

Prediction accuracy of genomic selection models for earliness in tomato

Aurelio Hernández-Bautista¹, Ricardo Lobato-Ortiz^{1*}, J. Jesús García-Zavala¹,
Serafín Cruz-Izquierdo¹, José Luis Chávez-Servia², Mario Rocandio-Rodríguez³,
Yolanda Del Rocío Moreno-Ramírez³, Enrique Hernandez-Leal⁴, Martha Hernández-Rodríguez¹,
and Delfino Reyes-Lopez⁵

¹Colegio de Postgraduados, Campus Montecillo, km 36.5 Carr. México-Texcoco, 56230, Texcoco, México.

*Corresponding author (rlobato@colpos.mx).

²Instituto Politécnico Nacional, Centro Interdisciplinario de Investigación para el Desarrollo Integral Regional (CIIDIR), Santa Cruz Xoxocotlán, 71230, México.

³Universidad Autónoma de Tamaulipas, Instituto de Ecología Aplicada, Ciudad Victoria 87019, México.

⁴Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias, Av. José Santos Valdez No. 1200 Pte. 27440, Col. Centro Matamoros, Coahuila, México.

⁵Benemérita Universidad Autónoma de Puebla, Facultad de Ingeniería Agrohidráulica, Teziutlán, Puebla 00000, México.

Received: 10 March 2020; Accepted: 2 May 2020; doi:10.4067/S0718-58392020000400505

ABSTRACT

Genomic selection is considered to be an important tool in plant breeding programs. However, its application in the earliness of tomato (*Solanum lycopersicum* L.) has not been studied. The objective of the present study was to evaluate the prediction performance of six statistical models for six quantitative characteristics related to earliness in tomato. The study used phenotypic and genotypic data belonging to an F₂ population consisting of 172 tomato plants. Simple sequence repeat (SSR) markers were obtained using genotypic information, and the genomic values were estimated by the following six different statistical models: Bayesian Lasso (BL), Bayesian ridge regression (BRR), BayesA, BayesB, BayesC π , and reproducing kernel Hilbert spaces (RKHS) regression. The correlation values ranged from 0.17 to 0.57. The highest association values were found in days to flowering of the third inflorescence and 1000-seed weight, which were greater than 0.5. In general, all the models performed in a similar manner because only slight differences were observed among the correlation values. Specifically, BL, BayesB, and RKHS exhibited the highest Pearson correlation values for most traits. According to the results, genomic selection could be a useful tool to support tomato breeding focused on earliness.

Key words: Genetic gain, genomic selection, *Solanum lycopersicum*, statistical models.

INTRODUCTION

Yield, fruit quality, and disease resistance are the main characteristics that breeders seek to increase in commercial tomato (*Solanum lycopersicum* L.) varieties (Scott et al., 2013). However, earliness has become an important characteristic that makes new tomato varieties very attractive in the seed market. Earliness in tomato indirectly increases yield and offers more supply opportunities by increasing the production window. Specifically, the tomato production cycle for each region in Mexico depends on the biotic challenges found in the production area. For example, the production period in Sinaloa State is shorter compared with other production areas due to the high number of viruses transmitted by whiteflies (SAGARPA, 2020). Farmers from Sinaloa must therefore keep their production up to 17 trusses from October to May. In contrast, there are other areas where a total of 35 trusses per plant can be harvested by farmers during the whole

year because there are no biotic problems. Therefore, high-yielding tomato varieties over a short period and with early flowering are needed for Mexican tomato production.

Currently, the selection of the best individuals in the tomato breeding is carried out by combining phenotypic and marker-assisted selection (Foolad, 2007; Sabatini et al., 2013). This two-step scheme first selects individuals resistant to viruses, fungi, and nematodes based on genotypic information, followed by a visual selection for yield, fruit quality, precocity, and amount of foliage.

The development of new technologies to support the improvement of tomato has gained considerable interest. These technologies aim to increase selection precision and reduce costs and time in the generation cycle (Cobb et al., 2019). The purpose of the selection methods for self-pollinating species is to generate fixed lines at the lowest cost and in the least amount of time. Methodologies such as single seed descent and double haploids allow developing fixed inbred lines in the shortest possible time (Courtois, 1993). The development of several rapid generational advancement (RGA) schemes has emerged as another option. According to Collard et al. (2017), using RGA methodologies has successfully reduced the cost of obtaining inbred rice lines by the single seed descent method under greenhouse conditions. Recently, genomic selection (GS) has become another efficient tool to select the best individuals. Recent studies have shown that GS provides a positive genetic gain for each selection cycle under different breeding schemes (Beyene et al., 2015; Gezan et al., 2017; Zhang et al., 2017). Genomic selection has become a very attractive tool for breeding programs given the availability of many molecular markers and high throughput genotyping technology (Crossa et al., 2017). One of the key points of GS involves estimating the genomic estimated breeding values (GEBV). Various parametric and non-parametric statistical models have therefore been developed to estimate GEBV. For tomato, Yamamoto et al. (2016) studied the performance of the following models: genomic best linear unbiased prediction (GBLUP), Bayesian Lasso (BL), weighted Bayesian shrinkage regression (wBSR), and reproducing kernel Hilbert spaces regression (RKHS). These authors concluded that the performance of each statistical model depended on the type of population and evaluated trait. Hernández-Bautista et al. (2016a) found that GBLUP and BL predicted GEBV better than multiple regression (fixed effects model) for traits related to fruit size and weight.

Although several studies have evaluated the efficiency of GS for fruit size and quality, there is limited information about the efficiency of GS for traits related to earliness in tomato. Therefore, the objective of the present study was to evaluate the performance of the BL, BRR, BayesA, BayesB, BayesC π , and RKHS models in the prediction of six quantitative characteristics related to earliness in tomato.

