Received: 21 November 2019 | Accepted: 11 February 2020

DOI: 10.1002/tpg2.20014

The Plant Genome

ORIGINAL RESEARCH

Genomic prediction of maternal haploid induction rate in maize

Vinícius Costa Almeida[1,2] | **Henrique Uliana Trentin**[1] | **Ursula Karoline Frei**[1] | **Thomas Lübberstedt**[1]

[1]Department of Agronomy, Iowa State University, 716 Farm House Ln, Ames, IA, 50011–1051, USA

[2]General Biology Department, Federal University of Viçosa, Peter H Rolfs Avenue, Viçosa, MG, 36570–900, Brazil

Correspondence
Vinícius Costa Almeida and Ursula Karoline Frei, Department of Agronomy, Iowa State University, 716 Farm House Ln, Ames, IA, 50011–1051, USA.
Email: viniciosallmeida@gmail.com; ufrei@iastate.edu

Funding information
National Institute of Food and Agriculture, Grant/Award Numbers: IOW01018, IOW04314, IOW05510, NIFA award 2018-51181-28419

Abstract

Genomic prediction (GP) might be an efficient way to improve haploid induction rate (HIR) and to reduce the laborious and time-consuming task of phenotypic selection for HIR in maize (*Zea mays* L.). In this study, we evaluated GP accuracies for HIR and other agronomic traits of importance to inducers by independent and cross-validation. We propose the use of GP for cross prediction and parental selection in the development of new inducer breeding populations. A panel of 159 inducers from Iowa State University (ISU set) was genotyped and phenotyped for HIR and several agronomic traits. The data of an independent set of 53 inducers evaluated by the University of Hohenheim (UOH set) was used for independent validation. The HIR ranged from 0.61 to 20.74% and exhibited high heritability (0.90). High cross-validation prediction accuracy was observed for HIR (r = 0.82), whereas for other traits it ranged from 0.36 (self-induction rate) to 0.74 (days to anthesis). Prediction accuracies across different sets were higher when the larger panel (ISU set) was used as a training population (r = 0.54). The average HIR of the 12,561 superior predicted progenies (μ_{SP}) ranged from 1.00–18.36% and was closely related to the corresponding midparent genomic estimated breeding value (GEBV). A predicted genetic variance (V_G) of reduced magnitude was observed in the twenty crosses with highest midparent GEBV or μ_{SP} for HIR. Our results indicate that although GP is a useful tool for parental selection, decisions about which cross combinations should be pursued need to be based on optimal trade-offs between maximizing both μ_{SP} and V_G.

1 | INTRODUCTION

In modern maize breeding, doubled haploid (DH) technology has allowed breeders to speed up breeding cycles by the rapid development of inbred lines (De La Fuente, Frei, & Lübberstedt, 2013). The success of the DH technology lies on the efficient production of haploid seeds from segregating populations. For the *in vivo* induction of maternal haploids, pollen of the haploid inducer is used to pollinate a maternal donor from which DH lines are to be developed (Prigge, Schipprack, Mahuku, Atlin, & Melchinger, 2012a). Therefore, availability of haploid inducers with high haploid induction rate (HIR) is a key factor to reduce the cost of DH line production.

It was recently discovered that a 4-bp insertion in the last exon of a pollen-specific phospholipase, named

Abbreviations: DH, doubled haploid; DTA, days to anthesis; GBLUP, genomic best linear unbiased prediction; GDU, growing degree units; GEBV, genomic estimated breeding value; GP, genomic prediction; HIR, haploid induction rate; ISU, Iowa State University; QTL, quantitative trait loci; SIR, self-induction rate; SNP, single nucleotide polymorphism; UOH, University of Hohenheim.

MATRILINEAL (*MTL*), substantially increases the HIR of maize maternal haploid inducers (Gilles et al., 2017; Kelliher et al., 2017; Liu et al., 2017). A single nucleotide polymorphism (SNP) in the *ZmDMP* gene, which encodes for a DUF679 domain protein, enhances the HIR of inducers fixed for the *mtl* allele by 2- to 3-fold (Liu et al., 2015; Zhong et al., 2019). However, multiple genes are known to affect HIR in cross-pollinations with inducers (Deimling, Röber, & Geiger, 1997; Dong et al., 2013; Kelliher et al., 2017; Prigge et al., 2012a; Trentin, Almeida, Frei, & Lubberstedt, unpublished data). Apart from *MTL* and *ZmDMP*, which are located on chromosomes 1 and 9, respectively, quantitative trait loci (QTL) on chromosomes 3 (*qhir2* and *qhir3*), 4 (*qhir4*), 5 (*qhir5* and *qhir6*), 7 (*qhir7*) have been identified. Through breeding, the accumulation of beneficial alleles resulted in an increase of HIR from 3.2% (Stock 6, Coe, 1959) to 14.5% (PHI-3, Rotarenco, Georgeta, & Fuia, 2010). Improving HIR can significantly decrease the cost of DH line production by reducing the amount of land and labor required to produce and select the desirable amount of haploid seeds. A significant improvement towards the automation of haploid selection was recently achieved with the development of high oil content inducers and of machines capable of discriminating haploid and hybrid seeds based on their differences in oil content (Melchinger, Schipprack, Wurschum, Chen, & Technow, 2013; Wang et al., 2016). Since both oil content and HIR are governed by multiple QTL, genomic selection might be a better approach for their simultaneous improvement than the introgression of a limited number of QTLs by marker-assisted backcrossing or selection.

Meuwissen, Hayes, and Goddard (2001) proposed the simultaneous use of all molecular markers across the genome to capture small and large effect QTLs and to predict the breeding values of selection candidates. This genome-wide selection approach, commonly referred to as genomic prediction (GP), relies on the establishment of a prediction model based on phenotypic and genotypic information of a training set to predict the genetic merit of individuals with only genotypic information in a breeding population. We believe that GP will optimize resource-allocation of inducer breeding programs by the genotyping of a great number of selection candidates and field-testing of only the most promising ones.

One of the challenges in an inducer development program is identifying suitable parents among several possible crosses to derive new populations. According to Bernardo (2014), the progeny of desirable crosses should have high mean and large genetic variance, which would give breeders greater opportunities to select superior individuals for more generations of breeding. To address the trade-offs between decisions about selecting lines to intermate based on high genotypic values and maintaining potential genetic variability, Schnell and Utz (1975) proposed the usefulness concept (U_P), which is a function of the population mean, genetic variance, heritability and selection intensity predicted or applied to a given bi-parental cross. A modification to U_P was proposed by Zhong and Jannink (2007), which estimates the achievable genetic gain from a cross by focusing on the mean and genetic variance of progenies while disregarding trait heritability.

Core Ideas

- Genomic prediction can be successfully implemented to select superior haploid inducers
- Midparent GEBV played the major role in the prediction of superior inducer progenies
- Induction rate can be improved without negatively impacting agronomic performance

Unlike population mean, which can be easily predicted based on midparent values, estimating genetic variance relies on knowledge of population parameters that are not readily known (Bernardo, 2010). Recent studies have used GP to evaluate the potential of crosses (Bernardo et al., 2014; Lado et al., 2017; Mohammadi, Tiede, & Smith, 2015; Tiede, Kumar, Mohammadi, & Smith, 2015). Based on the predicted progeny means and genetic variances of crosses involving 75 lines, Yao, Zhao, Chen, Zhang, and Wang (2018) were able to identify potential crosses for improving both yield and quality traits in wheat. In an inducer development program, it would be of interest to improve the agronomic performance of inducers besides increasing HIR. Poor pollen production, plant vigor, lodging tolerance and disease susceptibility are common disadvantages of existing inducers (Kebede et al., 2011; Rotarenco et al., 2010). There has been an undocumented concern that HIR and agronomic performance might be negatively correlated. Therefore, cross prediction would be a great approach to identify promising parents to generate progenies with satisfactory agronomic performance and high HIR before crosses are made in the field.

