ARTHUR MARTINS ALMEIDA BERNARDELI

GENOME WIDE SELECTION OPTIMIZATION IN MAIZE BREEDING

Thesis submitted to the Genetics and Breeding Graduate Program of the Universidade Federal de Viçosa in partial fulfillment of the requirements for the degree of *Doctor Scientiae*.

Adviser: Aluízio Borém de Oliveira

VIÇOSA - MINAS GERAIS
2023

Ficha catalográfica elaborada pela Biblioteca Central da Universidade Federal de Viçosa - Campus Viçosa

T
B522g
2023

Bernardeli, Arthur Martins Almeida, 1993-
Genome wide selection optimization in maize breeding / Arthur Martins Almeida Bernardeli. – Viçosa, MG, 2023.
1 tese eletrônica (73 f.): il. (algumas color.).

Texto em inglês.
Orientador: Aluízio Borém de Oliveira.
Tese (doutorado) - Universidade Federal de Viçosa, Departamento de Agronomia, 2023.
Referências bibliográficas: f. 36-42.
DOI: https://doi.org/10.47328/ufvbbt.2023.173
Modo de acesso: World Wide Web.

1. Milho - Melhoramento genético. 2. Milho - Seleção. 3. Marcadores genéticos. 4. Milho híbrido. I. Oliveira, Aluízio Borém de, 1959-. II. Universidade Federal de Viçosa. Departamento de Agronomia. Programa de Pós-Graduação em Genética e Melhoramento. III. Título.

CDD 22. ed. 631.153

Bibliotecário(a) responsável: Bruna Silva CRB-6/2552

ARTHUR MARTINS ALMEIDA BERNARDELI

GENOME WIDE SELECTION OPTIMIZATION IN MAIZE BREEDING

Thesis submitted to the Genetics and Breeding Graduate Program of the Universidade Federal de Viçosa in partial fulfillment of the requirements for the degree of *Doctor Scientiae*.

APPROVED: March 31, 2023.

Assent:

Documento assinado digitalmente
ARTHUR MARTINS ALMEIDA BERNARDELI
Data: 13/04/2023 16:58:53-0300
Verifique em https://validar.iti.gov.br

Arthur Martins Almeida Bernardeli
Author

Documento assinado digitalmente
ALUIZIO BOREM DE OLIVEIRA
Data: 13/04/2023 17:14:31-0300
Verifique em https://validar.iti.gov.br

Aluízio Borém de Oliveira
Adviser

To God, to my wife, parents, sister, grandparents, aunts, and beloved ones.

ACKNOWLEDGEMENTS

To God.

To my wife.

To my parents.

To my sister.

To my grandparents, aunts, and the other beloved ones.

To the Universidade Federal de Viçosa, for the opportunity to complete my PhD. research.

To Aluízio Borém, for believing in and encouraging my career.

To Diego Jarquin for providing means of improving my career.

To EMBRAPA, and fellows from EMBRAPA - Maria Marta, Lauro Guimarães, José Henrique Soller Guilhen, and Isadora Oliveira -, for the research mentoring.

To UFV Genetics and Breeding Program and fellows, for the friendship and scientific contributions.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES and CAPES-PRINT program). The scholarship was granted by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

“Ask, and it shall be given you; seek, and you shall find;
knock, and it shall be opened unto you.”

Matthew 7:7

ABSTRACT

BERNARDELI, Arthur Martins Almeida, D.Sc., Universidade Federal de Viçosa, March, 2023. **Genome Wide Selection Optimization in Maize Breeding**. Adviser: Aluízio Borém de Oliveira.

Maize is a staple crop and the most grown cereal worldwide. The expansion of this crop was possible due to efforts in management and breeding. In the breeding standpoint, advances were achieved in the release of hybrids presenting heterosis, field experimental design and analyses, establishment of heterotic patterns, and effective seed production and marketing. From the last decade on, advances in data analyses benefited from the surge of genotypic data, allowing the prediction of hybrids without being tested through genomic selection approaches. This study aims to convert a high-density SNP data set and use it in a genomic selection or predicting non-tested hybrids and non-observed environments, and for indicating most promising mating parent material for obtaining hybrids and inbred lines for ASI, EPP, FFT, GY, and MFT maize traits. For that, we ranked the SNPs according to their effects from a ME analyses and selected the minimum portion of markers that reached the plateau of prediction accuracy per chromosome, followed by eliminating the repeated markers between traits, and removing the ones tightly linked according to LD analyses. For the GS of hybrids and environments, three methods that comprised GCA and SCA main and interaction effects were fitted, and the prediction accuracy was assessed. The step of selecting parent material was performed according to PS, GS, and GM. The GM methods used the marker effects predicted in the previous GS step, and the 40 top- and bottom-performing crosses and their respective parent lines were selected for each trait. The selected SNPs maintained the accuracy for all traits under drought or well-watered conditions when compared to using full SNP set. For GWS of hybrids, Model 3 performed better for all traits when cross validation schemes had information of all environments (CV1 and CV2) in terms of prediction accuracy, and Model 2 performed better when there was missing information about environments (CV0 and CV00). The mating parents chosen for positive selection were different than the ones from negative selection, ensuring maximization of gains for hybrid and inbred lines development. The highest coincidences of selected parent lines occurred in GS-based methods (Methods 1, 3, 5, 7, 9, 11, 13, and 15), where parents were directly selected based on means or GCA/SCA (and interaction) values of their respective hybrids. The methods based on crosses simulations (Methods 2, 4, 6, 8, 10, 11, 12, 14, and 16) had moderate to low coincidences, but were consistent in indicating the best parent materials overall. GS- and GM-based parent selection results must be further compared to Method 17 (observed crosses) for an effective validation. PS, GS, and GM methods together must help in the decision making of selecting parent material for future crosses. These approaches must be further performed using other training populations.

Keywords: Cross-validation. Hybrids. Inbred Lines. Prediction Accuracy. SNP Markers.

RESUMO

BERNARDELI, Arthur Martins Almeida, D.Sc., Universidade Federal de Viçosa, março de 2023. **Genome Wide Selection Optimization in Maize Breeding**. Orientador: Aluízio Borém de Oliveira.

O milho é o cereal mais cultivado em todo o mundo. A expansão dessa cultura foi possível devido aos esforços de manejo e melhoramento. Do ponto de vista do melhoramento, avanços foram alcançados na liberação de híbridos com heterose, delineamento e análises experimentais em campo, estabelecimento de padrões heteróticos e produção e comercialização efetiva de sementes. A partir da última década, avanços na análise de dados foram beneficiados pelo surgimento de dados genotípicos, permitindo a predição de híbridos sem serem testados por meio de abordagens de seleção genômica. Este estudo tem como objetivo converter um conjunto de dados de SNP de alta densidade e usá-lo em seleção genômica para predizer híbridos não testados e ambientes não observados, e indicar o material parental de acasalamento mais promissor para a obtenção de híbridos e linhagens para os caracteres ASI, EPP, FFT, GY e MFT em milho. Para isso, SNPs foram classificadas de acordo com seus efeitos estimados, e uma porção mínima de marcadores que atingiram o platô de acurácia de predição por cromossomo foram selecionados, seguida de uma eliminação de marcadores repetidos entre as características e remoção daqueles ligados de acordo com uma análise de LD. Para a GWS de híbridos e ambientes, três modelos que compreenderam os efeitos principais e de interação de GCA e SCA foram ajustados e a acurácia de predição foi avaliada. A etapa de seleção do material de parental foi realizada de acordo com PS, GS e GM. Os métodos de GM usaram os efeitos de marcadores preditos na etapa anterior de GS, e os 40 cruzamentos de desempenho superior e inferior e suas respectivas linhagens parentais foram selecionados para cada característica. Os SNPs selecionados mantiveram a acurácia de predição para todas as características em condição de déficit hídrico e irrigação, quando comparado ao uso de todos os marcadores disponíveis. Para GS de híbridos, o Modelo 3 teve melhor desempenho para todas as características quando os esquemas de validação cruzada tinham informações de todos os ambientes (CV1 e CV2) em termos acurácia, e o Modelo 2 teve melhor desempenho quando faltavam informações sobre os ambientes (CV0 e CV00). Os genitores escolhidos para seleção positiva foram diferentes dos genitores para seleção negativa, garantindo a maximização dos ganhos para o desenvolvimento de híbridos e linhagens endogâmicas. As maiores coincidências de linhagens parentais selecionadas ocorreram em métodos baseados em GS (Métodos 1, 3, 5, 7, 9, 11, 13 e 15), onde os pais foram selecionados diretamente com base em médias ou valores de GCA/SCA (e interação) de seus respectivos híbridos. Os métodos baseados em simulações de cruzamentos (Métodos 2, 4, 6, 8, 10, 11, 12, 14 e 16) tiveram coincidências moderadas a baixas, mas foram consistentes em indicar os melhores materiais de parentais em geral. Os resultados da seleção de pais baseados em GS e GM devem ser comparados com o Método 17 (cruzamentos observados) para uma validação eficaz. Os métodos PS, GS e GM juntos devem ajudar na tomada de decisão de seleção de material parental para futuros cruzamentos.

Palavras-chave: Acurácia de Predição. Híbridos. Linhagens Endogâmicas. Marcadores SNP. Validação Cruzada.

LIST OF ACRONYMS AND ABBREVIATIONS

ASI	anthesis-silking interval
BLUE	best linear unbiased estimation/estimators
BLUP	best linear unbiased prediction/predictor
CV	cross-validation
EPP	ears per plot
FFT	female flowering time
G-BLUP	genomic best linear unbiased prediction
GCA	general combining ability
GCAxE	general combining ability by environment interaction
GE	genotype by environment interaction
GY	grain yield
GWAS	genome wide association study
GWS	genome wide selection
J10D	Janaúba location in 2010, dought stress condition
J10W	Janaúba location in 2010, well-watered stress condition
J11D	Janaúba location in 2010, dought stress condition
J11W	Janaúba location in 2010, well-watered stress condition
LD	linkage disequilibrium
ME	marker effect estimation
MFT	male flowering time
MSE	mean square error
PCoA	principal coordinate analysis
RFLP	restriction fragment length polymorphism
RR-BLUP	ridge-regression best linear unbiased prediction
SCA	specific combining ability
SCAxE	specific combining ability by environment interaction
SD	standard deviation
SNP	single nucleotide polymorphism

LIST OF SYMBOLS

$r_{y\hat{y}}$	prediction accuracy
$\hat{y}$	predicted value

SUMMARY

1 INTRODUCTION

Sharp increments in maize yield were first observed right after the elucidation of heterosis and the release of the first single cross hybrids (Shull, 1948). The proposed genetic designs in corn breeding as the diallel crosses, test crosses, and other family-based tests assisted in estimating the genetic effects of inbred lines evaluated through their hybrids, facilitating the process of parental screening. Added to that, statistical analyses were adapted to the routine of large-scale breeding programs allowing to predict genetic effects and estimate variances even with the existence of missing data, keeping selection accuracy similar or superior to the former least-squares based analyses that were commonly performed. Among those, we can cite the Best linear unbiased prediction (BLUP; Henderson, 1975) of genetic effects and restricted most-likelihood method (REML; Thompson, 2008) for assessing variances. These analytical progresses ensured positive selection gains over decades in maize breeding, which was boosted using genome-based approaches as assistive tool to hybrid selection.

Genomic selection (GS) has gained space ever since its advantages were shown by Meuwissen et al. (2001). Years before, Bernardo showed an adaption of the common BLUP with the replacement of the pedigree-based relationship matrix for the genome-based one using RFLP markers Bernardo (1994). By doing that, breeders would be able to predict the performance of yet-to-test or non-observed hybrids preserving an adequate accuracy and allowing to relocate resources to other steps of the hybrid development pipeline. These have become more evident once the genotyping processes have become cost- and time-efficient. To date, GS studies in maize have focused on several subjects besides predicting non-observed hybrids, such as testing different settings of training sets according to (un)observed parent lines (Technow, 2019), or cross validations schemes that assess accuracy of predicting a non-observed environments or non-tested hybrids (Jarquin et al., 2021) . Most of these studies have used moderate to high number of molecular markers, as in the studies performed by (Dias et al., 2018). However, using a dense-genotyping dataset can compromise the efficacy of the genetic analyses, which can become more time consuming in computing the relationship matrix and the multi-environment phenotypic data, especially if the prediction model is not parsimonious, or if the purpose of the research involves obtaining simulated crosses (e.g., genomic mating; Akdemir and Sánchez, 2016).

In this regard, other studies have been performed in plant and animal breeding to reduce the dimensionality of the genotypic data maintaining the accuracy of results. Some of these studies were based on variance and levels of linkage (dis)equilibrium (Lado et al., 2017), ranking subset of markers according to their effect estimated through RR-BLUP (Resende et al., 2012; Sousa et al., 2019; DoVale et al., 2021, unpublished). These studies can also be applied in the maize case to make the genotyping step more efficient, and to make possible the exploit of other areas of research in maize, such as simulating crosses based on a previous empirical dataset.

It is known that crosses can be chosen according to the phenotypic or genomic estimated breeding values (GEBV), or through the covariance among parents. In all these cases, restrictions can be deployed based on pedigree to control inbreeding and to preserve gains and genetic diversity in the long-term. Akdemir and Sánchez (2016) proposed a look-ahead method of indicating mating pairs for a certain trait in major crops based on the combination of the marker covariates of two-parent lines to generate progenies in a certain generation after cycles of recombination. The crosses were indicated according to the means of the progenies for a certain trait, which is the product of the combined marker covariates and the effects of each locus previously estimated. Through this, it was stablished another assistive method for setting up nurseries besides the ones mentioned above, and a more elaborate list of parent material can be prepared. This is particularly important in the case of maize breeding, once crosses are numerous within heterotic groups (HG) for generating self of double haploid (DH) lines, and between HG for obtaining hybrids. Even though a previous screening of parent lines is performed, the remaining potential inbred lines are numerous, and it is sometimes impracticable to perform crosses two-by-two in every paired combination.

Considering the contributions from PS and GS to maize breeding and the current bottlenecks in improving this crop, this study aims to assess the efficiency of the conversion of a high-density to a low-density genotyping SNP data in predicting hybrids. Using the same genotyping dataset, we seek to elucidate the prediction of hybrids in a range of CV schemes and recommend crosses for obtaining hybrids and inbred lines using phenotypic, GS, and genomic mating approaches.

2 CORE IDEAS

i. Similar prediction accuracies can be achieved using reduced and full SNP datasets in corn breeding.

ii. Observed mean, SCA, and GCA as well as simulated crosses can assist the choice of mating pairs selection to obtain promising single-cross hybrids and inbred lines.

iii. Crosses prediction is paramount for nursery management in a corn breeding program.

3 MATERIALS AND METHODS

a. PLANT MATERIAL

188 female inbred lines representing Dent, Flint, and Common (unrelated to Dent and Flint) heterotic groups were crossed with two (2) male testers, resulting in 308 test cross hybrids. There were 85 female inbred lines belonging to the Dent group, 86 to the Flint group, and 17 to the Common group. Male testers 1 and 2 belonged to Dent and Flint groups, respectively. All the parent lines belong to the EMBRAPA corn breeding program in Brazil. The distinct heterotic groups are presented in the Principal Coordinated Analyses (PCoA) plot in Figure 1, and the mating pairs to obtain each of the test-cross hybrids are shown in Figure 2.

b. FIELD TRIALS

The 308 test-cross hybrids were tested in field trials in Janaúba (state of Minas Gerais, Brazil, coordinates, and elevation) in two subsequent years (2010 and 2011). The combination of location, year and watering treatment was considered as one distinct environment, as it follows: Janaúba in 2010 and 2011 under drought stress and well-watered conditions (J10D and J10W, respectively), and Janaúba in 2010 and 2011 under drought stress and well-watered conditions (J11D and J11W, respectively). The experimental unit consisted of a 4 m single row plot with final plant density of 16 seedlings, spaced 0.8 m apart to the adjacent plot. Each test was split in 6 sets due large number of hybrids, which were carried out under randomized complete block design with three (drought stress tests) and two (well-watered tests) replicates. Sets 1, 2, and 3 comprised the crosses made with Flint male tester; and 4, 5, and 6 the ones with crosses made with Dent male tester.

For the drought stress tests, the irrigation was interrupted days before the first flowering, and water stress treatment was performed during whole flowering and seed filling. For the well-watered tests, irrigation was performed to meet the water loss due to evapotranspiration based on Penman-Monteith equation according to each phonologic stage of the crop. The soil water content was monitored through DIVINER 2000® probe (Sentek Sensor Technologies, Australia) up to 0.7 m depth.

Notes on anthesis-silking interval (ASI, interval between male and female flowering in days), number of ears per plot (EPP), days to female flowering time (FFT), grain yield (GY, in metric tons per hectare adjusted to 13% grain moisture), and days to male flowering time (MFT) were taken. FFT and MFT notes were taken when at least 50% of the plants in the experimental unit have emerged silk or shedded pollen. More detailed information about plant material and field trials can be found in the study published by Dias et al. (2018).

c. GENOMIC DATA

Tissue samples of young leaves of inbred lines were collected, and genomic DNA was extracted according to the cetyl trimethylammonium bromide method (CTAB; Saghai-Maroof et al., 1984). Fluorometer Qubit® 2.0 (Life Technologies^{TM}, USA) was used to quantify DNA samples. After diluting DNA to desirable concentration, samples were run on a 1% agarose gel using tris-acetate-EDTA buffer and GelRed^{TM} (Biotium, USA) stain, and recorded through the Imager Gel Doc L-PIX (Loccus Biotecnologia, Brazil). DNA samples were sent to Cornell University Genomic Diversiy Facility (Ithaca, NY, USA) to undergo genotyping-by-sequencing (GBS; Elshire et al., 2011) process by the use of ApeKI restriction enzyme. Sequence tags were aligned to the B73 maize reference genome (AGPv3) USING Burrows-Wheeler alignment tool (Li and Durbin 2009). Tassel version 5 (Glaubitz et al., 2014) was used for SNP calling through GBS pipeline. For quality control of genomic data, SNPs were discarded if minor allele frequency (MAF) was inferior to 5%, missing genotypes per SNP were more than 20%, or if heterozygous genotypes were superior to 5%. Following, NPUTE (Roberts et al., 2007) software was used to impute missing data. 57,294 was the final number of SNPs after filtering. The genotypic data of test-cross hybrids was obtained after synthetic virtual crosses between female inbred lines and male testers. More detailed information about genomic data can be found in the study published by Dias et al. (2018).

d. STATISTICAL ANALYSES

Field data quality control and estimation of genetic parameters

The statistical model for randomized complete block design within environments was $Y_{ij} = \mu + \beta_j + g_i + \varepsilon_{ij}$ (Model 1). Y_{ij} was the observed data for hybrids i (i=1,2,…,I) within blocks j (j=1,2,…,J), μ is the overall mean of observations, β_j is the random effect of blocks, $\beta_j \sim N(0, \sigma_\beta^2)$; g_i is the random effect of hybrids, $g_i \sim N(0, \sigma_g^2)$; and ε_{ij} is the random effect of experimental errors, $\varepsilon_{ij} \sim N(0, \sigma^2)$.

The quality control was performed to remove the outlier grain yield datapoints to ensure a tolerable coefficient of variation, defined as $CV = \frac{\sigma_p}{\mu}.100$ (%), with σ_p being the standard deviation of the phenotypic or observed data, and μ the overall mean. Grain yield data was adjusted for field experimental design using Model 1 through Best Linear Unbiased Estimation (BLUE; Henderson, 1975), considering g_i as fixed effect. The best linear unbiased estimates (BLUEs) of grain yield served as input data for further analyses. The variance components of random effects was estimated through restricted maximum likelihood (REML; Thompson, 2008), and computed to the broad-sense heritability on an entry mean basis formula, considering g_i as random effect as originally set for Model 1, being e and r the number of environments and replications, respectively, as follows:

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_{gE}^2}{er} + \frac{\sigma^2}{r}}$$

SNP markers selection

SNP markers from a high-density panel (57,294 SNPs) were selected to obtain a low-density panel through a marker effect estimation (ME) analysis based on a GWAS procedure, followed by a genome wide selection (GWS) analysis. First, the 308 hybrids were divided into 4 subpopulations of 77 individuals (subpopulations 1, 2, 3, and 4). ME analyses were performed over 231 individuals each time, corresponding to all combinations of 3 subpopulations, using the statistical model that accounts for relationship between hybrids.

$$Y_{ij} = \mu + \alpha + \alpha\gamma_i + g_j + g\gamma_{ij} + e_{ij}$$

being Y_{ij} the observed values for hybrids adjusted for field design (BLUEs); μ is the overall mean; α is the fixed effect of markers associated to hybrids; $\alpha\gamma_i$ is the fixed interaction term between markers (α) and years (γ_i); g_j is the random effect of hybrids, being $g_j \sim N(0, G\sigma_g^2)$; and $g\gamma_{ij}$ is the interaction term between hybrids and years, being $g\gamma_{ij} \sim N(0, (Z_g G Z_g') \circ (Z_\gamma Z_\gamma')\sigma_{g\gamma}^2)$. e_{ij} is the random effect of experimental errors, $e_{ij} \sim N(0, \sigma^2)$.

Following the estimation of each parameter, the SNP effects were calculated according to the formula $\widehat{m} = | \frac{\alpha}{\alpha + \alpha\gamma} |$. The SNPs were ranked in descending order according to $\widehat{m}$ estimates per chromosome for each subpopulation set. Posteriorly, GWS analyses were performed using each subpopulation set as training individuals and the remining subpopulation as testing individuals. This procedure ensured independence between the marker effects estimation in the ME step and the GWS cross-validation step.

