.05)	<b>1</b>	1.32(0.10)
	<b>F-4H</b>	0.88(0.23)	1.54(0.32)	0.09(0.10)	<b>4</b>	0.66 (0.11)
	<b>T-D</b>	0.62(0.23)	-	0.01(0.07)	<b>2</b>	0.84(0.12)
	<b>T-F</b>	-	1.80(0.44)	0.10(0.14)	<b>5</b>	0.77(0.14)

<b>LIGN (x10<sup>2</sup>)</b>	<b>F-1H</b>	0.04(0.01)	0.02(0.00)	0.00	<b>1</b>	0.02(0.00)
	<b>F-4H</b>	0.07(0.01)	0.05(0.01)	0.00	<b>2</b>	0.02(0.00)
	<b>T-D</b>	0.06(0.01)	-	0.00	<b>0</b>	0.03(0.00)
	<b>T-F</b>	-	0.06(0.01)	0.00	<b>5</b>	0.03(0.00)
<b>CELL (x10<sup>2</sup>)</b>	<b>F-1H</b>	0.09(0.02)	0.08(0.02)	0.00	<b>0</b>	0.15(0.01)
	<b>F-4H</b>	0.33(0.09)	0.51(0.13)	0.00	<b>0</b>	0.50(0.05)
	<b>T-D</b>	0.25(0.10)	-	0.06(0.06)	<b>19</b>	0.37(0.06)
	<b>T-F</b>	-	0.24 (0.10)	0.04(0.05)	<b>14</b>	0.39(0.06)
<b>HCELL (x10<sup>2</sup>)</b>	<b>F-1H</b>	0.23(0.04)	0.17(0.03)	0.00	<b>0</b>	0.23(0.01)
	<b>F-4H</b>	0.50(0.12)	0.65(0.16)	0.00	<b>0</b>	0.50(0.05)
	<b>T-D</b>	0.37(0.12)	-	0.13(0.10)	<b>26</b>	0.35(0.07)
	<b>T-F</b>	-	0.36 (0.13)	0.05(0.06)	<b>12</b>	0.46(0.07)

<sup>a</sup> Percentage of SCA variance computed as  $\frac{\sigma_{SCA}^2}{\sigma_{GCA_d}^2 + \sigma_{GCA_f}^2 + \sigma_{SCA}^2} \times 100$

<sup>b</sup> Standard error in brackets

**Table S3** Predictive abilities obtained for the F-4H and each of the tester designs using the F-1H as TRS in scenario 1a.

<b>Validation Set</b>	<b>F-4H</b>		<b>T-D</b>	<b>T-F</b>
Predicted hybrid value	GCA <sub>f</sub> + GCA <sub>d</sub>	GCA <sub>f</sub> + GCA <sub>d</sub> + SCA	GCA <sub>d</sub>	GCA <sub>f</sub>
<b>DMY</b>	0.78	0.77	0.69	0.80
<b>DMC</b>	0.78	0.78	0.84	0.74
<b>DtSilk</b>	0.79	0.79	0.84	0.76
<b>PH</b>	0.87	0.87	0.75	0.87
<b>DINAG</b>	0.86	0.86	0.82	0.85
<b>DINAGZ</b>	0.79	0.79	0.77	0.78
<b>MFU</b>	0.82	0.82	0.79	0.79
<b>NDF</b>	0.69	0.69	0.76	0.57
<b>LIGN</b>	0.87	0.87	0.78	0.91
<b>CELL</b>	0.72	0.72	0.58	0.73
<b>HCELL</b>	0.80	0.80	0.61	0.83

**Table S4** Predictive abilities obtained for the F-1H using the F-4H (using a model including or not the SCA effect) or the tester designs as TRS in scenario 2a.

Training set	F-4H		Tester designs
Predicted hybrid	GCA <sub>f</sub> + GCA <sub>d</sub>	GCA <sub>f</sub> + GCA <sub>d</sub> + SCA	GCA <sub>f</sub> + GCA <sub>d</sub>
<b>DMY</b>	0.56	0.56	0.54
<b>DMC</b>	0.64	0.64	0.65
<b>DtSilk</b>	0.55	0.55	0.56
<b>PH*</b>	0.69	0.69	0.66
<b>DINAG*</b>	0.78	0.78	0.75
<b>DINAGZ*</b>	0.73	0.73	0.70
<b>MFU</b>	0.67	0.67	0.65
<b>NDF</b>	0.51	0.51	0.50
<b>LIGN</b>	0.80	0.80	0.81
<b>CELL</b>	0.64	0.64	0.63
<b>HCELL</b>	0.73	0.73	0.72

\* Significant difference assessed by Williams tests ( $\alpha=0.05$ ) between the predictive abilities obtained when training the GS model on the F-4H or on the tester designs

**Table S5** Correlations between the GCA BLUPs of the F-1H predicted using the F-4H or the tester designs as TRS in scenario 2a.

BLUPs correlated	DMY	DMC	DtSilk	PH	DINAG	DINAGZ	MFU	NDF	LIGN	CELL	HCELL
<b>GCA<sub>f</sub></b>	0.85	0.95	0.94	0.809	0.94	0.92	0.93	0.90	0.89	0.86	0.84
<b>GCA<sub>d</sub></b>	0.90	0.88	0.86	0.94	0.94	0.91	0.92	0.76	0.95	0.94	0.96

**Table S6** Predictive abilities obtained by training the GS model on the F-4H or on the tester designs to predict T0, T1 or T2 hybrids in the F-1H (VS) in scenario 2a.

Training set	F-4H <sup>a</sup>				Tester designs			
Hybrid type in the VS <sup>b</sup>	T1D	T1F	T0	All	T1D	T1F	T0	All
Number of hybrids in the VS	106	95	742	<b>951</b>	107	95	741	<b>951</b>
<b>DMY</b>	0.61	0.74	0.49	<b>0.56</b>	0.60	0.75	0.49	<b>0.54</b>
<b>DMC</b>	0.72	0.81	0.61	<b>0.64</b>	0.73	0.78	0.62	<b>0.65</b>
<b>DtSilk</b>	0.61	0.72	0.50	<b>0.55</b>	0.63	0.73	0.52	<b>0.56</b>
<b>PH</b>	0.75	0.78	0.65	<b>0.69</b>	0.74	0.75	0.62	<b>0.66</b>
<b>DINAG</b>	0.79	0.84	0.77	<b>0.78</b>	0.76	0.84	0.73	<b>0.75</b>
<b>DINAGZ</b>	0.70	0.81	0.72	<b>0.73</b>	0.66	0.80	0.69	<b>0.70</b>
<b>MFU</b>	0.70	0.78	0.64	<b>0.67</b>	0.68	0.75	0.63	<b>0.65</b>
<b>NDF</b>	0.63	0.70	0.47	<b>0.51</b>	0.60	0.65	0.47	<b>0.50</b>
<b>LIGN</b>	0.83	0.83	0.79	<b>0.80</b>	0.84	0.84	0.79	<b>0.81</b>
<b>CELL</b>	0.59	0.71	0.62	<b>0.64</b>	0.62	0.73	0.61	<b>0.63</b>
<b>HCELL</b>	0.71	0.77	0.72	<b>0.73</b>	0.73	0.78	0.71	<b>0.72</b>