MATERIALS AND METHODS

Population, phenotyping, and genotyping

This study was performed on a tomato (*Solanum lycopersicum* L.) F₂ population consisting of 172 plants grown under greenhouse and hydroponic conditions in Texcoco (19°30' N, 98°53' W; 2250 m a.s.l.), State of Mexico. This site has an average annual temperature of 27 °C and approximately 147 mm yearly rainfall. The F₂ population was obtained from hybrid plants derived from the cross between lines LOR82 and 11904 (Hernández-Bautista et al., 2016b). Line LOR82 was obtained from a landrace from Puebla State, Mexico. This line exhibits an indeterminate growth habit, large-sized fruits (similar to Saladette type tomato), high firmness, late flowering, and high plant vigor. Line 11904 was obtained from an accession of *S. pimpinellifolium*. It has small round red fruits, a large number of flowers per truss, high resistance to diseases, weak plants, and early flowering. The F₂ seeds were sown in trays with 128 holes and grown under an average temperature of 25 °C. The F₂ plants were evaluated during the spring-summer cycle in 2012. Plants were harvested 80 and 100 d after transplanting.

The phenotypic evaluation considered six traits and was performed on single plants. The number of flowers per inflorescence was measured by the mean obtained from the third and fourth inflorescence. Days to emergence was calculated as the number of days elapsed from the sowing date to plant emergence. Days to first inflorescence were estimated as the number of days from sowing to anthesis of the first truss. Similarly, days to third inflorescence were measured as the number of days from sowing to anthesis of the third truss. Days to ripening were estimated as the number of days between sowing and the first ripe fruit. The 1000-seed weight (g) was measured as the mean weight of 100 seeds obtained from five red fruits from each F₂ plant. This score was then transformed to obtain the 1000-seed weight.

Fertilization was applied using the Steiner nutrient solution (Steiner, 1984). Concentrations were modified according to the phenological stages, 50% during the month and 100% for the rest of the cycle. Four irrigation events were applied each day throughout the cycle. Insects were controlled by periodic sprayings of imidacloprid (Confidor, Bayer) at 0.5 mL L⁻¹. To prevent late blight (*Phytophthora infestans* Mont. de Bary), two sprayings of metalaxil M + clorotalonil (Ridomil Gold, Syngenta) were applied monthly as 250 g 100 L⁻¹ water. Pruning of the leaves, branches, and training plants was performed weekly.

Genotypic data were obtained from 61 simple sequence repeat (SSR) markers. The SSR markers were located on 11 chromosomes, except chromosome 12. Some of these markers have been previously associated with earliness quantitative trait loci (QTLs) in tomato (Hernández-Bautista et al., 2016b). The linkage map spanned 815.71 cM of the tomato genome with mean marker spacing of 13.37 cM. A genotypic matrix was obtained and coded as -1 for a homozygous *S* genotype from *S. pimpinellifolium* L., 0 for a heterozygous genotype, and 1 for a homozygous genotype from *S. lycopersicum*.

Prediction models

Six statistical models with different approaches were used to estimate the marker effects. The predictive models were Bayesian Lasso (BL), Bayesian ridge regression (BRR), BayesA, BayesB, BayesC π , and reproducing kernel Hilbert spaces (RKHS) regression.

For all Bayesian models, the marker effects were estimated by the following model:

$$y = 1\mu + X\beta + e$$

where y is the vector of phenotypic values, μ is an intercept, X is the matrix of genotype indicators, β is a vector of a random marker, and e is the vector of residual effects. The conditional distribution of marker effects for each model differs in the assumptions of the prior distribution of marker effects, which determines the type of shrinkage or variable selection imposed on the estimates. The BL model implements a double exponential (DE) density. This density has higher mass at zero and thicker tails than the normal density and induces shrinkage estimates that depend on the size of the effect (Park and Casella, 2008). For the BRR model, the extent and type of shrinkage estimates are controlled by a Gaussian prior, which applies the same shrinkage to all the marker effects. The BayesA model uses a scaled-t density, which induces all the markers to have an effect but a different variance (Meuwissen et al., 2001). Two finite mixture prior densities were implemented for the BayesB and BayesC π models. A point mass at zero and a scaled-t slab were used for BayesB (Meuwissen et al., 2001), while a combination of a point mass at zero and a Gaussian slab was implemented in BayesC π (Habier et al., 2011). Both models assume that each marker has either a zero or non-zero effect with probabilities of π and $1 - \pi$, respectively. In addition, for markers with non-zero effects, their variance is different in BayesB and similar in BayesC π . In BayesB, we set the parameter as $\pi = 0.95$, and the remaining hyperparameters were set as described in Pérez and de los Campos (2014).

The RKHS regression uses the reproducing kernel (RK) that maps from pairs of points or markers $K(x_i, x'_i)$ in input space into the real line and which must satisfy $\sum_i \alpha_i \alpha'_i K(x_i, x'_i) \geq 0$ for any non-null sequence of coefficients α_i (Gianola and van Kaam, 2008). According to de los Campos et al. (2009), Bayesian RKHS regression can be represented as follows:

$$\begin{cases} y = 1\mu + u + \varepsilon \text{ with} \\ p(\mu, u, \varepsilon) \propto N(u | 0, K \alpha_i^2) N(\varepsilon | 0, I \alpha_i^2) \end{cases}$$

where $K(x_i, x'_i)$ is the kernel matrix whose entries are the evaluations of the RK at pairs of points in input space. This K matrix replaces the observed numerator relationship matrix (A_g) implemented in the standard animal model (Quaas and Pollak, 1980) with a Gaussian prior evaluated by the squared Euclidean distance between markers.

