

# Genomic Selection for Late Blight and Common Scab Resistance in Tetraploid Potato (*Solanum tuberosum*)

Felix Enciso-Rodriguez,<sup>\*,1</sup> David Douches,<sup>\*,1</sup> Marco Lopez-Cruz,<sup>\*,1</sup> Joseph Coombs,<sup>\*,1</sup> and Gustavo de los Campos<sup>†,‡,§,1,2</sup>

<sup>\*</sup>Department of Plant, Soil, and Microbial Sciences, <sup>†</sup>Department of Epidemiology & Biostatistics, <sup>‡</sup>Department of Statistics & Probability, and <sup>§</sup>Institute for Quantitative Health Science and Engineering, Michigan State University, East Lansing, Michigan, 48824

ORCID ID: 0000-0002-3490-6586 (F.E.-R.)

**ABSTRACT** Potato (*Solanum tuberosum*) is a staple food crop and is considered one of the main sources of carbohydrates worldwide. Late blight (*Phytophthora infestans*) and common scab (*Streptomyces scabies*) are two of the primary production constraints faced by potato farming. Previous studies have identified a few resistance genes for both late blight and common scab; however, these genes explain only a limited fraction of the heritability of these diseases. Genomic selection has been demonstrated to be an effective methodology for breeding value prediction in many major crops (e.g., maize and wheat). However, the technology has received little attention in potato breeding. We present the first genomic selection study involving late blight and common scab in tetraploid potato. Our data involves 4,110 (Single Nucleotide Polymorphisms, SNPs) and phenotypic field evaluations for late blight (n=1,763) and common scab (n=3,885) collected in seven and nine years, respectively. We report moderately high genomic heritability estimates ( $0.46 \pm 0.04$  and  $0.45 \pm 0.017$ , for late blight and common scab, respectively). The extent of genotype-by-year interaction was high for late blight and low for common scab. Our assessment of prediction accuracy demonstrates the applicability of genomic prediction for tetraploid potato breeding. For both traits, we found that more than 90% of the genetic variance could be captured with an additive model. For common scab, the highest prediction accuracy was achieved using an additive model. For late blight, small but statistically significant gains in prediction accuracy were achieved using a model that accounted for both additive and dominance effects. Using whole-genome regression models we identified SNPs located in previously reported hotspots regions for late blight, on genes associated with systemic disease resistance responses, and a new locus located in a WRKY transcription factor for common scab.

## KEYWORDS

Potato  
Late blight  
Common scab  
Disease  
resistance  
Genome-Wide  
Regression  
Genomic  
Selection  
Polyploidy  
Bayesian  
BGLR

The potato (*Solanum tuberosum* L.) is considered the sixth most important agricultural commodity worldwide after sugar cane, maize, rice,

wheat and milk. In 2014, the global production of potatoes exceeded 385 million tons, positioning China as the largest producer with more than 66 million tons, followed by Russia, India and the United States (FAOSTAT 2016). As a staple food, this crop represents one of the main sources of carbohydrates, fiber, minerals and vitamins, providing essential nutrients and energy needed for healthy body development and function (Kolasa 1993; Drewnowski and Rehm 2013).

Despite its great economic and food security importance, potatoes face high production losses caused mainly by biotic factors. Among them, pathogens such as late blight (*Phytophthora infestans* (Mont.) de Bary), represent the most devastating disease for potato worldwide. Late blight infects vegetative tissues, typically killing the entire plant, within 7 to 10 days. This pathogen accounts for annual losses of 16% of the global potato production (Haverkort *et al.*, 2009). Under increasingly variable

Copyright © 2018 Enciso-Rodriguez *et al.*

doi: <https://doi.org/10.1534/g3.118.200273>

Manuscript received March 27, 2018; accepted for publication May 19, 2018; published Early Online May 24, 2018.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Supplemental material available at Figshare: <https://doi.org/10.25387/g3.6262214>.

<sup>1</sup>All authors contributed equally.

<sup>2</sup>Corresponding author: Gustavo de los Campos. 775, Woodlot Dr, East Lansing, MI 48823. [gdeloscampos@epi.msu.edu](mailto:gdeloscampos@epi.msu.edu)

weather conditions, late blight incidence is expected to escalate worldwide, affecting mainly highlands in developing countries (Sparks *et al.* 2014).

Soil-borne pathogens such as common scab (*Streptomyces scabies* Thaxter), reduces the potato quality and marketability by causing superficial lesions on the tuber surface (Dees and Wanner 2012). Susceptibility to common scab is dependent upon genotype, time and environmental conditions (Wanner 2006; Wanner and Kirk 2015), having a negative impact mainly in underground tissues in development, such as stolons and tubers. This pathogen has spread worldwide and due to its saprophyte nature, (common scab can survive in winter), thus becoming a permanent source of inoculum for the next planting seasons, causing losses up to \$100/Ha (Wanner and Kirk 2015).

Pathogen infection can be controlled by using protectants or systematic fungicides; however, there methods can be ineffective if the environmental conditions favor pathogen dispersion (Nowicki *et al.* 2011) or the emergence of fungicide-resistant genotypes (Pomerantz *et al.* 2014). The most effective way to control the incidence of late blight and common scab in potatoes is through the generation of resistant varieties (Ahn and Park 2013). However, breeding for resistant varieties via phenotypic selection can take up to 15 years, making traditional breeding time-consuming and sometimes ineffective against fast-evolving pathogens (The Potato Genome Sequencing Consortium 2011; Lozano *et al.* 2012).

Marker-assisted (Barone 2004) and genomic selection (GS) strategies (Meuwissen *et al.* 2001) can accelerate the process of breeding disease resistance. Several studies on late blight and common scab resistance have reported variants conferring resistance to these pathogens; however, most of the genomic research has focused on late blight (Gebhardt *et al.* 2004; Malosetti *et al.* 2007; Muktar *et al.* 2015; Mosquera *et al.* 2016; Braun *et al.* 2017a) and are largely based on phenotype-single marker association analyses. To the best of our knowledge, no study so far has considered the use of GS for breeding resistance to late blight and common scab in potato. Therefore, in this article, we use Whole-Genome Regression methods commonly used in GS to: (i) study important features of the genetic architecture of resistance to late blight and common scab (including trait heritability, extent of genetic-by-environment interactions (G×E) and the importance of non-additive effects), (ii) identify large-effect variants contributing to resistance to late blight and scab, and (iii) assess the prediction accuracy of GS for resistance to those two pathogens.

