Peer-reviewed version available at *Int. J. Mol. Sci.* **2019**, *20*, 359; doi:10.3390/ijms20020359

Article

Evaluation of Genomic Prediction for Pasmo Resistance in Flax

Liqiang He [1,2], Jin Xiao [2], Khalid Y. Rashid [3], Gaofeng Jia [4], Pingchuan Li [3], Zhen Yao [3], Xiue Wang [2], Sylvie Cloutier [1,*], and Frank M. You [1,2,*]

[1] Ottawa Research and Development Centre, Agriculture and Agri-Food Canada, Ottawa, ON K1A 0C6, Canada; liqiang.he@canada.ca (L.H.); sylvie.cloutier@canada.ca (S.C.); frank.you@canada.ca (F.M.Y.);
[2] State Key Laboratory of Crop Genetics and Germplasm Enhancement, College of Agriculture, Nanjing Agricultural University/JCIC-MCP, Nanjing Jiangsu210095, China; xiaojin@njau.edu.cn (J.X.); xiuew@njau.edu.cn (X.W.)
[3] Morden Research and Development Centre, Agriculture and Agri-Food Canada, Morden, MB R6M 1Y5, Canada; khalid.rashid@canada.ca (K.Y.R.); zhen.yao@canada.ca (Z.Y.); lipingchuan@gmail.com (P.L.)
[4] Crop Development Centre, University of Saskatchewan, Saskatoon, SK S7N 5A8, Canada; gaofeng.jia@usask.ca (G.J.);

* Correspondence: frank.you@canada.ca (F.M.Y.); sylvie.cloutier@canada.ca (S.C.); Tel.: +1-613-759-1539 (F.M.Y.); +1-613-759-1744 (S.C.)

Abstract: Pasmo (*Septoria linicola*) is a fungal disease causing major losses in seed yield and quality, and stem fibre quality in flax. Pasmo resistance (PR) is quantitative and has low heritability. To improve PR breeding efficiency, the accuracy of genomic prediction (GP) was evaluated using a diverse worldwide core collection of 370 accessions. Four marker sets, including three defined by 500, 134, and 67 previously identified quantitative trait loci (QTL) and one of 52,347 PR-correlated genome-wide single nucleotide polymorphisms, were used to build ridge regression best linear unbiased prediction (RR-BLUP) models using pasmo severity (PS) data collected from field experiments performed during five consecutive years. With five-fold random cross-validation, GP accuracy as high as 0.92 was obtained from the models using the 500 QTL when the average PS was used as the training dataset. GP accuracy increased with training population size, reaching values >0.9 with training population size greater than 185. Linear regression of the observed PS with the number of positive-effect QTL in accessions provided an alternative GP approach with an accuracy of 0.86. The results demonstrate the GP models based on marker information from all identified QTL and the 5-year PS average is highly effective for PR prediction.

Keywords: genomic selection; genomic prediction; genotyping by sequencing; pasmo resistance; pasmo severity; quantitative trait loci; single nucleotide polymorphism; *Septoria linicola*; flax

1. Introduction

Flax (*Linum usitatissimum L.*) is an important food and fibre crop cultivated and grown in cooler regions of the world, such as Canada [1]. Pasmo, elicited by the fungus *Septoria linicola*, is one of the most widespread diseases of flax, causing reductions in seed and oil yield, as well as fibre quality and durability [2]. Developing resistant cultivars is the most viable and effective option to control this disease that has become widespread in all flax production areas of North America and other parts of the world. Resistance to pasmo has a low heritability [3] and is quantitatively inherited [4]. Large variations in pasmo disease severity were observed in the flax core collection, which can be capitalized upon to develop resistant cultivars [3]. Phenotypic recurrent selection is typically used to develop cultivars with improved resistance, and selection is usually carried out based on phenotypic assessments of resistance in field conditions [5]. However, field assessment of pasmo severity (PS) in

germplasm and breeding lines is costly and, is heavily influenced by the environments due to strong genotype × environment (G × E) interactions [3,4].

With the advancements in molecular marker development over the last decade, efforts to use marker-assisted breeding strategies have been pursued. One such strategy involves identifying quantitative trait loci (QTL) in biparental mapping populations and using markers to efficiently backcross QTL into elite breeding materials [6]. This so-called marker-assisted recurrent selection (MARS) or simply marker-assisted selection (MAS) characterizes many breeding programs that employ molecular markers to select non-phenotyped individuals for crossing and downstream selection of segregating populations [7]. This method is suitable for the selection of monogenic or oligo-genic architectures, but has limited use for quantitative traits controlled by many genes of smaller effects [8]. Genomic selection (GS) or prediction (GP) is an alternative marker-assisted breeding strategy better suited to polygenic quantitative traits, especially those with low heritability, because it makes use of all marker effects across the entire genome to calculate genomic estimated breeding values (GEBVs) [9] for individual plant selection [9,10].

In GP, a training population (TP) is genotyped with genome-wide markers and phenotyped for the trait(s) under selection; statistical models that best predict the breeding values from the marker data are then applied to select non-phenotyped germplasm. GP has been used to select for disease resistance in several crops such as *Fusarium* head blight (FHB) in wheat, a typically quantitatively inherited trait with predominantly additive genetic variation, where GP had a significantly higher accuracy than pedigree-based information alone [11]. GP feasibility has also been studied for selection of wheat rust resistance, and was found particularly effective when validation lines had at least one which is close to the reference lines [12]. The implementation of GP on northern leaf blight, a complex genetic architecture trait in maize, resulted in superior gains and reduced breeding cycle time to ≤ 80% of the phenotypic cycle [13]. Despite the many successful examples, the use of GP to improve disease resistance in crops has been challenging for two reasons: (i) selection for major resistance genes can be ephemeral due to changes in pathogen races; and (ii) breeding for minor resistance genes with small effects may face the remarkable complexities encountered in GP [14].

The fast-evolving genotyping platforms have been a game-changer in the implementation of GP, allowing the production of large numbers of genome-wide markers, whereas progresses in phenotyping were not associated with similar cost reduction or quantum leaps in throughput. Given the number of markers (p) and sample size (n) in a given population, there are many more p effects to be estimated than the n, leading to an infinite number of possible marker effect estimates [15], i.e., the so-called "large p, small n problem" ($p >> n$) when applying markers to predict phenotypes [11]. Several GP statistical models have been proposed to address this issue [16]. For example, the ridge-regression best linear unbiased prediction (RR-BLUP) is a mixed linear model that considers markers as random effects. Covariance between markers is considered to be zero, and the marker variance is assumed to be the total genetic variance divided by the number of markers. The variance is assumed to be equal for all markers, allowing many more marker effects to be estimated than there are phenotypic records [17]. The Bayesian LASSO (BL) assumes markers to have equal variances and, performs continuous shrinkage and variable selection simultaneously, with small-effect markers shrinking more severely than larger-effect loci. In the $p >> n$ setting, LASSO will select at most n variables and set the effects of the remaining predictors at zero [18]. Although the problem is solved statistically in these models, improving the accuracy and efficiency of GP by reducing the number of genome-wide markers would be advantageous because any increment in the TP size comes at a cost [19-22]. Genome-wide association study (GWAS) is an approach to identify genome-wide markers linked to QTL, resulting in a limited number of favorable genetic loci responsible for traits of interest [23]. For example, GP of crown rust resistance in *Lolium perenne* demonstrated GWAS' ability to identify and rank markers, which enabled the identification of a small subset of single nucleotide polymorphisms (SNPs) that could achieve predictive abilities close to that attained using the complete marker set [24]. Utilization of GWAS removes a large proportion of unrelated markers and in the construction of prediction models.