For the GWS step, the G-BLUP model was used:

$$y_{ij} = \mu + \gamma_i + g_j + g\gamma_{ij} + e_{ij}$$

being y_{ij} the observed values for hybrids adjusted for field design (BLUEs); μ is the overall mean; γ_i is the fixed effect of years; g_j is the random effect of hybrids, being $g_j \sim N(0, G\sigma_g^2)$; and $g\gamma_{ij}$ is the random interaction term between hybrids and years, being $g\gamma_{ij} \sim N(0, (Z_g G Z_g') \circ (Z_\gamma Z_\gamma')\sigma_{g\gamma}^2)$. e_{ij} is the random effect of experimental errors, $e_{ij} \sim N(0, \sigma^2)$. The GWS analyses initiated using no kinship and was repeated over every 0.5% of ranked SNP markers per chromosome, until reaching the total number of markers, and the prediction accuracy was computed each run as the Pearson Correlation between predicted and observed values of hybrids according to the formula $r_{\hat{y}y} = cor(\hat{y}, y)$. The markers were selected per trait when the prediction accuracy from GWS that used the portion of ranked SNPs presented similar values to when the full set of markers was used to construct the additive kinship. One SNP was retained in cases that there were multiple selected SNPs every 10Kbp interval to shrink even more the number of selected SNPs. This physical distance was determined after performing a linkage disequilibrium (LD) analyses per chromosome (Supplemental File 1). Repeated selected SNPs for distinct traits were merged. A similar approach was performed by Sousa et al. (2019) for soybean breeding purposes, however, the authors estimated marker effects and cross-validated the predictions through ridge-regression best linear unbiased prediction (RR-BLUP) analyses.

For the steps involved in the SNP selection process, the package ASReml-R version 4.0 (Butler et al., 2017) was used to perform the analyses, and the additive genomic relationship matrices were calculated using the AGHmatrix package (Amadeu et al., 2016) in R software (R Core Team, 2022), following the formula $G = \frac{XX'}{p}$.

GWS model fitting and cross validation schemes

After obtaining the low-density SNP genotypic data, more complex GWS analyses in terms of statistical models and cross-validations schemes were performed for grain yield using the R package BGLR (Pérez and de los Campos, 2014). The models contain general and specific combining abilities (GCA and SCA, respectively) of parent inbred lines, as in the studies performed by Bernardo (1994), Technow et al. (2014), Kadam et al. (2016), Acosta-Pech et al. (2017). The GCA and SCA by environment interaction terms (GCAxE and SCAxE, respectively) were modeled according to the reaction norm procedure as shown by Jarquin et al. (2021). The following statistical models were used for GWS analyses:

Model 1: general combining ability model for main effects

This model includes the information about the parent inbred lines (P1 and P2) that resulted in the test-cross hybrids y_j for the environment i, corresponding to the general combining ability (GCA) of the parents, as follows:

$$y_{ij} = \mu + E_i + g_{P1_j} + g_{P2_j} + e_{ij}$$

being y_{ij} the observed values for hybrids; E_i is the random effect of environments, where $E_i \sim N(0, \sigma_E^2)$; g_{P1_j} and g_{P2_j} are the scores derived from each test-cross for inbred parent lines, where $g_{P1_j} = \Sigma_m^p x_{P1_{jm}} b_{mP1}$ and $g_{P2_j} = \Sigma_m^p x_{P2_{jm}} b_{mP2}$, $p = (i = 1,2, \dots, m)$, $X_{P1} = \{X_{P1_j}\}$ and $X_{P2} = \{X_{P2_j}\}$ are the marker incidence matrices for inbred parent lines, and b_{mP1} and b_{mP2} are the marker effects for the m^{th} marker such that $b_{mP1} \sim N(0, \sigma_{b\,P1}^2)$ and $b_{mP2} \sim N(0, \sigma_{b\,P2}^2)$, being $\sigma_{b\,P2}^2$ and $\sigma_{b\,P2}^2$ the variance components for the parent inbred lines, $g_{P1} = \{g_{P1_j}\} \sim N(0, G_{P1}\sigma_{P1_g}^2)$ and $g_{P1} = \{g_{P1_j}\} \sim N(0, G_{P1}\sigma_{P1_g}^2)$ with $G_{P1} = \frac{X_{P1}X'_{P1}}{p}$ and $G_{P2} = \frac{X_{P2}X'_{P2}}{p}$ and G_{P2}. $\sigma_{P1_g}^2 = p\ x\ \sigma_{b\,P1}^2$ and $\sigma_{P2_g}^2 = p\ x\ \sigma_{b\,P2}^2$ are the variance components of the inbred parent line effects, and e_{ij} is the random effect of experimental errors, $e_{ij} \sim N(0, \sigma^2)$.

Model 2: general and specific combining ability model for main effects

This model includes the information about the parent inbred lines (P1 and P2) and their crossing that resulted in the test-cross hybrids y_j for the environment i, corresponding to the general and specific combining abilities (GCA and SCA) of the parents and hybrids. SCA effect was modeled using the cell-by-cell product of the covariance structures from parent inbred lines, as follows:

$$y_{ij} = \mu + E_i + g_{P1_j} + g_{P2_j} + g_{P1_j x P2_j} + e_{ij}$$

being y_{ij} the observed values for hybrids; E_i is the random effect of environments, where $E_i \sim N(0, \sigma_E^2)$; g_{P1_j} and g_{P2_j} are the scores derived from each test-cross for inbred parent lines, where $g_{P1_j} = \Sigma_m^p x_{P1_{jm}} b_{mP1}$ and $g_{P2_j} = \Sigma_m^p x_{P2_{jm}} b_{mP2}$, $p = (i = 1,2, \dots, m)$, $X_{P1} = \{X_{P1_j}\}$ and $X_{P2} = \{X_{P2_j}\}$ are the marker incidence matrices for inbred parent lines, and b_{mP1} and b_{mP2} are the marker effects for the m^{th} marker such that $b_{mP1} \sim N(0, \sigma_{b\,P1}^2)$ and $b_{mP2} \sim N(0, \sigma_{b\,P2}^2)$, being $\sigma_{b\,P2}^2$ and $\sigma_{b\,P2}^2$ the variance components for the parent inbred lines, $g_{P1} = \{g_{P1_j}\} \sim N(0, G_{P1}\sigma_{P1g}^2)$ and $g_{P2} = \{g_{P2_j}\} \sim N(0, G_{P2}\sigma_{P2g}^2)$ with $G_{P1} = \frac{X_{P1}X'_{P1}}{p}$ and $G_{P2} = \frac{X_{P2}X'_{P2}}{p}$

and G_{P2}. $\sigma^2_{P1_g} = p\ x\ \sigma^2_{b\ P1}$ and $\sigma^2_{P2_g} = p\ x\ \sigma^2_{b\ P2}$ are the variance components of the inbred parent line effects, $g_{P1\ x\ P2} = \{g_{P1_j\ x\ P2_j}\} \sim N(0, G_{P1\ x\ P2}\sigma^2_{P1g\ x\ P2g})$, where $G_{P1\ x\ P2} = G_{P1} \circ G_{P2}$, being $\circ$ the Hadamard (cell-by-cell) product of two covariate matrices, and e_{ij} is the random effect of experimental errors, $e_{ij} \sim N(0, \sigma^2)$.

Model 3: general and specific combining ability model for main and interaction effects

Besides addressing SCA and GCA main effects, this model allows the prediction of hybrids performance by environment through modeling the interaction of GCA and SCA with environments using the reaction norm procedure, as follows:

$$y_{ij} = \mu + E_i + g_{P1_j} + g_{P2_j} + g_{P1_jxP2_j} + gE_{P1_j} + gE_{P2_j} + gE_{P1_jxP2_j} + e_{ij}$$

being y_{ij} the observed values for hybrids; E_i is the random effect of environments, where $E_i \sim N(0, \sigma^2_E)$; g_{P1_j} and g_{P2_j} are the scores derived from each test-cross for inbred parent lines, where $g_{P1_j} = \Sigma^p_m x_{P1_{jm}} b_{mP1}$ and $g_{P2_j} = \Sigma^p_m x_{P2_{jm}} b_{mP2}$, $p = (i = 1, 2, \ldots, m)$, $X_{P1} = \left\{X_{P1_j}\right\}$ and $X_{P2} = \left\{X_{P2_j}\right\}$ are the marker incidence matrices for inbred parent lines, and b_{mP1} and b_{mP2} are the marker effects for the m^{th} marker such that $b_{mP1} \sim N\left(0, \sigma^2_{b\ P1}\right)$ and $b_{mP2} \sim N\left(0, \sigma^2_{b\ P2}\right)$, being $\sigma^2_{b\ P2}$ and $\sigma^2_{b\ P2}$ the variance components for the parent inbred lines, $g_{P1} = \{g_{P1_j}\} \sim N(0, G_{P1}\sigma^2_{P1g})$ and $g_{P2} = \{g_{P2_j}\} \sim N(0, G_{P2}\sigma^2_{P2g})$ with $G_{P1} = \frac{X_{P1}X'_{P1}}{p}$ and $G_{P2} = \frac{X_{P2}X'_{P2}}{p}$ and G_{P2}. $\sigma^2_{P1_g} = p\ x\ \sigma^2_{b\ P1}$ and $\sigma^2_{P2_g} = p\ x\ \sigma^2_{b\ P2}$ are the variance components of the inbred parent line effects, $g_{P1\ x\ P2} = \{g_{P1_j\ x\ P2_j}\} \sim N(0, G_{P1\ x\ P2}\sigma^2_{P1g\ x\ P2g})$, where $G_{P1\ x\ P2} = G_{P1} \circ G_{P2}$, being $\circ$ the Hadamard (cell-by-cell) product of two covariate matrices. For the interaction terms, $gE_{P1} = \{gE_{P1_j}\} \sim N(0, (Z_{gP1}G_{P1}Z'_{gP1}) \circ (Z_E Z'_E)\sigma^2_{P1gE})$, $gE_{P2} = \{gE_{P2_j}\} \sim N(0, (Z_{gP2}G_{P2}Z'_{gP2}) \circ (Z_E Z'_E)\sigma^2_{P2gE})$, and $gE_{P1\ x\ P2} = \{gE_{P1_j\ x\ P2_j}\} \sim N(0, (Z_{gP1}G_{P1}Z'_{gP1}) \circ (Z_{gP2}G_{P2}Z'_{gP2}) \circ (Z_E Z'_E)\sigma^2_{P1\ x\ p2gE})$, where $\sigma^2_{P1\ x\ p2gE}$, $\sigma^2_{P1\ x\ p2gE}$, and $\sigma^2_{P1\ x\ p2gE}$ are the variance components of the GCA (P1 and P2) and SCA (P1 x P2) interaction with environments, Z_{gP1}, Z_{gP2}, and Z_E are the incidence matrices for associating inbred parent lines and environments with observed performance of hybrids (y_{ij}). e_{ij} is the random effect of experimental errors, $e_{ij} \sim N(0, \sigma^2)$.

Hybrids were predicted in environments according to the differential cross-validation (CV) schemes CV00, CV0, CV1, and CV2 through all prediction models (1, 2, and 3). Tested and non-tested hybrids were predicted in non-tested environments (CV0 and CV00, respectively), and in tested environments (CV2 and CV1, respectively). The illustration of each CV scheme can be found in Figure 2, based on a random assignment of hybrids to 4 folds.

Crosses selection methods

The procedures for selecting mating parents served the purposes of obtaining inbred lines within heterotic groups and single cross hybrids between heterotic groups, considering that some hybrids were tested previously (training population), but others were not, which reflects CV00, CV0, CV1, and CV2. The male and female inbred lines were directly selected based on either the predicted hybrid performance ($\hat{y}$), general combining ability, or specific combining abilities of hybrids; or based on simulated crosses. The simulated crosses were performed in the R package Genomic Mating (Akdemir and Sánchez, 2016) and considered the selection of 40 crosses per run for the top- and bottom-performing crosses. The simulations used the previously predicted marker effects of either predicted hybrid performance ($\hat{y}$), general combining ability, or specific combining abilities, obtained according to the item *2.3.4 GWS model fitting and cross validation schemes*. The detailed information for mating parents' selection for corn hybrid pipeline and Dent/Flint inbred line pipelines is presented below.

Hybrid pipeline

Method 1

Direct selection of hybrids based on predicted specific combining ability (SCA) effect for cross validation 0 (CV0) scheme.

Method 2

Selection of hybrids based on simulated crosses using specific combining ability (SCA) marker effects for cross validation 0 (CV0) scheme.

Method 3

Direct selection of hybrids based on predicted mean ($\hat{y}$) for cross validation 0 (CV0) scheme.

Method 4

Selection of hybrids based on simulated crosses using predicted mean ($\hat{y}$) marker effects for cross validation 0 (CV0) scheme.

Method 5

Direct selection of hybrids based on predicted specific combining ability (SCA) effect for cross validation 00 (CV00) scheme.

Method 6

Selection of hybrids based on simulated crosses using specific combining ability (SCA) marker effects for cross validation 00 (CV00) scheme.

Method 7

Direct selection of hybrids based on predicted mean ($\hat{y}$) for cross validation 0 (CV00) scheme.

Method 8

Selection of hybrids based on simulated crosses using predicted mean ($\hat{y}$) marker effects for cross validation 00 (CV00) scheme.

Method 9

Direct selection of hybrids based on predicted specific combining ability (SCA) effect for cross validation 1 (CV1) scheme.

Method 10

Selection of hybrids based on simulated crosses using specific combining ability (SCA) marker effects for cross validation 1 (CV1) scheme.

Method 11

Direct selection of hybrids based on predicted mean ($\hat{y}$) for cross validation 1 (CV1) scheme.

Method 12

Selection of hybrids based on simulated crosses using predicted mean ($\hat{y}$) marker effects for cross validation 1 (CV1) scheme.

Method 13

Direct selection of hybrids based on predicted specific combining ability (SCA) effect for cross validation 2 (CV2) scheme.

Method 14

Selection of hybrids based on simulated crosses using specific combining ability (SCA) marker effects for cross validation 2 (CV2) scheme.

Method 15

Direct selection of hybrids based on predicted mean ($\hat{y}$) for cross validation 2 (CV2) scheme.

Method 16

Selection of hybrids based on simulated crosses using predicted mean ($\hat{y}$) marker effects for cross validation 2 (CV2) scheme.

Method 17

Observed crosses. For all CV00 and CV0 schemes, $SCA = GFM$. For all CV1 and CV2 schemes, $SCA = GFM + GFM.E$.

Inbred line pipelines - Dent and Flint heterotic groups

To obtain genomic mating derived inbred lines, it was considered that each combination should result in S5 inbred lines according to the following methods:

Method 1

Direct selection of inbred lines based on predicted general combining ability (GCA) effect for cross validation 0 (CV0) scheme.

Method 2

Selection of inbred lines based on simulated crosses using specific combining ability (GCA) marker effects for cross validation 0 (CV0) scheme.

Method 3

Direct selection of inbred lines based on predicted mean ($\hat{y}$) for cross validation 0 (CV0) scheme.

Method 4

Selection of inbred lines based on simulated crosses using predicted mean ($\hat{y}$) marker effects for cross validation 0 (CV0) scheme.

Method 5

Direct selection of inbred lines based on predicted specific combining ability (GCA) effect for cross validation 00 (CV00) scheme.

Method 6

Selection of inbred lines based on simulated crosses using specific combining ability (GCA) marker effects for cross validation 00 (CV00) scheme.

Method 7

Direct selection of inbred lines based on predicted mean ($\hat{y}$) for cross validation 0 (CV00) scheme.

Method 8

Selection of inbred lines based on simulated crosses using predicted mean ($\hat{y}$) marker effects for cross validation 00 (CV00) scheme.

Method 9

Direct selection of inbred lines based on predicted specific combining ability (GCA) effect for cross validation 1 (CV1) scheme.

Method 10

Selection of inbred lines based on simulated crosses using specific combining ability (GCA) marker effects for cross validation 1 (CV1) scheme.

Method 11

Direct selection of inbred lines based on predicted mean ($\hat{y}$) for cross validation 1 (CV1) scheme.

Method 12

Selection of inbred lines based on simulated crosses using predicted mean ($\hat{y}$) marker effects for cross validation 1 (CV1) scheme.

Method 13

Direct selection of inbred lines based on predicted specific combining ability (GCA) effect for cross validation 2 (CV2) scheme.

Method 14

Selection of inbred lines based on simulated crosses using specific combining ability (GCA) marker effects for cross validation 2 (CV2) scheme.

Method 15

Direct selection of inbred lines based on predicted mean ($\hat{y}$) for cross validation 2 (CV2) scheme.

Method 16

Selection of inbred lines based on simulated crosses using predicted mean ($\hat{y}$) marker effects for cross validation 2 (CV2) scheme.

Method 17

Observed crosses. For all CV00 and CV0 schemes, $GCA = GF + GM$. For all CV1 and CV2 schemes, $GCA = GF + GM + GF.E + GM.E$.

4 RESULTS

a. FIELD TRIALS

Well-watered tests delivered as double yield as drought stress tests, reaching 3.38 ton/ha in J11D and 7.1 ton/ha in J11W as lowest and highest grain yield overall means, respectively. Table 1 also shows that male flowering time did not change severely within each year regardless watering treatment, ranging from 63.25 in J10D to 67.84 in J11D. The same happened to FFT, with 64.31 and 70.61 days to silking in J10D and J11D. In addition, there was no distinct pattern of how well watered or drought stress treatments affected FFT and MFT. However, well-watered treatment shortened the difference between MFT and FFT, which means that silks would be readily available to receive shredded pollen from tassels water was available for uptake. The ASI difference between well-watered and drought stress treatments in 2010 was 0.6 days. For 2011, it was higher, corresponding to 2.21 days. For EPP, corn plants were more affected in drought stress regime in 2010 than in 2011, presenting 12.35 ears per plot, in contrast to 17.84 in the same year. Therefore, yield related traits (GY and EPP) were clearly affected by water treatment, but time-related traits, which can ultimately affect yield, did not have sharp changes in estimates except for ASI.

Overall, the heritability estimates were above 0.6 for all traits, except for GY and ASI in J11W, denoting genetic variance plays a big role in the inheritance of these features. Genetic correlation estimates between pairs of traits were affected by watering treatment (Table 2a and Table 2b) in comparison to when all the dataset was evaluated jointly or filtering the well-watered datapoints (Table 2b and 2c). In drought stress tests, traits were more positively or negatively related to each other than when well-watered or full data were considered. The highest correlation estimate was 0.81 between EPP and GY for drought stress conditions, and the lowest was 0.02 between EPP and ASI in well-watered conditions, which is reasonable to attribute the hypothesis that the plenty availability pf water synch silking and pollen shredding in such a way that the adequate numbers of ears are observed in each plot. From table 2, it is shown that grain yield is directly affected by all the other mentioned traits regardless of water regime. These correlation estimates provide some insights that the strongly related traits probably share high-effect SNPs in common, being a useful information for the cut off in SNP marker selection step of this research.

b. SNP MARKER SELECTION

Hybrids were predicted with G-BLUP method using cumulative portions of 0.01% of selected markers according to the previous marker effect estimation through GWAS for all traits. The increasing SNP subsets were plotted against the prediction accuracy (Figure 4). The pattern of the slope was similar to all traits, and in general the datapoint where the slopes performed asymptotically corresponded to 4,69% of SNP portion, corresponding to 2,689 markers out of the total for each trait. In general, the prediction accuracy was slightly higher when the extra removal of markers based on LD analyses (Strategy 1) was performed considering the 4.69% highest effect markers subset. However, the number of markers did not decrease significantly after performing LD analyses (Table 1).

To ensure that the genome-based prediction would perform better than the phenotype based one, a baseline prediction was carried out. The absence of no kinship (zero markers) delivered negative correlations to all traits. Using the total set of markers, the predictions were approximately 0.6, 0.2, 0.6, 0.25 and 0.55 for ASI, EPP, FFT, GY and MFT, respectively, for drought stress condition. For well-watered condition, the prediction estimates were 0,6, 0.09, 0.5, 0.2, and 0.5 for ASI, EPP, FFT, GY and MFT, respectively. After selection of SNPs, GBLUP analyses were run for all traits using the selected markers for all traits jointly, removing the repeated ones across all traits, and the accuracies remained like when the full set was used, with minor changes. For well-watered and drought stress tests, the prediction accuracies remained similar for the same traits, except for EPP where a significant reduction was observed. Overall, prediction accuracy for time-related traits (ASI, FFT, and MFT) were higher to yield-related traits (EPP and GY) regardless of the watering treatment.

Using randomly selected markers (Sup. Figure 1), a distinct pattern of the accuracy slopes was observed when compared to when the selection followed the marker effect criteria for all traits. Along the prediction slopes, prediction did not reach a clear asymptotic pattern, and in some cases, smaller marker subsets delivered greater accuracies and decreased with the addition of more markers, and vice-versa.

Based on the results, low-density markers are representative for genomic prediction purposes in terms of accuracy of the studied trats once the estimates remained mostly unchanged with just minor changes. In some cases, the prediction accuracies were slightly greater when compared to the full set of SNPs. Therefore, 2,689 markers provided the same reliability to predict test-cross corn hybrids for time- and yield-related traits. We reiterate that the dominant relationship matrix associated to the dominant effect was not included to predict the test-cross hybrids once unclear responses were obtained in terms asymptotic patterns of prediction accuracies over selected SNP portions, which were like the ones presented in Sup. Figure 1.

Some papers address that the reduction of markers has negative effects over long-term breeding purposes. However, we cannot state that once the long-term variance and genetic diversity between original and selected group of lines or hybrids selected through phenotypic and genome-based selection (full and reduced SNP dataset) were not evaluated, neither that variance nor genetic diversity would shrink more quicky in genomic prediction using low-density SNP panels. We state that the purpose of this section of the research was to provide a low-density maker set that would be used to simulate crosses between inbred lines based on the real marker effects, which originally requires high computational demand. Therefore, the reduction of the SNP dataset addressed in this study would ease the process of developing inbred lines and single-cross hybrids by the choice of the best mating pairs in the routine of a nursery in a corn breeding program.

c. PREDICTION ACCURACIES

It was assessed the efficiency of three different models to predict test-cross hybrids through three different prediction models and four different cross-validation schemes. Model three, the most complete, which includes the effects of GCA, SCA, and their interactions with environments, presented the highest accuracies with moderate mean square error (MSE) when attempting to predict tested hybrids in tested environments (CV2) and untested hybrids in tested environments (CV1). The highest accuracy estimates were 0.73 and 0.64 for male flowering time and the lowest 0.41 and 0.27 for ears per plot for CV2 and CV1, respectively. The inclusion of the specific combination effect of inbred lines (SCA) and the interaction terms enhanced the accuracy in at least 7.3% (for MFT ON CV2) and up to 158% (for GY on CV1) to the baseline model (Model 1). Therefore, we noticed benefits of including SCA and GE terms in the model for hybrid prediction when considering the cross-validation schemes that address observed environments, regardless of if the hybrids were previously tested.