Our data involves (up to) nine years of field evaluations for late blight and common scab at two Michigan State University's (MSU) research centers. We considered models that accounted for additive effects and various forms of dominance and evaluated two different statistical methods. Our results suggest that sizable fraction of the inter-individual differences in disease resistance (~46% for late blight and 45% for common scab) can be captured by the SNP set used in the study. The extent of G×E was low for common scab and high for late blight. We found that additive models can capture more than 90% of the genetic variance. We report large-effect SNPs contributing to late blight resistance in chromosomes V and IX, that have been previously reported to harbor resistance genes to this pathogen. We also report the first SNP associated with common scab resistance, located on chromosome IX, and positioned in a transcription factor known for its role in systemic defense and resistance responses. Our results demonstrate that genomic selection can yield moderately accurate prediction of disease resistance for genotypes that have been not evaluated in field trials. Thus, GS could be used for rapid cycling selection for resistance to both late blight and common scab in tetraploid potato.

## MATERIALS AND METHODS

### Data

Data were collected from early generation and advanced tetraploid potato genotypes derived from bi-parental crosses at the MSU potato breeding program. Additional advanced breeding genotypes from other United States breeding programs and reference varieties were also included. The available genotypes (n = 381) represent different market classes for fresh market, chip-processing, and russet-type fresh market and processing varieties. These genotypes were evaluated in field trials that included annual selections from MSU's potato breeding program, where each year poorly performing genotypes were replaced with new genotypes, while maintaining control genotypes during consecutive years.

**Late blight field resistance trials** (273 genotypes and a total of 1,763 disease records) were conducted in inoculated foliar field trials during seven years (2010-2015 and 2017) at the MSU's Clarksville Research Center (Clarksville, MI). Potato seed tubers were hand planted early- to mid-June as four-plant hills in 1.5 m plots in a randomized complete block design with one to three replicates. Late blight-susceptible rows and plots were planted around the perimeter and between blocks to promote an even late blight distribution in the field. After approximately 60 days, all plots were inoculated with a zoospore suspension of late blight at  $3 \times 10^6$  spores/mL at the end of July or beginning of August of each year. Over the 7-year period, different isolates were used to infect the trial depending on the prominent isolate in the region. The *P. infestans* strain (clonal lineage) detected in each year in the trial can be found in Table S1 in File S4. Following inoculation, plots were rated visually for the percentage of foliar area affected by late blight. Ratings were taken at 3 to 7-day intervals, based on the rate of disease progression during 35-50 days post inoculation (DPI). Finally, the percent defoliation data were used to calculate the relative area under the disease progress curve - RAUDPC (Fry 1978).

**Common scab field resistance trials** (370 genotypes and a total of 3,885 disease records) were conducted under field conditions during nine years (2009-2017) in a disease nursery at the MSU's Montcalm Research Center (Lakeview, MI). The field was inoculated with common scab from aggressive Michigan isolates, and has been cultivated for high disease pressure for the past nine years. The trials were planted in a randomized complete block design consisting of one to four replications of five-hill plots. After harvesting, mature tubers in plots were assessed for their overall plot disease rating scale of 0-5. The rating was based on a combined score for common scab coverage and lesion severity in which a rating of 0 indicates zero infection and 3.0 or greater scores represent highly susceptible genotypes with >50% infection and severe pitted lesions (Driscoll *et al.* 2009).

**SNP genotypes** were obtained using the Infinium 8303 Potato Array. Plant DNA was isolated from young potato leaves or tubers using the Qiagen DNeasy Plant Mini Kit (Qiagen, Germany), following manufacturer's instructions. DNA was quantified using the Quant-iT PicoGreen dsDNA Assay kit (Invitrogen, San Diego, CA). Genotype scoring was performed using the GenomeStudio software (Illumina, San Diego, CA). The tetraploid SNP calling was performed as per Hirsch *et al.* (2013), using a custom tetraploid genotype calling based on theta values from the Illumina GenomeStudio (Illumina, San Diego, CA) and subsequently filtered, removing poor quality markers. SNPs were coded by counting the number of copies of a reference allele (*e.g.*, B) where 0 denotes fully homozygous allele (AAAA), 1-3 represent heterozygous genotypes (AAAB, AABBB, ABBBB, respectively) and 4 the other homozygous genotype (BBBB). The genotype file was filtered by retaining SNPs with minor allele frequency (MAF) >0.05 and missing rate <0.15. The

remaining missing SNP-based genotypes were imputed with the SNP means. The final number of SNPs that passed the quality filtering were 4,110.

We compared the observed and expected rates of heterozygous loci, the later derived under the assumption of Hardy-Weinberg (HW) equilibrium. Averaged across loci, the observed rates of heterozygosity (0.663) was only slightly higher than the one predicted from estimated allele frequencies (0.647). The regression of the observed and expected frequency of heterozygous loci had an estimated slope of 0.98 (SE = 0.0025) and a  $R^2$  of 0.974. Moreover, we did not identify any clear outlier SNP that may have indicated a significant deviation of the observed frequency of heterozygous relative to the one predicted from the estimated allele frequency of the locus.

**Genomic relationships** were computed from centered and scaled SNP-based genotypes according to VanRaden (2008):  $GRM = \frac{\tilde{X}\tilde{X}'}{ncol(\tilde{X})}$ . Here,  $GRM$  is a matrix describing genomic relationships between genotypes,  $\tilde{X} = \{[X_{im} - mean(X_{im})]/sd(X_{im})\}$  is a matrix of centered and scaled SNP-based genotypes ( $X_{im} \in \{0, 1, 2, 3, 4\}$  counts the number of copies of the reference allele at the  $m^{th}$  loci. Subtracting the  $mean(X_{im})$ , centers the SNP-based genotypes to a null mean and dividing by the SNP standard deviation,  $sd(X_{im})$ , scale SNP-based genotypes to unit variance). Finally, division by the number of SNP-based genotypes,  $ncol(\tilde{X})$ , makes the average diagonal value of  $GRM$  equal to one. We use this matrix to quantify genomic relationships and to derive principal components, the later were computed by applying the *eigen()* R-function to  $GRM$ .

## Statistical analyses

We use whole-genome regression models (Meuwissen *et al.* 2001; de los Campos *et al.* 2013) for estimation of marker effects and variance component analyses and for assessment of prediction accuracy. The general form of the statistical model used was as follows:

$$y_{ijk} = \mu + \sum_{h=1}^5 PC_{hi}\gamma_h + b_j + g_i + ge_{ij} + \varepsilon_{ijk} \quad [1]$$

where  $y_{ijk}$  is a phenotypic score (for either late blight or common scab) of the  $k^{th}$  replicate of the  $i^{th}$  genotype collected in year  $j$ ,  $\mu$  is an intercept,  $\sum_{h=1}^5 PC_{hi}\gamma_h$  is a regression on the first five SNP-derived principal components,  $b_j$  are year effects,  $g_i$  is the main effect of the  $i^{th}$  genotype (alternative specifications of this effect are discussed below),  $ge_{ij}$  represents a genotype-by-year interaction and  $\varepsilon_{ijk}$  are error terms, which were treated as normal, independently and identically distributed (*iid*) with year-specific variances, that is  $\varepsilon_{ijk} \sim N(0, \sigma_j^2)$ .