All predictive models were fitted with BGLR (Pérez and de los Campos, 2014) in the R statistical software (R Core Team, 2013). We set a total of 50 000 iterations for all the models, a thinning equal to 5, and the first 5000 cycles discarded as burn-in. A convergence test was performed according to the Gelman-Rubin convergence test for all parameters.

Cross-validation and prediction accuracy

The prediction accuracy of each model was evaluated by a 10-fold process. The population was divided into two sets; the training population consisted of 103 plants (60% individuals), whereas the validation population included 69 plants (40%). The individuals from the training population were randomly selected for each fold. Models were trained with the

training population. Based on the effects estimated in the training population, the models were run in the cross-validation population (CVP) to predict the genomic estimated breeding value (GEBV) of each plant from the validation population. This process was repeated 10 times. The predictive accuracy of each model was evaluated by the Pearson correlation between GEBV and the observed phenotypic value of individuals belonging to the CVP.

Estimation of selection differential

The selection differential is a parameter that evaluates if the predictive model effectively estimates the values of superior individuals. It was estimated at three different levels of selection intensity (5%, 10%, and 20%). The selection of superior individuals was based on the GEBV, but the differential was estimated based on the values of the higher GEBVs. This procedure was performed for each cross-validation. A mean was calculated based on the 10 cross-validation replicates. The estimate of the selection differential for each cross-validation was calculated by the following expression:

$$\text{Selection differential} = \frac{(\text{Mean of selected individuals in the validation population}) - (\text{Mean of the validation population})}{(\text{Mean of validation population})} \times 100$$

RESULTS

Performance of parents, F₁, and cross-validation population (CVP)

The distribution of individuals belonging to the CVP is shown in Figure 1. The CVP ranged from 6 to 13 flowers per inflorescence, with a mean of 11.17 flowers. This value was higher and lower than means for LOR82 (5.7 flowers) and 11904 (25.7 flowers), respectively. For days to emergence, the CVP mean was 7.61 d, and exhibited maximum and minimum values of 6 and 13 d, respectively. Similarly, parents and F₁ had values close to the mean of the CVP. For days to first inflorescence, the CVP values were lower than 11904 and higher than LOR82, demonstrating transgressive segregation. The lowest and highest values in the CVP were 55 and 79 d, respectively. As for days to third inflorescence, the population varied between 72 and 101 d. However, only some CVP values were lower than those of 11904. For days to ripening, there was a phenotypic range from 110 to 134 d, with a mean of 120.90 d. The F₁ and 30% of the individuals belonging to CVP were earlier for days to ripening than both parents. Finally, the CVP exhibited a mean of 2.42 g and minimum and maximum values of 1.18 and 3.55 g, respectively, for 1000-seed weight. This mean was close to the value observed in F₁.

Correlations between genomic estimated breeding values (GEBV) estimated in the training population

The correlations between GEBV estimated in the training population by different methods ranged from 0.9427 to 0.9993 (Table 1). Most of the pairs showed high values, indicating high similarity between the GEBV from the different models. The pairs with the lowest correlations (less than 0.95) were for days to flowering of the third inflorescence. These pairs were BayesB with BRR and BayesB with RKHS.

Performance of the statistical models in the cross-validation population (CVP)

The prediction accuracy values for the six statistical models are presented in Table 2. Considering all the traits, correlation values ranged from 0.17 to 0.57. The highest association values occurred for days to flowering of the third inflorescence and 1000-seed weight with a value greater than 0.5. In contrast, the lowest correlation values were detected for days to emergence and days to ripening.

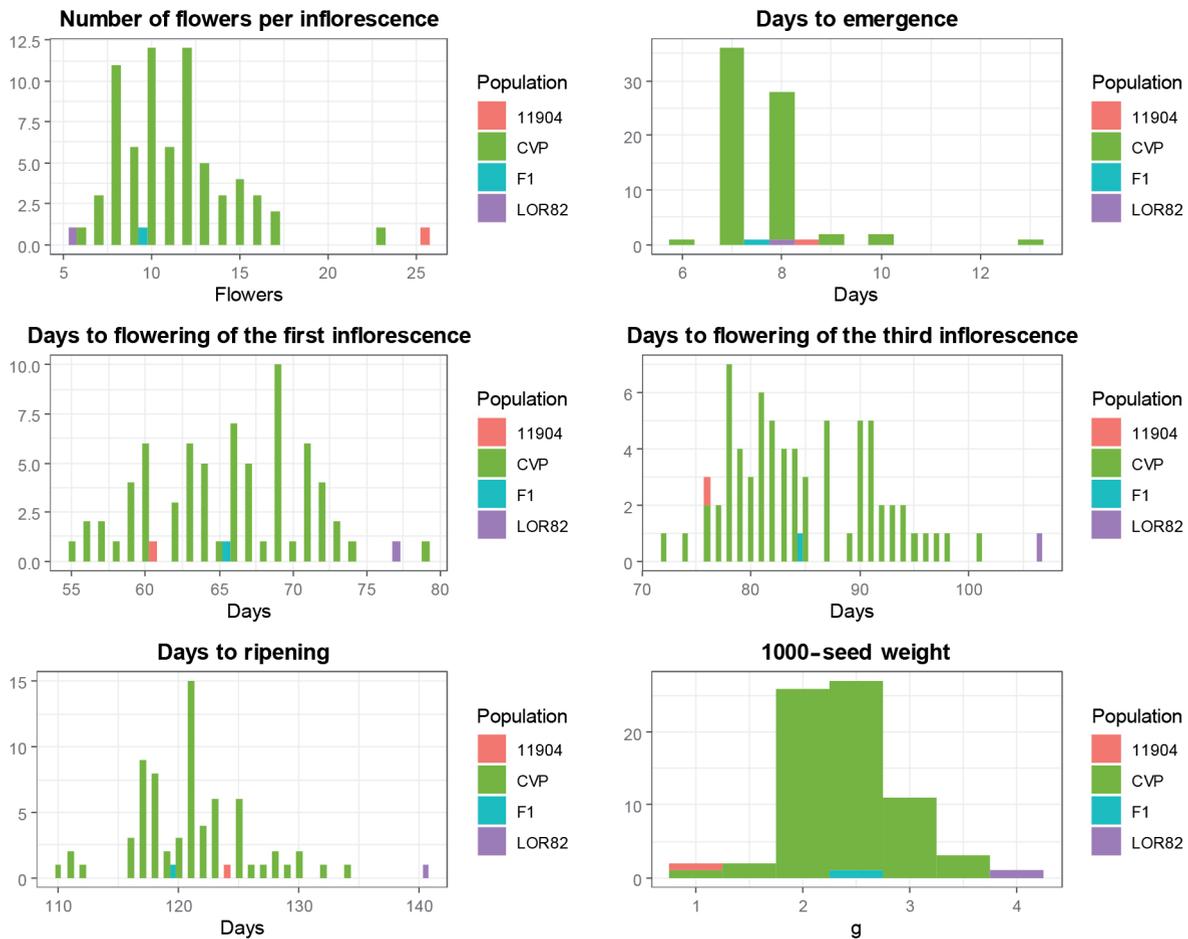
In general, all models performed in a similar manner since only slight differences were observed among the correlation values. However, the BL, BayesB, and RKHS models exhibited the highest Pearson correlation values for most of the traits.