Model two, the intermediate one complete, which includes the effects of GCA and SCA, presented the highest accuracies when attempting to predict untested hybrids in untested environments (CV00) and tested hybrids in untested environments (CV0). The highest accuracy estimates were 0.72 and 0.70 for male flowering time and the lowest 0.44 and 0.40 for ears per plot for CV00 and CV0, respectively. The inclusion of the specific combination effect of inbred lines (SCA) enhanced the accuracy in at least 2.8% (for MFT on CV00), and up to 57% (for GY trait on CV0) to the baseline model (Model 1). For ASI on CV00, the contribution was no contribution for including either SCA effect or interactions. Therefore, we noticed benefits of including only SCA besides GCA effects for hybrid prediction when considering the cross-validation schemes that address non-observed environments, regardless of if the hybrids were previously tested. However, the mean square error in predicting hybrids in non-observed environments were higher than when the prediction occurred in observed environments regardless of the model (best fitted one or not) for all traits. This fact is plausible due to lack of information of a whole environment (CV0), or a whole environment for a yet-to-test hybrid (CV00).

d. CROSSES SELECTION

The mating between inbred lines was performed within Dent and Flint heterotic groups for the development of the inbred line pipeline, and across Dent and Flint groups for hybrid pipelines. We used the marker effects obtained through the best-fitted model in the previous item 3.3 for all traits and all cross-validation scheme. All cross-validation schemes were addressed so this research can draw recommendations of crossing designs for all the cases of (un)tested hybrids in (un)tested environments, and not only to a single cross-validation scheme/prediction model that delivered greatest accuracies. Due to the abundance of results for recommendations of mating pairs, the outcomes for ASI, EPP, FFT, and MFT were included in the Supplemental Materials section, and GY will be discussed in the main text solely. For inbred line and hybrid pipelines, each round of crosses simulations selected the 40 top- and bottom-performing crosses, in a pool of thousands of possible ones.

The pipeline used to develop inbred lines through crosses between lines within the same heterotic group considered that the inbreeding should be decreased to ensure that a certain genetic gain would be achieved when the developed lines became new parent material for new hybrids. This ensured that the additive variance of the within-heterotic groups progenies was increased. It increased the probability that new lines be picked as a mating pair to the opposite heterotic group, thus delivering greater genetic gains to new hybrids. We can also consider that inbred lines derived from superior crosses would persist longer as parent materials EMBRAPA's corn breeding program.

For Dent pipeline, groups of crossing methods 1, 3, 5, 7, 9, 11, 13, and 15 coincided more in terms of having at least one recommended parent line per simulated cross in common for positive (40 top-performing) as well as negative (40 bottom-performing) crosses (Figure 6). These methods directly selected parent lines based on GCA or $\hat{y}$ marker effects from the genomic prediction step, meaning that no crosses were simulated. The simulation-based methods (Methods 2, 4, 6, 8, 10, 12, 14, and 16) reached moderate to high coincidence estimates for Dent and Flint positive selection pipeline, but low ones for the negative selection. Table 4 shows the detailed information of the occurrence of best and worst crosses and their respective inbred lines. For dent line pipeline, the parent lines 131 and 34 were present in one or multiple crossing combinations for the positive selection in cases of non-observed environments (CV0 and CV00) using GCA or $\hat{y}$ marker effects to simulate crosses (Methods 2, 4, 6, and 8). For the cases of observed environments (CV1 and CV2) and their respective cross simulation methods (10, 12, 14, and 16), the results were not consistent with the previous case, indicating that line

117 would be highly recommended. For the negative selection, line 67 is the only one that appears in all cases but methods 10 and 12, and that should be discarded as a potential parent line for developing new Dent lines. For Flint pipeline, the parent lines 90 and 143 were frequent in multiple combinations for the positive selection in cases of non-observed and observed environments (all CV schemes) using GCA or $\hat{y}$ marker effects to simulate crosses. For the negative selection, line 110 derived inferior crosses in all cases except for Method 4 and should not be forwarded as a promising parent material for flit pipeline.

Regarding hybrid pipeline, the group of methods 1, 3, 5, 7, 9, 11, 13, and 15 coincided also coincided more in terms of selected parent lines similarly to Dent and Flint pipelines. In this case, these methods directly selected parent lines based on SCA or $\hat{y}$ estimates from the genomic prediction step, meaning that no crosses were simulated. The simulation-based methods (Methods 2, 4, 6, 8, 10, 12, and 14) reached low to moderate coincidence estimates when compared to Dent and Flint pipelines regardless of positive or negative selection. These former methods were based on crosses using SCA or $\hat{y}$ marker effects estimated in the genomic prediction step, meaning that crosses were simulated. In Table 4, the parent lines 161, 27, 55, 117 showed to derive outperforming hybrids using SCA or $\hat{y}$ marker effects to simulate crosses (Methods 2, 4, 6, 8, 10, 12, 14, and 16). As previously mentioned, line 117 was also recommended as parent for the development of new Dent lines for CV1 and CV2 according to the crosses simulations. For the negative selection, lines 58 and 153 is the only appear most cases of simulated crosses as poor parent lines, and thus needed to be removed.

5 DISCUSSION

a. EFFICIENCY OF THE LOW-DENSITY SNP PANEL

In theory, high-density genotyping allows a better prediction of the breeding values of selection candidates due to a better coverage and resolution of the whole genome (Meuwissen et al., 2001). However, a small quantity of genotyping markers is adequate in many circumstances once it leads to a reduction of genotyping costs (Tayeh et al., 2015; Ma et al., 2016; Li et al., 2018; Gorjanc et al., 2018; e Sousa et al., 2019). Particularly in the case of this study, using a subset of SNP markers was beneficial to perform in silico crosses to obtain new inbred lines (*Dent* and *Flint* pipelines) and hybrids (Hybrid pipeline), which became more computational efficient when compared with the use of the full SNP set. It was proved that the prediction accuracy was maintained or slightly improved in all cases of traits and environments.

The computational efficiency due to SNP set reduction was observed not only in the crosses selection step, but also in the calculation of the genetic relationship matrices.

Over generations, QTLs recombine, and the linkage disequilibrium decreases in normal circumstances. Selection tends to create new LD, thus shrinking genetic diversity. However, we cannot state that the use of the subset of SNPs in our GS and crosses pipelines will lead to decreases in genetic diversity one this was not the goal of this study, but it may compromise further calibrations of GS models as we consider the selection of SNP subsets as population dependent. The drift of alleles not covered by low-density genotyping may confound the interpretation of genetic diversity estimates over GS cycles. Therefore, it is recommended that new SNP selection is performed once in a while prior to training GS models with new genotypes. DoVale et al. (2021, unpublished) stated the differences in costs of both high- and low-density genotyping procedures are neglected. However, if the letter is slightly more cost effective, it may be beneficial to small-budget breeding programs or other programs that are starting off to merge the GS in their breeding pipeline.

In simulation studies, high-density genotyping panels are advantageous when using large training datasets with adequate effective population size (Meuwissen et al., 2001; Meuwissen, 2009; Ma et al., 2016) . However, our study used elite germplasm data, where elite inbred lines were genotyped, and elite hybrids were tested. These justifies using SNP subsets once or elite populations usually contains small effective population sizes (St. Martin, 1982), and the prediction accuracy should not be hampered in this case, just as Ma et al. (2016) presented for sampling SNPs based on haplotype blocks approach for soybean GS.

The benefits of our SNP sampling methodology were similar to (Resende et al., 2012; e Sousa et al., 2019). Both reported similar accuracies and bias in their GS analyses using select SNP sets for Loblolly pine, rice, and maize through (re-)estimating marker effects. Other authors addressed the benefits of SNP reduction methodologies. Some of them also estimated SNP effects through GWAS prior to assign them to the relationship matrices (Moser et al., 2010; Resende et al., 2012; Szyda et al., 2013; Tayeh et al., 2015; Spindel et al., 2016; Hoffstetter et al., 2016; Li et al., 2018).

Our prediction accuracy estimates using the reduced SNP set were maintained when compared to the original set. The reduction in SNP quantity was approximately 95%. Heffner et al. (2011) and Tayeh et al. (2015) reported maintenance or minor reduction in prediction accuracy when using selected SNPs for wheat and pea GS, respectively. Both studies kept around 10% of the total number of SNPs.

Sousa et al. (2019) stated that the research they presented just addressed a single trait that resulted in a trait-specific marker selection and suggested that more extensive research should be performed using several traits. Therefore, our research meets broader genome-based objectives in breeding programs for multiple-traits selection, where adequate genotyping arrays can be deployed from a high-density initial genotyping. Besides, our research selected the same number of makers evenly thought the 10 maize chromosomes, which can lead to less mean square error of prediction accuracies. However, our sampled markers account for variations of genome regions that are in LD with the marker-QTL loci. The variations of regions that might recombine and lose LD with the marker-QTL loci can be attributed to other polygenic effects and non-genotyped regions, as reported by Sousa et al. (2019). We also state that genetic architecture, population structure and type (elite germplasm or collection), and population effective size must be considered prior to selecting SNPs.

b. PREDCITION ACCURACY HYBRIDS AND ENVIRONEMENTS

Several researches elucidated the contribution of different prediction models, cross validation schemes, presence/absence of parent material information, and optimal training sets to the area of genomic selection of maize single crosses (Schrag et al., 2009, 2010; Massman et al., 2013; Technow et al., 2014; Kadam et al., 2016, 2021; Jarquin et al., 2021). The present study showed that the inclusion of at least SCA main effect can enhance the prediction accuracy, however the accuracy estimates vary depending on the cross-validation scheme. Jarquin et al. (2021) stated that the inclusion SCA effects result in improvements in the prediction models When the study population contains inter- and intra-heterotic group hybrid combinations, as in this research. In this research, we did not estimate the contributions of main and interaction effects to the overall variance, therefore we cannot draw conclusions if the SCA term holds most of variation in the prediction model for all traits and cross-validation schemes. However, we state that SCA effect captured the combining abilities of every paired cross, especially because each cross was replicated in all environments, contributing to the train the prediction model, and providing a better fit as outcome. This was especially the case of Model 3 in tested environments (CV1 and CV2), where larger improvements were noticed when GCAxE and SCAxE were included. In addition, the interaction terms in Model 3 contributed more to the prediction accuracy compared to baseline model than Model 2 did for non-tested environments (CV0 and CV00) for all traits.

Similarly, the addition of the interaction terms to Model 3 did not contribute to enhance the prediction accuracy when compared to the inclusion of main effects solely (Model 2). In some cases, Model 3 could not even outperform the baseline model in terms of prediction accuracy (ASI and FFT in CV00, and MFT and FFT in CV0). Therefore, predicting already tested or yet to test hybrids in non-observed environments was not contributed by modeling the differential performance of GCA and SCA between locations. A hypothesis is that there were few locations with unsignificant environmental differences between them. It would be expected that previously tested hybrids in observed environments (CV2) would deliver greatest accuracies for its best fitted models than the other cv schemes, once more information of environments and hybrids would be borrowed to train the prediction model. This was not the case of this research, in accordance with Jarquin et al. (2021) findings, except for EPP, where prediction accuracies in CV00 outperformed all other cases of cross validating schemes. However, the borrowing of environmental information in CV1 did not outperform the prediction accuracies of CV0 and CV00 in many cases for most of traits. Therefore, we cannot state that abundant availability of abundant information of hybrids and environments (CV2) would enhance prediction accuracy, or that predicting non-tested hybrids in non-observed environments (CV00) would be the worst prediction case. Contrary to Jarquin et al. (2021), prediction accuracy of Model 2 in CV00 outperformed Model 3 in CV1 for most traits (EPP, FFT, GY, and MFT). This fact can be attributed to the genetic architecture of this trats (moderate to high heritability in our tests, Table 1), or that modeling GCAxE and SCAxE borrowing the information of environments without borrowing information of hybrids (CV1) would be a confounding factor for the prediction process. For the main trait in corn breeding program, gran yield, the prediction accuracies ranged from 0.38 (CV1) to 0.50 (CV2) for the best-fitted models, which were similar to the prediction accuracies for grain yield found in Kadam et al. (2016) and Jarquin et al. (2021) results, being 0.27-0.53 and below 0.5, respectively. For the former study, we considered the models equivalent to the ones addressed in this research. For the latter, we considered the case in which no parent lines were tested previously. However, GY prediction accuracies found here were lower than the ones obtained by Massman et al. (2013) and Technow et al. (2014). According to Windhausen et al. (2012), this can be attributed to population structure, and when it is not accounted for, accuracies can vary due to differences in performances of groups of small subpopulations. These former studies were probable to have a greater number of diverse subpopulations than in our research, and more their parent lines participated in more hybrid combinations.

The way this research addressed the contribution of each parent solely (GCA) and their interaction (SCA) to the predictions of hybrid performances is better than to predict those through modeling the covariance among male and female parent lines in terms of prediction accuracy estimates. Covariance among parents can confound the genetic relationship in sets of tested hybrids if not properly estimated (Kadam et al., 2016). This process was eased by the reaction norm approach where the incidence matrices can accommodate the general and specific contribution of parents to the respective hybrid (Jarquin et al., 2014a).

Overall, there was no clear pattern of either the accuracies were lower or greater in predicting hybrids in not-yet observed environments considering the best-fitted model, than in already-tested environments for all the addressed traits, differently from Jarquin et al. (2021).

The general conclusion is that models that include SCA effects were more efficient than the baseline model, and the one that did not include G x E terms were efficient to predict hybrids in non-observed environments. On the other hand, the one that included interaction terms were effective to predict hybrids in observed environments. It is important to remark that regardless of the cross-validation schemes, the prediction accuracies for the main trat in corn breeding, grain yield (GY), for the best-fitted models were moderate. A future perspective of this subject research is to include environmental covariates to assess the accuracy of prediction models in different cross-validation schemes for the mentioned traits.

Potential research following this topic would be to investigate the inclusion of environmental covariates (e.g., water potential in the soil for drought and irrigated watering treatments during crucial phenological stages). However, the contribution of environmental covariates to the prediction accuracy is not clear once contrasting results were found by Jarquin et al. (2014a) and Jarquin et al. (2021).

c. EFFECTIVENESS OF CROSSING METHODS

Genomic selection of hybrid performance was performed by several authors (Technow et al., 2014; Xu et al., 2014; Kadam et al., 2016; Guo et al., 2019). To date, there are two consolidated forms of selecting parent material: phenotypic selection from field trials data and from genomic estimated breeding values. The latter was facilitated after cost-efficient genotyping technologies as genotyping by sequencing (GBS), where all potential crosses between lines can be made, like a diallel, being more representative for selecting specific combination than a test-cross (He et al., 2014). Another strategy was proposed by Akdemir and Sánchez (2016), defined as genomic mating, where crosses were simulated using marker effects obtained by previous simulation or empirical studies. However, few studies indicated best-performing crosses based on direct selection of GEBV or simulated crosses. Similar to other commercial hybrid breeding programs, our research comprised the generation of a large population of test-cross hybrids. Our best-fitted models provided the marker effects for simulating crosses considered GCA effect and SCA effect, and in some cases their interaction with the environment (CV1 and CV2). In this regard, we assessed which was the best crosses simulation strategy for inbred line development within each heterotic group based on either direct selection or simulated crosses using $\hat{y}$ or GCA estimates, or their respective marker effects.

For hybrid development between heterotic groups, SCA and its maker effects were also used. For direct selection on $\hat{y}$, the performance of lines in inter-heterotic group crosses were assessed on the sum of the line GCA effect and SCA effects of two candidate lines once dominant variance accounts more for F1 development. Intra-heterotic groups, the average effects of GCA of both parents were used, once additive variance accounts more for line development. Our strategy could identify elite single crosses for hybrid and inbred line development, facilitating the nursery management, shortening cycle for cultivar release, and relieving research costs once poor crosses to select lines could be discarded previously to being performed Contrary, this research could also indicate superior crosses that would never be indicated by conventional test-cross scheme, as pointed by Kadam et al. (2021) when predicting hybrid performance through GS. This can be justified when the SCA variance is low or when the heterotic groups are not well stablished (Reif et al., 2007; Zhao et al., 2015). Through our strategy, we maximized the merit of parent lines out of each selected crosses for inbred line and hybrid development, for both direct selection on GEBV or simulated crosses once the frequency of favorable alleles was increased on both inbred and hybrid pipelines regardless of the selection methods.

All these methodologies also helped to elucidate how the design of cross validation schemes can contribute to selection of mating parents. Overall, there was a moderate to high coincidences of selected parents for all CV schemes that involved direct selection of $\hat{y}$ or GCA and SCA (and interactions) values, but moderate to low in the simulation-based ones. We recommend that the selection method, with its respective CV scheme, that correlates most with observed crosses (Method 17) be indicated for the purpose of predicting best-performing crosses. Another remarkable point was that positive and negative selection methods were divergent in selecting parent lines, securing that just the pest-performing lines would be combined to result in superior progenies (F1 or new inbred lines).

In our study, we performed every possible two-by-two combination between lines for Dent (88 lines), Flint (90 lines) and hybrid pipelines. However, due to the dimensionality of results, just the 40 top and bottom crosses were shown. We highlight that genetic architecture can compromise the recommendation of inbred lines to be crossed. For example, inbred line 161 was highly indicated for GY superior crosses for most selection methods, but not well recommended for ASI (Table 4c and Sup. Table 1c), where low to null genetic correlations or very distinct magnitudes of heritabilites among 2 or more traits may lead to lack of coincidence of selected parents between different methods (Table 2). A reduced number of lines might be recommended for crossing after observing which inbred lines overlap for different traits, ensuring that a feasible number of crosses will be performed in a nursery, instead of an infinitesimal amount. We reinforce the idea that predicting superior crosses according to a previously conducted GS study does not configure a validation of selected crosses. That is the reason why Method 17 (observed crosses) was proposed.

So far, most of GS studies have assessed the performance of non-tested hybrids. It can also predict not-yet-obtained hybrids (as in CV1 and CV00). However, our approach goes beyond, and recommend not-yet-performed crosses within and between heterotic groups for all parent lines. Here, we ensured that gains were maximized, and inbreeding (within heterotic group) were minimized, once these were secured by Akdemir and Sánchez (2016) in the genomic mating purpose. Similarly, to the reports from these authors, our results indicated discretely what parent lines should be crosses, which is simpler than in interpreting a continuous performance of lines across a range of crosses. Although corn is naturally a cross-pollinated crop species, in Genomic Mating we could control the type of cross in our simulation according to our breeding purposes (self for dent and fling inbred lines, and cross-pollination for hybrids), and who mates with whom (inter and intra HG) as pointed out by Akdemir and Sánchez (2016). The Phenotypic selection (Method 17), GS, and SC methods should be merged to lead breeders to more precise parent mating selection, and not using one of these approaches solely.

6 CONCLUSION

The conversion from a 50k to a low-density SNP genotyping dataset was efficient to capture hybrid performance for different traits and cross validation scenarios, maintaining the accuracy, and brought benefits to our cross-selection pipeline. We showed how prediction accuracy can clearly be affected by the different prediction models, CV schemes for corn key traits. We reaffirm that the crosses can be indicated directly from GWS, using GEBV. Our look ahead approach performed further crosses through the simulation-based crosses. However, they did not coincide and their coincidence with Method 17 must be checked out. The assumption here is a broad idea of how to conduct a GS study in maize, with a range of CV schemes, starting with dense genotyping and finishing with a low coverage-genotyping, and how it can be deployed to select crosses. Adaptions should be made to this, since the assumptions depend on population size and type relatedness of individuals, genetic architecture of traits, crop species and breeding purposes. As future contributions, we suggest a similar study be performed in extra environments, once marker selection and effects for cross selection step are environment dependent. We also suggest that our selected SNP set undergo a reverse process than the one in this study, with the imputation of 2K set to check if accuracies are maintained, with the possibility of inclusion of other testing individuals for training set.

7 ACKNOWLEDGEMENTS

The support and assistance from the Universidade Federal de Viçosa, Brazil, and its Graduate Program in Genetics and Breeding, Department of Agronomy, Department of Biology, and CEPET/UEPE-UFV Experimental Stations (Capinópolis-MG and Viçosa-MG) is highly appreciated. The support and assistance from the Department of Agronomy and Horticulture of the University of Nebraska-Lincoln, USA is acknowledged. The efforts of EMBRAPA Maize and Sorghum Breeding Program/Biometry Lab in Brazil for carrying out the field work and making this dataset available is highly appreciated. This research received financial support from EMBRAPA, CNPq, CAPES, and CAPES/PRINT program.

8 STATEMENTS

There is no conflict of interest in the conceptualization of this research and production of the manuscript.

9 REFERENCES

Acosta-Pech, R., J. Crossa, G. de los Campos, S. Teyssèdre, B. Claustres, et al. 2017. Genomic models with genotype × environment interaction for predicting hybrid performance: an application in maize hybrids. Theor Appl Genet 130(7): 1431–1440. doi: 10.1007/s00122-017-2898-0.

Akdemir, D., and J.I. Sánchez. 2016. Efficient Breeding by Genomic Mating. Front. Genet. 7. doi: 10.3389/fgene.2016.00210.

Amadeu, R.R., C. Cellon, J.W. Olmstead, A.A.F. Garcia, M.F.R. Resende, et al. 2016. AGHmatrix: R Package to Construct Relationship Matrices for Autotetraploid and Diploid Species: A Blueberry Example. The Plant Genome 9(3). doi: 10.3835/plantgenome2016.01.0009.

Bernardo, R. 1994. Prediction of Maize Single-Cross Performance Using RFLPs and Information from Related Hybrids. Crop Sci. 34(1): 20–25. doi: 10.2135/cropsci1994.0011183X003400010003x.

Butler, D. G., Cullis, B.R., A. R. Gilmour, Gogel, B.G. and Thompson, R. 2017. ASReml-R Reference Manual Version 4. VSN International Ltd, Hemel Hempstead, HP1 1ES, UK.

Dias, K.O.D.G., S.A. Gezan, C.T. Guimarães, A. Nazarian, L. da Costa e Silva, 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(1): 24–37. doi: 10.1038/s41437-018-0053-6.

DoVale, J.C., H.F. Carvalho, F. Sabadin, and R. Fritsche-Neto. 2021. Reduction of genotyping marker density for genomic selection is not an affordable approach to long-term breeding in cross-pollinated crops. BioRxiv: 2021–03.

Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, et al. 2011. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species (L. Orban, editor). PLoS ONE 6(5): e19379. doi: 10.1371/journal.pone.0019379.

Glaubitz, J.C., T.M. Casstevens, F. Lu, J. Harriman, R.J. Elshire, et al. 2014. TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline (N.A. Tinker, editor). PLoS ONE 9(2): e90346. doi: 10.1371/journal.pone.0090346.

Gorjanc, G., R.C. Gaynor, and J.M. Hickey. 2018. Optimal cross selection for long-term genetic gain in two-part programs with rapid recurrent genomic selection. Theor Appl Genet 131(9): 1953–1966. doi: 10.1007/s00122-018-3125-3.

Guo, T., X. Yu, X. Li, H. Zhang, C. Zhu, et al. 2019. Optimal Designs for Genomic Selection in Hybrid Crops. Molecular Plant 12(3): 390–401. doi: 10.1016/j.molp.2018.12.022.

He, J., X. Zhao, A. Laroche, Z.-X. Lu, H. Liu, et al. 2014. Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. Front. Plant Sci. 5. doi: 10.3389/fpls.2014.00484.

Heffner, E.L., J. Jannink, and M.E. Sorrells. 2011. Genomic Selection Accuracy using Multifamily Prediction Models in a Wheat Breeding Program. Plant Genome 4(1): plantgenome2010.12.0029. doi: 10.3835/plantgenome2010.12.0029.

Henderson, C.R. 1975. Best Linear Unbiased Estimation and Prediction under a Selection Model. Biometrics 31(2): 423. doi: 10.2307/2529430.

Hoffstetter, A., A. Cabrera, M. Huang, and C. Sneller. 2016. Optimizing Training Population Data and Validation of Genomic Selection for Economic Traits in Soft Winter Wheat. G3 Genes|Genomes|Genetics 6(9): 2919–2928. doi: 10.1534/g3.116.032532.

Jarquin, D., J. Crossa, X. Lacaze, P. Du Cheyron, J. Daucourt, et al. 2014a. A reaction norm model for genomic selection using high-dimensional genomic and environmental data. Theor Appl Genet 127(3): 595–607. doi: 10.1007/s00122-013-2243-1.

Jarquin, D., K. Kocak, L. Posadas, K. Hyma, J. Jedlicka, et al. 2014b. Genotyping by sequencing for genomic prediction in a soybean breeding population. BMC Genomics 15(1): 740. doi: 10.1186/1471-2164-15-740.

Jarquin, D., N. de Leon, C. Romay, M. Bohn, E.S. Buckler, et al. 2021. Utility of Climatic Information via Combining Ability Models to Improve Genomic Prediction for Yield Within the Genomes to Fields Maize Project. Front. Genet. 11: 592769. doi: 10.3389/fgene.2020.592769.

Kadam, D.C., S.M. Potts, M.O. Bohn, A.E. Lipka, and A.J. Lorenz. 2016. Genomic Prediction of Single Crosses in the Early Stages of a Maize Hybrid Breeding Pipeline. G3 Genes|Genomes|Genetics 6(11): 3443–3453. doi: 10.1534/g3.116.031286.

Kadam, D.C., O.R. Rodriguez, and A.J. Lorenz. 2021. Optimization of training sets for genomic prediction of early-stage single crosses in maize. Theor Appl Genet 134(2): 687–699. doi: 10.1007/s00122-020-03722-w.

Lado, B., S. Battenfield, C. Guzmán, M. Quincke, R.P. Singh, et al. 2017. Strategies for Selecting Crosses Using Genomic Prediction in Two Wheat Breeding Programs. The Plant Genome 10(2). doi: 10.3835/plantgenome2016.12.0128.

Li, B., N. Zhang, Y.-G. Wang, A.W. George, A. Reverter, et al. 2018. Genomic Prediction of Breeding Values Using a Subset of SNPs Identified by Three Machine Learning Methods. Front. Genet. 9: 237. doi: 10.3389/fgene.2018.00237.

Ma, Y., J.C. Reif, Y. Jiang, Z. Wen, D. Wang, et al. 2016. Potential of marker selection to increase prediction accuracy of genomic selection in soybean (Glycine max L.). Mol Breeding 36(8): 113. doi: 10.1007/s11032-016-0504-9.

Massman, J.M., A. Gordillo, R.E. Lorenzana, and R. Bernardo. 2013. Genomewide predictions from maize single-cross data. Theor Appl Genet 126(1): 13–22. doi: 10.1007/s00122-012-1955-y.

Meuwissen, T.H. 2009. Accuracy of breeding values of “unrelated” individuals predicted by dense SNP genotyping. Genet Sel Evol 41(1): 35. doi: 10.1186/1297-9686-41-35.

Meuwissen, T.H.E., B.J. Hayes, and M.E. Goddard. 2001. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. Genetics 157(4): 1819–1829. doi: 10.1093/genetics/157.4.1819.

Moser, G., M.S. Khatkar, B.J. Hayes, and H.W. Raadsma. 2010. Accuracy of direct genomic values in Holstein bulls and cows using subsets of SNP markers. Genet Sel Evol 42(1): 37. doi: 10.1186/1297-9686-42-37.

Pérez, P., and G. de los Campos. 2014. Genome-Wide Regression and Prediction with the BGLR Statistical Package. Genetics 198(2): 483–495. doi: 10.1534/genetics.114.164442.

Persa, R., M. Grondona, and D. Jarquin. 2021. Development of a Genomic Prediction Pipeline for Maintaining Comparable Sample Sizes in Training and Testing Sets across Prediction Schemes Accounting for the Genotype-by-Environment Interaction. Agriculture 11(10): 932. doi: 10.3390/agriculture11100932.

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Reif, J.C., F.-M. Gumpert, S. Fischer, and A.E. Melchinger. 2007. Impact of Interpopulation Divergence on Additive and Dominance Variance in Hybrid Populations. Genetics 176(3): 1931–1934. doi: 10.1534/genetics.107.074146.

Resende, M.F.R., P. Muñoz, M.D.V. Resende, D.J. Garrick, R.L. Fernando, et al. 2012. Accuracy of Genomic Selection Methods in a Standard Data Set of Loblolly Pine (*Pinus taeda* L.). Genetics 190(4): 1503–1510. doi: 10.1534/genetics.111.137026.

Roberts, A., L. McMillan, W. Wang, J. Parker, I. Rusyn, et al. 2007. Inferring missing genotypes in large SNP panels using fast nearest-neighbor searches over sliding windows. Bioinformatics 23(13): i401–i407. doi: 10.1093/bioinformatics/btm220.

Saghai-Maroof, M.A., K.M. Soliman, R.A. Jorgensen, and R.W. Allard. 1984. Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics. Proc. Natl. Acad. Sci. U.S.A. 81(24): 8014–8018. doi: 10.1073/pnas.81.24.8014.

Schrag, T.A., M. Frisch, B.S. Dhillon, and A.E. Melchinger. Marker-based prediction of hybrid performance in maize single-crosses involving doubled haploids.

Schrag, T.A., J. Möhring, A.E. Melchinger, B. Kusterer, B.S. Dhillon, et al. 2010. Prediction of hybrid performance in maize using molecular markers and joint analyses of hybrids and parental inbreds. Theor Appl Genet 120(2): 451–461. doi: 10.1007/s00122-009-1208-x.

Shull, G. H. (1948). What is" heterosis"?. Genetics, 33(5), 439.

Sousa, M.B., G. Galli, D.H. Lyra, Í.S.C. Granato, F.I. Matias, et al. 2019. Increasing accuracy and reducing costs of genomic prediction by marker selection. Euphytica 215(2): 18. doi: 10.1007/s10681-019-2339-z.

Spindel, J.E., H. Begum, D. Akdemir, B. Collard, E. Redoña, et al. 2016. Genome-wide prediction models that incorporate de novo GWAS are a powerful new tool for tropical rice improvement. Heredity 116(4): 395–408. doi: 10.1038/hdy.2015.113.

St. Martin, S. 1982. Effective Population Size for the Soybean Improvement Program in Maturity Groups 00 to IV 1. Crop Science 22(1): 151–152.

Szyda, J., K. Żukowski, S. Kamiński, and A. Żarnecki. 2013. Testing different single nucleotide polymorphism selection strategies for prediction of genomic breeding values in dairy cattle based on low density panels. Czech J. Anim. Sci. 58(3): 136–145. doi: 10.17221/6670-CJAS.

Tayeh, N., A. Klein, M.-C. Le Paslier, F. Jacquin, H. Houtin, et al. 2015. Genomic Prediction in Pea: Effect of Marker Density and Training Population Size and Composition on Prediction Accuracy. Front. Plant Sci. 6. doi: 10.3389/fpls.2015.00941.

Technow, F. Use of F2 Bulks in Training Sets for Genomic Prediction of Combining Ability and Hybrid Performance.

Technow, F., T.A. Schrag, W. Schipprack, E. Bauer, H. Simianer, et al. 2014. Genome Properties and Prospects of Genomic Prediction of Hybrid Performance in a Breeding Program of Maize. Genetics 197(4): 1343–1355. doi: 10.1534/genetics.114.165860.

Thompson, R. 2008. Estimation of quantitative genetic parameters. Proc. R. Soc. B. 275(1635): 679–686. doi: 10.1098/rspb.2007.1417.

Windhausen, V.S., G.N. Atlin, J.M. Hickey, J. Crossa, J.-L. Jannink, et al. 2012. Effectiveness of Genomic Prediction of Maize Hybrid Performance in Different Breeding Populations and Environments. G3 Genes|Genomes|Genetics 2(11): 1427–1436. doi: 10.1534/g3.112.003699.

Xu, S., D. Zhu, and Q. Zhang. 2014. Predicting hybrid performance in rice using genomic best linear unbiased prediction. Proc. Natl. Acad. Sci. U.S.A. 111(34): 12456–12461. doi: 10.1073/pnas.1413750111.

Zhao, Y., Z. Li, G. Liu, Y. Jiang, H.P. Maurer, et al. 2015. Genome-based establishment of a high-yielding heterotic pattern for hybrid wheat breeding. Proc. Natl. Acad. Sci. U.S.A. 112(51): 15624–15629. doi: 10.1073/pnas.1514547112.

10 FIGURES

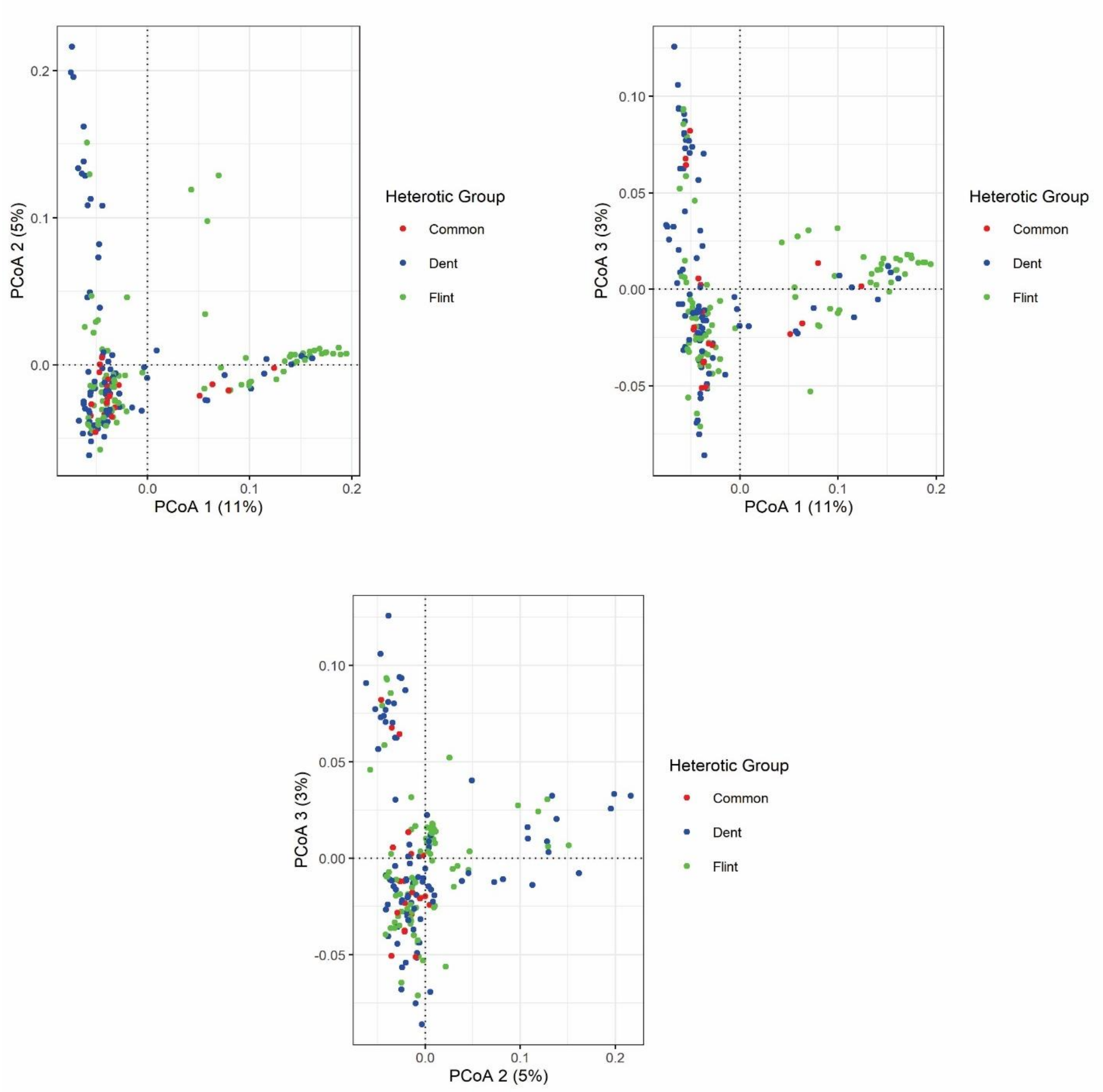


Figure 1. Principal coordinate analyses (PCoA) of 190 inbred lines according to Common, Dent, and Flint heterotic groups in corn, as well as male testers.

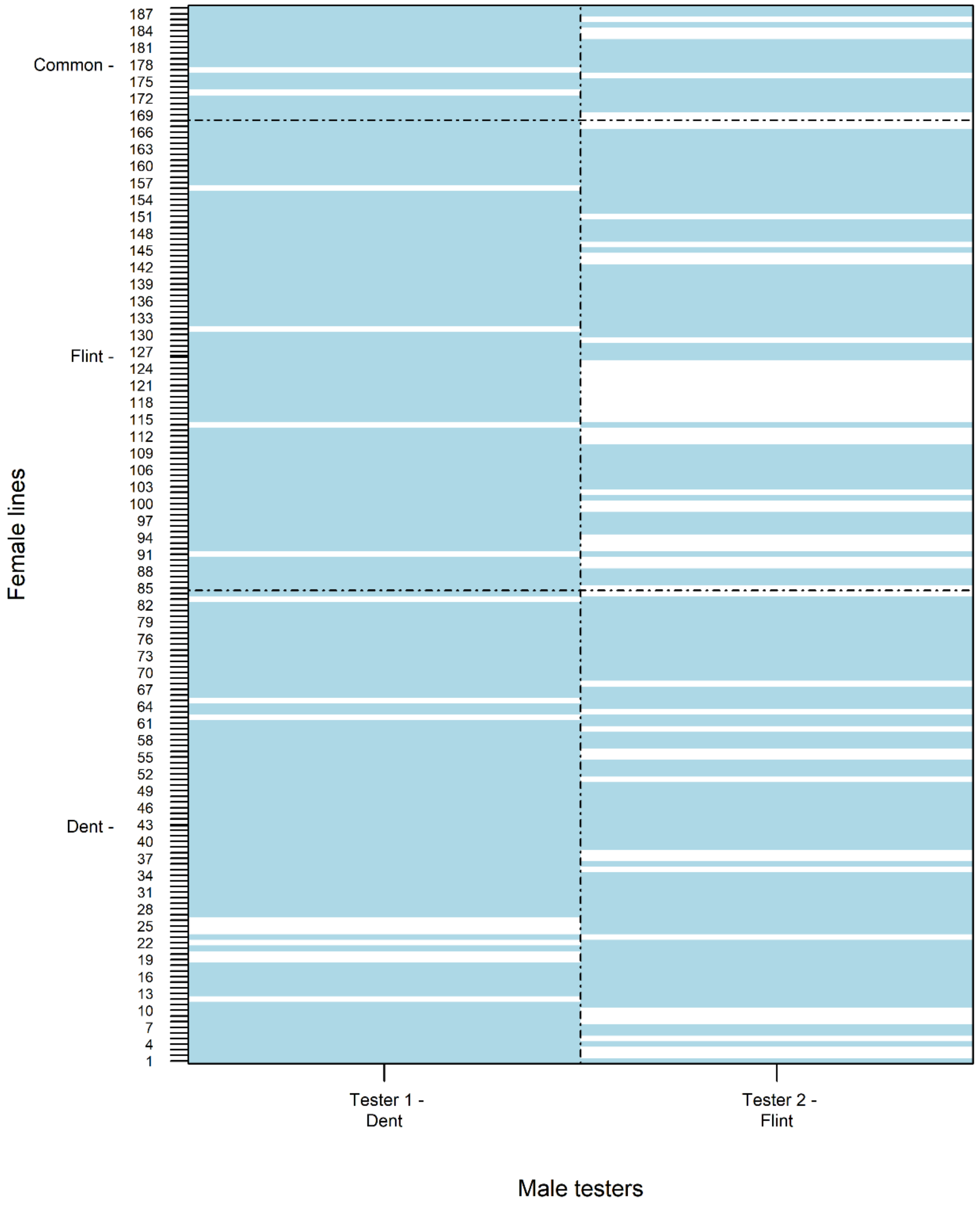


Figure 2. Test-cross scheme between 188 female inbred lines (85 Dent, 86 Flint, and 17 Common) and 2 male testers to obtain 308 single-cross hybrids.

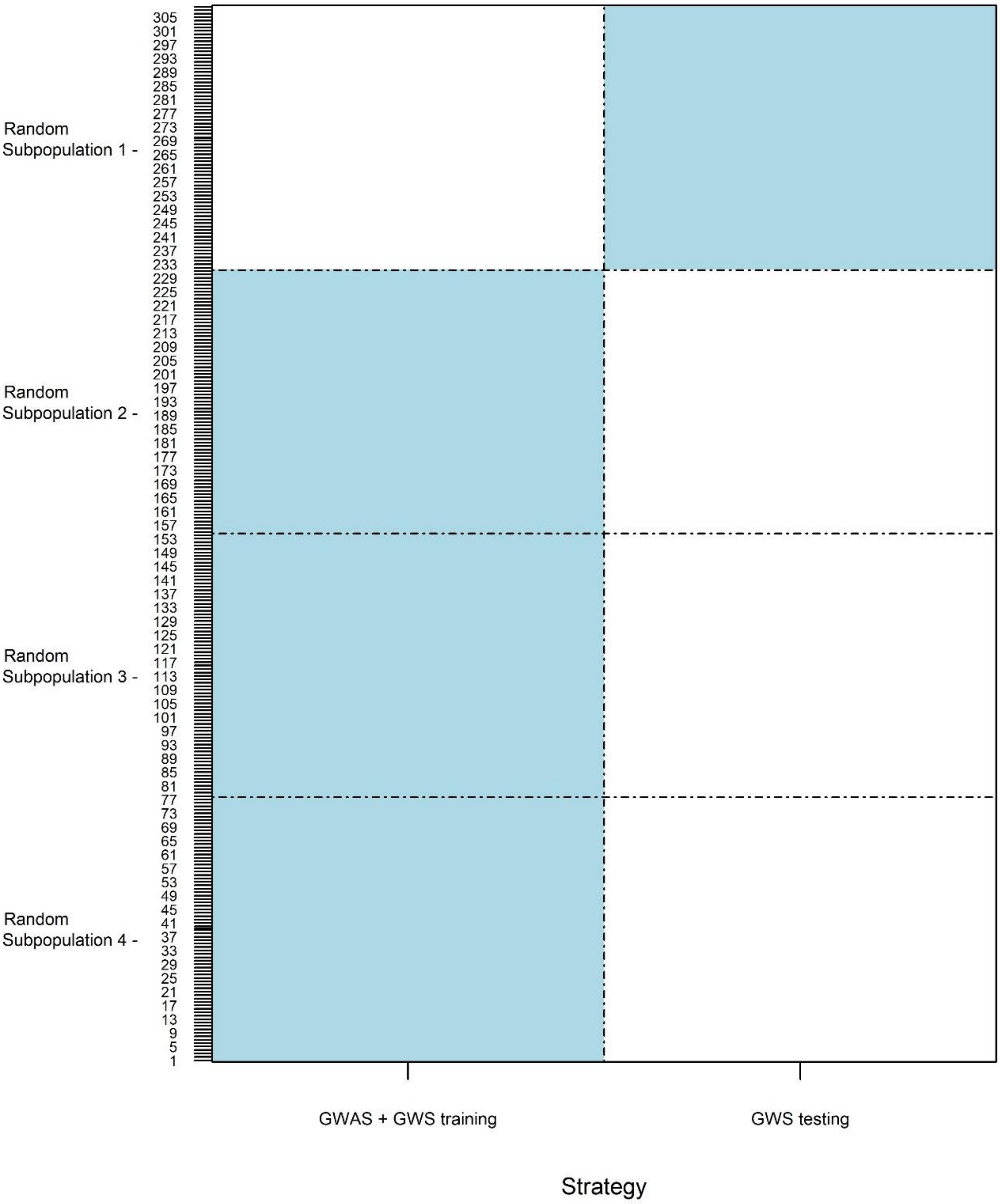


Figure 3. Illustration of hybrids assigned to differential subpopulations for independent maker effect estimation (ME) and genome wide selection analyses (GWS, cross validation of predicted hybrids using the selected markers in the ME). Each subpopulation was submitted at least once to ME, GWS training step, and GWS testing step. This was useful to convert the full SNP dataset into reduced array.