Year had seven levels for late blight and nine levels for common scab (2009, 2010,..., 2017) and was treated as a random effect. Genetic and genetic-by-year interactions were also modeled as random effects. We considered four specifications for modeling the main effect of genotypes:

- **Genotype effect.** In this specification we assumed that the main effects of the genotypes where *iid* draws from normal distributions  $g_i \sim N(0, \sigma_g^2)$ . In this specification, no genetic information (SNPs) was used and no assumptions about gene action (additive, dominance, epistasis) were made. This specification was used as a baseline for a model that could be fitted without having genomic information. The other three specifications included genotypes as inputs.
- **Additive model (A).** Here, the main effect of the genotype was represented using a linear combination of the marker genotypes, that is  $g_i = \sum_{m=1}^{4110} \tilde{X}_{im}\alpha_m$  where  $\tilde{X}_{im} = [X_{im} - mean(X_{im})]/sd(X_{im})$  were centered and scaled genotypes code at the  $m^{th}$  SNP in the  $i^{th}$  genotype and  $\alpha_m$  is the additive effect of the markers.

- **Additive + Dominance (A+D).** In this case, the main effects of genotypes have an additive component plus one that accounted for dominance; therefore in this model  $g_i = \sum_{m=1}^{4110} \tilde{X}_{im}\alpha_m + \sum_{m=1}^{4110} \tilde{D}_{im}d_m$  where  $\tilde{D}_{im} = [D_{im} - mean(D_{im})]/sd(D_{im})$  are (centered and standardized) dummy variables for heterozygous loci, here  $D_{im}=1$  (=0) indicates that the  $m^{th}$  SNP of the  $i^{th}$  genotype was in heterozygous (homozygous) state and  $d_m$  is the dominant effect of the markers.
- **General model (G).** Here,  $g_i = \sum_{m=1}^{4110} \sum_{n=0}^4 \tilde{W}_{imn}\gamma_{mn}$ , where  $\tilde{W}_{imn}$  are (centered and standardized) dummy variables for genotypes carrying  $n$  copies of the reference allele and  $\gamma_{mn}$  is the general effect of the markers. Since there are up to five distinct genotypes (0,1,...,4) this model includes up to four degree of freedom per locus. This parameterization allows for any form of interactions of alleles within locus; thus, it can be considered the most general specification for a model accounting for additive and dominance effects.

## Prior distributions for effects

**Marker effects** (including both additive, dominance and those of the G model) were treated as random. We considered two prior distributions of effects: (i) treating SNP effects as draws from normal distributions with null mean and model-specific variances (*i.e.*, there were separate variances for additive and dominance), this approach was implemented using the Bayesian Ridge Regression (BRR) specification in the Bayesian Generalized Linear Regression (BGLR) R-package (Pérez and de los Campos 2014), and (ii) a Bayesian shrinkage-variable selection method (BayesB, Meuwissen *et al.* 2001). As with BRR, in BayesB different regularization parameters (probabilities of non-null effects and scale parameters) were assigned to effects in additive and dominance. BayesB was implemented in BGLR using the “BayesB” keyword for the model argument of the linear predictor.

**Genotype-by-year effects** ( $ge_{ij}$ ) were treated as IID normal with mean zero and variance common to all the interactions, that is,  $ge_{ij} \sim N(0, \sigma_{ge}^2)$ .

**Sequence of models:** Using the specifications described above, we produced a sequence of models designed to quantify the amount of variance explained (and the contribution to prediction accuracy) of each of the terms entering in the model of expression [1]. The sequence of models considered is summarized in Table 1.

We used the whole-genome regression models described above for three purposes: (1) estimation of variance components, (2) identification of variants with high contribution to additive variance and (3) assessment of prediction accuracy in cross-validation.

**Variance components:** The amount of variance accounted for by each of the terms included in the model was estimated using the methods described in de los Campos *et al.* (2015) and Lehermeier *et al.* (2017) (see Supplementary File S1 for further details). We use these methods to decompose the total phenotypic variance into components due to year, genetics factors, genotype-by-year interactions (G×E) and within-year error variance. We also use this approach to assess the relative contribution of SNP-additive and dominance effects.

**Identification of SNPs With a sizable contribution to additive variance:** Response to selection is directly proportional to additive variance (Falconer and Mackay 1996). Thus, in GS, the single-loci additive variance represents a natural metric to assess the relative importance of individual loci from a breeding perspective. Under linkage equilibrium, the contribution of each locus to additive variance is given by  $Var(X_{im}\alpha_m) = Var(X_{im})\alpha_m^2$ . In our case, genotypes were standardized

■ Table 1 Sequence of models

Model # (label) <sup>a</sup>	Effects Included							
	Year	Genotype <sup>b</sup>	PC <sup>c</sup>	Additive <sup>d</sup>	Dominance <sup>e</sup>	General <sup>f</sup>	Genotype-by-Year <sup>g</sup>	Error
M1	×							×
M2	×	×						×
M3	×	×					×	×
M4	×	×	×				×	×
M5 (A)	×		×	×			×	×
M6 (A+D)	×		×	×	×		×	×
M7 (G)	×		×			×	×	×

<sup>a</sup>M1-M7 are model numbers. <sup>b</sup>Random effect of the genotype (no SNPs used, no assumption about gene action are made).

<sup>c</sup>Principal components, <sup>d</sup>linear regression on allele content (0/1/2/3/4), <sup>e</sup>Simple dominance (1 degree of freedom per locus representing heterozygous) and <sup>f</sup>General model for additive + dominance (with up to 4 degrees of freedom per locus). <sup>g</sup>Genotype-by-year interaction. An 'x' indicates that the effects was included in the model.

to unit variance; therefore,  $\text{Var}(X_{im}\alpha_m) = \alpha_m^2$ . We used samples from the posterior distribution of SNP effects from the A model to assess the contribution of individual loci to additive variance. (Further details about how these quantities were computed can be found in the script provided with the Supplementary File S1). It is important to note that this measure does not account for linkage disequilibrium; thus it can only be taken as a proxy of the contribution of a locus to additive variance.

**Prediction accuracy evaluation:** We implemented two cross-validation schemes. First, we used a fivefold cross-validation, assigning genotypes to folds. When using this approach all the phenotypic records of a genotype are assigned to either training or testing populations. Thus, this approach yields an estimate of the prediction accuracy that can be achieved predicting the performance of genotypes that have not been evaluated in field trials (*i.e.*, prediction based on genotype data only) and is equivalent to the method labeled as Cross-Validation one (CV1) in Burgueño *et al.* (2012). For this scheme, genotypes were assigned to folds completely at random and the fivefold Cross Validation (CV) was repeated 100 times to obtain accurate estimates of the average prediction correlation and its standard deviation.