<sup>a</sup> TRS used to calibrate the Model (5.1) including the SCA effect

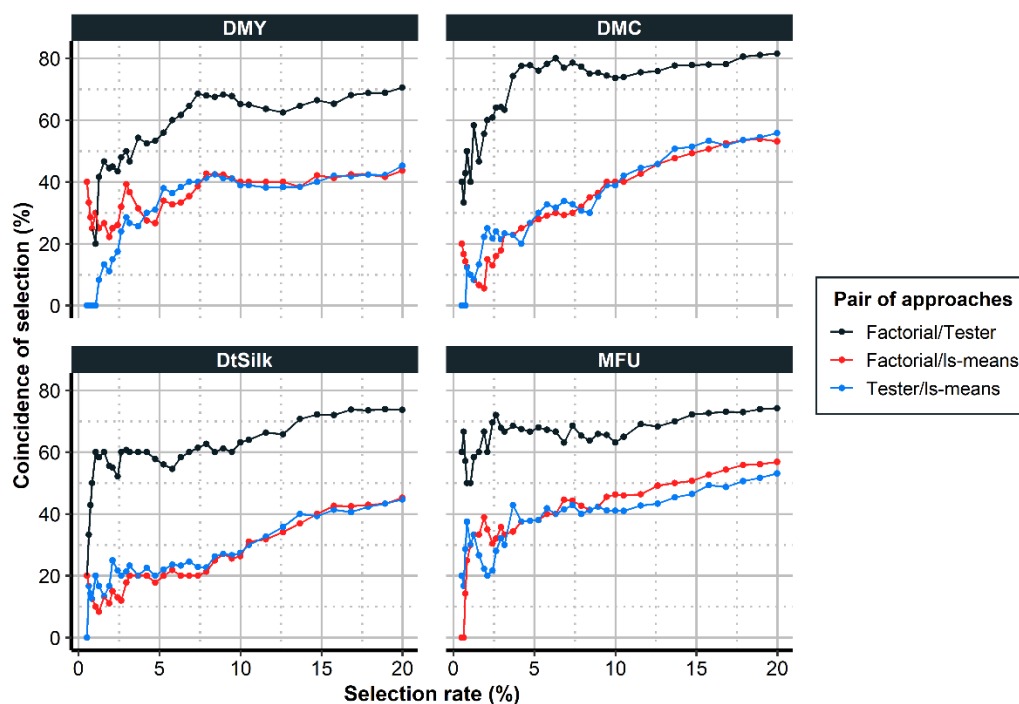
<sup>b</sup> Different hybrids types in the VS: T0 hybrids where none of the parental lines was in the TRS, T1F hybrids where only the flint line was in the TRS and T1D where only the dent line was in the TRS

**Table S7** Predictive abilities obtained when training the GS model (i) on 216 hybrids of the F-1H design to predict 216 hybrids of the F-4H design and (ii) on 216 hybrids of the F-4H design to predict 216 hybrids of the F-1H in scenario 3.

TRS	F-1H		F-4H	
	Mean <sup>a</sup>	Sd <sup>b</sup>	Mean	Sd
<b>DMY</b>	0.55	0.04	0.50	0.05
<b>DMC</b>	0.62	0.03	0.60	0.03
<b>DtSilk</b>	0.57	0.04	0.44	0.05
<b>PH</b>	0.71	0.03	0.62	0.04
<b>DINAG</b>	0.77	0.02	0.76	0.02
<b>DINAGZ</b>	0.71	0.03	0.70	0.03
<b>MFU</b>	0.71	0.03	0.65	0.03
<b>NDF</b>	0.55	0.04	0.46	0.05
<b>LIGN</b>	0.80	0.02	0.78	0.02
<b>CELL</b>	0.58	0.03	0.60	0.03
<b>HCELL</b>	0.69	0.03	0.70	0.03

<sup>a</sup> Mean of the predictive abilities over the 100 repetitions

<sup>b</sup> Standard deviation of the 100 repetitions



**Fig.S1** Coincidence of selection of the F-1H single-cross hybrids computed for different selection rates (%) in scenario 2a. The coincidence of selection was computed in black between the factorial and the tester approaches (GBLUP), in red between the factorial approach (GBLUP) and the phenotypic values (hybrid ls-means) and in blue between the tester approach (GBLUP) and the phenotypic values (hybrid ls-means). The BLUPs estimated in scenario 2a were used to compute the coincidence of selection.

### File S2 Benchmark model used in Scenario 1a

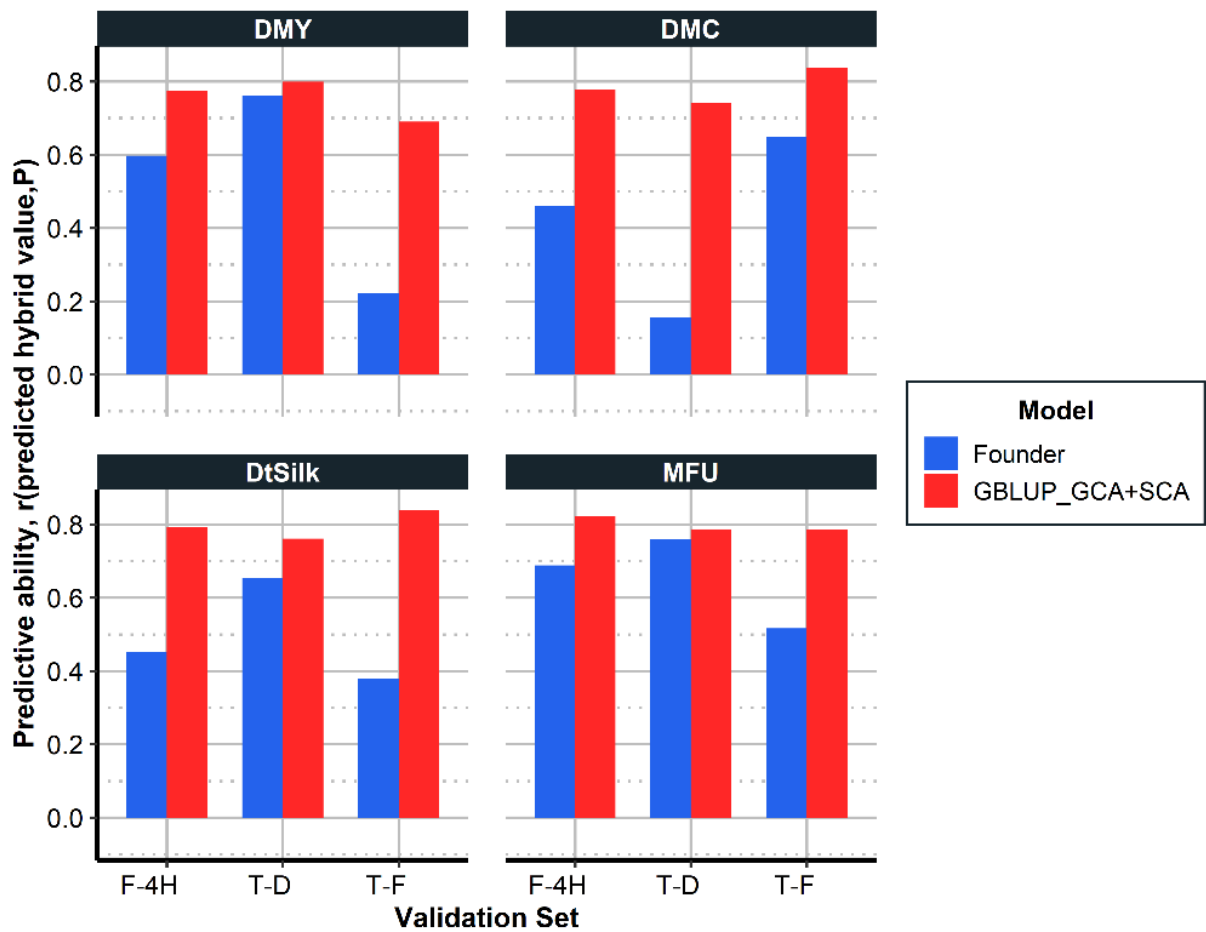
The model considered as a benchmark was implemented on the ls-means and was defined as:

$$Y_{ijj'j'kk'} = \mu + \varphi_i + \varphi_j + \delta_{i'} + \delta_{j'} + U_k + U'_{k'} + S_{kk'} + E_{ijj'j'kk'}, (S.3)$$

where  $Y_{ijj'j'kk'}$  is the phenotypic value of the hybrid produced by crossing flint parental line  $k$  (issued from crossing founder lines  $i$  and  $j$ ) and dent parental line  $k'$  (issued from founder lines  $i'$  and  $j'$ ),  $\varphi_i$  and  $\varphi_j$  are the fixed effect of the flint founders  $i$  and  $j$ ,  $\delta_{i'}$  and  $\delta_{j'}$  are the fixed effects of the dent founders  $i'$  and  $j'$ .  $U_k$  ( $U'_{k'}$ ) is the random effect of the GCA effect of flint line  $k$  (dent line  $k'$ ) with  $U_k \sim \mathcal{N}(0, \sigma_{GCA_f}^2)$  iid ( $U'_{k'} \sim \mathcal{N}(0, \sigma_{GCA_d}^2)$  iid) where  $\sigma_{GCA_f}^2$  and  $\sigma_{GCA_d}^2$  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  $S_{kk'} \sim \mathcal{N}(0, \sigma_{SCA}^2)$  iid with  $\sigma_{SCA}^2$  being the SCA variance.  $E_{ijj'j'kk'}$  is the error term. We assume that the errors follow  $E_{ijj'j'kk'} \sim \mathcal{N}(0, \sigma_E^2)$  iid. The different random effects of the model are assumed to be independent.

For the prediction of the F-4H hybrids, the predictive ability was computed as the correlation between the observed phenotypes of the VS (ls-means of the hybrids) and the sum of the BLUEs estimated in the model for the founder parents of the hybrid's parental lines. When the tester designs were used as VS, the predictive ability was computed as the correlation between the mean of the ls-means for a given line over the two testers and the sum of the BLUEs estimated for the founder parents of the line.

The predictive abilities of the Founder model were compared to the ones obtained with the GBLUP model in the same prediction scenario (scenario 1a) (**Fig.S2**). When predicting the F-4H, the Founder model (benchmark model) showed moderate to high predictive abilities ranging from 0.45 (DtSilk) to 0.69 (MFU) and the GBLUP model predictive abilities ranged from 0.78 (DMY and DMC) to 0.82 (MFU). When predicting the tester designs, the Founder model predictive abilities ranged from 0.15 (T-D for DMC) to 0.76 (T-D for DMY) and the GBLUP model predictive abilities ranged from 0.76 (T-D for DtSilk) to 0.84 (T-F for DMC and DtSilk). The GBLUP model always outperformed the Founder model.



**Fig.S2** Predictive abilities obtained when using the F-1H as TRS to predict the F-4H and the tester designs. Two types of prediction models were implemented: the “Founder” model corresponding to the benchmark model (S.3) defined in *File S1* and the “GBLUP\_GCA+SCA” model corresponding to the GBLUP GS model (5.1).

## Supplementary materials. Chapter 2

---

### Chapter 2. Portability of genomic predictions trained on sparse factorial designs across two maize silage breeding cycles

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

<sup>1</sup> Université Paris-Saclay, INRAE, CNRS, AgroParisTech, Génétique Quantitative et Evolution - Le Moulon, 91190 Gif-sur-Yvette, France

<sup>2</sup> UE 0394 SMH, INRAE, 2297 Route de l'INRA, 40390, Saint-Martin-de-Hinx, France

<sup>3</sup> Maïsadour Semences SA, F-40001 Mont-de-Marsan Cedex, France

<sup>4</sup> RAGT2n, Genetics and Analytics Unit, 12510 Druelle, France

<sup>5</sup> MIA, INRAE, AgroParisTech, Université Paris-Saclay, 75005, Paris, France

#### File S1-Spatial corrections

The BLUPs of spatial effects were estimated on raw performance using model (S1) on the (G0S+G1)\_F-1H design. The model implemented was:

$$Y_{hii'lxyz} = \mu + \lambda_l + (\tau_h + \rho_{lh}) \times t_h + [H_{h(ii')} + H\lambda_{lh(ii')}] \times (1 - t_h) + (R_{x(l)} + C_{y(l)}) \times (1 - d_l) + B_{z(l)} \times d_l + E_{hii'lxyz}, \text{ (S1)}$$

where  $Y_{hii'lxyz}$  is the phenotypic value of hybrid  $h$  produced by crossing the parental lines  $i$  and  $i'$ , evaluated in environment,  $l$  located at row  $x$ , column  $y$  and in block  $z$ .  $\mu$  is the intercept,  $\lambda_l$  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),  $\tau_h$  is the fixed effect of control hybrids with 19 levels (2 for commercial hybrids + 16 for founder hybrids + one for non-control hybrids) and  $\rho_{lh}$  is the effect of the interaction between trial  $l$  and control hybrid  $h$ .  $H_{h(ii')}$  is the random genetic effect of experimental hybrid  $h$ , produced by crossing the flint line  $i$  and the dent line  $i'$  and follows a normal distribution:  $H_{h(ii')} \sim \mathcal{N}(0, \sigma_G^2)$ .  $H\lambda_{lh(ii')}$  is the genotype by trial interaction with  $H\lambda_{lh(ii')} \sim \mathcal{N}(0, \sigma_{G \times E_l}^2)$  iid within trial and independent between trials.  $R_{x(l)}$  is the random effect of row  $x$  in environment  $l$ ,  $C_{y(l)}$  is the random effect of column  $y$  in environment  $l$  and  $B_{z(l)}$  is the random effect of block  $z$  in environment  $l$ , with  $R_{x(l)} \sim \mathcal{N}(0, \sigma_{R_l}^2)$ ,  $C_{y(l)} \sim \mathcal{N}(0, \sigma_{C_l}^2)$  and



$B_{z(l)} \sim \mathcal{N}(0, \sigma_{B_l}^2)$  which are assumed to be independent and identically distributed (iid).  $d_l$  indicates the spatial effect to consider in environment  $l$ , it is set to 1 if the spatial effect is modeled by a block effect and by 0 if modeled by row and column effects. To construct the vector  $d$ , for each trait and each trial, two types of spatial effects were tested the line and column effects or the block effect. The best spatial effect was chosen based on the AIC criterion.  $E_{hiiltxyz}$  is the error term of the model, with  $E_{hiiltxyz} \sim \mathcal{N}(0, \sigma_{E_l}^2)$  iid within trial and independent between trials. The different random effects of the model are assumed to be independent.