Selection differential

Table 3 shows the selection differential obtained by each model for different levels of selection intensity. For the number of flowers per inflorescence, RKHS exhibited the highest values at the 5% and 10% selection intensity and BayesB at 20%. For days to emergence, the best statistical models were BL at 5% selection intensity, BRR at 10%, and BayesB at 20%. As for days to flowering of the first inflorescence, RKHS had the highest selection differential value at 5% and 20%, BL at 10%, and BayesA at 20%. For days to flowering of the third inflorescence, the selection differential values were

lower compared with days to flowering of the first inflorescence. At 5% selection intensity, all the models performed in a similar manner; however, RKHS was the only consistent model for the different levels of selection intensity. For days to ripening, BL, BayesA, BayesC π , and RKHS were the best models at 5%, while Bayes B outperformed all other models at 10% and 20%. Finally, for 1000-seed weight, high values at different levels of selection intensity were observed. These values ranged from 24.628 to 30.083 (BRR and RKHS) at 5% selection intensity, from 21.116 to 23.182 at 10%, and from 15.248 to 16.715 at 20%.

Figure 1. Mean of parents and F₁ and distribution of individuals belonging to the cross-validation population.



11904: Male line; CVP: F₂ cross-validation population; F₁: F₁ hybrid obtained crossing LOR82 and 11904; LOR82: female line.

Predictive model efficiency for superior individuals

Another way to assess predictive model efficiency is to analyze the predictive ability of the model, specifically in superior individuals. Therefore, we estimated the correlation between the values of superior individuals and their respective GEBV at different levels of selection intensity (Figure 2). The models were able to efficiently predict the values for the best individuals for days to emergence and 1000-seed weight, and values were greater than 0.5. However, for the remaining variables, the models showed poor predictive ability, especially for days to ripening and days to first and third inflorescence.

The predictive efficiency of the models changed under certain levels of selection intensity. For days to emergence, 1000-seed weight, and days to flowering of the first inflorescence, the prediction improved at 5% selection intensity. For days to ripening and number of flowers per inflorescence, the prediction was higher at 10% selection intensity. Regarding days to third inflorescence, the predictive ability of the models improved when selection intensity increased to 20%.

Table 1. Pearson correlations for genomic estimated breeding value (GEBV) estimated in the training population by different genomic selection (GS) models.

	Number of flowers per inflorescence						Days to flowering of the third inflorescence				
	BRR	BayesA	BayesB	BayesCπ	RKHS		BRR	BayesA	BayesB	BayesCπ	RKHS
BL	0.9940	0.9984	0.9963	0.9973	0.9937	BL	0.9755	0.9988	0.9920	0.9863	0.9762
BRR		0.9859	0.9859	0.9973	0.9959	BRR		0.9744	0.9427	0.9969	0.9972
BayesA			0.9951	0.9987	0.9935	BayesA			0.9924	0.9862	0.9750
BayesB				0.9862	0.9862	BayesB				0.9608	0.9443
BayesCπ					0.9942	BayesCπ					0.9932
	Days to emergence						Days to ripening				
	BRR	BayesA	BayesB	BayesCπ	RKHS		BRR	BayesA	BayesB	BayesCπ	RKHS
BL	0.9769	0.9838	0.9930	0.9753	0.9865	BL	0.9842	0.9962	0.9887	0.9872	0.9884
BRR		0.9911	0.9758	0.9901	0.9922	BRR		0.9920	0.9642	0.9957	0.9920
BayesA			0.9827	0.9850	0.9950	BayesA			0.9877	0.9955	0.9876
BayesB				0.9763	0.9818	BayesB				0.9789	0.9609
BayesCπ					0.9890	BayesCπ					0.9846
	Days to flowering of the first inflorescence						1000-seed weight				
	BRR	BayesA	BayesB	BayesCπ	RKHS		BRR	BayesA	BayesB	BayesCπ	RKHS
BL	0.9834	0.9921	0.9938	0.9830	0.9945	BL	0.9980	0.9979	0.9986	0.9976	0.9980
BRR		0.9962	0.9814	0.9973	0.9876	BRR		0.9993	0.9979	0.9993	0.9990
BayesA			0.9905	0.9965	0.9912	BayesA			0.9981	0.9992	0.9983
BayesB				0.9856	0.9824	BayesB				0.9979	0.9975
BayesCπ					0.9828	BayesCπ					0.9983

BL: Bayesian Lasso; BRR: Bayesian Ridge Regression; RKHS: Reproducing Kernel Hilbert Spaces Regression.