We are unaware of the use of GP for parental selection and for the improvement of HIR or agronomic performance of maize maternal haploid inducers. Phenotypic selection was employed in the few cases in which the breeding schemes of inducer development were described (Chaikam et al., 2015; Kebede et al., 2011; Rotarenco et al., 2010). The objectives of this study were to (1) evaluate the GP accuracy for HIR and agronomic traits important to maize maternal haploid inducers, to (2) validate GP procedures for HIR by using different sets of inducers as training and validations populations, and to (3) evaluate the utility of GP in cross prediction and parental selection.

2 | MATERIALS AND METHODS

2.1 | Germplasm

We used a diverse panel of North American and European haploid inducers to identify QTL affecting HIR. The North American panel contained inducers in the background of different public and Ex-PVP lines, such as A637, B73, DK78004, LH82, Mo17, PHG50 and Va35, and ranged from the F_4 to mostly homozygous (Supplemental Table S1). All lines from the North American panel had the cross between RWS and RWK as their source of *qhir1*. The European panel included lines such as, MHI, PHI-3, RWS, RWK and eight F_7 progeny derived from the cross of RWS and PHI-3, which were selected for high HIR. The HIR of all inducers was evaluated in crosses with the commercial hybrid Viking 60-01N, from Albert Lea Seed Company (Minnesota, USA). This hybrid was chosen as a donor because it possesses good inducibility and allows clear expression of the *R1-nj* marker. Inducibility is the ability of the donor parent in generating haploid seeds, and multiple studies indicated that the source germplasm has an impact on HIR (Chase, 1952; De La Fuente et al., 2018; Eder & Chalyk, 2002; Lashermes & Beckert, 1988; Randolph, 1940).

2.2 | Field plot design and data collection

The inducers and donor used for this experiment were sown at the Iowa State University Agronomy and Agricultural Engineering Farm, located in Boone, IA, during the summer of 2018. The trial was grown under rainfed conditions, following the same practices used for maize production in Central Iowa. Pre and post-emergence herbicides along with hoeing were used for weed control. Urea ammonium nitrate was applied in the area before sowing. Two blocks of inducer and donor genotypes were sown at different planting dates to ensure enough seeds would be produced to obtain reliable estimates of HIR. The first block of donor was sown on 8 May 2018 and the first block of inducers was sown on 21 May 2018. The second block of donor was sown along with the first block of inducers, while the second block of inducers was sown on 31 May 2018. Sowing of inducer blocks were delayed because most inducers have a significantly shorter maturity than the donor used.

Due to the differences in sowing dates, the first donor block was pollinated only by the first inducer block, and the second donor block was only pollinated by the second inducer block. In this study, the blocking term of the statistical model was confounded with planting date. Therefore, all environmental variability, whether caused by planting date, different physiological development, soil variability or some other factor, was captured by the blocking term. The weather condition surrounding the peak of pollen shed of inducers from the first block (15 July 2018) were considerably warmer and less humid (avg. max. 30.7°C, avg. 25.1°C, avg. min. 18.9°C, avg. RH 78.1%) than the same period for the inducer block (24 July 2018; avg. max. 25.5°C, avg. 21.8°C, avg. min. 20.9°C, avg. RH 88.4%). Given limited resources, conducting the same experiment in an additional location or planting date, or with an additional donor, would not have been possible due to logistics and the extensive time required for pollination and haploid selection. Each inducer and donor blocks were composed of subblocks containing 16 plots. Inducers were not randomized within subblocks because the great difference in vigor among them would adversely affect other traits for which data were collected for a companion study. For example, if hybrid and inbred inducers were randomized, shading caused by differences in plant height would be detrimental to the development of inbred lines. Plots were 5.5 m long and 0.75 m wide and were sown with 25 seeds. Inducer and donor blocks were sown side-by-side, and pollen from inducers in a given subblock was carried to the adjacent donor subblock. Multicolored tags with easy-to-match codes were used to ensure that the pollen from each inducer plot was placed in the corresponding donor plot. For the two planting dates, days to anthesis (DTA, GDUs) was recorded when 50% of the plants of a plot were shedding pollen. Plant height (cm), ear height (cm), and tassel size (cm) were recorded on three representative plants of each plot. Plant and ear heights were measured as the distance between the soil surface to the node of insertion of the flag leaf and upper ear, respectively. Tassel size was measured as the distance between the first tassel branch and the top of the tassel. Self-induction rate (SIR, %) was calculated as the ratio of haploid plants over the total number of plants in an inducer plot.

2.3 | Haploid induction and selection

At each planting date, bulk pollen from each inducer plot was collected in tassel bags and used to pollinate at least ten ears of the donor, which were covered before silk emergence using wax bags. Due to nicking issues, we were not able to pollinate the targeted number of ears for all inducers. Ears were harvested when seeds reached the black layer stage and were air-dried for one week. Visual haploid selection was performed using the *R1-nj* phenotypic marker, and the number of putative haploid and diploid seeds per ear was recorded. Putative haploid seeds from each inducer were saved in a single envelope, whereas putative diploid seeds were discarded. The ploidy of the putative haploid seeds was verified by cutting them in halves and checking the presence of anthocyanin pigmentation in the embryonic region. Subsequently, the actual number of haploid and diploid seeds within each planting date were summed to calculate the HIR. Embryo and endosperm

abortion, which are correlated with HIR (Prigge et al., 2012a; Xu et al., 2013; Zhao, Xu, Xie, Chen, & Jin, 2013), occur at different stages of seed development (Xu et al., 2013). This makes the identification of embryo and endosperm aborted seeds quite subjective, and for this reason, these two classes of seeds were not included in the formula to calculate HIR.

In total, 3,396 ears set seed and were evaluated, corresponding to an average of 9.3 ears per inducer per planting date. Two out of the targeted ten ears pollinated by each inducer at each planting date were evaluated by the second author of this manuscript, while the remaining ears were evaluated by a team of trained undergraduate and graduate students. For each inducer and planting date, the HIR data of the second author's sample was compared to the HIR data of other evaluators. If, for a given inducer in a given planting date, an underestimation of 3.0% or more in HIR data of the team of evaluators was noticed, and such underestimation contributed to a reduction of 2.0% or more in the overall HIR, the data from such inducer were excluded from further analysis. Overestimated HIR data did not represent a problem, since the ploidy of the putative haploid seeds was subsequently verified by cutting the seeds and checking for anthocyanin pigmentation in the embryonic region. Data from inducers that did not produce at least 300 seeds on a given planting date were also excluded from further analysis. After this culling, HIR data of 159 inducers (ISU inducer set) across the two planting dates remained.