In a second prediction scheme (CV2), we assigned years to folds (*i.e.*, there as many folds as years). Thus, when analyzing the  $j^{\text{th}}$  fold, data from the  $j^{\text{th}}$  year was assigned to testing and data from all the other years was used for training. This CV approach yields an estimate of the prediction accuracy that can be achieved when attempting to predict future year performance based on past data. Note that in this case, unlike CV1, when predicting data for the  $j^{\text{th}}$  genotype on the  $j^{\text{th}}$  year all the data from the  $i^{\text{th}}$  genotype collected in other years was part of the training dataset.

In both CV schemes prediction accuracy was evaluated by computing the within-year correlation between phenotypes and CV predictions.

**Software:** All the analyses were conducted using R (The R Development Core Team 2010). Models were fit using the BGLR-R package. For each model, we ran the Gibbs sampler algorithm for a total of 33,000 cycles, discarding the first 3,000 samples for burn-in; one of every five of the remaining samples was saved and used to estimate posterior means and standard deviations.

## Data availability

Scripts demonstrating how each of the models were implemented in BGLR are given in File S1. The genotype and phenotype data are provided in Files S2 and S3 for late blight and common scab, respectively. File S4 contains Tables S1-S5. Supplemental material is available at Figshare: <https://doi.org/10.6084/m9.figshare.6336911> and <https://doi.org/10.25387/g3.6262214>.

## RESULTS

The distribution of late blight and common scab infection varied substantially between years (Figure 1). In general, RAUDPC median values decreased from 2010 to 2012, with US-22 as the prevalent late blight strain on infected plants. In subsequent years, a differential response for late blight resistance was observed when US-23 was the prevalent strain. Disease pressure changes, together with the environment fluctuations between years contribute to explain the phenotypic variation observed for the late blight resistance response. Similarly, for common scab, a reduced frequency of resistant genotypes (0-1 score) was observed from 2009 until 2013, having at the same time an increasing number of intermediate susceptible genotypes (2-3 scores). Since 2013 and until 2017, an increased frequency of common scab resistant genotypes was observed (Figure 1).

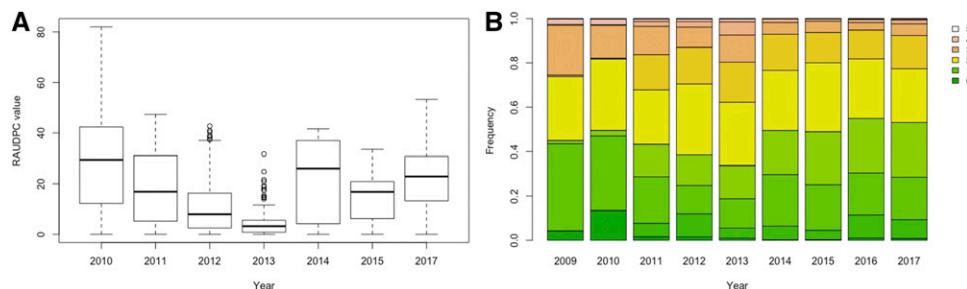
A principal component (PC) analysis showed that potato genotypes clustered in two groups, one involving 391 genotypes, and a small one including 22 genotypes (Figure 2). The eigenvalues associated to the first two PCs explained about 8% of the total genotypes variance (Figure 2). A cluster analysis using a correlation matrix derived from SNP markers supports the PC-analysis results (Figure 3). The heatmap also reveals that the strength of genomic relationships among the different materials is relatively small (the clear majority of the genotypes have genomic relationships with other genotypes smaller than 0.1, with only a few genotypes showing relationships comparable to parent-offspring or full-sib relations, *i.e.*, 0.5, Figure 3).

## Variance Components Estimates

The variance components analyses for late blight resistance (Table 2 and Table S2 in File S4) revealed that year explained roughly 25% of the variance in disease scores. For this trait, and taking as a reference the model M3, the main effect of genotype explained about 34% of the variance, genotype-by-year interactions explained 25% of the variance and the error term explained roughly 14% of the variance in late blight scores. These results suggest that a substantial proportion of within-year variance in late blight scores (roughly 70%, computed as  $0.34/[0.34+0.144]$ ) can be explained by main effects of genotypes. For *late blight*, the amount of genetic variance captured by the A model was roughly 94% of the variance captured by the G model (computed as  $0.330/0.352$ ).

For common scab (Table 2 and Table S2 in File S4) the main effect of genotype explained about 44% of the total variance, year and genotype-by-year effects explained only 3% and 6% of the total variance, respectively, and the error term accounted for almost one half (48%) of the variance in disease scores. For common scab we also observed that the amount of genetic variance captured by the A model was very similar to the one captured with the G model.





**Figure 1** Boxplot of late blight scores (A, relative area under the disease progress curve- RAUDPC) and bar plot for common scab scores (B, 0-5 rating scale).

The proportion of the total genetic variance that could be attributed to the first-5 PCs was substantial for late blight ( $\sim 30\%$ , computed as  $0.107/0.352$ ) and low for common scab ( $\sim 10\%$ ,  $0.051/0.451$ ).

For the A model (fitted using BayesB), we computed single-locus variances and used these estimates as proxies for the SNP relevance (Figures 4a and 4b). Additionally, we report in Figures S1 and S2, linkage disequilibrium (LD) plots for the 10 leading SNPs (*i.e.*, those with the larger single-SNP variance) for each trait. For both pathogens, there were a few regions with large single-SNP-variance. Specifically, for late blight, there were multiple SNPs distributed across the potato chromosomes (Figure 4a and Table S3 in File S4) with a sizable contribution to variance, suggesting that multiple genes contribute to the resistant phenotype. Conversely, for common scab, there was one SNP, located in chromosome IX (Figure 4b and Table S4 in File S4), that stands out for its contribution to variance and a few SNPs with a moderate contribution to phenotypic variance.

The results from the first cross-validation analysis (CV1) yielded an estimated prediction correlation of about 0.31 for late blight resistance using the G model. For this trait, there was a relatively small, albeit significant, increase in prediction correlation for the G model relative to the A model. Likewise, there was a slight superiority of BayesB over BRR (Table 3). In the case of common scab, the A model (with a prediction correlation of  $\sim 0.27$ ) outperformed the A+D (correlation  $\sim 0.26$ ) and G (correlation  $\sim 0.22$ ) models. These results agree with the variance component analyses results, where we also found evidence of a slightly higher relevance of non-additive effects in the case of late blight, compared to common scab.