The only GP empirical study published to date in flax, which used bi-parental populations for yield, oil content and fatty acid composition traits, indicated that GP could increase genetic gain per unit time in linseed breeding. The GP results significantly exceeded those from direct phenotypic selection, especially for traits with low broad-sense heritability [25]. Resistance to flax pasmo is polygenic. Our previous study reported 500 non-redundant QTL for PR from 370 diverse flax accessions of a core collection based on five-year pasmo field assessments; of those, 134 QTL were statistically stable in all five years and 67 had relatively stable and large effects [4]. The number of positive-effect QTL was also significantly negatively correlated with pasmo severity (PS), providing a way to predict PR. Taking advantage of the GWAS results for PR, we evaluated the potential of QTL markers in GP, and compared the GP efficiency affected by different markers, including genome-wide SNPs and QTL markers, to provide a realistic and highly accurate model for germplasm evaluation and parent selection in pasmo resistance breeding.

2. Results

2.1. Evaluation of Marker Sets Used in Genomic Prediction

Four marker sets were used for GP of pasmo resistance. The first marker set contained 52,347 genome-wide SNPs (SNP-52347) that were correlated to the five-year average PS and the PS of the five individual years at a 10^{-5} probability level [4]. The other three marker sets were the 500 unique QTL (SNP-500QTL), the 134 QTL statistically stable over five consecutive years (SNP-134QTL), and the 67 stable and relatively large-effect QTL (SNP-67QTL) sets previously identified for PR [4]. The SNP-500QTL dataset comprises markers for all small- or large-effects, including QTL stable across environments and environment-specific QTL identified using ten different statistical models and all six phenotypic datasets (Figure 1). The SNP-134QTL dataset is a subset of the SNP-500QTL dataset whereas SNP-67QTL is a subset of the former; all SNP-500QTL markers were included in SNP-52347. These four marker sets explained 54, 72, 27 and 29% of the phenotypic variation of the five-year PS average (PS-mean), respectively; these values exceeded those of the individual year PS data (Table 1). Although SNP-500QTL was a subset of SNP-52347, this marker set explained a greater percentage of the phenotypic variation for PS than SNP-52347 for all datasets.

Table 1. Phenotypic variation of pasmo severity (PS) ($h^2 \pm s$) explained by the four marker sets.

PS dataset	Marker set			
	SNP-500QTL	**SNP-134QTL**	**SNP-67QTL**	**SNP-52347**
PS-mean	0.72 ± 0.04	0.27 ± 0.05	0.29 ± 0.05	0.54 ± 0.07
PS-2012	0.64 ± 0.06	0.18 ± 0.05	0.16 ± 0.04	0.43 ± 0.08
PS-2013	0.63 ± 0.06	0.12 ± 0.04	0.12 ± 0.04	0.38 ± 0.08
PS-2014	0.65 ± 0.06	0.23 ± 0.05	0.20 ± 0.05	0.45 ± 0.08
PS-2015	0.56 ± 0.06	0.20 ± 0.05	0.17 ± 0.04	0.44 ± 0.09
PS-2016	0.53 ± 0.06	0.18 ± 0.05	0.18 ± 0.05	0.38 ± 0.07

Peer-reviewed version available at Int. J. Mol. Sci. 2019, 20, 359; doi:10.3390/ijms20020359

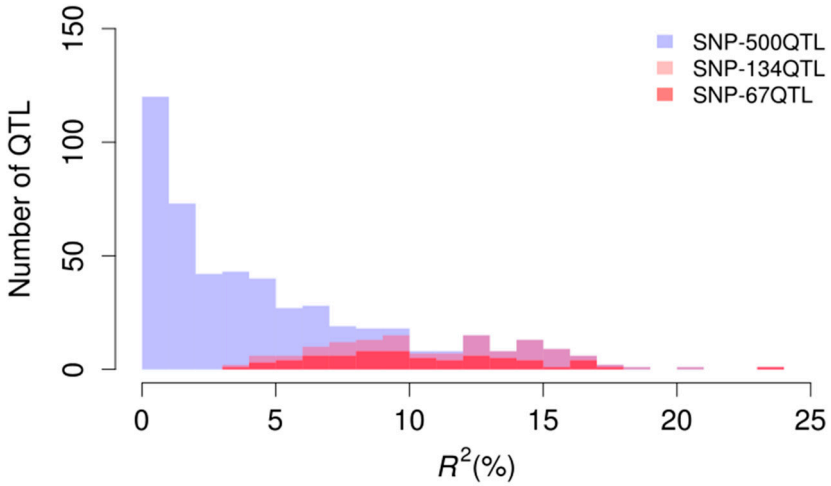

Figure 1. Distribution of R^2 (%) (phenotypic variation explained by individual QTL) in the three QTL marker sets.

2.2. *Accuracy of Genomic Prediction in Relation to Marker Sets and Pasmo Servrity Datasets*

Genomic prediction models were constructed using RR-BLUP with pairwise combinations of the four marker sets and the six PS datasets. Statistical models for the 24 combinations were generated and evaluated for their accuracy (*r*) and relative efficiency (*RE*) using a five-fold random cross-validation scheme (Table 2). *RE* represents the relative efficiency of GP over direct phenotypic selection which depends on the heritability of a selective trait. Direct phenotypic selection for a trait was considered to have a baseline efficiency of 1. Thus, RE values greater than 1 indicate GP models more efficient than direct phenotypic selection [25-27]. Analysis of variance (ANOVA) (Table S1) indicated that *r* and *RE* both significantly differed among the four marker sets and the six PS datasets; there was also a significant interaction effect between marker sets and PS datasets (Table S1). Owing to the significant marker × phenotype dataset interaction, multiple comparisons of the 24 combinations were performed. For all marker sets, the PS-mean models significantly outperformed those based on individual year datasets (Table 2). The SNP-500QTL marker set models generated significantly higher *r* and *RE* values than any other marker sets (Figure 2). Interestingly, the SNP-67QTL derived models produced slightly but significantly higher values of *r* and *RE* than SNP-134QTL models. The highest *r* and *RE* values were obtained for models combining the SNP-500QTL and PS-mean datasets (Table 2, Figure 2). Intriguingly, the SNP-52347 models yielded the lowest *r* and *RE* values despite including all QTL markers (Table 2, Figure 2); both BL and Bayesian ridge regression (BRR) corroborated this finding (Figure S1). No significant differences in *r* and *RE* values were observed among the three statistical models: RR-BLUP, BL and BRR (Figure S1).

Table 2. Accuracy (*r*) and relative efficiency (*RE*) values of the 24 combinations representing the four marker sets and six pasmo severity (PS) datasets using RR-BLUP obtained using a random five-fold cross-validation.