2.4 | Statistical analysis of phenotypic data

The phenotypic data of HIR, DTA, plant height, ear height, tassel size, and SIR were analyzed using a random model by likelihood methods (REML) which was part of the function *mmer* of the *sommer* R package (Covarrubias-Pazaran, 2016). Planting dates were considered as blocks. The statistical model was:

$$y_{ij} = \mu + B_i + G_j + e_{ij}$$

where μ is the overall mean; B_i ($i = 1, \ldots, I$) is the random effect of the ith block; $G_j = (1, \ldots, J)$ is the random effect of the jth inducer; and e_{ij} is the random error term. With the *sommer* package, we obtained best linear unbiased prediction (BLUP) adjusted means and variance components for each trait. Heritability on an entry-mean basis, was calculated based on variance estimates, through the following formula: $h^2 = \sigma_g^2/(\sigma_g^2 + \sigma_e^2/r)$, in which σ_g^2 and σ_e^2 correspond to genotypic and residual variance, respectively, and r corresponds the number of blocks (Hallauer, Carena, & Miranda, 2010). The BLUP adjusted means of the inducers were used to calculate the phenotypic correlation matrix, using the Pearson's correlation coefficient implemented in the CORR procedure in SAS 9.4 (SAS Institute, 2016).

2.5 | Genotyping and quality control

Leaf samples of each inducer were collected at seedling stage (V3), lyophilized and sent to the International Maize and Wheat Improvement Center (CIMMYT) for genotyping using the Diversity Arrays Technology platform (Kilian et al., 2012). Markers were aligned to version 4 of the B73 reference genome. From the 32,929 SNP markers detected by this platform, 16,639 remained after filtering for a call rate of at least 50% and minor allele frequency of 5%. The software Beagle 5.0 (Browning, Zhou, & Browning, 2018) was subsequently used for the imputation of missing data.

2.6 | Genomic prediction using the GBLUP model

The BLUP adjusted means of each trait were used in the prediction of the genomic estimated breeding values (GEBVs) through the GBLUP model:

$$\mathbf{y} = \mathbf{1}_n \mu + \mathbf{Zu} + \varepsilon$$

in which $\mathbf{y}$ is the vector of phenotypic values (BLUP adjusted means); $\mathbf{1}$ is the unit vector of length n (here n is the number of inducers); μ is the fixed model intercept; $\mathbf{u}$ is the estimated vector of random genetic effects of inducer with covariance matrix $\mathbf{K}\sigma_g^2$; $\mathbf{K}$ is the estimated genomic relationship matrix; and σ_g^2 is the estimated additive variance. The vector of random residuals ε has a covariance matrix $\mathbf{I}\sigma_e^2$, where $\mathbf{I}$ is an identity matrix and σ_e^2 is the residual variance. Term $\mathbf{Z}$ is the incidence matrix of $\mathbf{u}$. The R package rrBLUP (Endelman, 2011) was used for these analyses.

2.7 | Accuracy of prediction

To assess the predictive ability of GBLUP, we performed cross-validation for all measured traits and independent validation for HIR. Cross-validation was performed by randomly dividing the ISU inducer set into training and validation populations, which contained 60.0 and 40.0% of the individuals, respectively. Accuracy was estimated as the Pearson's correlation between the predicted GEBVs and the observed BLUP adjusted means. The average of fifty iterations was used as the measure of accuracy for each trait. Independent validation was performed using the phenotypic and genotypic information of a diverse set of 53 inducers evaluated by the University of Hohenheim (UOH set, Hu et al., 2016). This inducer set was genotyped with the Illumina MaizeSNP50 BeadChip (Ganal et al., 2011), which generated data of 40,572 SNPs. The genotypic data of this dataset was realigned from the version two to the version four of the B73's reference genome using

TABLE 1 Estimates of range, mean, genotypic variance and heritability on an entry-mean basis of haploid induction rate (HIR, %), days to anthesis (DTA, GDUs), plant height (PH, cm), ear height (ER, cm), tassel size (TS, cm) and self-induction rate (SIR, %) evaluated in the ISU inducer set

Traits	Average	Minimum	Maximum	Genotypic variance	Residual variance	Heritability
HIR	6.27	0.61	20.74	21.32	2.41	0.90
DTA	1327	1222	1461	2123.47	493.10	0.82
PH	150	98.45	188.89	352.73	82.42	0.81
EH	58.20	32.63	87.72	145.26	46.95	0.75
TS	30.84	24.01	39.17	8.32	3.93	0.68
SIR	12.66	5.84	27.66	36.37	62.23	0.37

annotation provided by Illumina (Susan Chambers, unpublished data). The same parameters used for quality control in the Diversity Arrays Technology platform's genotypic dataset were applied to the Illumina MaizeSNP50 BeadChip dataset. After this process, the remaining 27,031 markers of the UOH inducer set were joined with the 16,639 markers of the ISU inducer set, resulting in a combined dataset of 43,670 markers. Both datasets were connected by 4 common inducers (RWS, PHI-3, MHI, RWK-76) and 122 common SNPs. The merging process was performed as described by Li, Willer, Sanna, and Abecasis (2009), through the software Beagle 5.0 (Browning et al., 2018). For independent validation, phenotypic information of the UOH inducer set was masked and predicted using the model created with the information of ISU inducer set, and vice-versa. Accuracy was estimated as the Pearson's correlation between predicted and observed values in each inducer set. Cross-validation was also performed within the UOH and the merged (ISU+UOH) inducer sets, following the same procedures described above. Phenotypic data was used to classify part of the inducers as having superior and inferior performances. Subsequently, a coincidence index (Cantelmo, Von-Pinho, & Balestre, 2017) was used to evaluate how accurately the prediction model placed the same inducers in these two categories. Genetic diversity within each set was measured using nucleotide diversity analysis (Pi) (Nei & Li, 1979) implemented in the software TASSEL 5.0 (Bradbury et al., 2007).

2.8 | Cross prediction for parental selection

In order to select the best inducers to derive a new breeding population with high HIR and good agronomic performance, 12,561 biparental crosses were predicted using the data of the ISU set. For each possible crossing combination, 10 populations with 200 individuals were simulated using the R package *PopVar* (Mohammadi et al., 2015). For each trait, the population parameters midparent GEBV and predicted genetic variance (V_G) were calculated based on the marker effects. The superior progeny average (μ_{SP}) was calculated as the mean of the 10% of individuals with better performance simulated for each possible crossing combination.

TABLE 2 Phenotypic correlation coefficients of haploid induction rate (HIR, %), days to anthesis (DTA, GDUs), plant height (PH, cm), ear height (ER, cm), tassel size (TS, cm) and self-induction rate (SIR, %) observed in the ISU inducer set

	HIR	DTA	PH	EH	TS	SIR
HIR		−0.20**[a]	−0.28**	−0.21**	−0.19**	0.44**
DTA			0.43**	0.51**	0.29**	0.009ns[b]
PH				0.80**	0.30**	−0.05ns
EH					0.28**	−0.06ns
TS						−0.18**

[a]**significant at $P < .01$

[b]ns, non-significant

3 | RESULTS

3.1 | Inducer performance, heritability, and trait correlations

In the ISU inducer set, HIRs ranged from 0.61 to 20.74% and had an average of 6.27%, which was twice lower than the average SIR (Table 1). Plant and ear height values ranged from 99 to 189 cm (avg. 150 cm) and from 33 to 88 cm (avg. 59 cm), respectively. The average of days to anthesis and tassel size were 1,327 GDUs and 31 cm, respectively. In general, estimated heritability on an entry-mean basis was high for all traits except SIR. The highest value observed was for HIR (0.90), while for other traits it consistently exceeded 0.60, and ranging from 0.68 for tassel size to 0.82 for DTA.