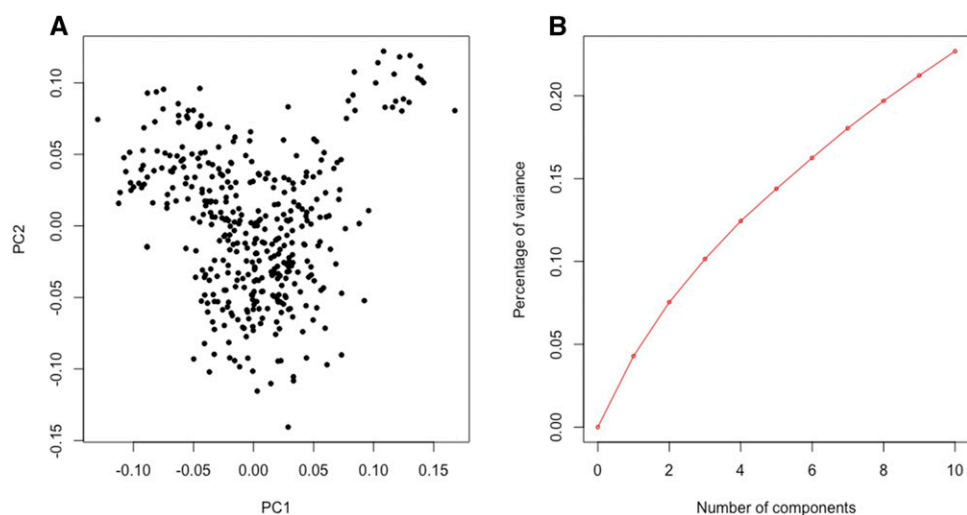
Note that in Table 3 we only included results from models using genotypes. Results from other models (*e.g.*, M2 and M3) are not presented because in CV1 they render zero within-year correlation. This happens because in CV1 predictions are entirely depending on borrowing of information between genotypes, a feature that is not possible in models that do not use genotype or pedigree information.

The results from the second cross-validation (*i.e.*, where years were assigned to folds, CV2) yielded higher estimates of prediction accuracy than those obtained in CV1 (Table 4 and Table S5 in File S4). This happens because in CV2 there is within-genotype borrowing of information across years. For late blight, prediction correlations ranged from 0.41 to 0.74, depending on the model and year. Likewise, for common scab, we obtained correlations ranging from 0.46 to 0.76. For both traits, the across-year average correlations showed small differences between models (with a slight superiority in favor of the G model).

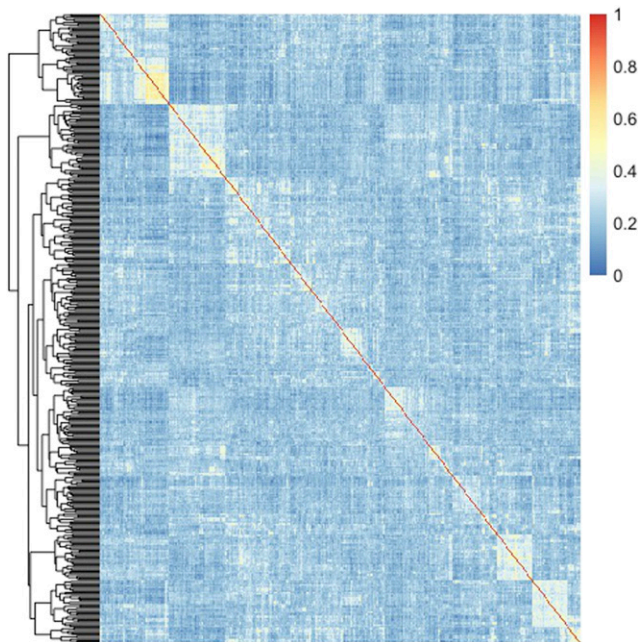
## DISCUSSION

Genomic selection has been quickly adopted for breeding in diploid species (Heffner *et al.* 2009; Daetwyler *et al.* 2013; de los Campos *et al.* 2013). However, the volume of research and the adoption of the GS technology for breeding of polyploid species has been much more limited (*e.g.*, Habyarimana *et al.* 2017; Sverrisdóttir *et al.* 2017). In this study, we demonstrate how genomic models commonly used in GS of diploid organisms can be applied for the analysis and prediction of disease susceptibility in autotetraploid potato.

Our results indicate that a sizable fraction of the within-year inter-individual differences in disease resistance (about  $0.46 \pm 0.04$  for late blight and  $0.45 \pm 0.02$  for common scab) can be explained using 4,110



**Figure 2** Principal component analysis of the Michigan State University's potato breeding genotypes derived from 4,110 SNPs: loadings on the first two marker-derived principal components (A) and proportion of variance explained by the top 10 principal components (B).



**Figure 3** Heatmap of the genomic relationship matrix (GRM) from the Michigan State University's potato breeding genotypes.

codominant SNPs from the Infinium 8303 Potato Array used in this study. These moderately high genomic heritability estimates for complex disease phenotypes indicates that, in principle, genomic prediction could be used successfully to select for resistance to late blight and common scab.

Previous studies have reported heritability estimates for these traits; however, differences in the nature of the genetic materials (diploid vs. tetraploid, hybrids vs. genotypes) and of the environmental conditions (natural vs. induced infection) makes the comparisons across studies difficult (Nelson 1978; Braun *et al.* 2017b). For instance, Haynes and Christ (1999) reported much higher heritability estimates for late blight resistance (0.8), but this study was based on diploid hybrids. For the same trait, estimates of heritability obtained using tetraploid genotypes are closer to the ones reported here (ranging from 0.31 to 0.69, Pajeroska-Mukhtar *et al.* 2009; Solano *et al.* 2014).

For common scab, previous heritability estimates are also highly variable, depending on the genetic material and the environmental conditions. For instance, using diploid potatoes derived from a cross between wild relatives (*S. phureja* × *S. stenotomum*) and cultivated potatoes (di-haploid *S. tuberosum* × *S. chacoense*), Haynes *et al.* (2009) and Braun *et al.* (2017b) reported broad sense heritability estimates ranging from 0.18 to 0.72 for different environments. However, studies involving tetraploid potatoes have reported higher heritability estimates with values ranging from 0.32 to 0.93 (Haynes *et al.* 1997; Bradshaw *et al.* 2008; Tai *et al.* 2009). More recently, 18 dedicated common scab and standard breeding program trials were conducted in fields with high disease pressure. The broad sense heritability estimates reported from these studies ranged from 0.75 to 0.90 for dedicated common scab trials and from 0.06 to 0.82 for standard breeding programs trials involving advanced breeding materials (Navarro *et al.* 2015).

The amount of variance in disease resistance that could be attributed to genotype-by-year interactions was high for late blight and very small for common scab. These differences are likely to be due to the different nature and characteristics of infection on the fields used to evaluate late

blight and common scab. Specifically, for late blight, the mean scores varied substantially between years (e.g., it was clearly low in 2013) reflecting changes on the late blight aggressiveness and late blight genotypes present in different years, resulting in a large extent of genotype-by-year interactions for this pathogen. On the other hand, our common scab data were generated in a nursery that has been used to evaluate common scab resistance in potato breeding genotypes for several years. Consequently, there was less variability in the mean scores across years and therefore we observed substantially less extent of G×E. A similar result was reported under comparable conditions by Murphy *et al.* (1995). Results based on fields trials performed in different locations for this pathogen have shown much higher variability over the years (Haynes *et al.* 2009).