Marker set	PS dataset	r ($\bar{x}$ ± s) [1]	RE ($\bar{x}$ ± s) [1]
SNP-500QTL	PS-mean	0.92 ± 0.02a	1.84 ± 0.04a
	PS-2012	0.84 ± 0.03b	1.68 ± 0.06b
	PS-2013	0.81 ± 0.04c	1.62 ± 0.07c
	PS-2014	0.82 ± 0.04c	1.63 ± 0.07c
	PS-2015	0.76 ± 0.05d	1.52 ± 0.09d

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 November 2018 doi:10.20944/preprints201811.0623.v1

Peer-reviewed version available at Int. J. Mol. Sci. 2019, 20, 359; doi:10.3390/ijms20020359

	PS-2016	0.76 ± 0.05d	1.52 ± 0.11d
SNP-134QTL	PS-mean	0.75 ± 0.06e	1.49 ± 0.11e
	PS-2012	0.68 ± 0.06f	1.36 ± 0.11f
	PS-2013	0.60 ± 0.07ij	1.19 ± 0.14ij
	PS-2014	0.60 ± 0.07i	1.21 ± 0.14i
	PS-2015	0.47 ± 0.09o	0.94 ± 0.18o
	PS-2016	0.56 ± 0.09l	1.12 ± 0.17l
SNP-67QTL	PS-mean	0.76 ± 0.05d	1.53 ± 0.1d
	PS-2012	0.67 ± 0.06g	1.35 ± 0.11g
	PS-2013	0.60 ± 0.07ij	1.20 ± 0.14ij
	PS-2014	0.60 ± 0.07ij	1.20 ± 0.14ij
	PS-2015	0.50 ± 0.09n	1.00 ± 0.17n
	PS-2016	0.59 ± 0.08k	1.17 ± 0.17k
SNP-52347	PS-mean	0.67 ± 0.07g	1.33 ± 0.14g
	PS-2012	0.63 ± 0.06h	1.27 ± 0.12h
	PS-2013	0.59 ± 0.07jk	1.19 ± 0.14jk
	PS-2014	0.53 ± 0.08m	1.06 ± 0.17m
	PS-2015	0.38 ± 0.09q	0.77 ± 0.17q
	PS-2016	0.46 ± 0.09p	0.93 ± 0.18p

[1] Different letters represent multiple test significance among the 24 combinations at the 0.05 probability level.

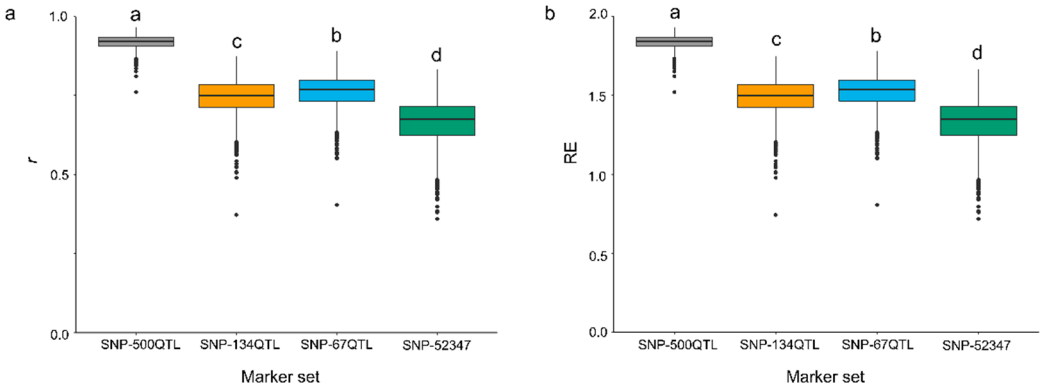

Figure 2. Accuracy (*r*) (**a**) and relative efficiency (*RE*) (**b**) of RR-BLUP prediction models built with combinations of four marker sets using the five-year average PS dataset (PS-mean) and random five-fold cross-validations. Letters above box plots indicated statistical significance ($P < 0.05$) for *r* and *RE* among marker sets.

2.3. Sample Size of Training Populations versus Genomic Prediction Accuracy

To find an optimal size for the TP, the relationship between TP size and prediction accuracy was analyzed. TPs of various sizes from 18 to 351, corresponding to 5 to 95% of the total 370 accessions, were used to build models with the SNP-500QTL marker set and the PS-mean phenotypic dataset. The prediction accuracy significantly increased for TP sizes up to 100, followed by smaller accuracy gains with every additional TP size increments (Figure 3). A GP accuracy >0.9 was obtained once the TP size reached 185.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 November 2018 doi:10.20944/preprints201811.0623.v1

Peer-reviewed version available at Int. J. Mol. Sci. 2019, 20, 359; doi:10.3390/ijms20020359

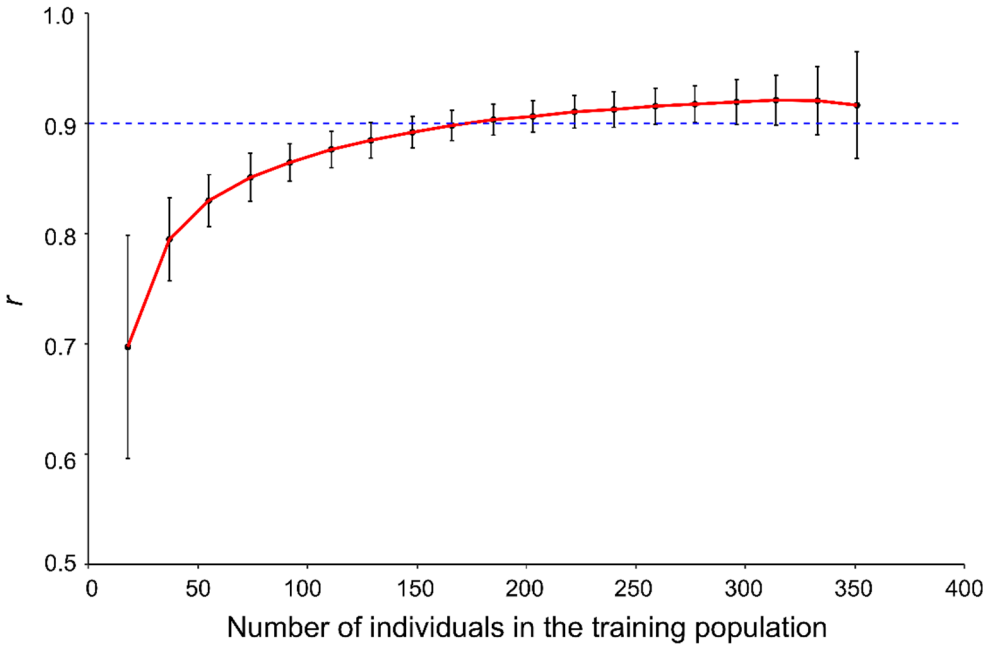

Figure 3. Relationship between the genomic prediction accuracy (*r*) and the size of the training population based on the SNP-500QTL marker set, the PS-mean dataset and the RR-BLUP models.

2.4. Prediction Models of Pasmo Resistance

All 370 accessions were used as a training population to build a prediction model using the SNP-500QTL genotypic dataset and the PS-mean phenotypic dataset because this combination outperformed all other models. The model was then employed to predict PS in each year (Table 3). Prediction accuracies (*r*) ranging from 0.71 to 0.81 and *RE* values of 1.42 to 1.62 were obtained when predicting PS for individual years (Table 3).

A prediction accuracy as high as 0.98 and a *RE* value of 1.96 were obtained when the model was used to predict PS-means of the 370 accessions (Table 3). A linear relationship was observed between the observed (y) and predicted PS (x): $y = 1.0522x - 0.3267$ ($R^2 = 0.96$) (Figure 4). Based on this equation, the average prediction interval between the two red dashed lines, representing the 95% confidence interval, was only less than 1 on the PS scale (Figure 4).

Positive-effect QTL (NPQTL) in the 370 accessions for the 500 QTL set were tallied. Significant linear correlation between PS-mean and NPQTL ($r = 0.86$ or $R^2 = 0.73$) was observed (Figure 5). This correlation was less than but close to the accuracy of the GP model with SNP-500QTL, and higher than the GP models using other marker sets (Table 2). However, the single linear regression equation ($y = -0.0262x + 11.934$) of the observed PS (y) to NPQTL (x) had a large standard deviation for each prediction value, with an average prediction interval width of 2.70, nearly three times the average prediction interval width of the GP model; that is, the NPQTL model had a higher prediction error than the GP model.