Phenotypic correlations between most traits were highly significant ($P < .01$) (Table 2). Non-significant correlations were only observed between SIR and DTA, and plant height and ear height. The HIR was slightly negatively correlated with DTA (−0.20), plant height (−0.28), ear height (−0.21) and tassel size (−0.19) and positively correlated with SIR (0.44).

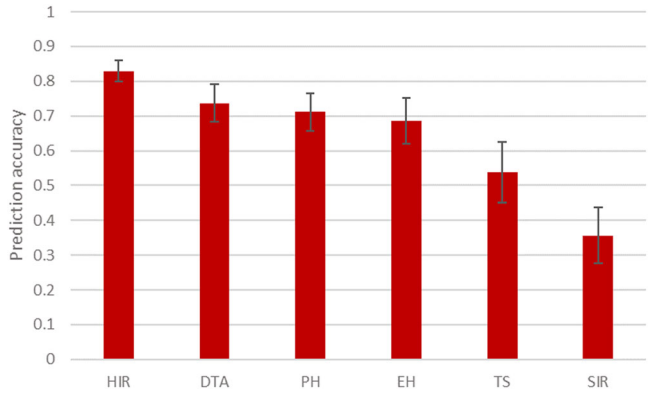

FIGURE 1 Average cross-validation prediction accuracies and standard deviation observed in the ISU inducer set, for the following traits: haploid induction rate (HIR, %); days to anthesis (DTA, GDUs); plant height (PH, cm); ear height (EH, cm); tassel size (TS, cm); self-induction rate (SIR, %)

3.2 | Genomic prediction

The results from the 60:40 cross-validation analysis showed that GBLUP is suitable to predict HIR. Estimated accuracies ranged from 0.75 to 0.88 (avg. 0.82), while for the other traits they ranged from 0.36 (SIR) to 0.74 (DTA) (Figure 1).

In order to evaluate the utility of prediction across independent sets, the ISU inducer set was used to predict the performance of the UOH set, and vice-versa. Population size played an important role in the accuracy of the prediction model. The accuracy across populations was higher when the ISU set, which was larger (159 vs. 53 individuals), was used as a training population and the UOH set as the prediction population (r = 0.54) (Table 3; Figure 2a). A principal component analysis clearly separated the inducers sets into two groups, confirming the presence of two different sub-populations (data not shown). Genetic diversity was higher in the UOH set (Pi = 0.34) than in the ISU set (Pi = 0.23).

Furthermore, training the model with ISU set was also effective to select the best (coincidence of 56%) and discard the worst inducers (coincidence of 62%) of the UOH set.

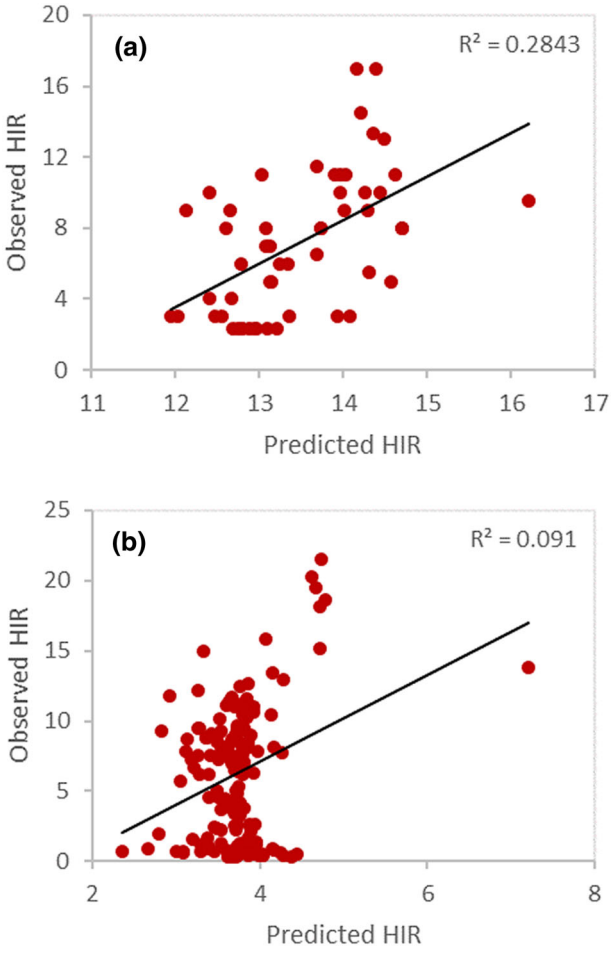

FIGURE 2 Linear regression of the 53 inducers genomic estimated breeding values with their observed values, using the ISU inducer set as training population (a). Linear regression of the 159 inducers genomic estimated breeding values with their observed values, using the UOH inducer set as training population (b)

When UOH was used as the training set, the accuracy was reduced (r = 0.30), as were the coincidence indexes to select the best (coincidence of 42%) and discard the worst inducers (coincidence of 29%) of the ISU set (Table 3; Figure 2b). The seemingly outlier in Figure 2b is RWS, inducer present in both sets and that, along with RWK, was used as a parent of most inducers of the ISU set. Cross-validation within the UOH set

TABLE 3 Prediction accuracies among different sets of inducers and coincidence index (%) of the 30% predicted inducers with higher and lower haploid induction rate (HIR, %). Sets are from Iowa State University (ISU) and University of Hohenheim (UOH)

			Coincidence index (%)	
Training set	**Validation set**	**Prediction accuracy**	**High HIR**	**Low HIR**
ISU	UOH	0.54	56.0	62.0
UOH	ISU	0.30	42.0	29.0
UOH[a]	-	0.70±0.08	-	-
ISU+UOH[b]	-	0.80±0.04	-	-

[a]Average 60:40 cross-validation prediction accuracy and standard deviation for the UOH inducer set

[b]Average 60:40 cross-validation prediction accuracy and standard deviation for the merged ISU+UOH inducer set

resulted in an average accuracy of 0.70, with values ranging from 0.45 to 0.89.

The increase in marker density and individuals gained by the integration of the ISU and UOH set did not influence the accuracy of the prediction model. The average prediction accuracy of a 60:40 cross-validation analysis in this merged population was comparable to the one observed for ISU set. For this new cross-validation scenario, accuracies ranged from 0.70 to 0.86 and had an average of 0.80 (Table 3).

3.3 | Cross prediction for HIR

Table 4 contains the population parameters, average and range of HIR of the 10 most promising crosses within the ISU set. The average HIR of the 12,561 superior predicted progenies (μ_{SP}) was 6.7%, ranged from 1.0 to 18.36% and was closely related to the corresponding midparent GEBV (avg. 6.2%, min. 0.46%, max 18.08%) (Table 4) and the average BLUP values for each possible crossing combination (Figure 3b). The predicted genetic variance (V_G) for HIR ranged from 0.01 to 0.14 and, in general, was of reduced magnitude in crosses involving inducers with high HIR (Figure 3a; Supplemental Table S2).