The comparison of the genomic variance estimates obtained with the A model and those obtained with the G model suggest that, for both pathogens, a sizable fraction of the total genetic variance (0.94 and 0.98, for late blight and common scab, respectively) can be captured by an additive model (Table 2 and Table S2 in File S4). The amount of genetic variance captured by the A model reflects an estimate of the variance that can be captured by regression on allele content (*i.e.*, by allele substitution effects). However, when dominance is included in the model (A+D), the estimated 'additive variance' no longer represents the variance explained by allele substitution effects; therefore, the additive component in the A+D model is smaller than the additive component estimated with the A model.

While our variance component estimates indicate that most of the genetic variance can be captured by an A model, our cross-validation analysis suggests that accounting for non-additive effects could improve prediction accuracy by a small but statistically significant margin in the case of late blight. These results agree with the theory that suggests that dominance and epistasis are expected to contribute to the expression of traits subjected to directional selection or those affecting the plant fitness such as late blight resistance (Killick and Malcolmson 1973). This may explain why the G model captured slightly more variance and predicted slightly more accurately late blight scores than the A model.

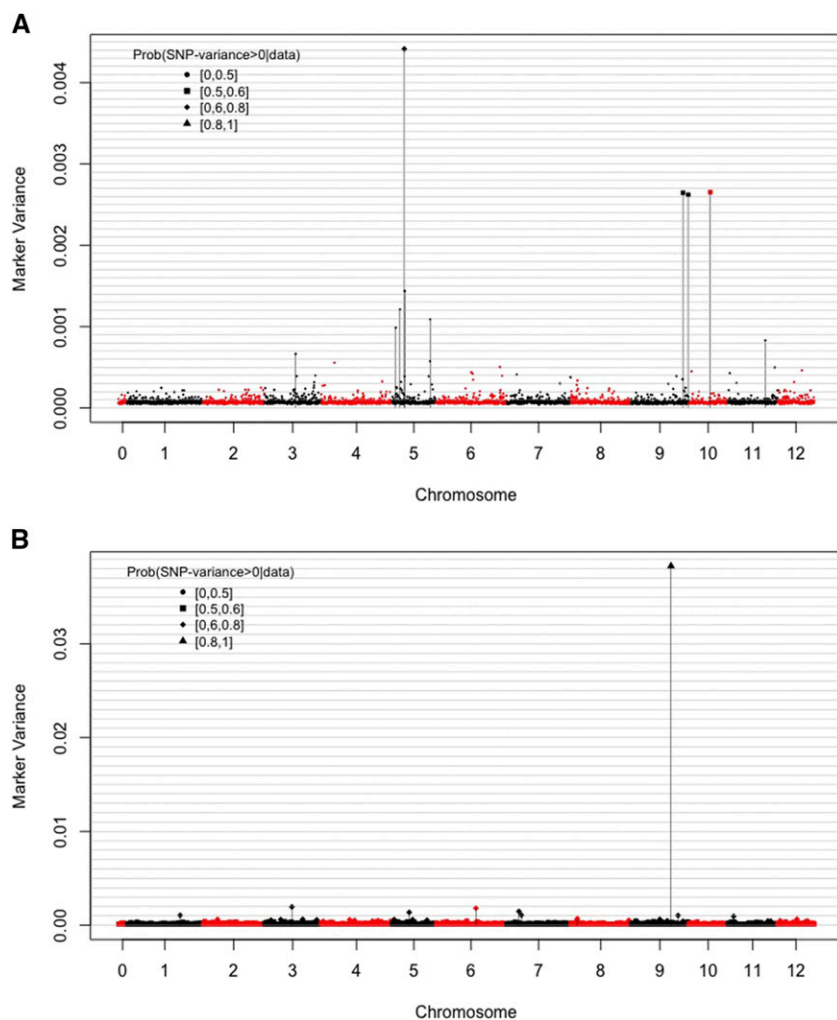
The presence of linkage disequilibrium (LD) between loci makes the partition of the total genetic variance into (orthogonal) locus-specific components not possible (de los Campos *et al.* 2015). However, it is worth looking at the relative size of estimated effects to explore features of the genetic architecture of the trait. We did this by inspecting the estimated SNPs variances (Figure 4). Overall, the proportion of variance explained by individual SNPs was low, reinforcing the idea that resistance to both common scab and late blight is polygenic. However, there were some SNPs with relatively large SNP-variances for both late blight (located mainly in chromosomes V and IX) and common scab (located in chromosome IX). (Figure 4, Table S3 and S4 in File S4). For late blight resistance, multiple quantitative trait loci (QTL) have been reported across the 12 potato chromosomes in tetraploid and diploid potato populations (Tiwari *et al.* 2013). Most of these major QTL are located in chromosomes III, IV, V, VII, XI and XII, characterized for harboring hotspot regions for resistance to late blight and other pathogens, not only for genes involved in quantitative resistance such as R genes, but also for genes involved in qualitative resistance (Malosetti *et al.* 2007; Pajeroska-Mukhtar *et al.* 2009; Álvarez *et al.* 2017). For instance, genes involved in carbohydrate metabolism such as sucrose synthase (Table S3 in File S4) play an active role in the defense response elicitation. Sucrose synthesis down-regulation has been described in the *Capsicum annuum* - *Phytophthora nicotianae* pathosystem, showing a decreasing concentration after challenging with beta-aminobutyric acid (BABA) and priming the synthesis of metabolites associated with the production of defense-related compounds (Stamler *et al.* 2015).

**Table 2** Variance components estimates (posterior standard deviation) derived from BayesB model for late blight and common scab resistance by model. Phenotypic scores were standardized to unit variance; hence estimates can be interpreted as the proportion of variance explained by each component. Results obtained with the fully Gaussian model (BRR) are presented in Table S2

Model # (label) <sup>a</sup>	Year	Genotype	Genetic					Genotype-by-year <sup>g</sup>	Error	
			Marker-derived			Total genetic <sup>f</sup>	Genotype-by-year <sup>g</sup>			
			PC <sup>b</sup>	Additive <sup>c</sup>	Dominance <sup>d</sup>					General <sup>e</sup>
M1	0.266 (0.021)								0.735 (0.025)	
M2	0.256 (0.014)	0.434 (0.018)					0.434 (0.018)		0.303 (0.011)	
M3	0.250 (0.027)	0.340 (0.031)					0.340 (0.031)	0.253 (0.019)	0.144 (0.006)	
M4	0.244 (0.026)	0.265 (0.028)	0.096 (0.024)				0.351 (0.032)	0.251 (0.018)	0.144 (0.006)	
M5 (A)	0.240 (0.027)		0.135 (0.069)	0.292 (0.051)			0.330 (0.035)	0.275 (0.020)	0.144 (0.006)	
M6 (A+D)	0.240 (0.027)		0.135 (0.062)	0.166 (0.063)	0.141 (0.049)		0.340 (0.034)	0.267 (0.020)	0.144 (0.006)	
M7 (G)	0.249 (0.028)		0.107 (0.043)			0.280 (0.034)	0.352 (0.033)	0.251 (0.019)	0.144 (0.006)	
Common Scab										
M1	0.033 (0.006)								0.971 (0.022)	
M2	0.030 (0.004)	0.456 (0.016)					0.456 (0.016)		0.523 (0.012)	
M3	0.029 (0.005)	0.440 (0.021)					0.440 (0.021)	0.059 (0.009)	0.483 (0.013)	
M4	0.029 (0.006)	0.419 (0.021)	0.030 (0.012)				0.447 (0.023)	0.056 (0.009)	0.485 (0.013)	
M5 (A)	0.031 (0.006)		0.132 (0.087)	0.507 (0.076)			0.443 (0.025)	0.061 (0.009)	0.485 (0.013)	
M6 (A+D)	0.031 (0.006)		0.107 (0.073)	0.356 (0.087)	0.151 (0.060)		0.448 (0.024)	0.057 (0.009)	0.485 (0.013)	
M7 (G)	0.031 (0.006)		0.051 (0.030)			0.442 (0.031)	0.451 (0.023)	0.056 (0.009)	0.484 (0.013)	