Table 3. Accuracy (*r*) and relative efficiency (*RE*) of genomic prediction for pasmo severity in different years using the RR-BLUP model built with the SNP-500QTL marker set and the PS-mean phenotypic data using all 370 accessions as training data set.

PS dataset for prediction	*r*	*RE*
PS-mean	0.98	1.96
PS-2012	0.73	1.46
PS-2013	0.71	1.42
PS-2014	0.81	1.62
PS-2015	0.71	1.43
PS-2016	0.77	1.55

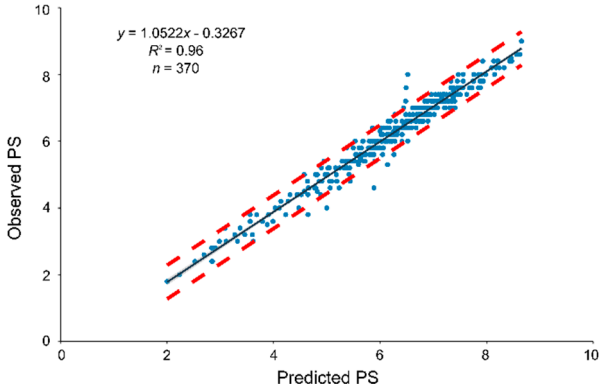

Figure 4. Linear regression of observed pasmo severity (PS) (y) to predicted PS (x) using the genomic prediction model built with the PS-mean dataset and the SNP-500QTL marker set of all 370 accessions as training data set. The red dashed lines represent upper and lower boundaries of the 95% prediction intervals. The average width of the intervals for all predicted values was 0.97.

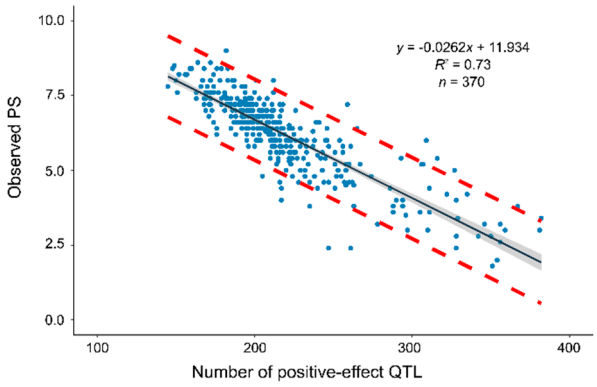

Figure 5. Linear regression of observed pasmo severity (PS) (y) to the number of positive-effect QTL (NPQTL) (x) in the 370 flax accessions. The grey band represents the 95% confidence interval, i.e., 95% of those intervals include the true value of the population mean. The red dashed lines represent the upper and lower boundaries of the 95% prediction interval, i.e., it is expected that the value of a sample lies within that prediction interval in 95% of the samples. The average width of the prediction interval for all predicted values is 2.70.

2.5. A Case Study of Genomic Prediction

To assess GP prediction accuracy, a training-testing partition was generated with random assignment of breeding lines to either training or testing subsets. Considering the different improvement status of accessions in the population (cultivars, breeding lines, landraces, or unknown types) and different levels of resistance, we randomly chose 20% of the 370 accessions in the population, i.e., 93 accessions (52 cultivars, 21 breeding lines, 3 landraces, and 17 unknown types) as validation dataset, i.e., a five-fold random cross-validation set. To predict the PS of these 93 accessions, a RR-BLUP model using the SNP-500QTL set and the PS-mean of the remaining 277 accessions as TP set was built and subsequently applied to the genomic data of the same 500 QTL of the 93 accession testing set to predict PS. The predicted results are shown in Figure 6a and Table S2. The prediction accuracy was as high as 0.95 (r between observed and predicted PS). Similarly, a linear regression model of observed PS (y) to NPQTL (x) of the 277 accessions (the same TP as GP) produced y = -0.026x + 11.902 (Figure S2), which was similar to the regression equation previously obtained with the complete accession set (Figure 5). Using this prediction model, predicted PS and intervals

were calculated (Figure 6b, Table S2). The prediction accuracy of 0.92 for NPQTL was slightly inferior to that of the GP model. The observed PS values all fell within prediction intervals (Table S2).

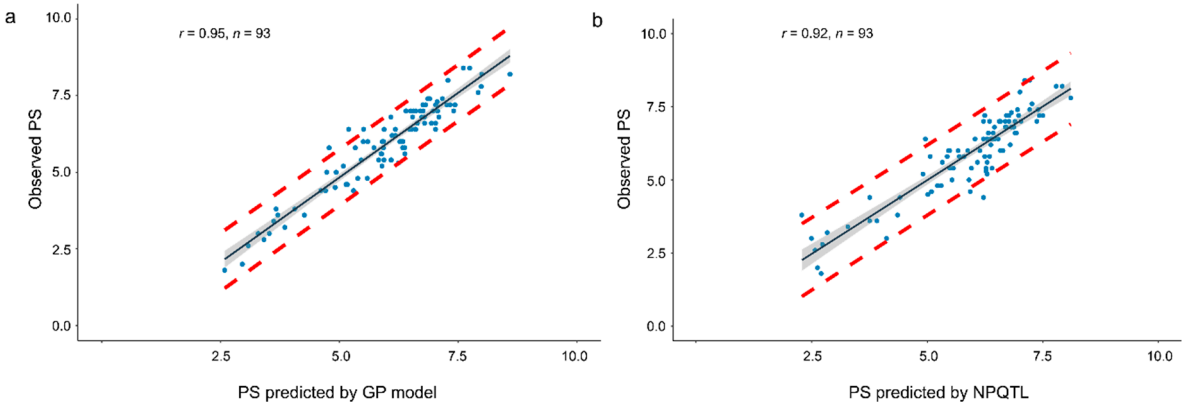

Figure 6. Relationship of observed pasmo severity (PS) of 93 randomly chosen accessions with the PS predicted by the genomic model constructed with the SNP-500QTL marker set and PS-mean dataset when a random subset of 277 accessions was used as training population (**a**) and with the PS predicted by the linear equation (Figure S2) of observed PS (y) to the number of positive-effect QTL (NPQTL) (x) of the same 277 accessions as training population (y = -0.026x + 11.902) (**b**). The grey band represents the 95% confidence interval. The red dashed lines represent the upper and lower boundaries of the 95% prediction interval.

3. Discussion

Cross-validation remains the most popular method to evaluate GP accuracy [14,28]. Our RR-BLUP model prediction accuracy of 0.92 for PR is the highest of all published GP models for plant disease resistance traits [14]. This model is especially valuable because PR has low heritability and high inheritance complexity [3,4]. The QTL markers, multi-year phenotypic data, and the genetic diversity and size of the population likely contributed positively to this high prediction accuracy [29].

3.1. All Detected QTL Used as Markers in Genomic Prediction

Three sets of QTL markers (SNP-500QTL, SNP-134QTL, and SNP-67QTL) and a genome-wide SNP marker set (SNP-52347) were evaluated here. GP models built using SNP-500QTL consistently outperformed models derived with any of the other three marker sets (Table 2, Figure S1), lending credence to the robustness and reliability of the QTL identified using multiple single-locus and multi-locus GWAS statistical methods [4]. Most GWAS aim to detect large-effect QTL, such as the SNP-67QTL set. While potentially useful in MAS, these tend to explain a reduced portion of the phenotypic variation compared to more comprehensive models (Table 1). Consequently, the GP models built with such marker sets have lower GP accuracies. Therefore, using all potential QTL associated with the selective trait to build GP models is advantageous because it greatly improves prediction accuracy. Prediction accuracies of models obtained with SNP-134QTL and SNP-67QTL data sets were comparable (Table 2, Figure S1) and they explained a similar proportion of the phenotypic variation for PS (Table 1), confirming the redundancy or overlap between the two datasets. Removal of redundant QTL from SNP-134QTL to produce SNP-67QTL produced slightly higher accuracy models (Figure 2). Simplifying GP models by removal of redundant and unrelated markers will ease the practical implementation of GP in breeding programs.