Table 2. Pearson correlations between observed values and genomic estimated breeding values (GEBV) of individuals from the cross-validation population.

Trait	BL	BRR	BayesA	BayesB	BayesCπ	RKHS
Number of flowers per inflorescence	0.34	0.32	0.34	0.33	0.32	0.35
Days to emergence	0.19	0.18	0.19	0.19	0.17	0.19
Days to flowering of the first inflorescence	0.40	0.39	0.39	0.39	0.38	0.41
Days to flowering of the third inflorescence	0.55	0.52	0.54	0.55	0.53	0.53
Days to ripening	0.26	0.26	0.25	0.25	0.25	0.26
1000-seed weight	0.57	0.56	0.57	0.57	0.56	0.56

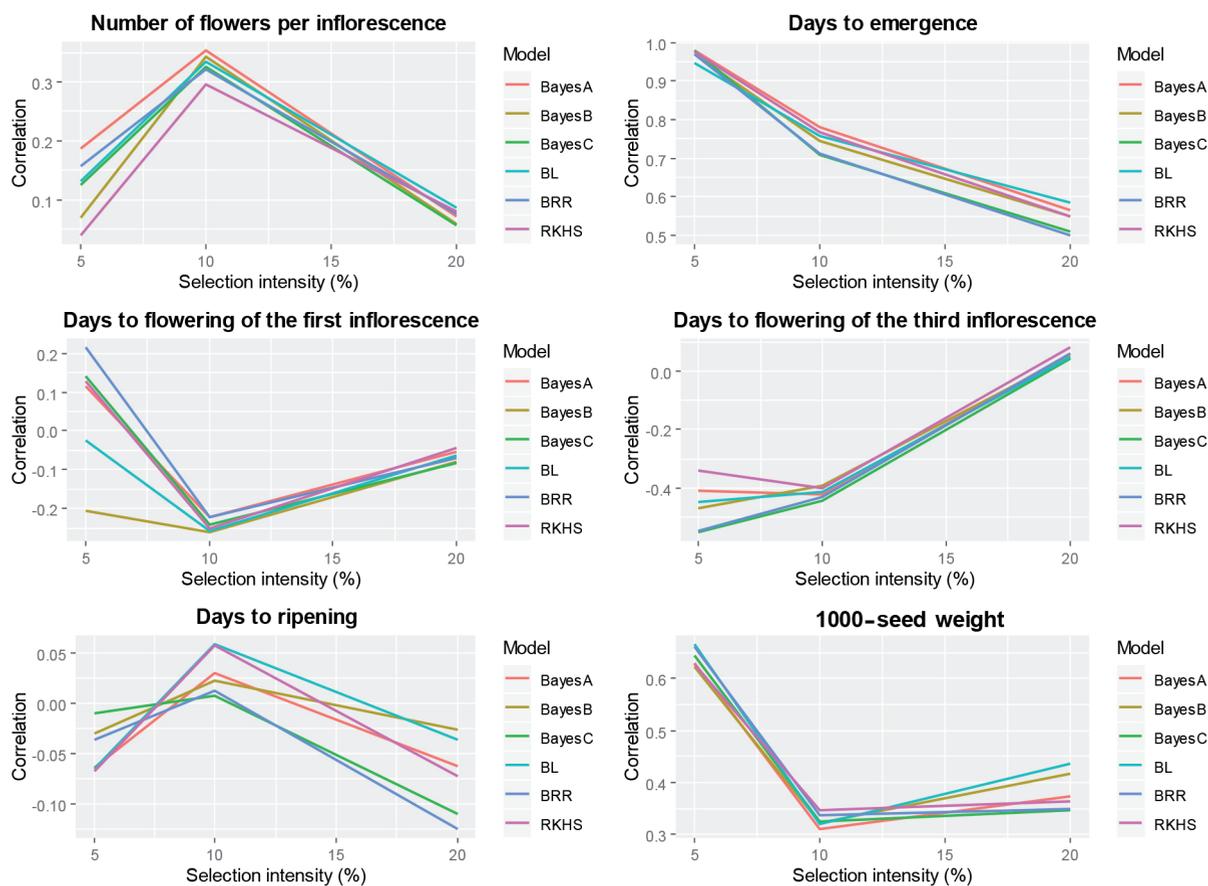
BL: Bayesian Lasso; BRR: Bayesian Ridge Regression; RKHS: Reproducing Kernel Hilbert Spaces Regression.

Table 3. Selection differential values for different levels of selection intensity (5%, 10%, and 20%) and genomic selection (GS) models.