<sup>a</sup>M1-M7 are model numbers (label). The effects included in each of them are described in the columns.

<sup>b</sup>Principal components, <sup>c</sup>linear regression on allele content (0/1/2/3/4), <sup>d</sup>Simple dominance (1 degree of freedom per locus representing heterozygous) and <sup>e</sup>General model for additive + dominance (with up to 4 degrees of freedom per locus), <sup>f</sup>Total genetic variance, <sup>g</sup>Genotype-by-year interaction.



**Figure 4** Estimated SNP-variances derived from BayesB model using the additive model for late blight (A) and common scab (B). (In both cases, phenotypes were disease scores standardized to a variance equal to one. Vertical lines indicate the positions of the top-10, according to estimated SNP-variance markers).

Additionally, these results validate earlier QTL reports obtained from MSU-derived populations using potato varieties carrying different late blight resistance genes coming from species previously used in resistance breeding such as *S. demissum* and *S. berthaultii* (Massa *et al.* 2015; N. Manrique-Carpintero, personal communication).

For common scab resistance, our results suggest an additive resistance effect with a clear major-effect SNP located on chromosome IX. This SNP is associated with a WRKY transcription factor known for their role in the modulation of the resistance responses in systemic and acquired plant resistance, activating or repressing the transcription of genes involved in the synthesis of defense related-proteins such as R proteins (Pandey and Somssich 2009; Buscaill and Rivas 2014). In addition to the loci discussed above, we were also able to identify additional SNPs with a sizable contribution to variance across the potato chromosomes (Figure 4). Interestingly, the SNP in the WRKY gene that appeared to have a sizable contribution to inter-individual differences in common scab resistance is located in a region where LD is relatively weak (see Figure S2).

For instance, we found one SNP located in chromosome III (Table S4 in File S4) associated with the primary metabolism-related protein fructokinase, whose concentration increases under pathogen attack as a mechanism to reduce the costs attributed to the defense response in soil-borne pathogens (Zimaro *et al.* 2011). Likewise, in chromosome V, we found one SNP related to the RNA synthesis-related protein DEAD-box

ATP-dependent RNA helicase, reported for its role in plant resistance by enhancing the defense response in both necrotrophic and biotrophic pathogens (Li *et al.* 2008). Overall, the evidence we found support the hypothesis that resistance to common scab involves multiples mechanisms of defense including the activation of genes related to systemic and R gene-mediated resistance.

There are few studies reporting QTL for common scab resistance. For instance, two QTL located in chromosome XI were detected in a diploid parental-derived population for the percentage of surface area infected and lesion type caused by common scab, explaining 21% and 18.2% of the total phenotypic variance, respectively (Braun *et al.* 2017a). For tetraploid populations, Amplified Fragment Length Polymorphisms (AFLPs) and Simple Sequence repeats (SSRs) markers have been used to establish an association between potato genotypes and the common scab resistance phenotype in a tetraploid bi-parental derived-population. Two copies of a dominant allele were detected in a QTL localized in chromosome II, explaining 8.1 and 7.1% of the phenotypic variance, respectively. A second QTL was localized in chromosome VI explaining 6.9% of the total phenotypic variance (Bradshaw *et al.* 2008). Therefore, the large-variance SNP detected in this study represents a new genomic region associated with common scab resistance, providing a framework for the development of molecular markers for marker-assisted selection and understand the genetics behind common scab resistance.



■ Table 3 Cross-validation correlations obtained with BRR and BayesB models by trait and model

Prior <sup>a</sup>	Model # (label) <sup>b</sup>	CV-Correlation		Proportion of times that the model in row gave a higher correlation than the model in columns					
		Average <sup>c</sup>	SD <sup>d</sup>	BRR <sup>a</sup>			BayesB <sup>a</sup>		
				M5 (A)	M6 (A+D)	M7 (G)	M5 (A)	M6 (A+D)	M7 (G)
Late Blight									
BRR	M5 (A)	0.258	0.023		0.96	0.00	0.33	0.91	0.00
	M6 (A+D)	0.241	0.023	0.04		0.00	0.04	0.57	0.00
	M7 (G)	0.312	0.017	1.00	1.00		1.00	1.00	0.5
BayesB	M5 (A)	0.260	0.024	0.67	0.96	0.00		0.94	0.00
	M6 (A+D)	0.240	0.024	0.09	0.43	0.00	0.06		0.00
	M7 (G)	0.313	0.017	1.00	1.00	0.50	1.00	1.00	
Common Scab									
BRR	M5 (A)	0.268	0.025		0.81	0.99	0.1	0.55	1.00
	M6 (A+D)	0.259	0.023	0.19		0.99	0.07	0.20	0.99
	M7 (G)	0.218	0.022	0.01	0.01		0.00	0.02	0.63
BayesB	M5 (A)	0.278	0.026	0.9	0.93	1.00		0.91	1.00
	M6 (A+D)	0.265	0.025	0.45	0.8	0.98	0.09		0.98
	M7 (G)	0.216	0.022	0	0.01	0.37	0.00	0.02	

<sup>a</sup> BRR uses a Gaussian prior for effects, BayesB uses a prior that has a point of mass at zero and a scaled-t slab.

<sup>b</sup> A: Additive model, A+D: additive+dominance, G: general model (with up to 4 degrees of freedom per locus).

<sup>c</sup> Average from 100 cross-validations.

<sup>d</sup> Standard deviation.

Our variance component estimates suggest that for both, late blight and common scab, a sizable amount of inter-individual differences in disease resistance can be captured using whole-genome regressions. However, the successful implementation of GS requires being able to predict future outcomes from past data. We assessed this using two CV analyses. Our results are based on genotypes derived from potato breeding programs. Some of these genotypes are related through pedigrees and there is some level of population stratification. Therefore, the prediction accuracies reported in our study should be considered representative of the prediction accuracy that one may be able to achieve when applying GS to breeding populations.