3.2 Superier Performance of Genomic Prediction Combined With GWAS

Surprisingly, the GP models built using SNP-52347 generated a lower prediction accuracy than the models with SNP-500QTL (Table 2, Figure S1), regardless of the statistical methods (Figure S1). Similarly, SNP-52347 explained a lower percentage of the phenotypic variation for PS than SNP-500QTL (Table 1). Besides interaction between SNPs, introduction of noise from genome-wide

markers [30], the low prediction accuracy may also be owing to some of the erroneously called SNPs and imputation of missing SNP data. SNP-500QTL includes all or nearly all QTL potentially associated with PS; additional markers, not only failed to increase, but actually reduced the prediction accuracy, further emphasizing the effectiveness of the QTL identification methodology adopted in our previously published GWAS study [4]. Similar findings were found for FHB in wheat where deoxynivalenol (DON) concentration QTL-linked markers significantly improve prediction accuracy compared to random genome-wide markers [30]. Markers linked to QTL underlying important traits are deemed more useful for prediction strategies because genome-wide markers may introduce noise, thereby reducing accuracy [30]. Using QTL for GP models may be beneficial to balance genetic backgrounds along with maximum gain of breeding value [31]. Genome-wide prediction models based on ~5000 SNPs from *de novo* GWAS for tropical rice improvement were as effective for prediction as the full marker set of 108,005 SNPs, indicating that the relationship between marker number and prediction accuracy is neither strict nor linear [32]. To sum up, combined applications of the QTL discovered via GWAS and the accelerated breeding cycles through GP facilitate the full use of genome-wide markers in crop disease resistance breeding [10,33]. Removal of redundant markers has the potential to alleviate the effect of the "large *p*, small *n*" issue.

3.3 Accuracy of GP Modeling by Environment, Training Population, and Statistical Methods

G × E interactions, which affects the accuracy of trait assessment, are common for plant traits. A strong G × E interaction was observed in flax PR in our previous study [4]. As a consequence, different PS QTL were identified for individual years and for the 5-year average [4]; similarly, GP efficiencies differed when individual yearly and average PS data sets were used as training sets (Table 2). The highest accuracies were obtained when the 5-year mean phenotypic data was used as training data (Table 3), suggesting that the average phenotypic data across multiple environments should be used for GP model construction. Theoretically, the means across multiple environments estimate or reflect the true breeding values of a trait.

Some studies report that prediction accuracy of GP is highly affected by the size of the TP. In general, the prediction accuracy increases with TP size [21,28,29,34-36]. In the GP of seed weight in soybean, for example, prediction accuracy was sensitive to changes in TP size, which may have led to changes of relatedness between training and validation sets [21]. Lorenzana and Bernardo observed that, in an Arabidopsis family, prediction accuracy improved by 0.10 when TP size increased from 48 to 96, by an additional 0.07 when TP size was increased to 192, and by a further 0.05 with a TP size of 332 [37]. Here GP accuracy >0.9 was observed when the TP size reached 185 which slightly increased to 0.921 with a TP size of 314 (Figure 3). Large TPs provide the statistical power needed to improve prediction accuracy [38], especially for traits with low heritability [34,39].When TP size is sufficiently large, even low heritability traits can be accurately predicted [28,40], including the low heritability PS studied therein. Diversity of the population also affect prediction accuracy [21,29,34,41-43]. A diverse TP may contain more QTL associated with selective traits and increase the correlation of the TP with validation populations (VPs) or test/prediction populations (PPs), resulting in a subsequent increase in prediction accuracy. Although some breeding lines [11,30,44] and bi-parental derived lines [25,41,45,46] are used for TPs, many studies have opted for a more diverse TP germplasm [29,41-43]. Our core collection TP preserves the variation present in the world collection of 3,378 accessions maintained by Plant Gene Resources of Canada (PGRC) and represents a broad range of geographical origins, different improvement statuses (landraces, historical and modern cultivars, breeding lines), and two morphotypes (linseed and fibre types) [1,3]. This collection also contains most parents of modern Canadian flax cultivars [25]. Therefore, diverse phenotypic and genetic variabilities within the flax core collection render it useful as a resource for breeding and as a TP for GP model construction.

A variety of statistical methods have been proposed for GP. In general, GP methods are based on additive models and their accuracies might vary slightly depending on the target traits. In practice, statistical models such as RR-BLUP, BRR, BL, BayesB, BayesA, wBSR, BayesCπ, E-Bayes, RKHS, RF, SVM, and NNet produce similar prediction accuracies [10,25]. Among them, RR-BLUP is most

commonly used because of some superior features [11,14,42,47-49]. For example, comparisons among RR-BLUP, Bayes-Cπ, and RKHSR showed no difference in accuracies in a wheat FHB study [19]. Similarly, no significant differences in accuracy among RR-BLUP, BRR, and BL were found for yield and seed quality traits in our previous flax study [25]. But, another wheat FHB study that evaluated RR-BLUP, LASSO and elastic net on a diverse set of breeding lines indicated that RR-BLUP outperformed other models [11]. RR-BLUP successfully recognized complex patterns with additive effects and delivered good GP in wheat disease resistance [50]. Furthermore, RR-BLUP has a clear-cut computational efficiency [11,49,51,52]. In this study, GP accuracy does not vary in three different models (Figure S1) and the RR-BLUP model with the 500 QTL markers and the 5-year mean PS produced high prediction accuracy and is therefore recommended for the prediction of PR in flax.

3.4. Pasmo Severity Prediction Using Number of Positive-effect QTL

A highly significant correlation ($r = 0.86$ or $R^2 = 0.73$) between NPQTL and PS (Figure 5) provides an alternative approach to directly predict PS phenotypes. The prediction accuracy using the linear regression equation of PS to NPQTL was inferior to the GP model (Figures 4, 5, and 6) because the QTL effects were variable (Figure 1), whereas the linear regression equation considered only the number of QTL but not their individual effects. However, NPQTL is advantageous because it can be readily calculated based on the genotyping by sequencing (GBS) or other genotyping data for the QTL markers [14] and the prediction accuracy based on the NPQTL is comparable to most GP models. Thus, the NPQTL-based prediction equation provides a simple alternative model for PS prediction.

4. Materials and Methods

4.1. Population

A total of 370 diverse flax accessions from the core collection [1] were used to evaluate different GP models. This subset of the core collection collected from 38 countries in 12 geographic regions has been used to identify the QTL associated with PS used in our models [4].

4.2. Pasmo Resistance Data

All flax accessions were assessed for PS from 2012 to 2016 at the Morden Research and Development Centre, Agriculture and Agri-Food Canada (AAFC), Morden, Manitoba, Canada [4]. The PS observed at late flowering or maturity was used for GP as previously described [4]. PS was assessed using a 0–9 scale (0 = no sign of infection and 9 = > 90% leaf and stem area infected) with scores of 0-2 as resistant (R), 3-4 as moderately resistant (MR), 5-6 as moderately susceptible (MS), and 7-9 as susceptible (S) [4]. Six sets of PS, including five individual year datasets and the 5-year average, were used for GP modeling.