**Table S1** 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 using (Model 1).

Trait	Hybrid	$\sigma_{GCA_f}^2$	$\sigma_{GCA_d}^2$	$\sigma_{SCA}^2$	$\sigma_{GCA_f \times E}^2$	$\sigma_{GCA_d \times E}^2$	$\sigma_{SCA \times E}^2$	$\sigma_E^2$ min-max <sup>a</sup>	%SCA <sup>b</sup>	H <sup>2</sup> <sup>c</sup>
<b>DMY</b>	<b>G0S</b>	0.50(0.24)	0.13(0.18)	0.14(0.18)	0.07(0.08)	0.12(0.09)	0.00	0.31(0.05)-1.40(0.12)	19	0.87
	<b>G1</b>	0.31(0.05)	0.25(0.05)	0.00	0.05(0.05)	0.07(0.04)	0.23(0.07)		0	0.80
<b>DMC</b>	<b>G0S</b>	0.46(0.28)	1.42(0.51)	0.00	0.00	0.39(0.13)	0.00	0.57(0.07)-3.62(0.27)	0	0.90
	<b>G1</b>	1.47(0.21)	0.73(0.21)	0.25(0.20)	0.11(0.05)	0.31(0.05)	0.00		10	0.92
<b>DtSilk</b>	<b>G0S</b>	1.63(0.71)	1.57(0.65)	0.03(0.27)	0.41(0.16)	0.05(0.12)	0.00	0.64(0.09)-2.33(0.20)	1	0.93
	<b>G1</b>	2.06(0.29)	1.41(0.27)	0.05(0.22)	0.17(0.07)	0.09(0.06)	0.24(0.11)		1	0.94
<b>PH</b>	<b>G0S</b>	13.30(25.69)	32.70(29.46)	45.20(33.71)	10.16(9.38)	15.22(10.15)	0.00	31.30(4.11)-108.33(9.08)	50	0.88
	<b>G1</b>	32.11(5.80)	51.45(7.13)	0.00	3.38(3.24)	0.00	15.81(4.78)		0	0.88
<b>DINAG</b>	<b>G0S</b>	0.71(0.37)	0.00	0.08(0.29)	0.32(0.45)	0.00	0.78(0.54)	1.47(0.20)-5.37(0.44)	10	0.56
	<b>G1</b>	0.48(0.12)	0.78(0.14)	0.00	0.16(0.15)	0.17(0.15)	0.55(0.26)		0	0.68
<b>DINAGZ</b>	<b>G0S</b>	0.51(0.31)	0.00	0.00	0.74(0.45)	0.00	0.42(0.43)	1.08(0.14)-3.88(0.35)	0	0.43
	<b>G1</b>	0.47(0.17)	0.65(0.16)	0.06(0.19)	0.14(0.15)	0.21(0.15)	0.44(0.25)		5	0.66
<b>MFU (x10<sup>2</sup>)</b>	<b>G0S</b>	0.44(0.24)	0.00	0.19(0.13)	0.27(0.17)	0.06(0.12)	0.00	0.12(0.02)-7.97(0.55)	30	0.56
	<b>G1</b>	0.49(0.09)	0.36(0.09)	0.00(0.09)	0.20(0.07)	0.13(0.07)	0.11(0.09)		0	0.62
<b>NDF</b>	<b>G0S</b>	0.46(0.22)	0.00	0.16(0.13)	0.01(0.13)	0.12(0.13)	0.00	0.11(0.00)-5.49(0.38)	25	0.62
	<b>G1</b>	0.52(0.09)	0.33(0.09)	0.06(0.09)	0.09(0.06)	0.05(0.06)	0.10(0.08)		6	0.70
<b>LIGN (x10<sup>2</sup>)</b>	<b>G0S</b>	0.33(0.67)	1.42(0.82)	0.00	2.79(1.31)	1.09(1.16)	0.21(0.78)	1.02(0.16)-23.93(1.84)	0	0.47
	<b>G1</b>	1.16(0.31)	0.77(0.27)	0.00	2.45(0.47)	1.75(0.46)	0.66(0.59)		0	0.48
<b>CELL</b>	<b>G0S</b>	3.81(3.08)	5.12(2.86)	0.00	0.00(2.00)	0.00	0.00	2.73(0.38)-505.18(32.48)	0	0.30

Supplementary materials. Chapter2

<b>(x10<sup>2</sup>)</b>	<b>G1</b>	1.63(0.99)	3.61(0.99)	1.28(1.10)	1.23(0.69)	0.00	0.00		20	0.24
<b>HCELL</b>	<b>G0S</b>	3.26(3.16)	6.65(3.62)	0.12(1.81)	1.32(2.78)	0.00	0.00		1	0.30
<b>(x10<sup>2</sup>)</b>	<b>G1</b>	6.18(1.61)	3.66(1.60)	1.25(1.61)	3.58(1.42)	1.28(1.22)	0.00	2.60(0.39)-418.84(27.37)	11	0.32

<sup>a</sup> Minimum and maximum residual variance across all environments

<sup>b</sup> Percentage of SCA variance computed as  $\frac{\sigma_{SCA}^2}{\sigma_{GCA_d}^2 + \sigma_{GCA_f}^2 + \sigma_{SCA}^2} \times 100$

<sup>c</sup> Broad-sense heritability

<sup>d</sup> Standard error in bracket

**Table S2** 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.

Trait	Ls-means												Genetic Gain <sup>b</sup>	
	Commercial <sup>a</sup>			Founder <sup>a</sup>			G0S <sup>a</sup>			G1 <sup>a</sup>			G0S <sup>a</sup>	G1 <sup>a</sup>
	Mean	Min-Max	Sd <sup>c</sup>	Mean	Min-Max	Sd	Mean	Min-Max	Sd	Mean	Min-Max	Sd		
<b>DMY</b>	<b>17.96</b>	17.18-18.74	1.10	<b>15.80</b>	14.27-17.47	0.84	<b>17.35</b>	14.82-18.89	0.92	<b>17.33</b>	14.33-19.72	0.85	1.55	1.52
<b>DMC</b>	<b>34.69</b>	34.37-35.00	0.45	<b>34.10</b>	30.53-37.48	1.94	<b>33.33</b>	30.54-37.13	1.40	<b>33.43</b>	29.39-39.18	1.63	-0.77	-0.67
<b>DtSilk</b>	<b>201.89</b>	200.27-203.50	2.28	<b>203.14</b>	199.80-206.16	1.50	<b>204.97</b>	201.80-208.79	1.82	<b>205.04</b>	197.98-211.40	1.97	1.83	1.90
<b>PH</b>	<b>246.31</b>	239.86-252.76	9.12	<b>245.11</b>	222.10-279.19	13.44	<b>259.52</b>	235.54-283.59	10.40	<b>257.50</b>	227.27-286.97	10.15	14.41	12.40
<b>DINAG</b>	<b>51.14</b>	50.71-51.57	0.61	<b>51.03</b>	47.48-54.84	2.09	<b>49.73</b>	47.54-52.63	1.35	<b>49.61</b>	45.08-53.35	1.37	-1.30	-1.42
<b>DINAGZ</b>	<b>43.13</b>	42.56-43.70	0.81	<b>42.98</b>	39.75-46.35	1.91	<b>42.22</b>	39.81-45.17	1.29	<b>42.03</b>	38.12-45.80	1.36	-0.75	-0.94
<b>MFU</b>	<b>95.54</b>	95.34-95.74	0.28	<b>95.28</b>	91.13-98.14	2.05	<b>93.17</b>	89.04-97.06	1.56	<b>93.13</b>	87.59-101.36	1.71	-2.11	-2.15
<b>NDF</b>	<b>41.10</b>	40.73-41.48	0.53	<b>41.34</b>	39.88-42.16	0.63	<b>42.07</b>	39.19-43.94	0.95	<b>42.11</b>	38.27-46.39	1.12	0.73	0.77
<b>LIGN</b>	<b>5.41</b>	5.37-5.46	0.07	<b>5.73</b>	5.27-6.41	0.32	<b>5.86</b>	5.20-6.26	0.23	<b>5.89</b>	5.28-7.00	0.23	0.13	0.17
<b>CELL</b>	<b>51.76</b>	51.51-52.01	0.35	<b>51.64</b>	50.85-52.40	0.50	<b>52.07</b>	50.92-53.34	0.48	<b>52.10</b>	50.55-53.91	0.46	0.44	0.46
<b>HCELL</b>	<b>42.90</b>	42.50-43.30	0.57	<b>42.70</b>	41.50-43.56	0.72	<b>42.04</b>	40.69-43.56	0.61	<b>42.01</b>	39.64-43.97	0.63	-0.66	-0.69