Trait	Selection intensity (%)	BL	BRR	BayesA	BayesB	BayesC π	RKHS	Phenotypic selection
Number of flowers per inflorescence	5	0.27	0.27	0.27	-3.31	0.27	5.64	59.36
	10	2.95	-0.63	-0.63	-0.63	2.95	4.74	47.72
	20	8.77	8.77	8.77	9.22	8.77	8.77	32.50
Days to emergence	5	-8.02	-2.76	-2.76	-5.39	-2.76	-0.13	-10.64
	10	-1.45	-2.76	-1.45	-1.45	-1.45	-1.45	-9.33
	20	-0.79	-0.13	-1.45	-2.76	-0.13	-2.10	-8.67
Days to flowering of the first inflorescence	5	-3.57	-3.57	-3.57	-3.57	-3.57	-4.49	-14.25
	10	-6.77	-4.33	-6.32	-6.16	-5.71	-5.71	-12.27
	20	-4.18	-4.49	-5.25	-4.64	-4.26	-5.25	-9.67
Days to flowering of the third inflorescence	5	-2.51	-2.51	-2.51	-2.51	-2.51	-2.51	-11.48
	10	-1.45	-2.63	-0.98	-1.45	-2.63	-2.63	-9.83
	20	-1.86	-2.28	-2.45	-2.81	-2.51	-4.05	-8.30
Days to ripening	5	-2.73	-2.07	-2.73	-1.24	-2.73	-2.73	-7.36
	10	-1.65	-1.65	-1.65	-2.07	-0.74	-1.65	-5.46
	20	-1.16	-1.32	-1.32	-1.41	-1.32	-1.28	-4.18
1000-seed weight	5	27.27	30.08	27.27	27.27	24.63	30.08	37.36
	10	21.12	23.18	23.18	21.12	21.12	21.12	31.36
	20	16.71	15.25	15.25	15.25	15.33	16.57	23.43

BL: Bayesian Lasso; BRR: Bayesian Ridge Regression; RKHS: Reproducing Kernel Hilbert Spaces Regression.

Figure 2. Relationship between the correlation of observed values and genomic estimated breeding values (GEBV) of superior individuals and selection intensity.



BL: Bayesian Lasso; BRR: Bayesian Ridge Regression; RKHS: Reproducing Kernel Hilbert Spaces Regression.

DISCUSSION

In the present study, we evaluated the prediction performance of six statistical models on tomato earliness related traits. Prediction performance was evaluated according to the relationship between GEBV and the observed phenotypic value belonging to the individual of the CVP. The prediction performance of the statistical models was similar. The correlations between GEBV and the phenotypic value ranged from 0.17 to 0.57, and the highest correlations occurred in BL, BayesB, and RKHS. Previous studies have reported the good performance of BL and BayesB to predict traits controlled by large-QTL effects (Thavamanikumar et al., 2015; Wang et al., 2016; Gezan et al., 2017). According to Thavamanikumar et al. (2015), the good performance of BL and BayesB is due to their ability to identify markers with large effects, allowing better prediction accuracy of GEBV. The good performance of RKHS is attributed to its capture of suitably unknown forms of interaction and non-additive effects in the prediction, which increases its predictive power (Gianola and van Kaam, 2008). In strawberry, RKHS performed slightly better than parametric models for mean weight, marketable yield, and soluble solid content. However, in other maize and wheat studies, there was a null difference between RKHS and GBLUP (Crossa et al., 2014; Juliana et al., 2017). These results suggest that the performance of RKHS largely depends on the underlying genetic architecture of the trait. Howard et al. (2014) point out that RKHS performs well when the traits are largely influenced by epistasis.

Training population size, relatedness between individuals, trait genetic architecture, and marker density are factors that affect the prediction accuracy of the models used for genomic selection (GS) (Meuwissen et al., 2001). In the present study, using a reduced number of markers was enough to obtain accuracy values greater than 0.5, demonstrating that the models with few markers had high predictive power to estimate GEBV for days to flowering of the third inflorescence and 1000-seed weight. The prediction values for days to flowering of the first inflorescence ranged from 0.38 to 0.41, whereas the range increased from 0.53 to 0.55 for days to flowering of the third inflorescence. In a previous study, we detected a major QTL at flowering called *dff1.1*, which increased its effect over time when measured from the first to third truss (Hernández-Bautista et al., 2016b). Therefore, the increasing range found in the present study could be explained by the increased effect of *dff1.1*, showing that this QTL increases the prediction accuracy of the statistical model.

In GS, it is commonly observed that the predictive models yield better for traits with high heritability than traits with low heritability. Duangjit et al. (2016) found that GBLUP improved its accuracy when the heritability of tomato quality traits also increased. This same pattern was observed by Hernández-Bautista et al. (2016a), who reported that the prediction accuracy of GBLUP and LASSO increased when it was evaluated for traits with heritability greater than 0.6. Previous classical and molecular studies classified days to ripening and emergence as traits with an intermediate or low heritability, and that they are controlled by small QTL effects (Grandillo and Tanksley, 1996; Haggard et al., 2015; Hernández-Bautista et al., 2016b). Therefore, the poor predictive ability found in the models for both traits occurred because both traits are controlled by many minor QTLs and the limited number of markers used in the present study.

Expected genetic gain is a function of selection differential and heritability of the trait to be improved (Molina, 1992). When one trait exhibited high heritability, the genetic gain largely depended on the selection differential. We transformed the selection differential in terms of percentage with respect to the F_2 mean to easily detect the genetic gain. For 1000-seed weight, GS obtained values similar to those of phenotypic selection at different levels of selection intensity, suggesting that the RKHS, BRR, and BL models effectively predicted the values of the superior individuals. Similarly, for days to emergence, the differential selection value obtained by BL at 5% selection intensity was very close to the value estimated by the phenotypic selection. This result suggests that BL performed poorly to predict the GEBV of individuals with low values, but it was able to efficiently predict the high values. For the remaining traits, the estimated selection differential was at least 40% lower than the phenotypic selection. This situation was caused in some cases by a moderate prediction of the best individuals, which allowed a drastic reduction of the selection differential (Figure 2). However, this moderate prediction could be improved by increasing the density of molecular markers evaluated in the training population (Zhang et al., 2019) and by incorporating markers linked to loci controlling the trait (Rutkoski et al., 2014). It is also known that GS reduces the time needed to develop a new variety. Therefore, low or intermediate values of selection differential obtained in the present study could be improved if several cycles were performed yearly or some steps of the classical breeding scheme were omitted.