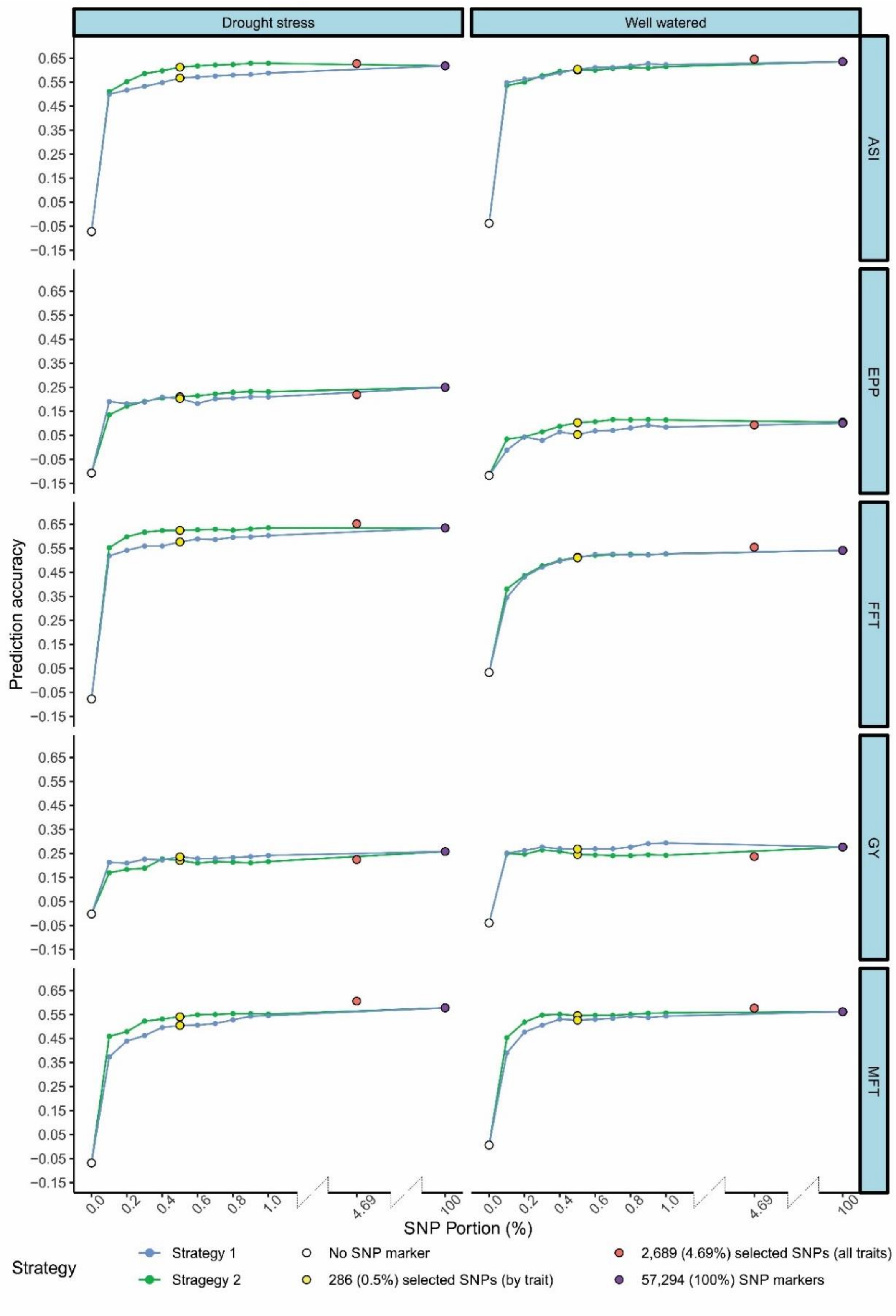


Figure 4. Prediction accuracies for each percentage of selected SNPs according to different marker selection strategies and watering conditions (drought stress and well-watered). ASI: anthesis-silking interval, EPP: ears per plot, FFT: female flowering time, GY: grain yield, MFT: male flowering time.

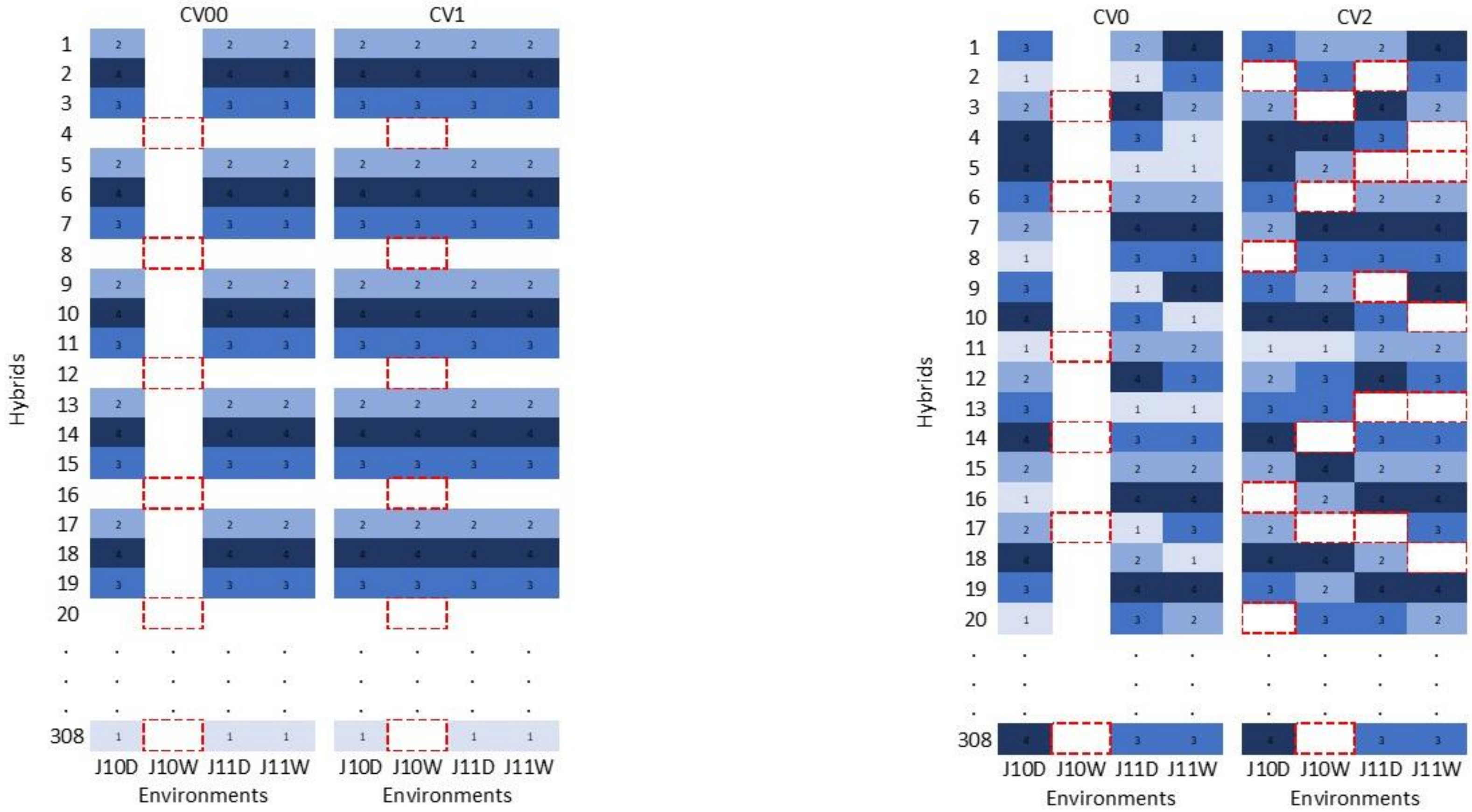


Figure 5. Illustration of hybrids assigned to different cross validation folds (1 to 4) to illustrate cross validation schemes (CV00, CV0, CV1, and CV2). Hybrids assigned to folds 1, 2, 3, and 4 were used for training the prediction model, and hybrids assigned to fold 1 were used for testing (red-dashed cells). Adapted from Persa et al. (2021).

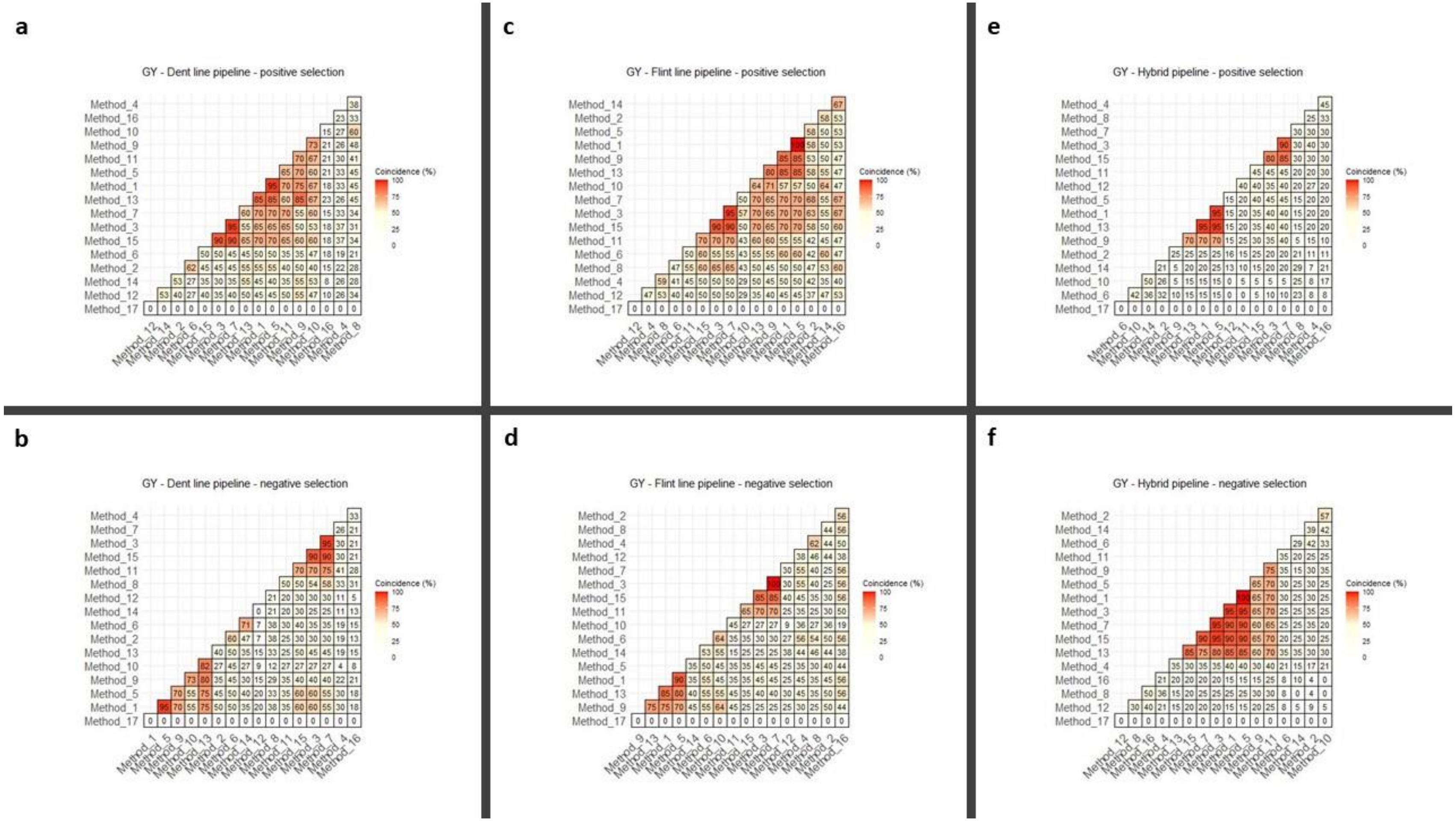


Figure 6. Coincidences ($C_{\%}$) between selected parent lines for each selection method considering Dent, Flint, and hybrid corn pipelines for grain yield (GY). Positive selection was carried out on the 40 top-performing crosses. Negative selection was carried out on the 40 bottom-performing crosses. See criteria for each selection method and for crosses simulations in the items under Material and Methods section.

11 TABLES

Table 1. Genetic parameters of 308 test-cross corn hybrids tested in four environments.

Environment	**Trait**	σ_g^2	σ^2	h_g^2	μ_g
J10D	ASI	1.03	1.71	0.63	1.06
	EPP	4.89	9.6	0.59	12.35
	FFT	3.51	3.3	0.75	64.31
	GY	0.62	1.16	0.6	3.47
	MFT	1.97	1.51	0.78	63.25
J10W	ASI	0.66	0.71	0.64	0.46
	EPP	4.81	8.85	0.51	17.84
	FFT	2.97	2.07	0.74	64.75
	GY	1.31	1.46	0.64	6.59
	MFT	2.06	1.53	0.72	64.29
J11D	ASI	1.85	2.78	0.65	2.85
	EPP	6.27	9.21	0.66	15.03
	FFT	5.03	5.46	0.72	70.68
	GY	0.56	0.91	0.63	3.38
	MFT	3.06	2.47	0.77	67.84
J11W	ASI	0.69	0.95	0.42	0.64
	EPP	9	3.85	0.7	17.08
	FFT	2.91	1.76	0.62	68.36
	GY	1.19	1.72	0.41	7.1
	MFT	3.16	0.8	0.79	67.73

J10D: Janaúba in 2010 under drought stress, J10W: Janaúba in 2010 under well-watered irrigation, J11D: Janaúba in 2011 under drought stress, J11W: Janaúba in 2011 under well-well-watered irrigation. ASI: anthesis-silking interval, EPP: ears per plot, FFT: female flowering time, GY: grain yield, MFT: male flowering time. σ_g^2: genetic variance, σ^2: residual variance, h_g^2: broad-sense heritability, μ_g: mean.

Table 2. Genetic correlation (r_g) between traits in corn under drought stress (Table 2a) and well-watered (Table 2b) conditions, and both datasets (Table 2c).

a

	EPP	FFT	GY	MFT
ASI	-0.32^{*}	0.74^{**}	-0.47^{*}	0.20^{*}
EPP		-0.43^{*}	0.81^{**}	-0.35^{*}
FFT			-0.56^{*}	-0.35^{*}
GY				-0.41^{*}

b

	EPP	FFT	GY	MFT
ASI	0.02	0.46^{*}	-0.22^{*}	-0.17^{*}
EPP		-0.14	0.68^{**}	-0.18^{*}
FFT			-0.36^{*}	-0.18^{*}
GY				-0.25^{*}

c

	EPP	FFT	GY	MFT
ASI	-0.15	0.60^{*}	-0.33^{*}	0.03
EPP		-0.32^{*}	0.77^{**}	-0.29^{*}
FFT			-0.50^{*}	-0.29^{*}
GY				-0.38^{*}

ASI: anthesis-silking interval, EPP: ears per plot, FFT: female flowering time, GY: grain yield, MFT: male flowering time. * significant at 0.05, ** significant at 0.01.

Table 3. Accuracies ($r_{y\hat{y}}$) of prediction models according to cross-validation schemes (CV00, CV0, CV1, and CV2) and traits. Values in bold have their respective models as the best-fitted ones.

Trait	$r_{y\hat{y}}$	CV00		CV0		CV1		CV2	
		$r_{y\hat{y}}$	MSE	$r_{y\hat{y}}$	MSE	$r_{y\hat{y}}$	MSE	$r_{y\hat{y}}$	MSE
ASI	Model 1	0.59	3.87	0.40	4.77	0.56	1.68	0.60	1.59
	Model 2	**0.59**	**3.99**	**0.57**	**3.99**	0.56	0.46	0.60	0.44
	Model 3	0.45	4.30	0.53	3.94	**0.60**	**1.53**	**0.66**	**1.37**
EPP	Model 1	0.38	17.20	0.32	17.05	0.13	11.05	0.30	9.77
	Model 2	**0.44**	**16.62**	**0.40**	**16.62**	0.22	0.70	0.40	0.61
	Model 3	0.45	16.67	0.37	16.91	**0.27**	**10.17**	**0.41**	**8.94**
FFT	Model 1	0.67	18.17	0.59	17.50	0.54	4.52	0.65	3.71
	Model 2	**0.69**	**18.78**	**0.65**	**18.78**	0.58	0.30	0.67	0.26
	Model 3	0.66	16.29	0.57	17.47	**0.63**	**3.66**	**0.71**	**3.02**
GY	Model 1	0.33	7.61	0.26	7.61	0.12	2.10	0.29	1.89
	Model 2	**0.45**	**7.63**	**0.41**	**7.63**	0.30	0.37	0.37	0.33
	Model 3	0.45	7.61	0.34	7.31	**0.38**	**1.80**	**0.50**	**1.61**
MFT	Model 1	0.70	9.91	0.68	9.61	0.56	2.57	0.68	2.01
	Model 2	**0.72**	**9.94**	**0.70**	**9.94**	0.61	0.30	0.71	0.25
	Model 3	0.72	8.95	0.64	9.41	**0.64**	**2.17**	**0.73**	**1.72**

ASI: anthesis-silking interval, EPP: ears per plot, FFT: female flowering time, GY: grain yield, MFT: male flowering time.

Model 1: $y_{ij} = \mu + E_i + g_{P1_j} + g_{P2_j} + e_{ij}$

Model 2: $y_{ij} = \mu + E_i + g_{P1_j} + g_{P2_j} + g_{P1_jxP2_j} + e_{ij}$

Model 3: $y_{ij} = \mu + E_i + g_{P1_j} + g_{P2_j} + g_{P1_jxP2_j} + gE_{P1_j} + gE_{P2_j} + gE_{P1_jxP2_j} + e_{ij}$

Table 4. Selected crosses for Dent, Flint, and hybrid corn pipelines for grain yield (GY) considering each selection method. Positive selection was carried out on the 40 top-performing crosses. Negative selection was carried out on the 40 bottom-performing crosses. See criteria for each selection method and for crosses simulations in the items under Material and Methods section.

Dent line pipeline

Simulated crosses - positive selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_2	34 x 131	14	Method_4	52 x 156	6	Method_6	34 x 131	10	Method_8	34 x 131	6
	95 x 148	10		12 x 44	4		11 x 95	6		11 x 46	4
	97 x 117	8		18 x 42	4		11 x 20	4		144 x 154	4
	11 x 128	2		99 x 144	4		95 x 148	4		1 x 75	2
	78 x 175	2		1 x 40	2		97 x 117	4		12 x 42	2
	99 x 131	2		2 x 46	2		105 x 144	4		12 x 149	2
	100 x 144	2		21 x 48	2		176 x 176	4		18 x 78	2
				32 x 97	2		21 x 148	2		34 x 167	2
				34 x 89	2		97 x 134	2		44 x 181	2
				34 x 100	2					94 x 91	2
				35 x 184	2					98 x 145	2
				46 x 117	2					99 x 167	2
				48 x 183	2					128 x 134	2
				93 x 166	2					159 x 40	2
				181 x 127	2					170 x 175	2
										176 x 185	2

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_10	11 x 117	14	Method_12	95 x 96	12	Method_14	46 x 117	10	Method_16	2 x 48	2
	96 x 119	6		99 x 175	12		97 x 117	8		6 x 117	2
	96 x 175	6		12 x 131	6		46 x 175	6		11 x 161	2
	176 x 176	4		46 x 117	2		94 x 128	4		12 x 127	2
	20 x 144	2		94 x 40	2		36 x 100	2		18 x 92	2
	34 x 131	2		100 x 159	2		96 x 119	2		21 x 150	2
	41 x 94	2		176 x 176	2		96 x 131	2		32 x 67	2
	46 x 117	2		177 x 181	2		99 x 167	2		33 x 134	2
	98 x 163	2					100 x 144	2		41 x 129	2
							100 x 183	2		48 x 80	2
										51 x 140	2
										75 x 128	2
										78 x 175	2
										84 x 85	2
										93 x 98	2
										105 x 116	2
										145 x 185	2
										147 x 154	2
										176 x 86	2
										182 x 148	2

Simulated crosses - negative selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_2	67 x 140	8	Method_4	20 x 100	4	Method_6	12 x 121	8	Method_8	84 x 140	8
	6 x 156	6		53 x 142	4		67 x 165	8		53 x 163	6
	49 x 67	6		67 x 168	4		85 x 152	8		80 x 152	4
	67 x 159	6		80 x 152	4		59 x 120	6		21 x 67	2
	6 x 41	4		151 x 86	4		49 x 116	4		33 x 116	2
	151 x 86	4		32 x 75	2		54 x 85	2		36 x 109	2
	54 x 109	2		32 x 153	2		67 x 140	2		41 x 154	2
	59 x 120	2		33 x 185	2		151 x 86	2		49 x 120	2
	121 x 91	2		36 x 76	2					53 x 183	2
				53 x 175	2					54 x 109	2
				54 x 145	2					120 x 164	2
				76 x 156	2					147 x 182	2
				79 x 101	2					151 x 86	2
				92 x 127	2					162 x 182	2
				150 x 161	2						

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_10	59 x 120	10	Method_12	53 x 142	18	Method_14	67 x 140	10	Method_16	2 x 48	2
	49 x 120	8		53 x 163	12		49 x 67	6		6 x 117	2
	116 x 23	8		75 x 147	6		59 x 85	6		11 x 161	2
	55 x 55	6		120 x 164	2		67 x 165	6		12 x 127	2
	24 x 150	4		147 x 147	2		12 x 121	4		18 x 92	2
	12 x 83	2					85 x 102	4		21 x 150	2
	55 x 89	2					151 x 86	4		32 x 67	2
										33 x 134	2
										41 x 129	2
										48 x 80	2
										51 x 140	2
										75 x 128	2
										78 x 175	2
										84 x 85	2
										93 x 98	2
										105 x 116	2
										145 x 185	2
										147 x 154	2
										176 x 86	2
										182 x 148	2

Observed crosses - positive selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_17			Method_17			Method_17			Method_17		

Observed crosses - negative selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_17			Method_17			Method_17			Method_17		

Table 4. (continued).