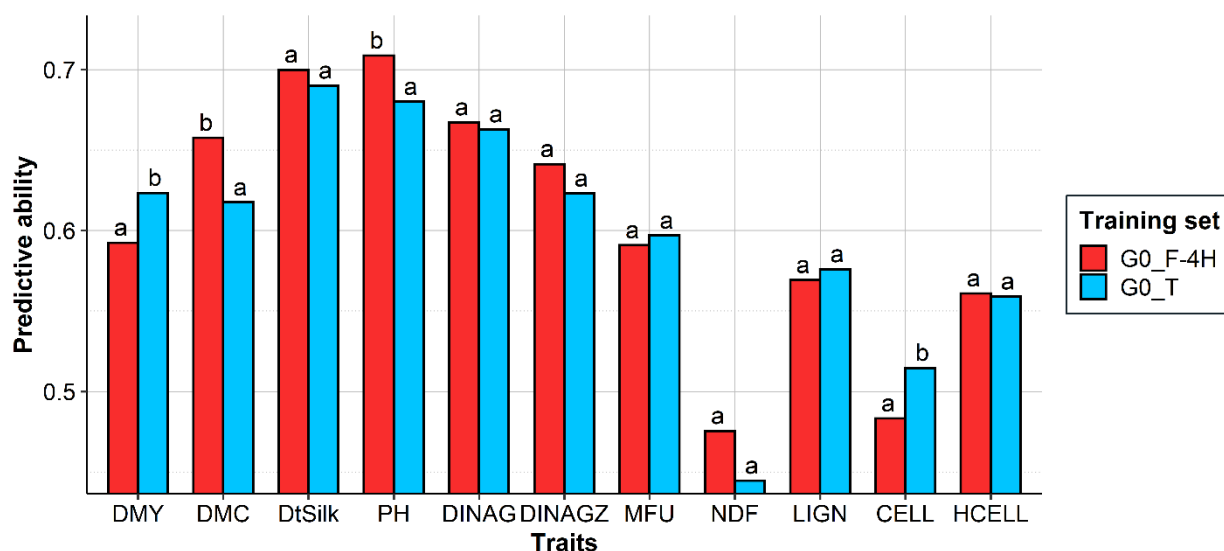
<sup>a</sup> Hybrid type in the experimental design comprise hybrids used as controls (commercial or founder hybrids) and experimental hybrids from the first generation (G0S) or the new generation (G1)

<sup>b</sup> Genetic gain computed as the difference between the mean performance of the experimental hybrids and the mean performance of the founder hybrids

<sup>c</sup> Standard deviation

**Table S3** Broad-sense heritability ( $H^2$ ), percentage of genetic variance assigned to SCA variance (%SCA) estimated for all experimental designs of the study.

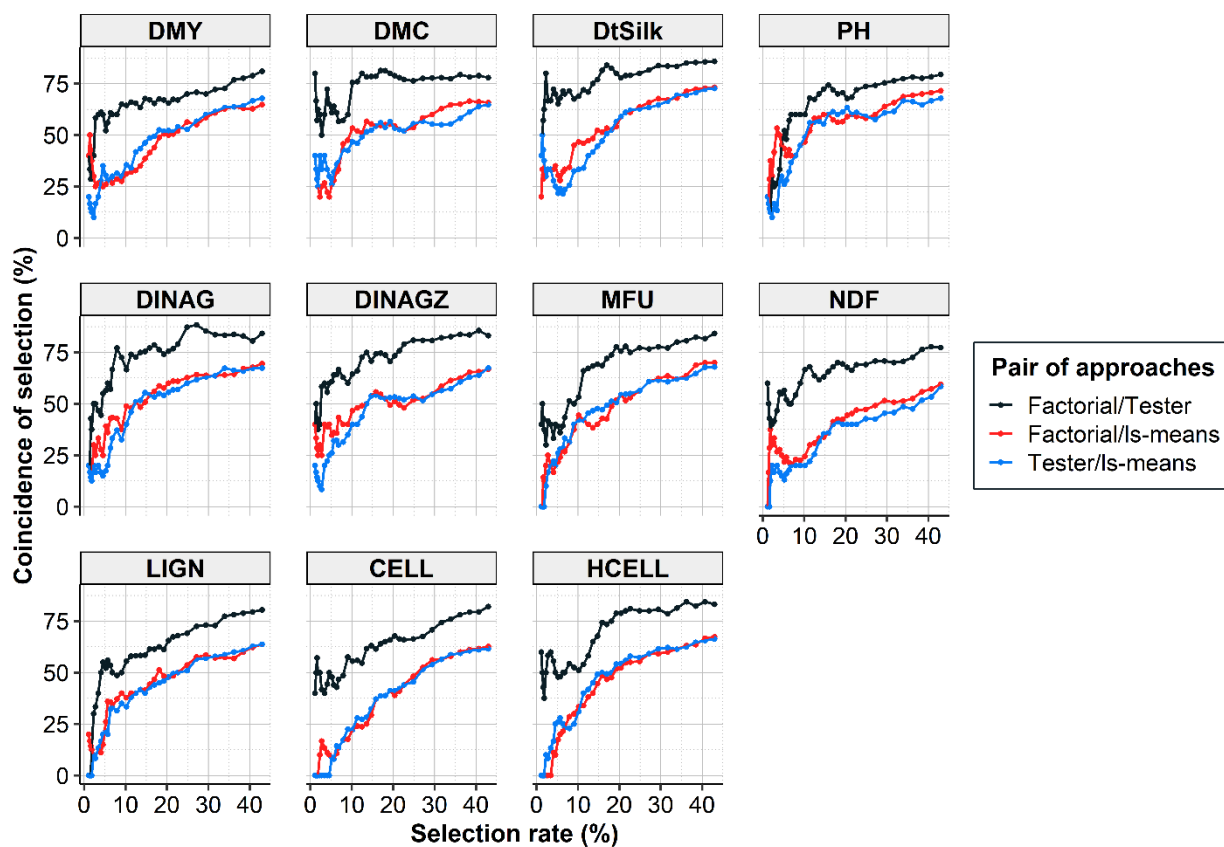
Experimental	Component	DMY	DMC	DtSilk	PH	DINAG	DINAGZ	MFU	NDF	LIGN	CELL	HCELL
<b>G0_F-1H</b>	$H^2$	0.84	0.91	0.91	0.90	0.85	0.79	0.79	0.72	0.89	0.78	0.84
	%SCA	20	15	17	15	0	0	10	13	0	3	0
<b>G0_F-4H</b>	$H^2$	0.89	0.93	0.93	0.92	0.87	0.78	0.86	0.80	0.86	0.57	0.73
	%SCA	9	6	6	3	2	4	3	6	0	3	3
<b>G0_T-D</b>	$H^2$	0.86	0.86	0.84	0.90	0.87	0.79	0.81	0.68	0.82	0.55	0.72
	%SCA	20	14	11	5	5	3	11	25	8	15	12
<b>G0_T-F</b>	$H^2$	0.85	0.93	0.87	0.87	0.77	0.67	0.75	0.76	0.69	0.36	0.44
	%SCA	21	13	17	31	14	19	14	18	0	13	11
<b>(G0S+G1)_F-1H</b>	$H^2$	0.81	0.92	0.94	0.88	0.68	0.65	0.62	0.70	0.48	0.24	0.32
	%SCA	0	7	1	0	0	1	7	9	0	17	9
<b>G1_F-1H</b>	$H^2$	0.80	0.92	0.94	0.88	0.69	0.66	0.62	0.70	0.48	0.23	0.32
	%SCA	0	10	1	0	0	4	0	6	0	19	11