During the prediction of the best individuals, model efficiency was better for certain levels of selection intensity and traits; this shows the influence of selection intensity and traits on the correlation of observed values and GEBV of superior individuals. Bhering et al. (2015), in a simulation study, found that the correlation between genetic and genomic

values was not constant for the different selection intensity values. These results suggest that the prediction of superior individuals is influenced by the genetic architecture of traits and the statistical model being used.

In tomato, the pedigree selection scheme is the most widely used by breeders to produce inbred lines (Hernández-Leal et al., 2019). However, this selection scheme is time-consuming and expensive. For example, if two selection cycles are carried out yearly, the generation of F₆ lines would take 3 yr. According to Cobb et al. (2019), the generation cycle is the easiest, cheapest, and most useful parameter to increase genetic gain. Its impact on genetic gain is much greater than increasing heritability by more than 0.6 or reducing selection pressure less than 5%. Based on our data, GS could help to increase genetic gain by rapidly advancing generations through phenotypic evaluation of the first cluster and breeding value predicted by GS. Tomato breeders usually evaluate many trusses to detect the consistency of fruit size, setting, and yield over time. One tomato generation cycle could therefore take at least 7 mo, but GS can accelerate the generation cycle and reduce the costs of maintaining large field trials over a long period. To date, studies in wheat (Watson et al., 2018), sorghum (Rizal et al., 2014), and rice (Tanaka et al., 2016) have reported that manipulating environmental variables allows producing seed in less time and increases the number of generations per year. Similarly, manipulating some environmental variables, such as water stress or temperature, can induce early flowering in tomato (Ewas et al., 2017). In this context, a combination of technologies such as a greenhouse with a controlled environment, marker-assisted selection for resistance, selection by the single seed descent method, and GS could be interesting and viable options for modern tomato breeding.

CONCLUSIONS

The Bayesian models and reproducing kernel Hilbert spaces regression (RKHS) model performed in a similar manner to predict tomato earliness traits. Specifically, Bayesian Lasso, BayesB, and RKHS exhibited the highest Pearson correlation values for most of the traits. Based on these results, genomic selection could be a useful tool to support some steps of tomato breeding focused on earliness.

ACKNOWLEDGEMENTS

This study was supported by the Consejo Nacional de Ciencia y Tecnología-Mexico (CONACyT) and the Colegio de Postgraduados-LPI-6.

REFERENCES

- Beyene, Y., Semagn, K., Mugo, S., Tarekegne, A., Babu, R., Meisel, B., et al. 2015. Genetic gains in grain yield through genomic selection in eight bi-parental maize populations under drought stress. *Crop Science* 55:154-163.
- Bhering, L.L., Junqueira, V.S., Peixoto, L.A., Cruz, C.D., and Laviola, B.G. 2015. Comparison of methods used to identify superior individuals in genomic selection in plant breeding. *Genetics and Molecular Research* 14:10888-10896.
- Cobb, J.N., Juma, R.U., Biswas, P.S., Arbelaez, J.D., Rutkoski, J., Atlin, G., et al. 2019. Enhancing the rate of genetic gain in public-sector plant breeding programs: lessons from the breeder's equation. *Theoretical and Applied Genetics* 132:627-645.
- Collard, B.C., Beredo, J.C., Lenaerts, B., Mendoza, R., Santelices, R., Lopena, V., et al. 2017. Revisiting rice breeding methods—evaluating the use of rapid generation advance (RGA) for routine rice breeding. *Plant Production Science* 20:337-352.
- Courtois, B. 1993. Comparison of single seed descent and anther culture-derived lines of three single crosses of rice. *Theoretical and Applied Genetics* 85:625-631.
- Crossa, J., Pérez, P., Hickey, J., Burgueño, J., Ornella, L., Cerón-Rojas, J., et al. 2014. Genomic prediction in CIMMYT maize and wheat breeding programs. *Heredity* 112:48-60.
- Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., de los Campos, G., et al. 2017. Genomic selection in plant breeding: methods, models, and perspectives. *Trends in Plant Science* 22:961-975.
- de los Campos, G., Naya, H., Gianola, D., Crossa, J., Legarra, A., Manfredi, E., et al. 2009. Predicting quantitative traits with regression models for dense molecular markers and pedigree. *Genetics* 182:375-385.
- Duangjit, J., Causse, M., and Sauvage, C. 2016. Efficiency of genomic selection for tomato fruit quality. *Molecular Breeding* 36:29.
- Ewas, M., Khames, E., Ziaf, K., Shahzad, R., Nishawy, E., Ali, F., et al. 2017. The Tomato DOF Daily Fluctuations 1, TDDF1 acts as flowering accelerator and protector against various stresses. *Scientific Reports* 7:1-16.
- Foolad, M.R. 2007. Genome mapping and molecular breeding of tomato. *International Journal of Plant Genomics* 2007:1-52.