Flint line pipeline

Simulated crosses - positive selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_2	45 x 124	6	Method_4	126 x 131	10	Method_6	45 x 172	10	Method_8	126 x 131	8
	5 x 22	4		10 x 88	6		10 x 22	6		10 x 22	6
	39 x 107	4		10 x 22	4		130 x 148	6		45 x 45	6
	77 x 123	4		10 x 139	4		19 x 146	4		19 x 133	4
	107 x 40	4		40 x 148	4		107 x 132	4		40 x 148	4
	126 x 131	4		45 x 45	2		130 x 141	4		118 x 139	4
	10 x 22	2		111 x 172	2		77 x 125	2		19 x 146	2
	37 x 107	2		124 x 137	2		132 x 148	2		26 x 88	2
	39 x 125	2		126 x 133	2		139 x 40	2		69 x 90	2
	45 x 23	2		160 x 133	2					123 x 139	2
	118 x 139	2		180 x 117	2						
	118 x 184	2									
	126 x 174	2									

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_10	3 x 118	12	Method_12	10 x 139	6	Method_14	45 x 139	6	Method_16	10 x 22	16
	45 x 40	6		19 x 107	6		139 x 40	6		118 x 139	8
	107 x 107	6		19 x 133	6		10 x 138	4		45 x 139	4
	3 x 25	4		88 x 131	6		19 x 146	4		8 x 174	2
	19 x 146	4		10 x 153	4		107 x 40	4		19 x 133	2
	22 x 113	2		118 x 118	4		118 x 139	4		39 x 107	2
	39 x 124	2		126 x 131	4		126 x 174	4		88 x 114	2
	40 x 185	2		45 x 172	2		22 x 113	2		88 x 183	2
	107 x 149	2		118 x 91	2		39 x 125	2		146 x 174	2
							45 x 45	2			
							118 x 138	2			

Simulated crosses - negative selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_2	61 x 106	14	Method_4	16 x 62	8	Method_6	15 x 57	12	Method_8	58 x 187	10
	58 x 187	8		106 x 112	8		56 x 110	6		61 x 106	8
	110 x 112	4		157 x 86	6		56 x 187	6		16 x 188	6
	8 x 157	2		16 x 110	4		50 x 157	4		8 x 17	4
	9 x 142	2		30 x 50	4		72 x 110	4		17 x 113	4
	56 x 187	2		58 x 187	4		5 x 58	2		157 x 86	4
	58 x 64	2		17 x 68	2		16 x 63	2		30 x 110	2
	63 x 64	2		31 x 108	2		17 x 68	2		110 x 110	2
	68 x 90	2		108 x 188	2		72 x 106	2			
	71 x 157	2									

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_10	72 x 106	20	Method_12	17 x 58	14	Method_14	61 x 106	18	Method_16	58 x 153	10
	50 x 65	8		58 x 187	10		50 x 157	6		17 x 58	6
	63 x 110	4		17 x 106	6		56 x 110	6		68 x 81	6
	14 x 106	2		8 x 64	2		56 x 187	4		58 x 187	4
	50 x 143	2		29 x 23	2		25 x 58	2		87 x 86	4
	57 x 65	2		31 x 108	2		56 x 108	2		8 x 58	2
	157 x 174	2		61 x 106	2		57 x 65	2		16 x 63	2
				73 x 90	2					56 x 158	2
										61 x 106	2
										110 x 110	2

Observed crosses - positive selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_17			Method_17			Method_17			Method_17		

Observed crosses - negative selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_17			Method_17			Method_17			Method_17		

Table 4. (continued).

Hybrid pipeline

Simulated crosses - positive selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_2	161 x 27	10	Method_4	48 x 131	16	Method_6	161 x 27	14	Method_8	55 x 117	16
	161 x 126	8		49 x 77	12		161 x 126	8		55 x 88	6
	155 x 23	4		52 x 148	8		155 x 23	6		144 x 126	6
	161 x 125	4		176 x 23	2		161 x 64	4		176 x 23	6
	12 x 188	2		181 x 45	2		12 x 143	2		11 x 130	2
	23 x 89	2					59 x 26	2		131 x 155	2
	40 x 50	2					83 x 26	2		175 x 88	2
	46 x 63	2					128 x 157	2			
	94 x 13	2									
	121 x 65	2									
	131 x 127	2									

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_10	161 x 27	18	Method_12	181 x 19	12	Method_14	161 x 27	16	Method_16	48 x 131	14
	23 x 117	6		40 x 3	6		117 x 9	6		55 x 117	12
	155 x 62	6		162 x 131	6		44 x 65	4		55 x 172	6
	12 x 169	4		55 x 88	4		55 x 157	4		99 x 22	4
	44 x 133	2		95 x 148	4		155 x 62	4		49 x 184	2
	117 x 9	2		185 x 118	4		34 x 62	2		181 x 23	2
	161 x 114	2		46 x 173	2		117 x 130	2			
				52 x 148	2		181 x 26	2			

Simulated crosses - negative selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_2	161 x 39	8	Method_4	99 x 58	12	Method_6	1 x 113	6	Method_8	79 x 58	12
	120 x 81	6		116 x 106	8		76 x 111	6		51 x 58	10
	76 x 111	4		79 x 58	4		116 x 106	4		32 x 58	6
	182 x 81	4		153 x 110	4		142 x 153	4		140 x 4	4
	185 x 110	4		182 x 158	4		80 x 89	2		83 x 106	2
	1 x 130	2		36 x 27	2		105 x 153	2		182 x 27	
	35 x 180	2		53 x 73	2		109 x 69	2			
	80 x 89	2		142 x 153	2		142 x 149	2			
	142 x 124	2		182 x 27	2		145 x 153	2			
	145 x 153	2					150 x 37	2			
	147 x 16	2					152 x 103	2			
	163 x 112	2					159 x 122	2			
							159 x 183	2			
							182 x 81	2			

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_10	142 x 153	8	Method_12	120 x 106	24	Method_14	66 x 86	6	Method_16	152 x 58	16
	53 x 73	6		51 x 58	10		76 x 111	6		67 x 106	6
	80 x 89	6		115 x 58	4		142 x 153	6		79 x 58	6
	145 x 153	4		43 x 153	2		119 x 129	4		51 x 58	4
	151 x 3	4					152 x 103	4		120 x 106	4
	76 x 111	2					163 x 112	4		83 x 106	2
	147 x 45	2					33 x 89	2		129 x 8	2
	153 x 81	2					53 x 126	2			
	161 x 39	2					84 x 16	2			
	163 x 112	2					142 x 185	2			
	167 x 73	2					152 x 74	2			

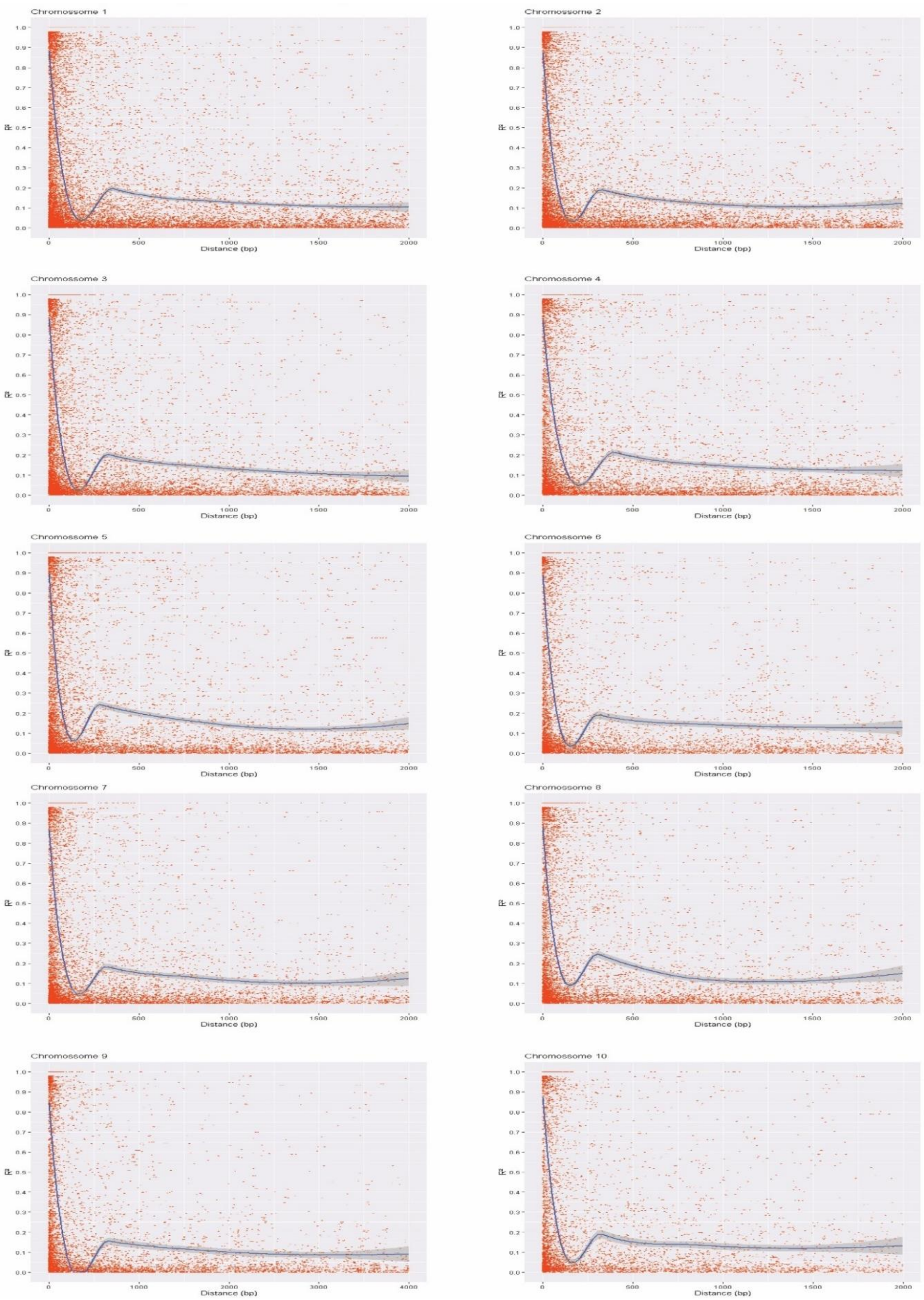
Observed crosses - positive selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_17			Method_17			Method_17			Method_17		

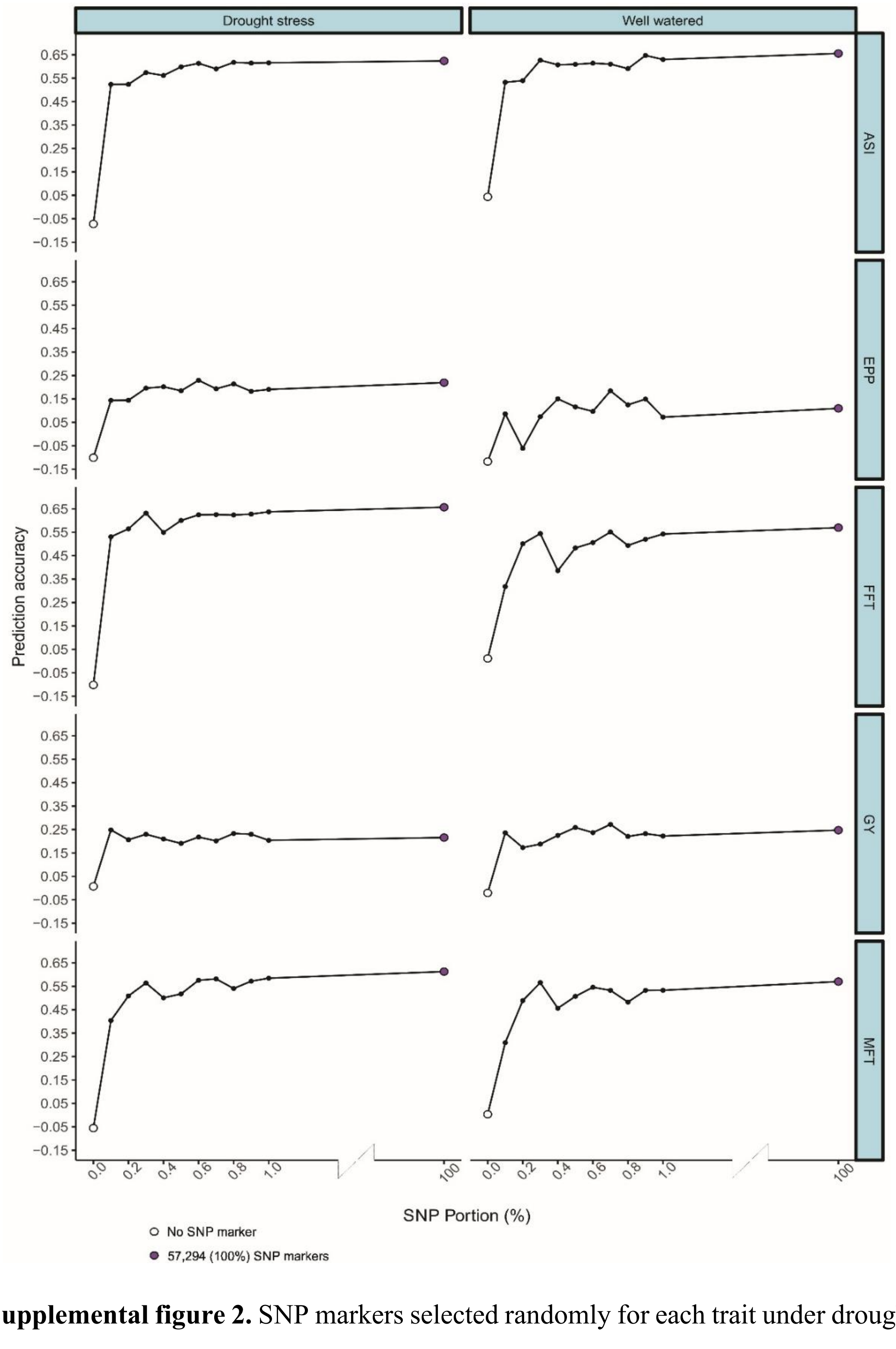
Observed crosses - negative selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_17			Method_17			Method_17			Method_17		

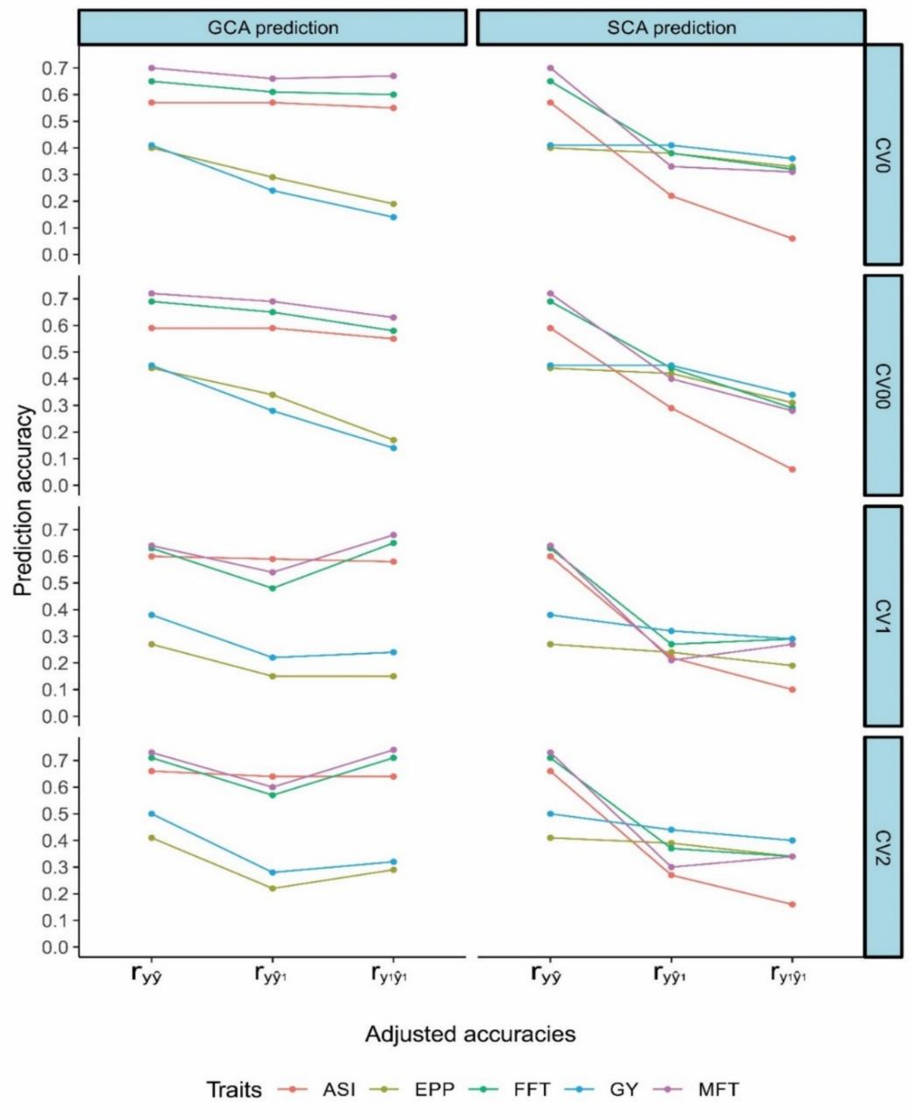
12 SUPPLEMENTAL MATERIAL



Supplemental figure 1. Linkage disequilibrium (LD) analyses of 57,294 SNPs in all corn chromosomes. R^2: coefficient of determination, bp: base pairs.



Supplemental figure 2. SNP markers selected randomly for each trait under drought stress and well-watered conditions. ASI: anthesis-silking interval, EPP: ears per plot, FFT: female flowering time, GY: grain yield, MFT: male flowering time.



Supplemental figure 3. Adjusted prediction accuracies for general (GCA) and specific (SCA) combining abilities obtained as $r_{y\hat{y}1}$ and $r_{y1\hat{y}1}$ for the best fitted models selected according to Table 3. $r_{y\hat{y}}$ is the usual prediction accuracy whose values can be found in Table 3.

For GCA prediction:

$\hat{y}1$: $\hat{y} - \hat{E} - \widehat{GFM} - \hat{\varepsilon}$ (Model 2), $\hat{y} - \hat{E} - \widehat{GFM} - \widehat{GFM.E} - \hat{\varepsilon}$ (Model 3)

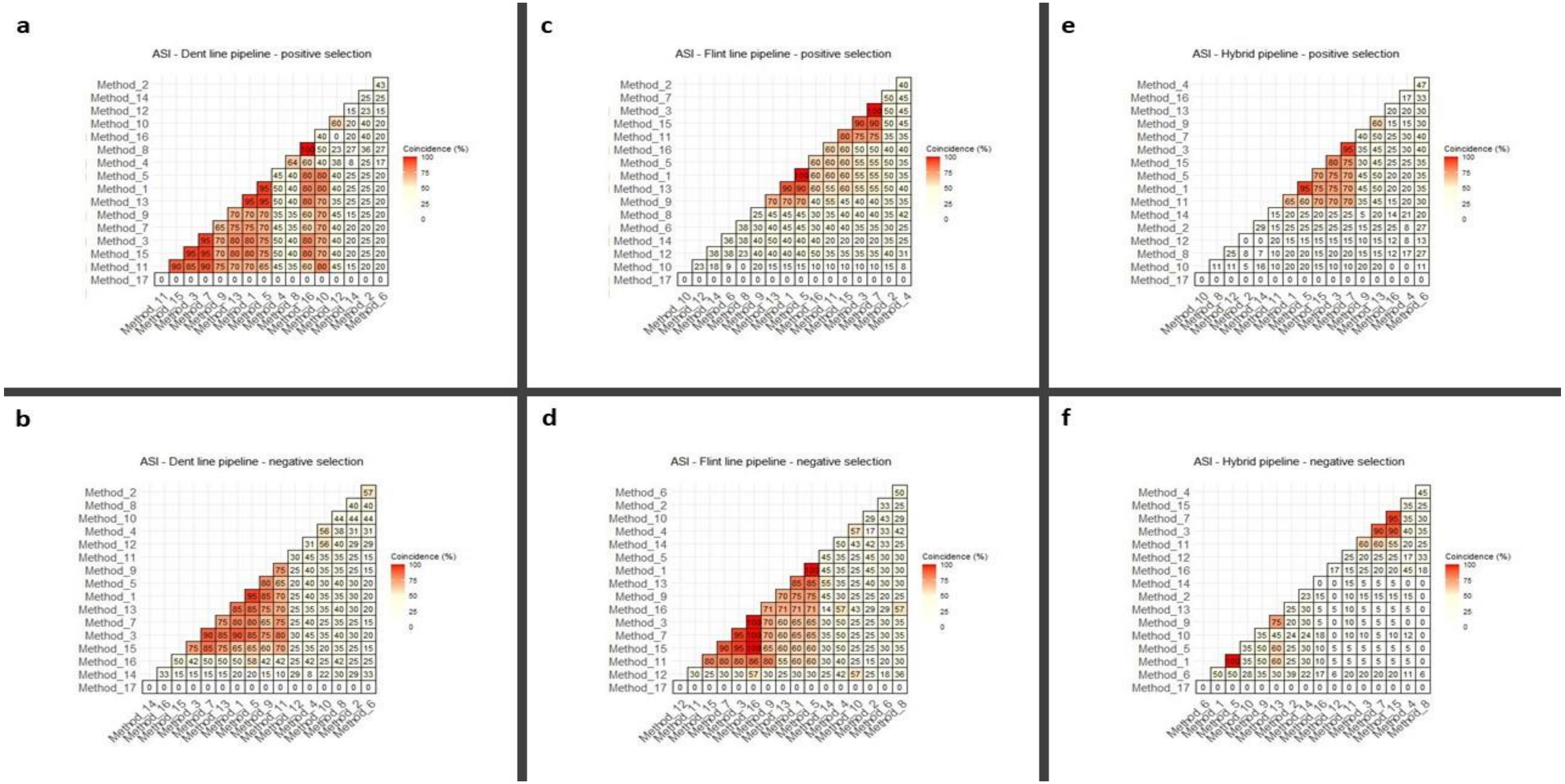
$y1$: $y - E - GFM - \varepsilon$ (Model 2), $y - E - GFM - GFM.E - \varepsilon$ (Model 3)

For SCA prediction:

$\hat{y}1$: $\hat{y} - \hat{E} - \widehat{GF} - \widehat{GM} - \hat{\varepsilon}$ (Model 2), $\hat{y} - \hat{E} - \widehat{GF} - \widehat{GM} - \widehat{GF.E} - \widehat{GM.E} - \hat{\varepsilon}$ (Model 3)

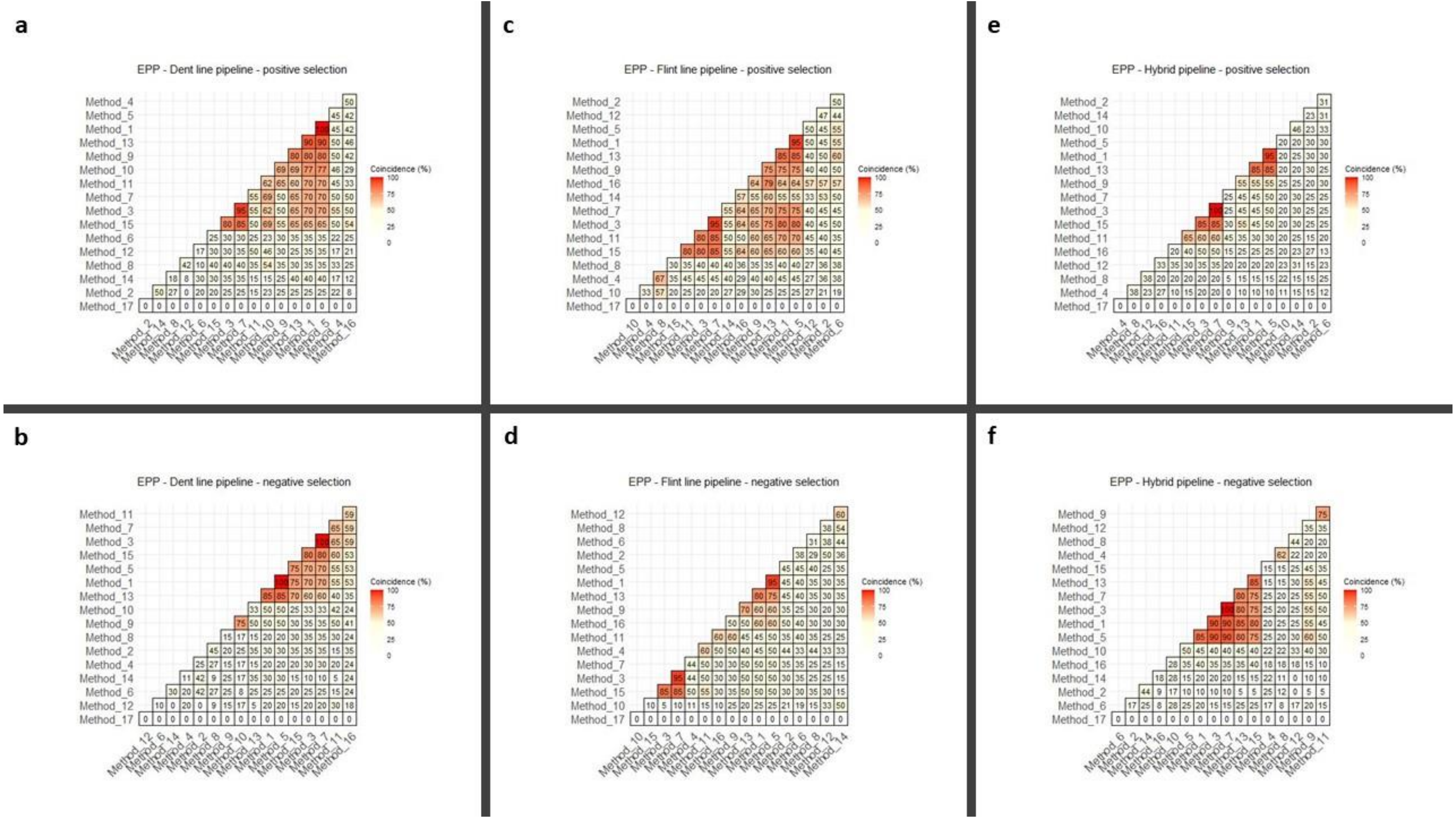
$\hat{y}1$: $y - E - GF - GM - \varepsilon$ (Model 2), $y - E - GF - GM - GF.E - GM.E - \varepsilon$ (Model 3)

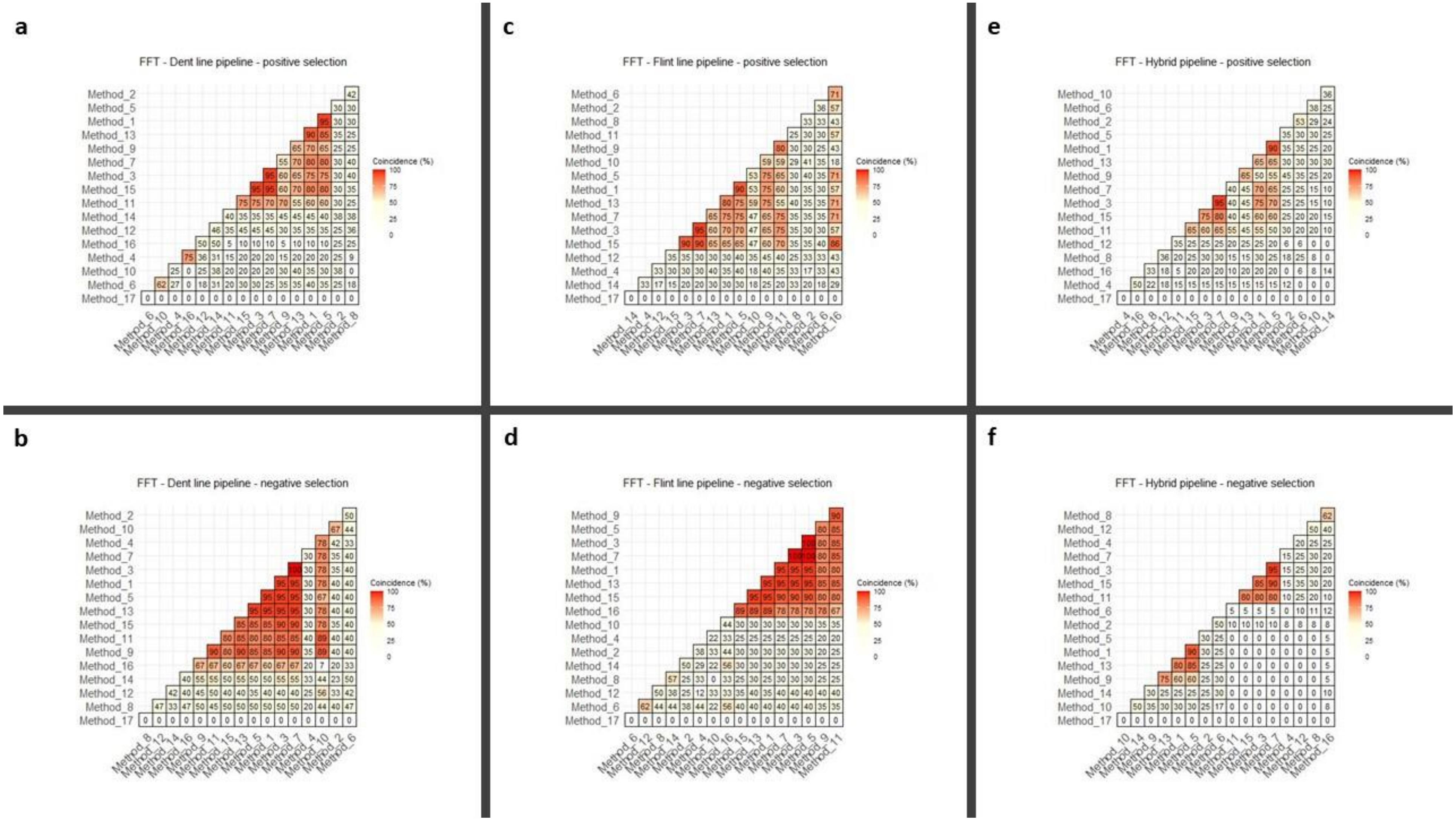
ASI: anthesis-silking interval, EPP: ears per plot, FFT: female flowering time, GY: grain yield, MFT: male flowering time.



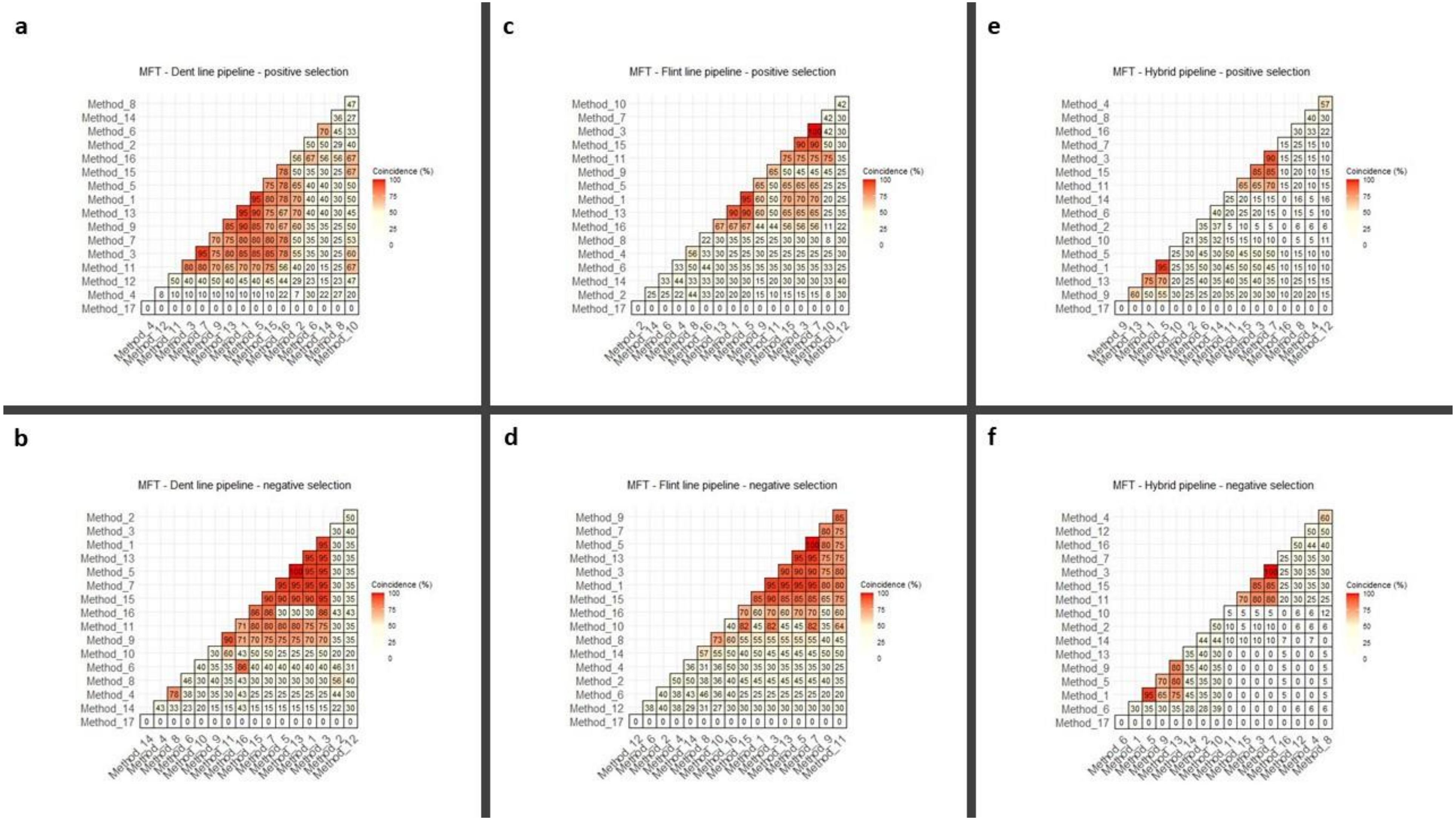
Supplemental figure 4. Coincidences ($C_{\%}$) between selected parent lines for each selection method considering Dent, Flint, and hybrid corn pipelines for anthesis-silking interval (ASI), ears per plot (EPP), female flowering time (FFT), and male flowering time (MFT). Positive selection was carried out on the 40 top-performing crosses. Negative selection was carried out on the 40 bottom-performing hybrid crosses. See criteria for each selection method and for crosses simulations in the items under Material and Methods section.

Supplemental figure 4. (continued).





Supplemental figure 4. (continued).



Supplemental figure 4. (continued).

Supplemental table 1a. Selected crosses for Dent, Flint, and hybrid corn pipelines for anthesis-silking interval (ASI) considering each selection method. Positive selection was carried out on the 40 top-performing crosses. Negative selection was carried out on the 40 bottom-performing crosses. See criteria for each selection method and for crosses simulations in the items under Material and Methods section.

Dent line pipeline

Simulated crosses - positive selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_2	67 x 165	32	Method_4	53 x 183	18	Method_6	67 x 165	30	Method_8	67 x 152	12
	140 x 163	4		2 x 102	12		67 x 152	6		53 x 163	6
	52 x 109	2		109 x 163	4		21 x 161	2		53 x 183	6
	67 x 115	2		67 x 152	2		52 x 184	2		43 x 53	4
				136 x 162	2					53 x 169	4
				147 x 168	2					109 x 163	4
										102 x 183	2
										151 x 165	2

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_10	67 x 140	22	Method_12	140 x 159	16	Method_14	67 x 165	34	Method_16	53 x 163	22
	75 x 165	6		67 x 140	12		43 x 159	2		53 x 183	16
	54 x 109	4		54 x 109	2		48 x 175	2		151 x 165	2
	102 x 183	4		67 x 83	2		79 x 105	2			
	154 x 142	4		120 x 168	2						
				134 x 159	2						
				136 x 152	2						
				154 x 142	2						

Simulated crosses - negative selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_2	36 x 94	26	Method_4	76 x 82	8	Method_6	20 x 41	20	Method_8	20 x 100	16
	20 x 41	6		94 x 164	8		94 x 164	14		20 x 41	8
	46 x 23	4		20 x 144	6		20 x 35	4		49 x 144	8
	94 x 164	4		94 x 40	6		20 x 101	2		94 x 164	4
				20 x 35	4					11 x 100	2
				36 x 145	4					121 x 166	2
				1 x 144	2						
				49 x 59	2						

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_10	36 x 100	18	Method_12	94 x 128	18	Method_14	20 x 41	24	Method_16	36 x 100	12
	20 x 128	10		20 x 20	10		20 x 100	14		20 x 41	6
	20 x 101	6		34 x 144	6		12 x 149	2		36 x 145	6
	94 x 164	4		20 x 100	4					20 x 35	4
	1 x 177	2		1 x 144	2					99 x 144	4
										11 x 95	2
										12 x 35	2
										20 x 95	2
										41 x 151	2

Observed crosses - positive selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_17			Method_17			Method_17			Method_17		

Observed crosses - negative selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_17			Method_17			Method_17			Method_17		

Supplemental table 1a. (continued).

Flint line pipeline

Simulated crosses - positive selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_2	47 x 143	6	Method_4	90 x 125	12	Method_6	90 x 183	14	Method_8	90 x 125	12
	69 x 90	6		57 x 81	10		47 x 184	12		81 x 155	10
	57 x 129	4		69 x 90	8		158 x 186	6		69 x 172	6
	81 x 174	4		122 x 187	4		137 x 171	4		81 x 142	6
	123 x 157	4		56 x 123	2		31 x 148	2		45 x 183	2
	186 x 185	4		122 x 143	2		69 x 90	2		122 x 185	2
	3 x 57	2		157 x 184	2					158 x 186	2
	30 x 47	2									
	56 x 135	2									
	69 x 172	2									
	137 x 169	2									
	158 x 186	2									

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_10	Method_10	47 x 143	Method_12	28 x 47	14	Method_14	158 x 186	14	Method_16	57 x 81	20
	Method_10	110 x 135		47 x 143	8		57 x 129	12		171 x 188	6
				71 x 125	6		122 x 143	4		122 x 143	4
				57 x 135	4		186 x 142	4		137 x 127	4
				90 x 125	2		17 x 142	2		57 x 137	2
				137 x 169	2		47 x 171	2		81 x 103	2
				137 x 171	2		186 x 185	2		143 x 183	2
				186 x 127	2						

Simulated crosses - negative selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_2	8 x 138	18	Method_4	138 x 180	12	Method_6	27 x 138	24	Method_8	73 x 138	18
	8 x 35	6		108 x 40	8		37 x 138	10		5 x 77	16
	14 x 37	6		58 x 138	6		5 x 22	4		5 x 22	2
	9 x 173	4		73 x 138	6		108 x 180	2		108 x 146	2
	5 x 22	2		37 x 130	2					126 x 146	2
	8 x 131	2		73 x 113	2						
	29 x 23	2		77 x 146	2						
				138 x 89	2						

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_10	73 x 138	20	Method_12	27 x 138	22	Method_14	15 x 138	10	Method_16	27 x 138	18
	58 x 138	8		73 x 138	10		71 x 40	8		73 x 138	14
	138 x 180	8		8 x 131	2		37 x 138	6		173 x 89	6
	8 x 138	2		16 x 185	2		9 x 138	4		126 x 146	2
	27 x 183	2		77 x 146	2		108 x 23	4			
				113 x 149	2		108 x 180	4			
							8 x 180	2			
							77 x 111	2			

Observed crosses - positive selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_17			Method_17			Method_17			Method_17		

Observed crosses - negative selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_17			Method_17			Method_17			Method_17		

Supplemental table 1a. (continued).

Hybrid pipeline

Simulated crosses - positive selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_2	53 x 73	14	Method_4	109 x 69	16	Method_6	53 x 39	12	Method_8	165 x 81	34
	129 x 81	14		165 x 90	16		53 x 73	8		165 x 90	4
	84 x 89	4		67 x 171	2		67 x 171	6		159 x 122	2
	147 x 39	4		92 x 22	2		109 x 69	4			
	43 x 153	2		147 x 143	2		53 x 126	2			
	183 x 135	2		156 x 31	2		147 x 172	2			
							155 x 90	2			
							159 x 122	2			
							165 x 90	2			

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_10	161 x 114	10	Method_12	165 x 127	24	Method_14	147 x 173	10	Method_16	109 x 69	20
	119 x 129	6		140 x 114	8		153 x 81	8		140 x 155	16
	159 x 122	6		105 x 50	6		147 x 130	6		53 x 39	2
	136 x 139	4		21 x 122	2		67 x 8	4		170 x 81	2
	163 x 107	4					119 x 129	4			
	11 x 130	2					96 x 155	2			
	18 x 148	2					102 x 81	2			
	32 x 180	2					142 x 153	2			
	168 x 131	2					147 x 143	2			
	182 x 130	2									

Simulated crosses - negative selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_2	21 x 26	8	Method_4	46 x 173	24	Method_6	12 x 68	6	Method_8	97 x 138	18
	94 x 13	8		178 x 5	6		20 x 71	6		46 x 173	8
	144 x 62	8		97 x 40	4		23 x 70	6		178 x 5	6
	12 x 188	6		12 x 126	2		94 x 13	6		36 x 179	4
	49 x 74	6		34 x 7	2		55 x 15	4		177 x 10	2
	6 x 70	2		144 x 126	2		144 x 179	4		177 x 118	2
	96 x 62	2					59 x 13	2			
							82 x 8	2			
							96 x 25	2			
							144 x 62	2			

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_10	12 x 188	8	Method_12	97 x 138	16	Method_14	49 x 71	28	Method_16	20 x 133	26
	52 x 157	6		34 x 146	8		52 x 15	4		20 x 7	6
	94 x 13	6		18 x 77	6		33 x 188	2		12 x 126	4
	145 x 23	6		20 x 133	6		55 x 88	2		94 x 9	2
	44 x 65	4		121 x 58	2		128 x 157	2		178 x 5	2
	55 x 157	4		177 x 118	2		181 x 62	2			
	24 x 64	2									
	161 x 126	2									
	175 x 63	2									

Observed crosses - positive selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_17			Method_17			Method_17			Method_17		

Observed crosses - negative selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_17			Method_17			Method_17			Method_17		

Supplemental table 1b. Selected crosses for Dent, Flint, and hybrid corn pipelines for ears per plot (EPP) considering each selection method. Positive selection was carried out on the 40 top-performing crosses. Negative selection was carried out on the 40 bottom-performing crosses. See criteria for each selection method and for crosses simulations in the items under Material and Methods section.

Dent line pipeline

Simulated crosses - positive selection

Method	Cross	Count
Method_2	79 x 82	22
	93 x 97	8
	94 x 91	4
	97 x 134	6
Method_4	6 x 41	2
	6 x 96	6
	12 x 42	8
	54 x 97	2
	55 x 23	2
	94 x 91	6
	94 x 184	4
	97 x 117	4
	98 x 119	2
	134 x 159	2
	156 x 23	2
Method_6	34 x 144	2
	89 x 117	2
	97 x 134	4
	97 x 178	4
	128 x 134	18
	134 x 184	6
	144 x 129	4
Method_8	12 x 42	2
	43 x 23	10
	93 x 97	2
	94 x 40	16
	95 x 96	4
	96 x 175	6
Method_10	1 x 42	2
	11 x 117	4
	41 x 94	2
	93 x 97	2
	94 x 40	6
	96 x 175	6
	96 x 183	2
	97 x 117	14
	97 x 129	2
Method_12	11 x 117	16
	23 x 40	2
	33 x 86	2
	40 x 185	2
	95 x 96	14
	96 x 175	2
	129 x 155	2
Method_14	34 x 144	2
	79 x 82	18
	94 x 165	2
	99 x 134	8
	99 x 175	2
	134 x 91	8
Method_16	6 x 98	2
	20 x 41	2
	21 x 133	4
	32 x 97	2
	42 x 161	2
	52 x 23	2
	55 x 144	4
	89 x 117	4
	96 x 104	4
	96 x 175	4
	96 x 181	2
	97 x 40	2
	134 x 184	4
	156 x 23	2

Simulated crosses - negative selection

Method	Cross	Count
Method_2	1 x 85	2
	24 x 66	20
	49 x 185	2
	67 x 167	8
	92 x 163	2
	147 x 163	2
	177 x 177	4
Method_4	1 x 150	2
	20 x 53	8
	55 x 66	2
	147 x 182	28
Method_6	1 x 150	2
	12 x 150	2
	24 x 145	16
	67 x 165	2
	67 x 167	4
	150 x 161	4
	177 x 177	10
Method_8	1 x 182	2
	49 x 67	4
	67 x 140	8
	75 x 147	2
	109 x 109	8
	147 x 182	8
	166 x 182	6
	177 x 117	2
Method_10	24 x 59	8
	44 x 52	2
	49 x 170	2
	55 x 55	4
	55 x 182	20
	85 x 102	2
	120 x 91	2
Method_12	36 x 102	2
	42 x 182	4
	53 x 142	10
	102 x 145	8
	162 x 182	2
	164 x 182	6
	182 x 183	8
Method_14	6 x 156	4
	24 x 66	14
	54 x 85	8
	67 x 167	4
	67 x 181	10
Method_16	32 x 67	8
	66 x 84	6
	67 x 167	4
	84 x 85	2
	92 x 120	6
	102 x 145	2
	109 x 178	4
	147 x 131	4
	147 x 182	2
	150 x 148	2

Observed crosses - positive selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_17			Method_17			Method_17			Method_17		

Observed crosses - negative selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_17			Method_17			Method_17			Method_17		

Supplemental table 1b. (continued).