The average HIR (13.0%) of the 10 superior predicted progenies was almost the double of the overall average (6.7%). HUT_005 appeared in 6 out of the 10 most promising crosses, indicating that this inducer is a good source of beneficial alleles for HIR.

3.4 | Cross prediction for agronomic traits and self-induction

Satisfactory agronomic performance was observed in 12,561 predicted crosses (Table 4). Plant height and ear height of the superior predicted progenies ranged from 113 to 184 cm (avg. 153 cm) and from 38 to 83 cm (avg. 60 cm), respectively. For tassel size and days to anthesis, the superior progenies averages were 31 cm and 1,327 GDUs, and values ranged from 26 to 37 cm and from 1,240 to 1,435 GDUs, respectively. The average self-induction rate was 2-fold larger than the one observed for HIR, and values ranged from 9.7 to 15.0%. The superior predicted progenies of the 10 most promising crosses also exhibited reasonable agronomic performance. We found plant height and DTA values ranging from 131 to 144 cm and from 1,281 to 1,331 GDUs (~57–64 days), respectively. Tassel size and SIR values ranged from 29.7 to 32.1 cm and from 12.8 to 14.7%, respectively.

4 | DISCUSSION

4.1 | Heritability and relationship between HIR and agronomic traits

Herein, we estimated high heritabilities for HIR. Previously, with other germplasm materials, HIR was shown to be a highly heritable trait, e.g., Prigge et al. (2012b), reported moderate to high heritabilities (0.32–0.80) in different filial generations of crosses between inducer UH400 and three non-inducer and one inducer line. A high correlation (92.1%) was observed between the HIR data of the first and second planting dates. For each inducer, an elevated number of ears and seeds were evaluated, fact that give us confidence in the heritability estimates obtained. The agronomic traits evaluated in this study exhibited high heritability and showed weak correlations with HIR. Prigge et al. (2012a) also observed weak correlations between HIR and DTA, plant height, and tassel size in a set of 190 tropical inducer candidates derived from crosses between temperate inducers and tropical CIMMYT lines. According to the authors, the weak correlations confirm the possibility of combining the high HIR of temperate inducers with the improved adaptation to tropical conditions found on CIMMYT lines. Altogether, these results demonstrate that HIR can be improved without negatively impacting agronomic performance.

4.2 | Genomic prediction for inducer development

In the current study, cross-validation results showed that HIR and most of the agronomic traits could be accurately predicted, suggesting that GP can be successfully implemented in inducer breeding. Several other studies exploring cross-validation suggested that GP is an efficient tool for improving polygenic traits in maize. Crossa et al. (2011), using multiple GP models, obtained prediction accuracies for DTA ranging 0.46–0.79. Zhang et al. (2015), when evaluating 19 biparental populations under stressing conditions, noted that accuracies for plant height ranged from 0.38 to 0.76. In both studies, prediction accuracies of grain yield were consistently lower than 0.50, which could be explained by the fact that this is a trait under direct and indirect control of many genes. The cross-validation prediction accuracies for DTA (0.58–0.83) and plant height (0.60–0.82) observed in this study were in the same range as those reported in the aforementioned studies, demonstrating the quality of our phenotypic data.

Although there are no studies on the application of GP for the improvement of HIR, Prigge et al. (2012b) suggested this method for increasing the frequency of favorable

TABLE 4 Ten potential predicted crosses for haploid induction rate (HIR, %), days to anthesis (DTA, GDUs); plant height (PH, cm), ear height (EH, cm), tassel size (TS, cm) and self-induction rate (SIR, %)

Predicted crosses[a]		HIR			DTA			PH			EH			TS			SIR		
P1	P2	MP	V_G	μ_{SP}	MP	V_G	μ_{SP}	MP	V_G	μ_{SP}	MP	V_G	μ_{SP}	MP	V_G	μ_{SP}	MP	V_G	μ_{SP}
HUT_005	HUT_138	14.62	0.083	15.13	1304.07	13.31	1310.37	128.58	3.66	131.94	47.18	0.85	48.77	29.34	0.058	29.77	14.47	0.011	14.66
HUT_005	HUT_017	14.58	0.073	15.06	1324.84	14.00	1331.38	140.08	3.69	143.52	49.85	0.80	51.41	31.71	0.053	32.11	13.36	0.013	13.56
HUT_008	HUT_138	14.44	0.083	14.96	1303.62	13.20	1310.01	127.67	3.60	131.02	47.29	0.84	48.85	29.28	0.056	29.69	14.26	0.012	14.45
HUT_005	HUT_161	14.41	0.076	14.88	1283.92	10.98	1289.76	128.12	3.16	131.14	46.34	0.70	47.80	29.96	0.059	30.38	13.61	0.012	13.81
HUT_005	HUT_100	11.99	0.101	12.55	1274.94	12.47	1281.13	138.36	4.29	141.92	49.07	0.93	50.67	31.26	0.069	31.71	13.74	0.014	13.96
HUT_005	HUT_167	11.67	0.100	12.22	1302.22	13.50	1308.57	137.08	3.20	140.17	52.24	0.87	53.84	30.93	0.061	31.36	14.51	0.011	14.70
HUT_008	HUT_167	11.49	0.103	12.06	1301.77	13.45	1308.13	136.18	3.14	139.20	52.35	0.82	53.95	30.87	0.059	31.29	14.30	0.011	14.48
HUT_004	HUT_167	10.99	0.102	11.55	1308.76	13.19	1315.18	136.60	3.35	139.76	54.32	0.82	55.90	30.84	0.058	31.27	14.19	0.011	14.38
HUT_007	HUT_052	10.77	0.101	11.32	1285.66	11.81	1291.56	134.34	28.4	144.89	55.52	0.97	57.28	29.84	0.063	30.27	12.63	0.012	12.83
HUT_005	HUT_180	10.23	0.112	10.82	1277.85	11.56	1283.59	123.46	3.52	126.63	46.41	0.84	47.96	30.68	0.060	31.10	12.85	0.013	13.05
Average[b]		6.24	0.092	6.77	1327.388	13.0	1333.64	149.42	5.45	153.29	58.18	0.88	59.81	30.84	0.063	31.27	12.66	0.013	12.86
Min		0.46	0.01	1.00	1240.14	2.53	1243.18	110.24	0.84	113.30	36.85	0.20	38.36	30.84	0.01	25.97	9.47	0.003	9.69
Max		18.08	0.14	18.36	1435.54	19.92	1441.84	179.90	37.4	183.66	81.74	1.42	83.31	36.64	0.09	37.14	15.24	0.019	15.43

[a]P1, parent 1; P2, parent 2; MP, midparent GEBV; V_G, genetic variance; μ_{SP}, superior progenies estimated values using an selected proportion of 10%

[b]Average and estimates of ranges of population parameters of 12,561 predicted crosses between 159 haploid inducer from the ISU set