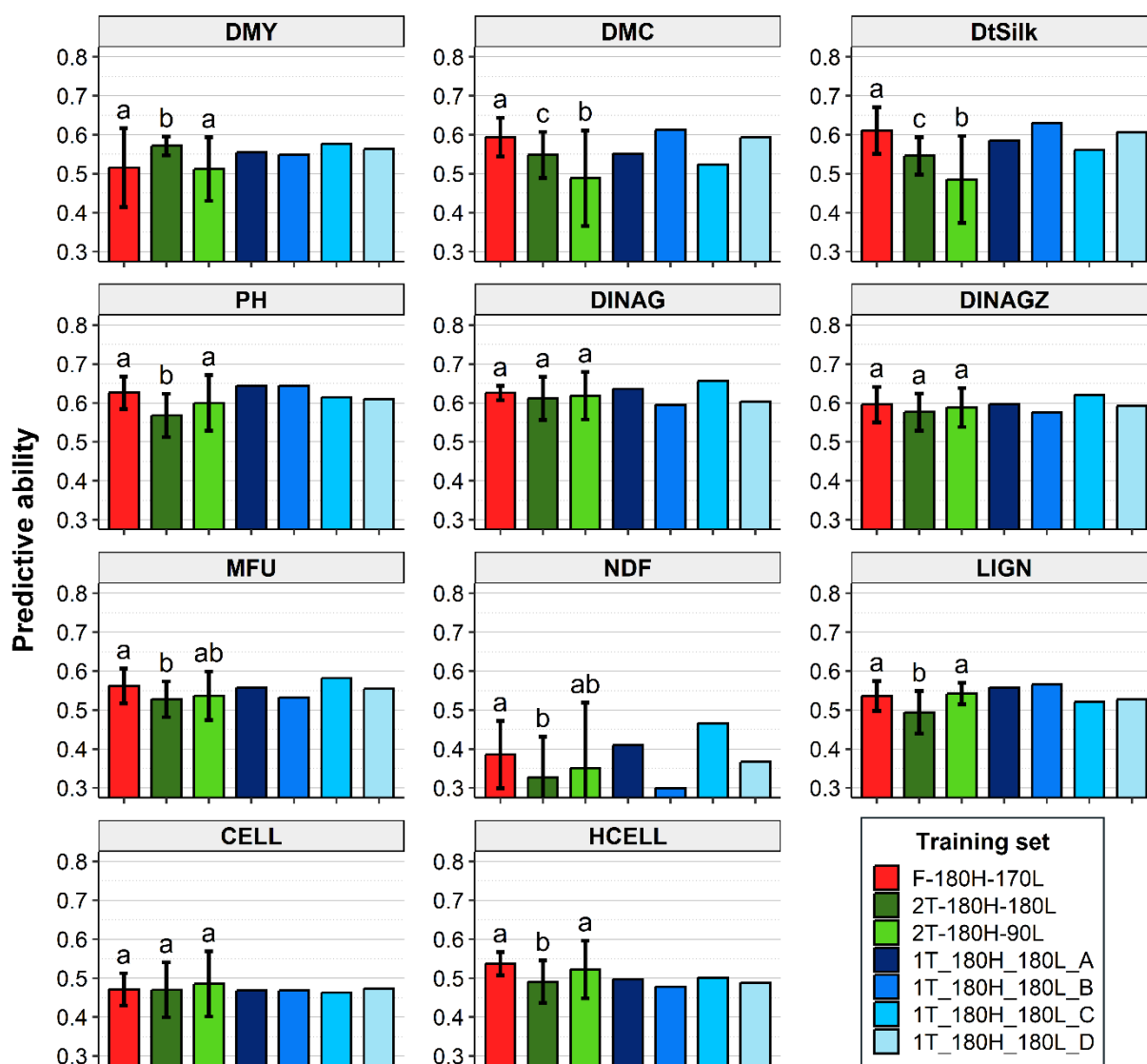
**Fig.S1** Predictive abilities obtained for the G1 hybrids (442) by training the GS model on the G0\_F-4H (363) or the G0\_T (360) TRSs in Scenario 2b. Williams tests were performed ( $\alpha=0.05$ ) and significant differences were indicated with letters: two different letters indicate a significant difference and at least one common letters indicate no significant difference.

**Table S4** Correlation between the GCA BLUPs predicted for the G1 single-cross hybrids from the (G0S+G1)\_F-1H design when using the G0\_F-4H or the tester designs as TRS.

Traits	Predicted GCA BLUPs		Predicted hybrid value
	Dent	Flint	
<b>DMY</b>	0.90	0.84	0.87
<b>DMC</b>	0.85	0.92	0.87
<b>DtSilk</b>	0.86	0.94	0.91
<b>PH</b>	0.90	0.77	0.85
<b>DINAG</b>	0.94	0.83	0.91
<b>DINAGZ</b>	0.93	0.79	0.90
<b>MFU</b>	0.94	0.85	0.89
<b>NDF</b>	0.86	0.78	0.81
<b>LIGN</b>	0.89	0.83	0.87
<b>CELL</b>	0.88	0.80	0.85
<b>HCELL</b>	0.91	0.81	0.88

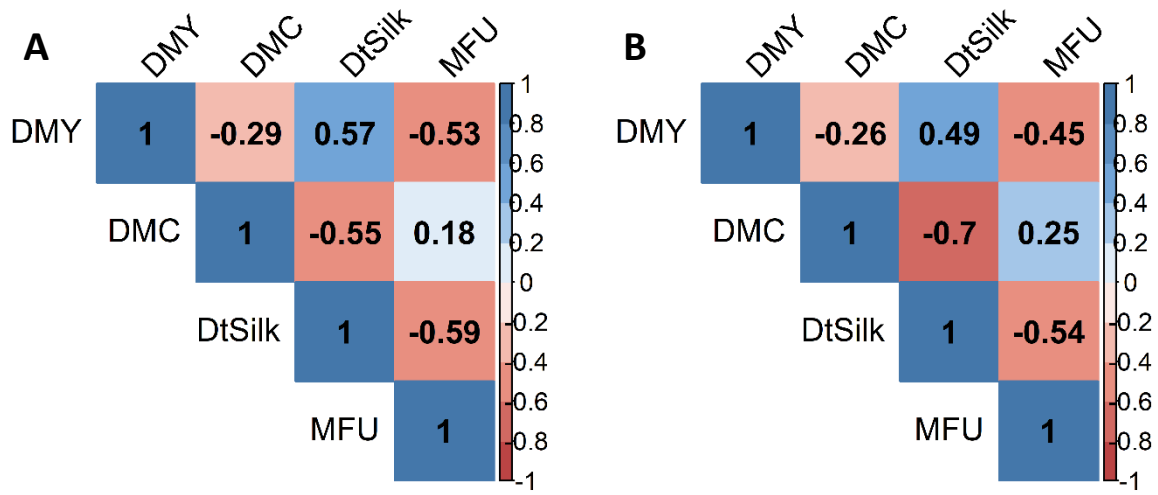


**Fig.S2** Coincidence of selection of the G1 single-cross hybrids (442) computed for different selection rates (%). The coincidence of selection was computed in black between the factorial and the tester approaches (GBLUP), in red between the factorial approach (GBLUP) and the phenotypic values (hybrid Is-means) and in blue between the tester approach (GBLUP) and the phenotypic values (hybrid Is-means). The BLUPs estimated in Scenario 2a (Figure S1) were used to compute the coincidence of selection.



**Fig.S3** 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 ( $\alpha=0.05$ ) 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.S4** Phenotypic correlation between traits in the G0\_F-1H design (A) or in the G1 hybrids (B).

## Supplementary materials. Chapter 3

### Chapter 3. Shortening maize hybrid breeding schemes using genomic predictions calibrated on factorial instead of tester designs improves cost efficiency over cycles.