Flint line pipeline

Simulated crosses - positive selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_2	3 x 124	8	Method_4	3 x 118	4	Method_6	4 x 39	6	Method_8	3 x 118	2
	4 x 135	4		10 x 45	6		5 x 22	2		10 x 139	8
	37 x 135	10		10 x 139	2		29 x 117	6		39 x 124	6
	39 x 125	2		19 x 90	4		37 x 135	4		45 x 45	2
	45 x 23	2		37 x 126	2		45 x 172	10		45 x 139	4
	45 x 131	4		45 x 45	2		118 x 184	2		118 x 139	12
	74 x 117	4		45 x 139	2		123 x 139	2		129 x 155	2
	118 x 138	6		118 x 139	8		124 x 179	8		172 x 129	4
				124 x 129	8						
				172 x 129	2						

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_10	3 x 118	20	Method_12	3 x 118	2	Method_14	34 x 144	2	Method_16	3 x 118	4
	3 x 130	16		4 x 135	4		79 x 82	18		4 x 135	6
	87 x 124	2		23 x 40	2		94 x 165	2		26 x 118	2
	139 x 40	2		74 x 129	2		99 x 134	8		39 x 125	2
				107 x 146	2		99 x 175	2		45 x 139	6
				123 x 139	24		134 x 91	8		117 x 184	16
				139 x 160	2					139 x 160	2
				179 x 117	2					141 x 40	2

Simulated crosses - negative selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_2	8 x 157	8	Method_4	10 x 58	2	Method_6	13 x 56	4	Method_8	5 x 58	2
	9 x 17	4		14 x 112	2		13 x 141	2		16 x 63	6
	13 x 56	4		15 x 27	2		30 x 110	2		25 x 58	4
	17 x 89	4		15 x 56	2		47 x 68	2		30 x 108	12
	58 x 64	12		17 x 142	2		50 x 157	4		56 x 60	6
	72 x 106	4		25 x 58	8		56 x 187	2		58 x 187	6
	110 x 112	2		30 x 108	2		58 x 64	8		112 x 157	2
	157 x 158	2		58 x 187	10		58 x 143	6		157 x 174	2
				71 x 157	4		62 x 70	4			
				72 x 106	2		157 x 86	6			
				110 x 110	4						

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_10	50 x 157	18	Method_12	9 x 56	2	Method_14	25 x 58	2	Method_16	8 x 173	2
	63 x 64	4		58 x 64	24		50 x 157	4		14 x 143	2
	71 x 157	6		58 x 187	2		56 x 60	6		17 x 58	32
	157 x 158	12		81 x 103	2		56 x 64	4		26 x 108	2
				110 x 112	2		56 x 187	6		56 x 64	2
				157 x 157	8		58 x 187	4			
							60 x 174	2			
							110 x 110	12			

Observed crosses - positive selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_17			Method_17			Method_17			Method_17		

Observed crosses - negative selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_17			Method_17			Method_17			Method_17		

Supplemental table 1b. (continued).

Hybrid pipeline

Simulated crosses - positive selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_2	6 x 72	2	Method_4	55 x 117	4	Method_6	12 x 23	22	Method_8	23 x 117	4
	12 x 23	28		55 x 172	2		49 x 129	2		42 x 138	2
	33 x 27	2		82 x 45	24		55 x 88	6		52 x 148	4
	55 x 88	2		82 x 77	6		96 x 155	10		55 x 117	28
	82 x 62	2		117 x 130	4					55 x 172	2
	94 x 13	2									
	161 x 27	2									
Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_10	12 x 23	26	Method_12	43 x 123	2	Method_14	12 x 68	8	Method_16	34 x 62	2
	18 x 9	4		55 x 172	2		18 x 9	4		55 x 88	2
	21 x 8	2		96 x 123	2		23 x 57	4		82 x 45	10
	117 x 9	4		117 x 9	22		23 x 117	6		97 x 138	2
	136 x 88	4		119 x 25	2		23 x 179	2		117 x 9	4
				144 x 138	4		75 x 23	2		128 x 70	8
				176 x 23	6		82 x 8	2		144 x 126	6
							96 x 155	4		175 x 70	2
							117 x 9	8		175 x 88	4

Simulated crosses - negative selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_2	40 x 132	2	Method_4	121 x 58	12	Method_6	51 x 114	2	Method_8	32 x 58	6
	76 x 31	2		152 x 58	20		120 x 81	2		53 x 73	2
	153 x 81	6		167 x 27	2		140 x 180	2		121 x 58	4
	182 x 81	28		182 x 17	2		163 x 112	2		152 x 58	18
	185 x 185	2		182 x 81	4		182 x 4	2		182 x 17	10
							182 x 81	20			
							182 x 108	6			
							182 x 130	4			
Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_10	46 x 141	4	Method_12	51 x 157	2	Method_14	46 x 81	16	Method_16	32 x 58	12
	51 x 149	2		53 x 73	2		127 x 81	14		46 x 81	4
	51 x 157	4		120 x 106	6		145 x 153	4		66 x 86	6
	53 x 39	2		150 x 157	8		182 x 108	2		66 x 143	2
	140 x 106	2		152 x 58	22		185 x 110	4		67 x 106	12
	147 x 130	4								147 x 16	4
	152 x 58	4									
	153 x 110	2									
	162 x 58	4									
	182 x 81	6									
	182 x 130	6									

Observed crosses - positive selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_17			Method_17			Method_17			Method_17		

Observed crosses - negative selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_17			Method_17			Method_17			Method_17		

Supplemental table 1c. Selected crosses for Dent, Flint, and hybrid corn pipelines for female flowering time (FFT) considering each selection method. Positive selection was carried out on the 40 top-performing crosses. Negative selection was carried out on the 40 bottom-performing crosses. See criteria for each selection method and for crosses simulations in the items under Material and Methods section.

Dent line pipeline

Simulated crosses - positive selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_2	12 x 156	10	Method_4	2 x 79	2	Method_6	49 x 67	10	Method_8	12 x 156	4
	51 x 170	4		43 x 53	2		54 x 145	2		51 x 170	10
	51 x 175	4		67 x 159	4		79 x 23	6		66 x 129	14
	67 x 116	4		67 x 167	2		79 x 156	18		92 x 163	8
	82 x 129	2		79 x 156	16		85 x 175	2		136 x 182	4
	151 x 129	6		79 x 165	10		136 x 162	2			
	152 x 170	4		156 x 127	2						
	156 x 131	6		156 x 131	2						
Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_10	54 x 85	2	Method_12	51 x 175	8	Method_14	36 x 51	2	Method_16	2 x 131	2
	54 x 145	6		53 x 163	8		53 x 163	2		53 x 163	38
	54 x 151	4		66 x 175	2		54 x 145	6			
	67 x 145	16		67 x 159	2		55 x 66	2			
	79 x 101	2		101 x 182	16		66 x 129	6			
	85 x 152	10		147 x 147	2		67 x 145	4			
				175 x 127	2		116 x 129	12			
							156 x 159	6			

Simulated crosses - negative selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_2	20 x 144	10	Method_4	1 x 169	2	Method_6	20 x 144	4	Method_8	2 x 46	2
	32 x 149	4		44 x 121	2		75 x 154	6		11 x 95	4
	76 x 98	2		44 x 181	2		94 x 164	2		20 x 41	2
	154 x 154	4		75 x 117	2		95 x 168	2		89 x 149	2
	154 x 169	2		76 x 98	8		144 x 154	4		94 x 40	2
	176 x 176	18		99 x 144	8		154 x 169	6		95 x 166	2
				144 x 176	14		176 x 176	16		144 x 176	24
				176 x 176	2					154 x 169	2
Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_10	20 x 144	12	Method_12	20 x 100	4	Method_14	36 x 148	4	Method_16	11 x 95	22
	44 x 98	4		40 x 185	2		44 x 183	6		36 x 100	4
	44 x 181	2		41 x 94	2		75 x 154	8		48 x 117	2
	76 x 98	4		92 x 92	4		94 x 40	12		75 x 154	2
	119 x 169	4		95 x 98	12		94 x 142	2		94 x 185	2
	176 x 176	14		98 x 119	4		100 x 144	6		104 x 149	2
				144 x 176	12		119 x 169	2		165 x 40	2
										176 x 176	4

Observed crosses - positive selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_17			Method_17			Method_17			Method_17		

Observed crosses - negative selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_17			Method_17			Method_17			Method_17		

Supplemental table 1c. (continued).

Flint line pipeline

Simulated crosses - positive selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_2	9 x 69	10	Method_4	17 x 58	26	Method_6	16 x 188	4	Method_8	10 x 58	4
	17 x 107	2		17 x 153	4		17 x 47	16		13 x 171	2
	56 x 171	16		25 x 58	2		17 x 153	6		16 x 62	2
	61 x 106	10		26 x 23	2		50 x 157	2		17 x 68	2
	71 x 157	2		50 x 129	2		61 x 187	10		56 x 187	2
				56 x 187	2		71 x 111	2		58 x 187	14

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_10	9 x 69	2	Method_12	8 x 58	2	Method_14	56 x 60	6	Method_16	17 x 58	28
	14 x 60	6		14 x 125	2		58 x 64	12		17 x 153	6
	17 x 68	2		17 x 47	20		61 x 187	22		61 x 106	2
	31 x 123	2		17 x 106	4					71 x 188	4
	47 x 171	2		47 x 171	4						
	50 x 157	16		50 x 157	4						
	56 x 123	2		58 x 64	2						
	56 x 158	2		124 x 171	2						
	61 x 158	2									
	157 x 184	2									
	157 x 188	2									

Simulated crosses - negative selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_2	37 x 130	8	Method_4	37 x 179	18	Method_6	37 x 179	16	Method_8	19 x 107	4
	45 x 45	2		73 x 138	12		40 x 148	6		37 x 179	22
	45 x 185	6		113 x 149	10		45 x 183	6		45 x 40	10
	112 x 174	22					73 x 138	4		45 x 45	4
	113 x 149	2					73 x 174	8			

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_10	22 x 174	10	Method_12	19 x 107	6	Method_14	19 x 31	2	Method_16	37 x 126	4
	73 x 174	2		35 x 35	4		37 x 179	26		37 x 179	8
	86 x 149	2		40 x 148	4		45 x 45	4		45 x 45	8
	103 x 149	8		45 x 40	12		112 x 174	8		73 x 174	16
	112 x 117	12		73 x 174	14					112 x 117	2
	174 x 35	6								172 x 174	2

Observed crosses - positive selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_17			Method_17			Method_17			Method_17		

Observed crosses - negative selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_17			Method_17			Method_17			Method_17		

Supplemental table 1c. (continued).

Hybrid pipeline

Simulated crosses - positive selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_2	51 x 113	2	Method_4	51 x 58	20	Method_6	67 x 171	6	Method_8	6 x 17	4
	53 x 4	2		79 x 58	14		80 x 125	4		67 x 171	6
	67 x 171	4		115 x 58	4		96 x 179	2		79 x 58	22
	80 x 125	4		183 x 106	2		142 x 153	8		80 x 64	4
	142 x 153	10					145 x 153	2		152 x 58	4
	145 x 153	4					152 x 91	2			
	147 x 39	6					153 x 74	4			
	153 x 74	2					159 x 107	2			
	159 x 183	2					182 x 3	10			
	181 x 153	4									

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_10	54 x 139	2	Method_12	6 x 17	2	Method_14	32 x 38	2	Method_16	79 x 58	6
	105 x 153	10		51 x 47	8		117 x 153	6		115 x 58	2
	129 x 171	4		51 x 50	22		119 x 129	2		129 x 81	10
	142 x 153	8		66 x 56	2		142 x 153	2		152 x 58	22
	147 x 45	6		152 x 58	4		145 x 153	2			
	153 x 74	2		162 x 14	2		147 x 3	6			
	182 x 3	8					153 x 81	4			

Simulated crosses - negative selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_2	12 x 103	2	Method_4	33 x 89	2	Method_6	12 x 4	2	Method_8	94 x 122	2
	12 x 188	6		95 x 37	34		12 x 188	8		95 x 37	26
	23 x 89	2		100 x 133	4		48 x 28	2		149 x 148	6
	85 x 26	12					121 x 65	22		176 x 19	6
	121 x 65	10					176 x 23	6			
	144 x 62	2									
	176 x 23	6									

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_10	6 x 70	6	Method_12	34 x 149	12	Method_14	6 x 70	4	Method_16	23 x 19	2
	12 x 68	2		92 x 149	8		12 x 68	8		95 x 37	10
	23 x 58	6		95 x 37	2		23 x 58	4		95 x 47	6
	24 x 64	12		176 x 148	6		121 x 14	20		117 x 149	2
	59 x 13	2		178 x 149	10		156 x 50	4		149 x 148	20
	85 x 68	10		183 x 40	2						
	181 x 23	2									

Observed crosses - positive selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_17			Method_17			Method_17			Method_17		

Observed crosses - negative selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_17			Method_17			Method_17			Method_17		

Supplemental table 1d. Selected crosses for Dent, Flint, and hybrid corn pipelines for male flowering time (MFT) considering each selection method. Positive selection was carried out on the 40 top-performing crosses. Negative selection was carried out on the 40 bottom-performing crosses. See criteria for each selection method and for crosses simulations in the items under Material and Methods section.

Dent line pipeline

Simulated crosses - positive selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_2	6 x 151	2	Method_4	11 x 120	2	Method_6	6 x 156	2	Method_8	18 x 51	14
	6 x 175	2		18 x 51	8		18 x 51	2		23 x 129	2
	12 x 83	2		51 x 142	2		24 x 59	4		34 x 105	2
	12 x 156	4		51 x 175	28		24 x 162	8		51 x 115	4
	24 x 66	6					51 x 175	22		79 x 101	14
	24 x 145	6					79 x 101	2		85 x 175	4
	24 x 162	4									
	79 x 101	10									
	82 x 83	2									
	85 x 175	2									

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_10	1 x 182	2	Method_12	6 x 101	22	Method_14	6 x 156	2	Method_16	24 x 162	2
	18 x 51	2		51 x 163	2		24 x 162	14		51 x 115	12
	51 x 115	2		52 x 150	2		51 x 175	4		51 x 175	22
	52 x 150	10		53 x 167	4		79 x 156	12		67 x 167	2
	66 x 129	2		54 x 145	6		85 x 175	6		79 x 101	2
	67 x 145	6		67 x 145	2		156 x 159	2			
	85 x 175	12		151 x 129	2						
	101 x 182	2									
	115 x 162	2									

Simulated crosses - negative selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_2	36 x 168	2	Method_4	42 x 169	2	Method_6	11 x 20	4	Method_8	42 x 154	6
	42 x 154	2		75 x 117	2		33 x 165	2		75 x 117	2
	75 x 154	4		75 x 154	26		36 x 168	2		109 x 119	2
	76 x 148	2		144 x 154	2		75 x 154	8		119 x 169	4
	119 x 169	30		176 x 117	4		119 x 169	4		144 x 154	18
				176 x 176	4		154 x 169	12		176 x 176	8
							176 x 117	4			
							176 x 185	4			

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_10	20 x 95	2	Method_12	21 x 148	2	Method_14	144 x 154	26	Method_16	33 x 165	4
	75 x 154	22		44 x 183	2		147 x 154	2		95 x 168	4
	94 x 91	2		98 x 119	4		154 x 127	4		154 x 169	28
	95 x 98	6		144 x 154	6		154 x 168	6		176 x 176	4
	136 x 177	2		154 x 168	6		176 x 176	2			
	176 x 176	6		154 x 169	20						

Observed crosses - positive selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_17			Method_17			Method_17			Method_17		

Observed crosses - negative selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_17			Method_17			Method_17			Method_17		

Supplemental table 1d. (continued).

Flint line pipeline

Simulated crosses - positive selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_2	17 x 58	30	Method_4	7 x 13	4	Method_6	13 x 171	4	Method_8	13 x 171	2
	29 x 62	2		9 x 17	6		16 x 61	2		17 x 58	6
	58 x 153	8		17 x 68	2		17 x 58	28		29 x 62	2
				17 x 106	4		25 x 106	2		58 x 187	28
				58 x 64	16		30 x 108	2		106 x 160	2
				58 x 187	8		47 x 62	2			

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_10	8 x 25	2	Method_12	8 x 157	4	Method_14	17 x 68	14	Method_16	7 x 39	2
	8 x 157	26		9 x 129	2		58 x 187	16		16 x 61	2
	9 x 17	4		17 x 58	20		61 x 187	4		17 x 58	26
	14 x 47	2		26 x 88	2		71 x 108	4		58 x 153	2
	17 x 68	2		26 x 108	2		71 x 157	2		69 x 77	8
	27 x 108	2		29 x 157	10						
	56 x 60	2									

Simulated crosses - negative selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_2	19 x 107	2	Method_4	37 x 179	20	Method_6	22 x 113	2	Method_8	4 x 45	2
	37 x 179	18		45 x 45	6		37 x 126	2		35 x 35	2
	73 x 174	14		103 x 149	4		45 x 45	6		37 x 126	2
	89 x 149	2		112 x 174	6		73 x 174	16		38 x 169	2
	112 x 117	4		173 x 174	4		107 x 107	14		40 x 183	2
										45 x 45	6
										73 x 174	10
										107 x 149	14

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_10	37 x 91	2	Method_12	45 x 40	4	Method_14	19 x 142	2	Method_16	37 x 130	2
	38 x 169	6		50 x 174	4		31 x 40	2		45 x 45	20
	38 x 174	6		73 x 174	22		37 x 126	2		45 x 183	6
	40 x 148	4		112 x 117	10		45 x 45	16		73 x 138	2
	45 x 45	14					73 x 174	2		103 x 149	6
	86 x 149	2					73 x 184	8		122 x 174	4
	107 x 40	6					89 x 149	2			
							107 x 107	2			
							173 x 174	4			

Observed crosses - positive selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_17			Method_17			Method_17			Method_17		

Observed crosses - negative selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_17			Method_17			Method_17			Method_17		

Supplemental table 1d. (continued).

Hybrid pipeline

Simulated crosses - positive selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_2	18 x 77	4	Method_4	51 x 180	8	Method_6	51 x 47	4	Method_8	23 x 58	8
	46 x 141	2		79 x 58	12		66 x 127	2		51 x 91	4
	53 x 39	2		102 x 160	6		67 x 127	4		79 x 58	22
	66 x 127	4		152 x 58	14		80 x 123	2		80 x 111	2
	80 x 3	2					93 x 143	4		150 x 157	2
	102 x 29	2					96 x 111	2		152 x 58	2
	102 x 113	4					129 x 171	2			
	142 x 127	4					145 x 153	2			
	142 x 153	6					151 x 3	6			
	145 x 153	2					153 x 74	6			
	151 x 3	8					162 x 111	2			
							181 x 153	4			

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_10	1 x 113	2	Method_12	46 x 129	4	Method_14	46 x 16	2	Method_16	78 x 9	2
	35 x 179	2		51 x 58	26		51 x 114	2		79 x 58	12
	51 x 113	6		51 x 180	2		53 x 4	2		152 x 58	22
	116 x 135	2		152 x 58	8		53 x 39	2		167 x 17	2
	129 x 171	2					80 x 3	6		168 x 60	2
	134 x 122	2					129 x 141	2			
	145 x 153	2					153 x 40	4			
	147 x 3	6					153 x 74	2			
	153 x 40	2					153 x 110	10			
	159 x 169	4					159 x 5	2			
	181 x 153	10					162 x 111	6			

Simulated crosses - negative selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_2	6 x 70	4	Method_4	95 x 37	10	Method_6	23 x 19	2	Method_8	42 x 60	2
	6 x 72	6		154 x 19	22		24 x 64	8		95 x 37	2
	12 x 188	2		176 x 148	2		59 x 13	2		154 x 19	20
	23 x 70	2		176 x 149	4		83 x 25	4		167 x 73	2
	117 x 61	8		185 x 118	2		85 x 103	2		176 x 149	14
	121 x 89	4					98 x 28	2			
	144 x 62	2					121 x 58	4			
	154 x 27	2					156 x 50	6			
	155 x 15	6					161 x 27	8			
	156 x 50	4					161 x 64	2			

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_10	12 x 23	2	Method_12	34 x 149	4	Method_14	6 x 72	12	Method_16	21 x 65	2
	12 x 68	10		95 x 107	2		12 x 23	4		95 x 37	6
	12 x 188	4		154 x 19	20		12 x 169	6		168 x 174	2
	24 x 64	8		154 x 127	6		12 x 188	2		176 x 149	30
	40 x 50	2		168 x 174	6		121 x 65	2			
	59 x 13	4		176 x 149	2		148 x 13	2			
	89 x 60	2					155 x 62	2			
	154 x 27	2					161 x 64	10			
	155 x 62	6									

Observed crosses - positive selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_17			Method_17			Method_17			Method_17		

Observed crosses - negative selection

Method	Cross	Count	Method	Cross	Count	Method	Cross	Count	Method	Cross	Count
Method_17			Method_17			Method_17			Method_17		