4.3. Genomic Data

A total of 258,873 SNPs were obtained from the 370 accessions after pruning by removing redundant SNPs. If two SNPs were located within a 200-kb window and had a pairwise correlation coefficient (r^2) greater than 0.8, only one was retained [4]. The missing data of SNPs were imputed using Beagle v.4.2 with default parameters [53]. From this GBS data [54], three QTL sets associated with PS were identified using GWAS [4]. These QTL included 500 unique QTL, 134 statistically stable QTL, and 67 stable and large-effect QTL. In addition, we performed Pearson's χ^2 test with Yate's continuity correction to detect all SNPs significantly associated with PS using a 10^{-5} probability level. The three QTL sets and the genome-wide SNP set were used to construct the GP models. Thus, GP models with the 24 combinations of the four marker sets and the six phenotypic datasets were built and compared.

4.4. Genomic Prediction Models

Three statistical methods RR-BLUP [9,17,20], Bayesian LASSO (BL) [20,25,33], and Bayesian ridge regression (BRR) [25,55] were used to build GP models for PS. These predictive models estimate marker effects by modelling markers as random effects. No fixed effects were fitted in the models. The statistical models and their computation procedures are described in details elsewhere [40,56]. The R package rrBLUP [51] was used to fit the RR-BLUP model and the R package BLR [57] was used to fit the BL and BRR models. The parameters used to fit BL and BRR were determined based on suggestions of de los Campos *et al*. [57]. Broad-sense heritability (0.25) of PS estimated in the population [3] was used. When preparing QTL marker data for model construction, the positive-effect allele of the tag SNP of a QTL was coded '1' and the alternative allele '-1'. Similarly for the SNP marker set, the reference allele of an SNP was coded '1' and the alternative allele '-1'. Missing data were coded '0'. The EM algorithm implemented in the R package rrBLUP [51] was used to impute the missing marker data because missing marker data were not allowed in the model construction.

4.5. Evaluation of Prediction Models

Two validation methods were used to evaluate prediction models generated from combinations of statistical models, marker sets, and PS datasets. The first method was a five-fold random cross-validation. The 370 flax accessions were randomly partitioned into five subsets. For a given partition, each subset was in turn used as validation or test data, and the remaining four subsets made the training dataset. This partitioning was repeated 500 times. In this manner, a total of 2,500 training data sets were created to build GP models and estimate marker effects. These were used to predict the breeding values of the individuals in the corresponding 2,500 test/validation datasets. The accuracy of the genomic predictions (r) was defined by the Pearson's simple correlation coefficient between the genetic values predicted by GP and the observed phenotypic values. The relative efficiency of genomic prediction over phenotypic selection (RE) was estimated using r/H^2 [26,27], where H^2 refers to the broad-sense heritability of PS, estimated to be 0.25 [3]. Means of r and RE of the 500 samplings for each marker set, GP model, and PS dataset were used to describe the prediction accuracy of GP and the efficiency of one GP cycle relative to one phenotypic selection cycle, respectively. To compare different marker and PS datasets, a joint analysis of variance was performed to test the statistical significance of differences in r and RE using R. As a case study, we randomly selected 20% of all 370 accessions as validation dataset and used the remaining 277 accessions as training dataset to build a GP model for genomic prediction of unknown germplasm.

The second cross-validation approach involved comparisons across different PS datasets, i.e., each of the six complete PS phenotypic datasets were used as training datasets to build GP models that were applied to itself and to the other five phenotypic datasets. The same set of markers for all 370 accessions was used for training and validation. This method tests the relevance of models built based on single year phenotypic data to predict phenotypes measured in different years.

4.6. Phenotypic Variation Explained by Markers

The phenotypic variation explained by all markers in various marker sets, denoted h^2_{SNP}, was estimated for all PS datasets based on the mixed linear model [58] implemented in the GCTA software [59]. The detailed calculation is described in [60].

5. Conclusions

Using a diverse worldwide flax core collection of 370 accessions as a training and test population with 500 QTL identified by GWAS, the 5-year average PS data, and the RR-BLUP statistical model, we developed a highly effective GP model with a prediction accuracy as high as 0.92 for pasmo, a low heritability and high inheritance complexity trait. This is the highest reported accuracy value of all GP models for plant disease resistance traits and comparable with previously published results. As an alternative, we developed a linear regression prediction model based on NPQTL that also produced a high prediction accuracy of 0.86. The GP model and the NPQTL-based regression

equation were validated and deemed to be applicable to the evaluation of flax germplasm including parent selection for PR. The use of all potential QTL associated with a target trait would be beneficial because the exclusion of a large proportion of unrelated markers would facilitate the construction of highly accurate GP models.

Supplementary Materials: Supplementary materials can be found online.

Author Contributions: Conceptualization, F.M.Y. and S.C.; Methodology, F.M.Y.; Software, F.M.Y.; Formal Analysis, F.M.Y., G.J., P.L. and L.H.; Resources, K.Y.R. (field phenotypic data), S.C., and F.M.Y.; Data Curation, F.M.Y., K.Y.R., and Z.Y.; Writing-Original Draft Preparation, F.M.Y., L.H., and J.X.; Writing-Review & Editing, F.M.Y., S.C., and X.W.; Visualization, Z.Y.; Supervision, F.M.Y., S.C., and X.W.; Funding Acquisition, S.C., K.R., and F.M.Y.

Funding: This work was part of the Total Utilization Flax GENomics (TUFGEN) project funded by Genome Canada and other stakeholders, the A-base project (J-001004) funded by Agriculture and Agri-Food Canada, and the flax cluster project funded by the Western Grains Research Foundation (WGRF) and the Canada-China science and technology and innovation action plan (2017ZJGH0106002).

Acknowledgments: We thank the China Scholarship Council for their financial support of Liqiang He for his research at Agriculture and Agri-Food Canada (AAFC).

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analysis, or interpretation of the data; in the writing of the manuscript; and in the decision to publish the results.

Abbreviations

ANOVA	Analysis of variance
BL	Bayesian LASSO
BRR	Bayesian ridge regression
DON	Deoxynivalenol
E-Bayes	Empirical Bayes
FHB	*Fusarium* head blight
G × E	Genotype by environment interaction
GBS	Genotyping by sequencing
GEBV	Genomic estimated breeding value
GP	Genomic prediction
GS	Genomic selection
GWAS	Genome-wide association study
MARS	Marker-assisted recurrent selection
MAS	Marker-assisted selection
MR	Moderately resistant
MS	Moderately susceptible
NNet	Neural network
NPQTL	Number of positive-effect QTL
PGRC	Plant Gene Resources of Canada
PP	Test/prediction population
PR	Pasmo resistance
PS	Pasmo severity
QTL	Quantitative trait locus/loci
QTN	Quantitative trait nucleotide

R	Resistant
RE	Relative efficiency
RF	Random forest
RKHS	Reproducing kernels Hilbert spaces regression
RR-BLUP	Ridge regression best linear unbiased prediction
S	Susceptible
SNPs	Single nucleotide polymorphisms
SVM	Support vector machine
TP	Training population
VP	Validation population
wBSR	Weighted Bayesian shrinkage regression