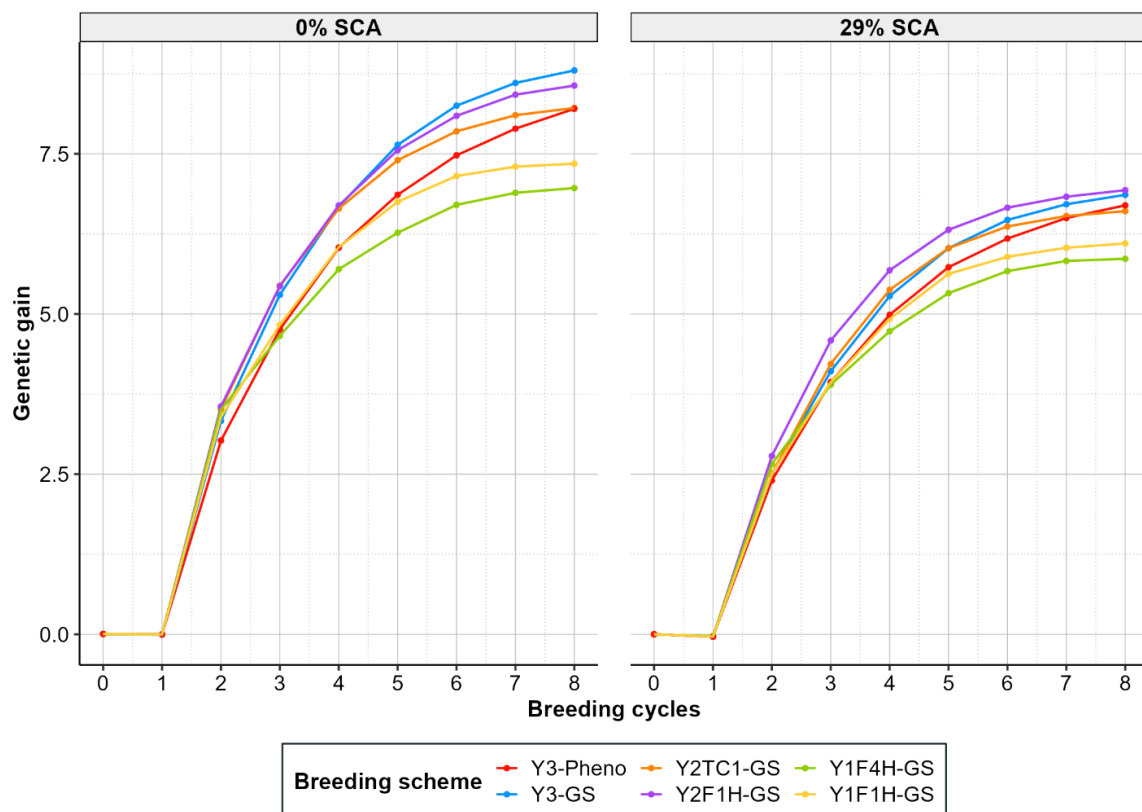
Alizarine Lorenzi<sup>1,2</sup>, Gaëtan Touzy<sup>2</sup>, Cyril Bauland<sup>1</sup>, Colin Guillaume<sup>3</sup>, Tristan Mary-Huard<sup>1,4</sup>, Alain Charcosset<sup>1</sup>, Laurence Moreau<sup>1</sup>

<sup>1</sup> Université Paris-Saclay, INRAE, CNRS, AgroParisTech, Génétique Quantitative et Evolution - Le Moulon, 91190 Gif-sur-Yvette, France

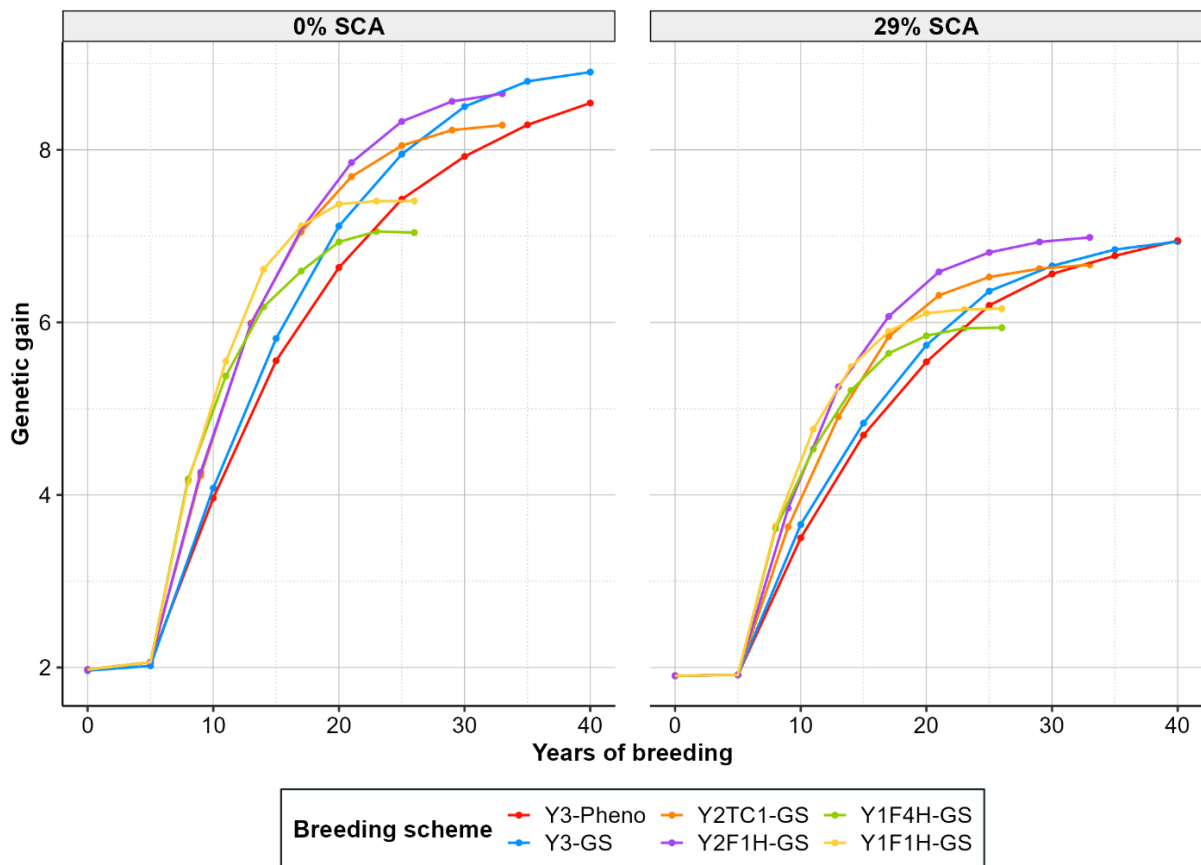
<sup>2</sup> RAGT2n, Genetics and Analytics Unit, 12510 Druelle, France