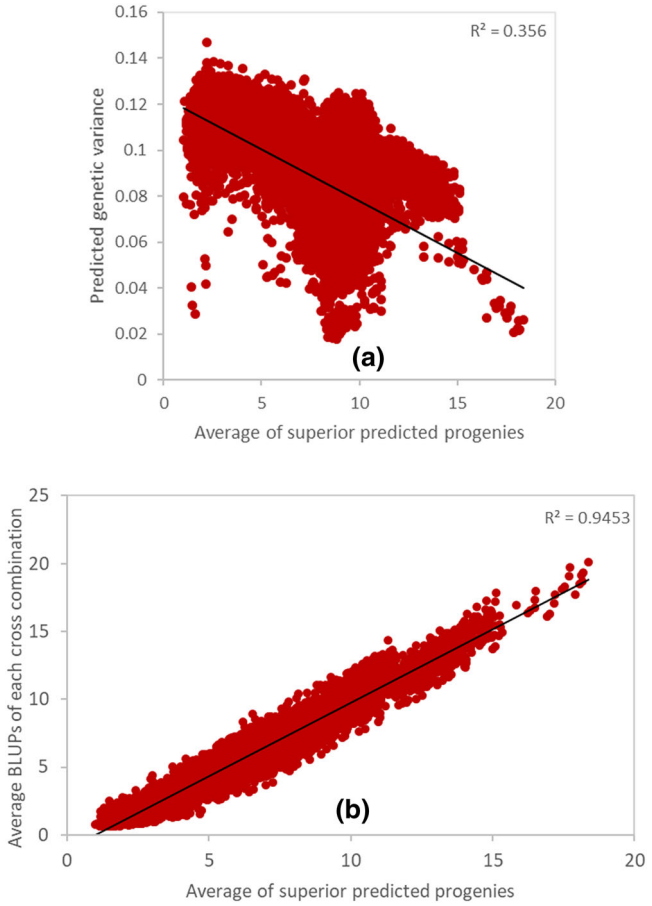

FIGURE 3 Relationship between the average of superior predicted progenies (μ_{SP}) and predicted genetic variance (V_G) (a) and average BLUPs of each possible crossing combination (b) for HIR (%) in each of the 12,561 predicted crosses

alleles in inducer development programs and to reduce the laborious and time-consuming task of phenotypic selection for HIR. Moreover, since HIR is governed by a major gene and affected by several other small-effect QTL, long-term genetic gain based on phenotypic selection would be extremely difficult to achieve once major genes become fixed.

Cross-validation is commonly used to evaluate the accuracy of GP. However, it can overestimate the model's prediction ability and hamper the breeder decision with to regards the real efficacy of the GP for selection purposes (Hofheinz, Borchardt, Weissleder, & Frisch, 2012). The evaluation of GP requires independent validation, using different experimental conditions and populations, and is a more accurate approach to draw useful inferences. In the present study, a satisfactory estimate of accuracy (0.54) for HIR was observed when independent validation was performed by predicting the GEBVs of inducers of the UOH set using a model trained with data of the ISU inducer set. This value was remarkably similar to the one reported by Albrecht et al. (2014), who observed an accuracy of 0.58 for grain yield in an independent validation study that used testcrosses of two different maize populations. Pace, Yu, and Lübberstedt (2015) were able to predict with an accuracy of 0.55 the total root length of 200 "extreme" (short or long) genotypes in an independent set of 2,431 maize inbred lines belonging to the Ames Panel (Romay et al., 2013). They concluded that GP is a useful approach for identifying the most informative genotypes. Cantelmo et al. (2017) also reported a satisfactory correlation (0.53) between the GEBVs of 402 maize hybrids predicted in the winter crop season and their mean values observed in the summer crop season for grain yield. They also reported that the coincidence index of genomic values of the winter and summer seasons was more efficient to discard the 20% lowest-yielding hybrids (coincidence of 89%) than to select the highest-yielding hybrids (coincidence of 43%). They concluded that genomic prediction might be useful for discarding the low-yielding hybrids during the winter, aiming the selection during the summer season. Considering a selection intensity of 30% (16 inducers), the coincidence index of the present study was efficient to both select the best (coincidence of 56%) and to discard the worst (coincidence of 62%) inducers. In other words, only ~6 out of 16 inferior or superior inducers would be erroneously selected or discarded in the UOH inducer set when the model was trained with the ISU inducer set. In a breeding program for haploid induction, it would be more reliable to use genomic prediction to discard families with lowest predicted HIR than to select families with highest predicted HIR, since that would give a chance for families that were predicted to have intermediate performance, but that in fact have good performance, to be field-tested. This would optimize resource allocation, since phenotyping efforts would be concentrated on the more promising genotypes.

Increasing the training population by adding the 53 inducers from the UOH set to the ISU set barely altered the accuracy of the GBLUP model (ISU = 0.82 vs ISU + UOH = 0.80). A similar observation was made in oat (*Avena sativa* L.), where the inclusion of historical data from older lines to the training population marginally changed the accuracy of prediction for five traits (Asoro, Newell, Beavis, Scott, & Jannink, 2011). In a GP for different maize root architectural traits, Pace et al. (2015) added the information of 200 selected lines from an independent prediction population to the original training population. They observed that increasing the number of individuals in the training population did not increase the prediction accuracy for some of the evaluated traits. Although no increase in prediction accuracy was observed by merging the two inducer sets, it seems beneficial to include the data of the UOH inducer set into the ISU inducer set, in order to increase genome coverage and the genetic diversity that the model can use for predictions. The former is composed by a set of diverse inducers genotyped with a different genotyping platform, and thus might contain useful information which

when combined with the latter set, could be used to efficiently select new breeding candidates.

4.3 | Cross prediction to improve the HIR and agronomic performance of inducers

A useful application of cross prediction to inducer breeding would be in the selection of parents to develop new breeding populations. The midparent value has been considered an important criterion to determine which crosses should be performed in plant breeding programs (Bernardo, 2014; Zhong & Jannink, 2007). In the current study, midparent GEBVs played the major role in the prediction of the superior progenies (μ_{SP}) for all traits investigated. For instance, high correlation was observed between these two criteria for all traits in the 12,561 possible crossing combinations. Our results are in agreement with those from Lado et al. (2017), who reported that crosses with the highest midparent values drove genetic gain for grain yield in wheat. They also observed that the relative importance of genetic variance was higher for quality traits. A similar observation was made by Yao et al. (2018), which noted that when selection intensity is large (e.g., 10%), genetic variance has less weight on the selection of superior progenies. Such observations indicate that the same parents would be chosen if usefulness or midparent GEBV were used as selection criteria, which in turn would equate the genetic gain achieved by both criteria. While in agreement, our results also indicate that employing midparent GEBV or μ_{SP} as the only criterion for parental selection can be risky. For instance, if HIR was the only trait of interest and if parental selection was solely based on one of these two criteria, the twenty parental combinations with highest midparent GEBV or μ_{SP} values would only include full-sibling, F_7 families derived from the cross between inducers RWS (Röber, Gordillo, & Geiger, 2005) and PHI-3 (Rotarenco et al., 2010) (Supplemental Table S2). The predicted genetic variance (V_G) for HIR among these twenty combinations varied from 0.021 to 0.048, whereas 0.092 and 0.147 were the average and highest V_G values of all 12,561 possible crossing combinations, respectively. It should be mentioned that these F_7 families were phenotypically selected for HIR (Trentin, Almeida, Frei, & Lubberstedt, unpublished data) and that their induction ability is considerably higher than the ones of other inducers included in the ISU panel (Supplemental Table S1). This might explain the fact that the cross combinations with higher midparent GEBV and μ_{SP} were composed of different combinations of these F_7 families. Unless for special reasons, it is very unlikely that a breeder would cross highly related individuals to develop new breeding populations, since that would limit the genetic gain that could be achieved. Therefore, high caution is recommended in the use of midparent GEBV or μ_{SP} as the solely selection criterion for parental selection. Maintaining genetic diversity is a key factor for the long-term success of breeding programs, and thus cross prediction and parental selection should be based in the optimal combination of midparent GEBV and V_G.