References

1. Diederichsen, A.; Kusters, P.M.; Kessler, D.; Bainas, Z.; Gugel, R.K. Assembling a core collection from the flax world collection maintained by Plant Gene Resources of Canada. *Genet. Resour. Crop Evol.* **2012,** 60, (4), 1479-1485.
2. Vera, C.L.; Irvine, R.B.; Duguid, S.D.; Rashid, K.Y.; Clarke, F.R.; Slaski, J.J. Pasmo disease and lodging in flax as affected by pyraclostrobin fungicide, N fertility and year. *Can. J. Plant Sci.* **2014,** 94, (1), 119-126.
3. You, F.M.; Jia, G.; Xiao, J.; Duguid, S.D.; Rashid, K.Y.; Booker, H.M.; Cloutier, S. Genetic variability of 27 traits in a core collection of flax (*Linum usitatissimum* L.). *Front. Plant Sci.* **2017,** 8, 1636.
4. He, L.; Xiao, J.; Rashid, K.Y.; Yao, Z.; Li, P.; Jia, G.; Wang, X.; Cloutier, S.; You, F.M. Genome-wide association studies for pasmo resistance in flax (*Linum usitatissimum* L.) *Preprints* **2018**, doi:10.20944/preprints201811.0336.v1.
5. Diederichsen, A.; Rozhmina, T.A.; Kudrjavceva, L.P. Variation patterns within 153 flax (*Linum usitatissimum* L.) genebank accessions based on evaluation for resistance to *fusarium* wilt, anthracnose and pasmo. *Plant Genet. Resour.* **2008,** 6, (1), 22-32.
6. Collard, B.C.Y.; Mackill, D.J. Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **2008,** 363, (1491), 557-572.
7. Heslot, N.; Jannink, J.L.; Sorrells, M.E. Perspectives for genomic selection applications and research in plants. *Crop Sci.* **2015,** 55, (1), 1-12.
8. Xu, Y.; Crouch, J.H. Marker-assisted selection in plant breeding: from publications to practice. *Crop Sci.* **2008,** 48, (2), 391-407.
9. Meuwissen, T.H.; Hayes, B.J.; Goddard, M.E. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **2001,** 157, (4), 1819-29.
10. Lipka, A.E.; Kandianis, C.B.; Hudson, M.E.; Yu, J.M.; Drnevich, J.; Bradbury, P.J.; Gore, M.A. From association to prediction: statistical methods for the dissection and selection of complex traits in plants. *Curr. Opin. Plant Biol.* **2015,** 24, 110-118.
11. Arruda, M.P.; Brown, P.J.; Lipka, A.E.; Krill, A.M.; Thurber, C.; Kolb, F.L. Genomic selection for predicting *Fusarium* head blight resistance in a wheat breeding program. *Plant Genome* **2015,** 8, (3).

12. Daetwyler, H.D.; Bansal, U.K.; Bariana, H.S.; Hayden, M.J.; Hayes, B.J. Genomic prediction for rust resistance in diverse wheat landraces. *Theor. Appl. Genet.* **2014,** 127, (8), 1795-803.
13. Technow, F.; Burger, A.; Melchinger, A.E. Genomic prediction of northern corn leaf blight resistance in maize with combined or separated training sets for heterotic groups. *G3 (Bethesda)* **2013,** 3, (2), 197-203.
14. Poland, J.; Rutkoski, J. Advances and challenges in genomic selection for disease resistance. *Annu. Rev. Phytopathol.* **2016,** 54, 79-98.
15. Gianola, D. Priors in whole-genome regression: the Bayesian alphabet returns. *Genetics* **2013,** 194, (3), 573-96.
16. Desta, Z.A.; Ortiz, R. Genomic selection: genome-wide prediction in plant improvement. *Trends Plant Sci.* **2014,** 19, (9), 592-601.
17. Whittaker, J.C.; Thompson, R.; Denham, M.C. Marker-assisted selection using ridge regression. *Genet. Res.* **2000,** 75, (2), 249-252.
18. Tibshirani, R. Regression shrinkage and selection via the Lasso. *J. Roy. Stat. Soc. SB-Method.* **1996,** 58, (1), 267-288.
19. Jiang, Y.; Zhao, Y.; Rodemann, B.; Plieske, J.; Kollers, S.; Korzun, V.; Ebmeyer, E.; Argillier, O.; Hinze, M.; Ling, J., *et al.* Potential and limits to unravel the genetic architecture and predict the variation of *Fusarium* head blight resistance in European winter wheat (*Triticum aestivum* L.). *Heredity* **2015,** 114, (3), 318-26.
20. Spindel, J.; Begum, H.; Akdemir, D.; Virk, P.; Collard, B.; Redona, E.; Atlin, G.; Jannink, J.L.; McCouch, S.R. Genomic selection and association mapping in rice (*Oryza sativa*): effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. *PLoS Genet.* **2015,** 11, (2), e1004982.
21. Zhang, J.; Song, Q.; Cregan, P.B.; Jiang, G.L. Genome-wide association study, genomic prediction and marker-assisted selection for seed weight in soybean (*Glycine max*). *Theor. Appl. Genet.* **2016,** 129, (1), 117-30.
22. Li, Y.; Ruperao, P.; Batley, J.; Edwards, D.; Khan, T.; Colmer, T.D.; Pang, J.; Siddique, K.H.M.; Sutton, T. Investigating drought tolerance in chickpea using genome-wide association mapping and genomic selection based on whole-genome resequencing data. *Front. Plant Sci.* **2018,** 9, 190.
23. Yu, J.; Buckler, E.S. Genetic association mapping and genome organization of maize. *Curr. Opin. Biotechnol.* **2006,** 17, (2), 155-60.
24. Arojju, S.K.; Conaghan, P.; Barth, S.; Milbourne, D.; Casler, M.D.; Hodkinson, T.R.; Michel, T.; Byrne, S.L. Genomic prediction of crown rust resistance in Lolium perenne. *BMC Genet.* **2018,** 19, (1), 35.
25. You, F.M.; Booker, H.M.; Duguid, S.D.; Jia, G.; Cloutier, S. Accuracy of genomic selection in biparental populations of flax (*Linum usitatissimum* L.). *Crop J.* **2016,** 4, (4), 290-303.
26. Dekkers, J.C. Prediction of response to marker-assisted and genomic selection using selection index theory. *J. Anim. Breed. Genet.* **2007,** 124, 331-341.
27. Ziyomo, C.; Bernardo, R. Drought tolerance in maize: indirect selection through secondary traits versus genomewide selection. *Crop Sci.* **2013,** 53, 1269-1275.
28. Crossa, J.; Perez-Rodriguez, P.; Cuevas, J.; Montesinos-Lopez, O.; Jarquin, D.; de Los Campos, G.; Burgueno, J.; Gonzalez-Camacho, J.M.; Perez-Elizalde, S.; Beyene, Y., *et al.* Genomic selection in plant breeding: methods, models, and perspectives. *Trends. Plant Sci.* **2017,** 22, (11), 961-975.