<sup>3</sup> Maisadour Semences SA, F-40001 Mont-de-Marsan Cedex, France

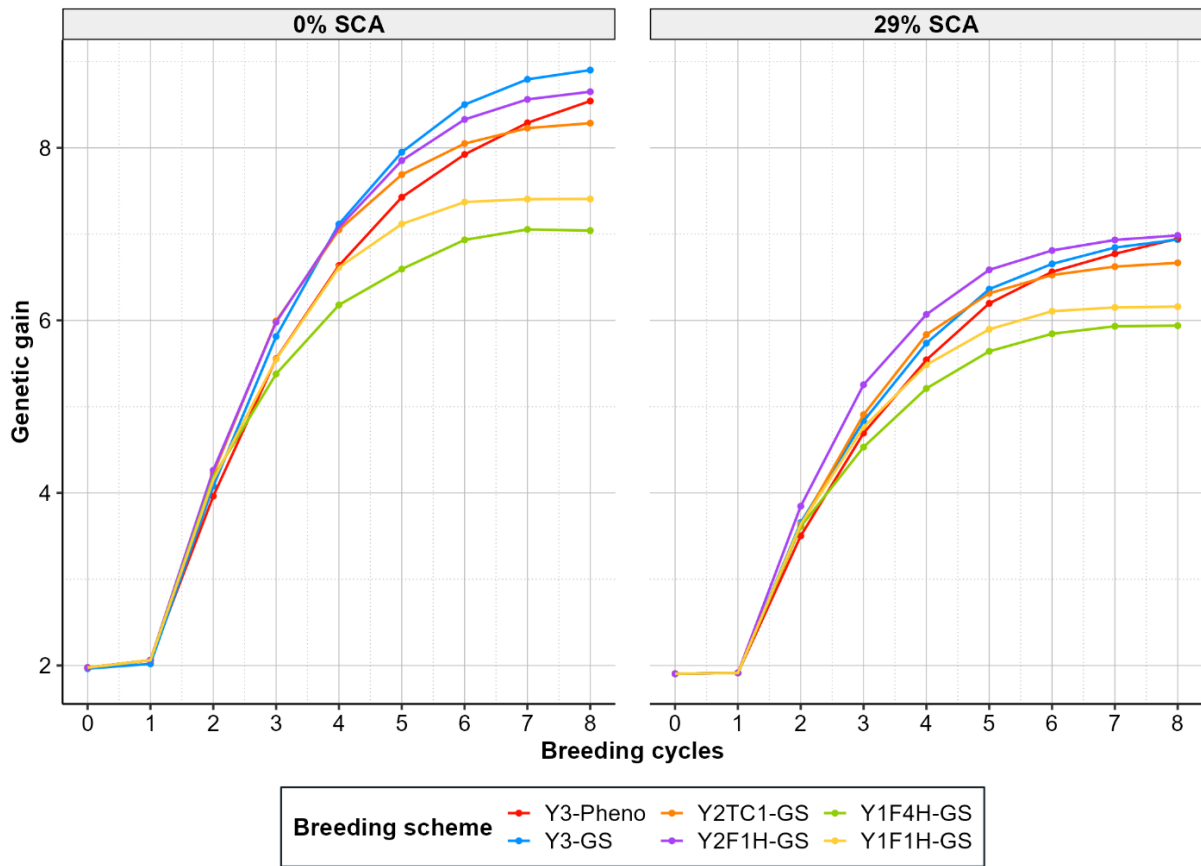
<sup>4</sup> MIA, INRAE, AgroParisTech, Université Paris-Saclay, 75005, Paris, France



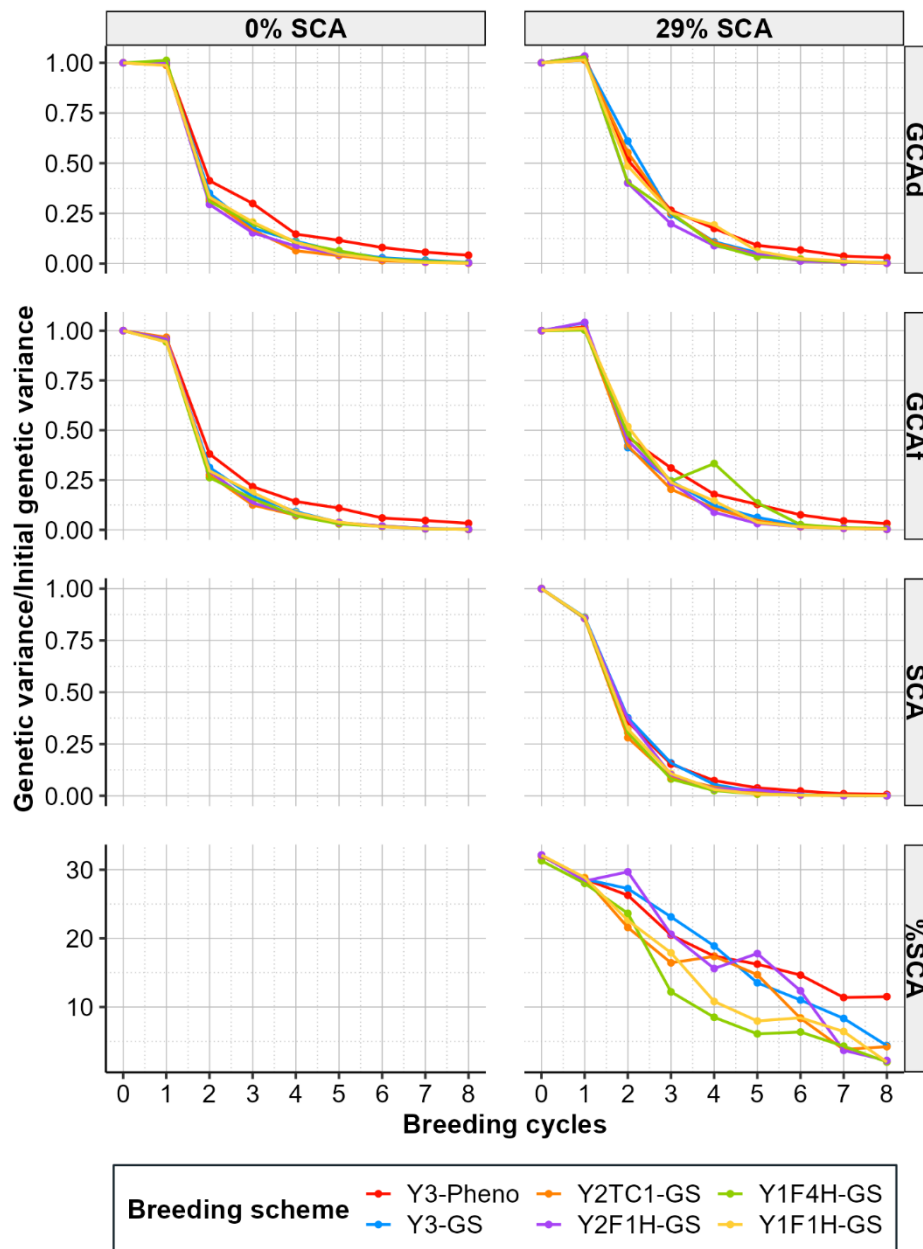
**Fig.S1** Evolution of the genetic gain along breeding cycles of all BSs for two SCA scenarios, with 0% or 29% SCA in the initial population. Genetic gain is expressed as the difference between the mean hybrid genetic value of the hybrids and the mean hybrid genetic value in the initial population divided by the standard deviation in the initial population. Each point represents the overall mean of 10 replicates. Year 0 corresponds to the initial population, year 5 corresponds to the initialization and the breeding starts after year 5.



**Fig.S2** Evolution of the genetic gain of the top ten hybrids over time of all BSs in two SCA scenarios, with 0% or 29% SCA in the initial population. Genetic gain is expressed as the difference between the mean hybrid genetic value of the hybrids and the mean hybrid genetic value in the initial population divided by the standard deviation in the initial population. Each point represents the overall mean of 10 replicates. Year 0 corresponds to the initial population, year 5 corresponds to the initialization and the breeding starts after year 5.



**Fig.S3** Evolution of the genetic gain of the top ten hybrids along breeding cycles of all BSs in two SCA scenarios, with 0% or 29% SCA in the initial population. Genetic gain is expressed as the difference between the mean hybrid genetic value of the hybrids and the mean hybrid genetic value in the initial population divided by the standard deviation in the initial population. Each point represents the overall mean of 10 replicates. Year 0 corresponds to the initial population, year 5 corresponds to the initialization and the breeding starts after year 5.



**Fig.S4** Evolution of the genetic variance decomposed into its GCA and SCA components and percentage of SCA variance over the total genetic variance for all breeding scheme strategies when the SCA variance in the initial population equals 0 or 29%. Genetic variance is expressed as the ratio between the genetic variance in the current population and the genetic variance in the initial population. Each point represents the overall mean of 10 replicates. The breeding cycle 0 corresponds to the initialization, and the breeding starts at breeding cycle 1.