- Gezan, S.A., Osorio, L.F., Verma, S., and Whitaker, V.M. 2017. An experimental validation of genomic selection in octoploid strawberry. *Horticulture Research* 4:16070.
- Gianola, D., and van Kaam, J.B. 2008. Reproducing kernel Hilbert spaces regression methods for genomic assisted prediction of quantitative traits. *Genetics* 178:2289-2303.
- Grandillo, S., and Tanksley, S.D. 1996. QTL analysis of horticultural traits differentiating the cultivated tomato from the closely related species *Lycopersicon pimpinellifolium*. *Theoretical and Applied Genetics* 92:935-951.
- Habier, D., Fernando, R.L., Kizilkaya, K., and Garrick, D. 2011. Extension of the Bayesian alphabet for genomic selection. *BMC Bioinformatics* 12:186.
- Haggard, J.E., Johnson, E.B., and Clair, D.A.S. 2015. Multiple QTL for horticultural traits and quantitative resistance to *Phytophthora infestans* linked on *Solanum habrochaites* chromosome 11. *G3: Genes, Genomics, Genetics* 5:219-233.
- Hernández-Bautista, A., Lobato-Ortiz, R., García-Zavala, J.J., López-Fortoso, F., Cruz-Izquierdo, S., Chávez-Servia, J.L., et al. 2016b. Quantitative trait locus mapping associated with earliness and fruit weight in tomato. *Scientia Agricola* 73:478-486.
- Hernández-Bautista, A., Lobato-Ortiz, R., García-Zavala, J.J., Parra-Gómez, M.A., Cadeza-Espinosa, M., Canela-Doñan, D., et al. 2016a. Implications of genomic selection for obtaining F_{2.3} families of tomato. *Scientia Horticulturae* 207:7-13.
- Hernández-Leal, E., Lobato-Ortiz, R., García-Zavala, J.J., Hernández-Bautista, A., Reyes-López, D., and Bonilla-Barrientos, O. 2019. Stability and breeding potential of tomato hybrids. *Chilean Journal of Agricultural Research* 79:181-189.
- Howard, R., Carriquiry, A.L., and Beavis, W.D. 2014. Parametric and nonparametric statistical methods for genomic selection of traits with additive and epistatic genetic architectures. *G3: Genes, Genomics, Genetics* 4:1027-1046.
- Juliana, P., Singh, R.P., Singh, P.K., Crossa, J., Huerta-Espino, J., Lan, C., et al. 2017. Genomic and pedigree-based prediction for leaf, stem, and stripe rust resistance in wheat. *Theoretical Applied Genetics* 130:1415-1430.
- Meuwissen, T.H., Hayes, B.J., and Goddard, M.E. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819-1829.
- Molina, J.D. 1992. Introducción a la genética de poblaciones y cuantitativa: Algunas implicaciones en genotecnia. 2ª ed. AGT Editor, Ciudad de México, México.
- Park, T., and Casella, G. 2008. The Bayesian LASSO. *Journal of the American Statistical Association* 103:681-686.
- Pérez, P., and de los Campos, G. 2014. Genome-wide regression and prediction with the BGLR statistical package. *Genetics* 198:483-495.
- Quaas, R.L., and Pollak, E.J. 1980. Mixed model methodology for farm and ranch beef cattle testing programs. *Journal of Animal Science* 51:1277-1287.
- R Core Team. 2013. R: A language and environment for statistical computing. Version 3.5.2. R Foundation for Statistical Computing, Vienna, Austria.
- Rizal, G., Karki, S.N., and Alcasid, M. 2014. Shortening the breeding cycle of sorghum, a model crop for research. *Crop Science* 54:520-529.
- Rutkoski, J.E., Poland, J.A., Singh, R.P., Huerta-Espino, J., Bhavani, S., Barbier, H., et al. 2014. Genomic selection for quantitative adult plant stem rust resistance in wheat. *The Plant Genome* 7:3.
- Sabatini, E., Beretta, M., Sala, T., Acciarri, N., Milc, J., and Pecchioni, N. 2013. Molecular breeding. p. 228-303. In Liedl, B.E., Labate, J.A., Stommel, J.R., Slade, A., and Kole, C. (eds.) *Genetics, genomics and breeding of tomato*. CRC Press, Boca Raton, Florida, USA.
- SAGARPA. 2020. Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación (SAGARPA), Ciudad de México, México. Available at <http://www.sagarpa.gob.mx> (accessed May 2020).
- Scott, J.W., Myers, J.R., Boches, P.S., Nichols, C.G., and Angell, F.F. 2013. Classical genetics and traditional breeding. p. 37-73. In Liedl, B.E., Labate, J.A., Stommel, J.R., Slade, A., and Kole, C. (eds.) *Genetics, genomics and breeding of tomato*. CRC Press, Boca Raton, Florida, USA.
- Steiner, A.A. 1984. The universal nutrient solution. *Proceedings 6th International Congress on Soilless Culture*, Lunteren. 29 April-5 May. International Society for Soilless Culture (ISOSC), Wageningen, The Netherlands.
- Tanaka, J., Hayashi, T., and Iwata, H. 2016. A practical, rapid generation advancement system for rice breeding using simplified biotron breeding system. *Breeding Science* 66:542-551.
- Thavamanikumar, S., Dolferus, R., and Thumma, B.R. 2015. Comparison of genomic selection models to predict flowering time and spike grain number in two hexaploid wheat doubled haploid populations. *G3: Genes, Genomics, Genetics* 5:1991-1998.
- Wang, X., Li, L., Yang, Z., Zheng, X., Yu, S., Xu, C., et al. 2016. Predicting rice hybrid breeding performance using univariate and multivariate GBLUP models based on North Carolina mating design II. *Heredity* 118:302-310.
- Watson, A., Ghosh, S., Williams, M.J., Cuddy, W.S., Simmonds, J., Rey, M.D., et al. 2018. Speed breeding is a powerful tool to accelerate crop research and breeding. *Nature Plants* 4:23-29.
- Yamamoto, E., Matsunaga, H., Onogi, A., Ohyama, A., Miyatake, K., Yamaguchi, H., et al. 2016. Efficiency of genomic selection for breeding population design and phenotype prediction in tomato. *Heredity* 118:1-8.
- Zhang, X., Pérez-Rodríguez, P., Burgueño, J., Olsen, M., Buckler, E., Atlin, G., et al. 2017. Rapid cycling genomic selection in a multiparental tropical maize population. *G3: Genes, Genomics, Genetics* 7:2315-2326.
- Zhang, H., Yin, L., Wang, M., Yuan, X., and Liu, X. 2019. Factors affecting the accuracy of genomic selection for agricultural economic traits in maize, cattle, and pig populations. *Frontiers in Genetics* 10:189.