ACKNOWLEDGMENTS

Authors would like to thank the USDA National Institute of Food and Agriculture (Project Numbers: IOW04314, IOW01018, IOW05510; NIFA award 2018-51181-28419), the ISU Plant Sciences Institute, the ISU R. F. Baker Center for Plant Breeding, the ISU K. J. Frey Chair in Agronomy and the Brazilian Federal Agency for Support and Evaluation of Graduate Education (CAPES) for support.

AUTHOR CONTRIBUTIONS

H.U.T., U.K.F., and T.L. designed this project, supervised the research and developed the ISU inducer set. V.C.A and H.U.T. performed most of the HIR phenotypic selection, genomic data analysis and writing of the manuscript, and both are considered first authors of this manuscript. All co-authors were involved in editing the manuscript.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

ORCID

Vinícius Costa Almeida
https://orcid.org/0000-0002-8886-9813
Henrique Uliana Trentin
https://orcid.org/0000-0001-8406-982X
Ursula Karoline Frei
https://orcid.org/0000-0001-5969-7306
Thomas Lübberstedt
https://orcid.org/0000-0002-0526-0798

REFERENCES

Albrecht, T., Auinger, H.-J., Wimmer, V., Ogutu, J. O., Knaak, C., Ouzunova, M., … Schön, C.-C. (2014). Genome-based prediction of maize hybrid performance across genetic groups, testers, locations, and years. *Theoretical and Applied Genetics*, *127*, 1375–1386. https://doi.org/10.1007/s00122-014-2305-z

Asoro, F. G., Newell, M. A., Beavis, W. D., Scott, M. P., & Jannink, J.-L. (2011). Accuracy and training population design for genomic selection on quantitative traits in elite North American oats. *Plant Genome*, *4*, 132–144. https://doi.org/10.3835/plantgenome2011.02.0007

Bernardo, R. (2010). *Breeding for quantitative traits in plants* (2nd ed.). Woodbury, MN: Stemma Press.

Bernardo, R. (2014). Genomewide selection of parental inbreds: Classes of loci and virtual biparental populations. *Crop Science*, *55*, 2586–2595. https://doi.org/10.2135/cropsci2014.01.0088

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: Software for association mapping

of complex traits in diverse samples. *Bioinformatics*, *23*, 2633–2635. https://doi.org/10.1093/bioinformatics/btm308

Browning, B. L., Zhou, Y., & Browning, S. R. (2018). A one-penny imputed genome from next-generation reference panels. *American Journal of Human Genetics*, *103*, 338–348. https://doi.org/10.1016/j.ajhg.2018.07.015

Cantelmo, N. F., Von-Pinho, R. G., & Balestre, M. (2017). Genome-wide prediction for maize single-cross hybrids using the GBLUP model and validation in different crop seasons. *Molecular Breeding*, *37*, 51. https://doi.org/10.1007/s11032-017-0651-7

Chaikam, V., Nair, S. K., Babu, R., Martinez, L., Tejomurtula, J., & Boddupalli, P. M. (2015). Analysis of effectiveness of R1-nj anthocyanin marker for in vivo haploid identification in maize and molecular markers for predicting the inhibition of R1-nj expression. *Theoretical and Applied Genetics*, *128*, 159–171. https://doi.org/10.1007/s00122-014-2419-3

Chase, S. S. (1952). Production of homozygous diploids of maize from monoploids. *Agronomy Journal*, *44*, 263–267

Coe, E. H. (1959). A line of maize with high haploid frequency. *American Naturalist*, *93*, 381–382.

Covarrubias-Pazaran, G. (2016). Genome assisted prediction of quantitative traits using the R package *sommer*. *PLOS ONE*, *11*. https://doi.org/10.1371/journal.pone.0156744

Crossa, J., Pérez, P., de los Campos, G., Mahuku, G., Dreisigacker, S., & Magorokosho, C. (2011). Genomic selection and prediction in plant breeding. *Journal of Crop Improvement*, *25*, 239–261. https://doi.org/10.1080/15427528.2011.558767

De La Fuente, G. N., Frei, U. K., & Lübberstedt, T. (2013). Accelerating plant breeding. *Trends in Plant Science*, *18*, 12. https://doi.org/10.1016/j.tplants.2013.09.001

De La Fuente, G. N., Frei, U. K., Trampe, B., Nettleton, D., Zhang, W., & Lübberstedt, T. (2018). A diallel analysis of a maize donor population response to in vivo maternal haploid induction: I. Inducibility. *Crop Science*, *58*, 1830–1837. https://doi.org/10.2135/cropsci2017.05.0285

Deimling, S., Röber, F. K., & Geiger, H. H. (1997). Methodology and genetics of in vivo haploid induction in maize. *Vortr. Pflanzenzüchtg*, *38*, 203–224.

Dong, X., Xu, X., Miao, J., Li, L., Zhang, D., Mi, X., … Chen, S. (2013). Fine mapping of *qhir1* influencing in vivo haploid induction in maize. *Theoretical and Applied Genetics*, *126*, 1713–1720. https://doi.org/10.1007/s00122-013-2086-9

Eder, J., & Chalyk, S. (2002). In vivo haploid induction in maize. *Theoretical and Applied Genetics*, *104*, 703–708. https://doi.org/10.1007/s00122-001-0773-4

Endelman, J. B. (2011). Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome*, *4*, 250–255. https://doi.org/10.3835/plantgenome2011.08.0024

Ganal, M. W., Durstewitz, G., Polley, A., Bérard, A., Buckler, E. S., Charcosset, A., … Falque, M. (2011). A large maize (*Zea mays* L.) SNP genotyping array: Development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. *PLOS ONE*, *6*. https://doi.org/10.1371/journal.pone.0028334

Gilles, L. M., Khaled, A., Laffaire, J., Chaignon, S., Gendrot, G., Laplaige, J., … Widiez, T. (2017). Loss of pollen-specific phospholipase NOT LIKE DAD triggers gynogenesis in maize. *EMBO Journal*, *36*, 707–717. https://doi.org/10.15252/embj.201796603

Hallauer, A. R., Carena, M., & Miranda, J. B. F. (2010). *Quantitative genetics in maize breeding* (3rd ed.). Ames, IA: Iowa State University.

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

Hu, H., Schrag, T. A., Peis, R., Unterseer, S., Schipprack, W., Chen S., … Melchinger, A. E. (2016). The genetic basis of haploid induction in maize identified with a novel genome-wide association method. *Genetics*, *202*, 1267–1276. https://doi.org/10.1534/genetics.115.184234

Kebede, A. Z., Dhillon, B. S., Schipprack, W., Araus, J. L., Bänziger, M., Semagn, K., … Melchinger, A. E. (2011). Effect of source germplasm and season on the in vivo haploid induction rate in tropical maize. *Euphytica*, *180*, 219–226. https://doi.org/10.1007/s10681-011-0376-3

Kelliher, T., Starr, D., Richbourg, L., Chintamanani, S., Delzer, B., Nuccio, M. L., … Martin, B. (2017). MATRILINEAL, a sperm-specific phospholipase, triggers maize haploid induction. *Nature*, *542*, 105–109. https://doi.org/10.1038/nature20827

Kilian, A., Wenzl, P., Huttner, E., Carling, J., Xia, L., Blois, H., … Uszynski, G. (2012). Diversity arrays technology: A generic genome profiling technology on open platforms. *Methods in Molecular Biology*, *888*, 67–89. https://doi.org/10.1007/978-1-61779-870-2_5

Lado, B., Battenfield, S., Guzman, C., Quincke, M., Singh, R. P., Dreisigacker, S., … Gutierrez, L. (2017). Strategies for selecting crosses using genomic prediction in two wheat breeding programs. *Plant Genome*, *10*, 1–12. https://doi.org/10.3835/plantgenome2016.12.0128

Lashermes, P., & Beckert, M. (1988). Genetic control of maternal haploidy in maize (*Zea mays* L.) and selection of haploid inducing lines. *Theoretical and Applied Genetics*, *76*, 405–410.