29. Gowda, M.; Das, B.; Makumbi, D.; Babu, R.; Semagn, K.; Mahuku, G.; Olsen, M.S.; Bright, J.M.; Beyene, Y.; Prasanna, B.M. Genome-wide association and genomic prediction of resistance to maize lethal necrosis disease in tropical maize germplasm. *Theor. Appl. Genet.* **2015,** 128, (10), 1957-68.
30. Rutkoski, J.; Benson, J.; Jia, Y.; Brown-Guedira, G.; Jannink, J.-L.; Sorrells, M. Evaluation of genomic prediction methods for Fusarium head blight resistance in wheat. *Plant Genome* **2012,** 5, (2).
31. Deshmukh, R.; Sonah, H.; Patil, G.; Chen, W.; Prince, S.; Mutava, R.; Vuong, T.; Valliyodan, B.; Nguyen, H.T. Integrating omic approaches for abiotic stress tolerance in soybean. *Front. Plant Sci.* **2014,** 5, 244.
32. Spindel, J.E.; Begum, H.; Akdemir, D.; Collard, B.; Redona, E.; Jannink, J.L.; McCouch, S. Genome-wide prediction models that incorporate de novo GWAS are a powerful new tool for tropical rice improvement. *Heredity (Edinb)* **2016,** 116, (4), 395-408.
33. Kayondo, S.I.; Pino Del Carpio, D.; Lozano, R.; Ozimati, A.; Wolfe, M.; Baguma, Y.; Gracen, V.; Offei, S.; Ferguson, M.; Kawuki, R., *et al.* Genome-wide association mapping and genomic prediction for CBSD resistance in *Manihot esculenta*. *Sci. Rep.* **2018,** 8, (1), 1549.
34. Wang, X.; Xu, Y.; Hu, Z.L.; Xu, C.W. Genomic selection methods for crop improvement: current status and prospects. *Crop J.* **2018,** 6, (4), 330-340.
35. Jarquin, D.; Kocak, K.; Posadas, L.; Hyma, K.; Jedlicka, J.; Graef, G.; Lorenz, A. Genotyping by sequencing for genomic prediction in a soybean breeding population. *BMC Genomics* **2014,** 15, 740.
36. Asoro, F.G.; Newell, M.A.; Beavis, W.D.; Scott, M.P.; Jannink, J.-L. Accuracy and training population design for genomic selection on quantitative traits in elite North American oats. *Plant Genome* **2011,** 4, (2).
37. Lorenzana, R.E.; Bernardo, R. Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. *Theor. Appl. Genet.* **2009,** 120, (1), 151-61.
38. Goddard, M. Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica* **2009,** 136, (2), 245-57.
39. Nielsen, H.M.; Sonesson, A.K.; Yazdi, H.; Meuwissen, T.H.E. Comparison of accuracy of genome-wide and BLUP breeding value estimates in sib based aquaculture breeding schemes. *Aquacult.* **2009,** 289, (3-4), 259-264.
40. Lorenz, A.J.; Chao, S.; Asoro, F.G.; Heffner, E.L.; Hayashi, T.; Iwata, H.; Smith, K.P.; Sorrells, M.E.; Jannink, J.L. Genomic selection in plant breeding. In *Advances in Agronomy*, 2011; Vol. 110, pp 77-123.
41. Cuevas, J.; Crossa, J.; Montesinos-Lopez, O.A.; Burgueno, J.; Perez-Rodriguez, P.; de Los Campos, G. Bayesian genomic prediction with genotype x environment interaction kernel models. *G3 (Bethesda)* **2017,** 7, (1), 41-53.
42. Dong, H.; Wang, R.; Yuan, Y.; Anderson, J.; Pumphrey, M.; Zhang, Z.; Chen, J. Evaluation of the potential for genomic selection to improve spring wheat resistance to Fusarium head blight in the Pacific Northwest. *Front. Plant Sci.* **2018,** 9, 911.
43. Isidro, J.; Jannink, J.L.; Akdemir, D.; Poland, J.; Heslot, N.; Sorrells, M.E. Training set optimization under population structure in genomic selection. *Theor. Appl. Genet.* **2015,** 128, (1), 145-58.
44. Rutkoski, J.E.; Poland, J.A.; Singh, R.P.; Huerta-Espino, J.; Bhavani, S.; Barbier, H.; Rouse, M.N.; Jannink, J.-L.; Sorrells, M.E. Genomic selection for quantitative adult plant stem rust resistance in wheat. *Plant Genome* **2014,** 7, (3).
45. McElroy, M.S.; Navarro, A.J.R.; Mustiga, G.; Stack, C.; Gezan, S.; Pena, G.; Sarabia, W.; Saquicela, D.; Sotomayor, I.; Douglas, G.M., *et al.* Prediction of cacao (*Theobroma cacao*) resistance to Moniliophthora spp. diseases via genome-wide association analysis and genomic selection. *Front. Plant Sci.* **2018,** 9, 343.

46. Enciso-Rodriguez, F.; Douches, D.; Lopez-Cruz, M.; Coombs, J.; de Los Campos, G. Genomic selection for late blight and common scab resistance in tetraploid potato (*Solanum tuberosum*). *G3 (Bethesda)* **2018,** 8, (7), 2471-2481.
47. Rutkoski, J.; Singh, R.P.; Huerta-Espino, J.; Bhavani, S.; Poland, J.; Jannink, J.L.; Sorrells, M.E. Genetic gain from phenotypic and genomic selection for quantitative resistance to stem rust of wheat. *Plant Genome* **2015,** 8, (2).
48. Gonzalez-Camacho, J.M.; Ornella, L.; Perez-Rodriguez, P.; Gianola, D.; Dreisigacker, S.; Crossa, J. Applications of machine learning methods to genomic selection in breeding wheat for rust resistance. *Plant Genome* **2018,** 11, (2).
49. Liabeuf, D.; Sim, S.C.; Francis, D.M. Comparison of marker-based genomic estimated breeding values and phenotypic evaluation for selection of bacterial spot resistance in tomato. *Phytopathology* **2018,** 108, (3), 392-401.
50. Ornella, L.; Singh, S.; Perez, P.; Burgueño, J.; Singh, R.; Tapia, E.; Bhavani, S.; Dreisigacker, S.; Braun, H.-J.; Mathews, K., *et al.* Genomic prediction of genetic values for resistance to wheat rusts. *Plant Genome* **2012,** 5, (3).
51. Endelman, J.B. Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome* **2011,** 4, (3), 250-255.
52. Piepho, H.P. Ridge regression and extensions for genomewide selection in maize. *Crop Sci.* **2009,** 49, (4), 1165-1176.
53. Browning, S.R.; Browning, B.L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* **2007,** 81, (5), 1084-97.
54. You, F.M.; Xiao, J.; Li, P.; Jia, G.; Yao, Z.; He, L.; Rashid, K.Y.; Duguid, D.S.; Booker, H.M.; Wang, X., *et al.* A first-generation haplotype map facilitates QTL identification and genomic prediction for complex traits in flax *Preprints* **2018**.
55. de los Campos, G.; Naya, H.; Gianola, D.; Crossa, J.; Legarra, A.; Manfredi, E.; Weigel, K.; Cotes, J.M. Predicting quantitative traits with regression models for dense molecular markers and pedigree. *Genetics* **2009,** 182, (1), 375-85.
56. de Los Campos, G.; Hickey, J.M.; Pong-Wong, R.; Daetwyler, H.D.; Calus, M.P. Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics* **2013,** 193, (2), 327-45.
57. de Los Campos, G.; Perez, P.; Vazquez, A.I.; Crossa, J. Genome-enabled prediction using the BLR (Bayesian Linear Regression) R-package. *Methods Mol. Biol.* **2013,** 1019, 299-320.
58. Yang, J.; Benyamin, B.; McEvoy, B.P.; Gordon, S.; Henders, A.K.; Nyholt, D.R.; Madden, P.A.; Heath, A.C.; Martin, N.G.; Montgomery, G.W., *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **2010,** 42, (7), 565-9.
59. Yang, J.; Lee, S.H.; Goddard, M.E.; Visscher, P.M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **2011,** 88, (1), 76-82.
60. You, F.M.; Xiao, J.; Li, P.; Yao, Z.; Jia, G.; He, L.; Kumar, S.; Soto-Cerda, B.; Duguid, S.D.; Booker, H.M., *et al.* Genome-wide association study and selection signatures detect genomic regions associated with seed yield and oil quality in flax. *Int. J. Mol. Sci.* **2018,** 19, (8), 2303.