Li, Y., Willer, C., Sanna, S., & Abecasis, G. (2009). Genotype imputation. *Annual Review of Genomics & Human Genetics*, *10*, 387–406. https://doi.org/10.1146/annurev.genom.9.081307.164242

Liu, C., Li, W., Zhong, Y., Dong, X., Hu, H., Tian, X., … Chen, S. (2015). Fine mapping of *qhir8* affecting in vivo haploid induction in maize. *Theoretical and Applied Genetics*, *128*, 2507–2515. https://doi.org/10.1007/s00122-015-2605-y

Liu, C., Li, X., Meng, D., Meng, D., Zhong, Y., Chen, C., … Chen, S. (2017). A 4-bp insertion at ZmPLA1 encoding a putative phospholipase a generates haploid induction in maize. *Molecular Plant*, *10*, 520–522. https://doi.org/10.1016/j.molp.2017.01.011

Melchinger, A. E., Schipprack, W., Wurschum, T., Chen, S., & Technow, F. (2013). Rapid and accurate identification of in vivo-induced haploid seeds based on oil content in maize. *Scientific Reports*, *3*, 2129. https://doi.org/10.1038/srep02129

Meuwissen, T. H. E., Hayes, B. J., & Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, *157*, 1819–1829.

Mohammadi, M., Tiede, T., & Smith, K. P. (2015). Popvar: A genome-wide procedure for predicting genetic variance and correlated response in biparental breeding populations. *Crop Science*, *55*, 2068–2077. https://doi.org/10.2135/cropsci2015.01.0030

Nei, M., & Li, W. H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences of the United States of America*, *76*, 5269–5273. https://doi.org/10.1073/pnas.76.10.5269

Pace, J., Yu, C. X., & Lübberstedt, T. (2015). Genomic prediction of seedling root length in maize (*Zea mays* L.). *Plant Journal*, *83*, 903–912. https://doi.org/10.1111/tpj.12937

Prigge, V., Schipprack, W., Mahuku, G., Atlin, G. N., & Melchinger, A. E. (2012a). Development of in vivo haploid inducers for tropical maize breeding programs. *Euphytica*, *185*, 481–490. https://doi.org/10.1007/s10681-012-0657-5

Prigge, V., Xu, X., Li, L., Babu, R., Chen, S., Atlin, G. N., & Melchinger, A. E. (2012b). New insights into the genetics of in vivo induction of maternal haploids, the backbone of doubled haploid technology in maize. *Genetics*, *190*, 781–793. https://doi.org/10.1534/genetics.111.133066

Randolph, L. F. (1940). (Note on haploid frequencies). *Maize Genetics Cooperation Newsletter*, *14*, 51.

Röber, F. K., Gordillo, G. A., & Geiger, H. H. (2005). In vivo haploid induction in maize - Performance of new inducers and significance of doubled haploid lines in hybrid breeding. *Maydica*, *50*, 275–283. https://doi.org/10.1534/genetics.111.133066

Romay, M. C., Millard, M. J., Glaubitz, J. C., Peiffer, J. A., Swarts, K. L., Casstevens, T. M., … Gardner, C. A. (2013). Comprehensive genotyping of the USA national maize inbred seed bank. *Genome Biology*, *14*, R55

Rotarenco, V. D., Georgeta, D. State, & Fuia, S. (2010). New inducers of maternal haploids in maize. *Maize Genetics Cooperative Newsletter*, *84*, 36–50

SAS Institute (2016). The SAS system for Windows, version 9.4. SAS Institute Inc, Cary

Schnell, F. W., & Utz, H. F. (1975). F1-leistung und elternwahl euphyder zu¨chtung von selbstbefruchtern. In Bericht über die Arbeitstagung der Vereinigung Österreichischer Pflanzenzüchter (pp. 243–248). BAL Gumpenstein, Gumpenstein, Austria.

Tiede, T., Kumar, L., Mohammadi, M., & Smith, K. P. (2015). Predicting genetic variance in biparental breeding populations is more accurate when explicitly modeling the segregation of informative genomewide markers. *Molecular Breeding*, *35*, 199. https://doi.org/10.1007/s11032-015-0390-6

Wang, H., Liu, J., Xu, X., Huang, Q., Chen, S., Yang, P., … Song, Y. (2016). Fully automated high-throughput NMR system for screening of haploid kernels of maize (corn) by measurement of oil content. *PLOS ONE*, *11*, e0159444. https://doi.org/10.1371/journal.pone.0159444

Xu, X., Li, L., Dong, X., Jin, W., Melchinger, A. E., & Chen, S. (2013). Gametophytic and zygotic selection leads to segregation distortion through in vivo induction of a maternal haploid in maize. *Journal of Experimental Botany*, *64*, 1083–1096. https://doi.org/10.1093/jxb/ers393

Yao, J., Zhao, D., Chen, X., Zhang, Y., & Wang, J. (2018). Use of genomic selection and breeding simulation in cross prediction for improvement of yield and quality in wheat (*Triticum aestivum* L.). *Crop Journal*, *6*, 353–365. https://doi.org/10.1016/j.cj.2018.05.003

Zhang, X., Pérez-Rodríguez, P., Semagn, K., Beyene, Y., Babu, R., López-Cruz, M. A., … Crossa, J. (2015). Genomic prediction in biparental tropical maize populations in water-stressed and well-watered environments using low-density and GBS SNPs. *Heredity*, *114*, 291–299. https://doi.org/10.1038/hdy.2014.99

Zhao, X., Xu, X., Xie, H., Chen, S., & Jin, W. (2013). Fertilization and uniparental chromosome elimination during crosses with maize haploid inducers. *Plant Physiology*, *163*, 721–731. https://doi.org/10.1104/pp.113.223982

Zhong, S., & Jannink, J. L. (2007). Using quantitative trait loci results to discriminate among crosses on the basis of their progeny mean and variance. *Genetics*, *177*, 567–576. https://doi.org/10.1534/genetics.107.075358

Zhong, Y., Liu, C., Qi, X., Jiao, Y., Wang, D., Wang, Y., … Chen, S. (2019). Mutation of ZmDMP enhances haploid induction in maize. *Nature Plants*, *5*, 575–580. https://doi.org/10.1038/s41477-019-0443-7

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Almeida VC, Trentin HU, Frei UK, Lübberstedt T. Genomic prediction of maternal haploid induction rate in maize. *Plant Genome*. 2020;e20014. https://doi.org/10.1002/tpg2.20014