We considered two different prediction problems and implemented different CV schemes to represent each prediction problem. Our first CV focused on the prediction of future scores from genotypes that were not evaluated in field trials (*i.e.*, prediction based on information from other genotypes). These analyses rendered moderately low CV-correlations ( $\sim 0.22$ - $0.31$  with some small differences between traits and models).

It is important to highlight that in CV1 the correlations reflect the prediction accuracy that can be achieved when predicting future phenotypes for genotypes that have not been evaluated in field trials. These predictions, although imperfect, could enable several rounds of rapid selection based on genotype data alone. The predictive correlation obtained in CV1 was about half of the correlation between phenotypes across years (compare results in Table 3 with those for M2 in Table 4). Thus, we conclude that with the array and sample size used in this study, the predictive accuracy for late blight and common scab scores obtained from a newly developed genotype that has been genotyped but not tested in the field is about half of the predictive power of a single phenotype record. If more than two selection cycles can be carried out per year, the reduction on generation interval that can be achieved with genomic prediction would overcome the lower accuracy and, eventually lead to faster yearly genetic gains.

Our second CV used years as folds; therefore, in this case, disease scores predictions for one-year data were obtained from the same genotypes over years. The results of the model based on year and genotype (M2), give a baseline estimates of the prediction accuracy that can be achieved with phenotypic prediction. In CV2, we obtained higher prediction correlations (0.56-0.61, Table 4) than with CV1. However, the performance of the genomic models was only slightly superior to predictions based on past phenotypes-only (*i.e.*, those that could be obtained with the M2 model). This result agrees with previous studies (*e.g.*, Crossa *et al.* 2010) that show that the benefits of genomic prediction are more important when predicting phenotypes of materials that have no (or very limited) data from previous field trials.

## Conclusions

We confirmed that a sizable fraction of inter-individual differences in late blight and common scab scores can be attributed to genetic factors and can be captured using whole-genome regressions. We found large genotype-by-year interactions for late blight and limited genotype-by-year interactions for common scab. For both late blight and common scab, we found that an additive model could account for a sizable ( $>90\%$ ) of the total genetic variance. However, for late blight, we found small (but statistically significant) gains in prediction accuracy when accounting for dominance. Our analyses confirm strong associations with disease resistance to SNPs in previously reported resistance hotspot regions for late blight and reported a novel locus that has a sizable contribution to common scab resistance. We demonstrated that prediction of disease resistance,

■ **Table 4** Year cross-validation correlations obtained with BayesB model by trait and model

Year	Model # (label) <sup>a</sup>				
	M2	M3	M5 (A)	M6 (A+D)	M7 (G)
<b>Late blight</b>					
2010	0.551	0.537	0.463	0.465	0.517
2011	0.652	0.658	0.608	0.611	0.642
2012	0.583	0.604	0.586	0.586	0.624
2013	0.422	0.415	0.484	0.485	0.492
2014	0.621	0.596	0.633	0.640	0.655
2015	0.719	0.730	0.678	0.696	0.745
2017	0.508	0.504	0.471	0.491	0.506
<b>Average</b>	0.579	0.578	0.560	0.568	0.597
<b>SD</b>	0.098	0.104	0.087	0.089	0.095
<b>Common scab</b>					
2009	0.459	0.460	0.472	0.471	0.466
2010	0.520	0.522	0.535	0.533	0.517
2011	0.610	0.611	0.622	0.625	0.618
2012	0.625	0.628	0.628	0.626	0.634
2013	0.750	0.759	0.731	0.737	0.750
2014	0.615	0.611	0.634	0.636	0.635
2015	0.649	0.653	0.666	0.671	0.659
2016	0.639	0.639	0.652	0.647	0.647
2017	0.508	0.510	0.519	0.520	0.515
<b>Average</b>	0.597	0.599	0.606	0.607	0.605
<b>SD</b>	0.088	0.090	0.082	0.083	0.089

<sup>a</sup>M2 includes year and genotype (no SNP information); M3: extends M2 with the addition of genotype-by-year interaction; M5 includes year, first 5 marker-derived PCs, additive effect of SNPs and genotype-by-year interaction; M6 expands M5 by adding the effects of dominance; M7 includes year, 5-PCs, genotype-by-year interactions and SNPs with up to 4 degrees of freedom per locus ('General' model).

using genomic prediction applied to autotetraploid potato, is feasible and can be implemented for SNP-based selection in potato breeding. Further research is needed to explore ways (larger sample size, more controlled environments, higher marker density) in which genomic prediction accuracy can be further improved.

## ACKNOWLEDGMENTS

The authors want to acknowledge valuable comments provided by the Editor and by two anonymous reviewers. FER received support from a Fulbright fellowship. GDLC received financial support from NSF grant 1444543. DD received financial support from USDA/ARS agreement 58-3655-4-042. The development and maintenance of BGLR has been supported by NIH-grant R01GM101219.

## LITERATURE CITED

Ahn, Y. K., and T.-H. Park, 2013 Resistance to common scab developed by somatic hybrids between *Solanum brevidens* and *Solanum tuberosum*. *Acta Agriculturae Scandinavica, Section B - Soil & Plant Sci.* 63: 595–603.

Álvarez, M. F., M. Angarita, M. C. Delgado, C. García, J. Jiménez-Gómez *et al.*, 2017 Identification of Novel Associations of Candidate Genes with Resistance to Late Blight in *Solanum tuberosum* Group Phureja. *Front. Plant Sci.* 8: 1040. <https://doi.org/10.3389/fpls.2017.01040>

Barone, A., 2004 Molecular Marker-assisted Selection for Potato Breeding. *Am. J. Potato Res.* 81: 111–117. <https://doi.org/10.1007/BF02853608>

Bradshaw, J. E., C. A. Hackett, B. Pande, R. Waugh, and G. J. Bryan, 2008 QTL mapping of yield, agronomic and quality traits in tetraploid potato (*Solanum tuberosum* subsp. *tuberosum*). *Theor. Appl. Genet.* 116: 193–211. <https://doi.org/10.1007/s00122-007-0659-1>

Braun, S. R., J. B. Endelman, K. G. Haynes, and S. H. Jansky, 2017a Quantitative trait loci for resistance to common scab and cold-induced sweetening in diploid potato. *Plant Genome* 10(3): 1–9. <https://doi.org/10.3835/plantgenome2016.10.0110>

Braun, S., A. Gevens, A. Charkowski, C. Allen, and S. Jansky, 2017b Potato Common Scab: a Review of the Causal Pathogens, Management Practices, Varietal Resistance Screening Methods, and Host Resistance. *Am. J. Potato Res.* 94: 283–296. <https://doi.org/10.1007/s12230-017-9575-3>

Burgueño, J., G. D. L. Campos, K. Weigel, and J. Crossa, 2012 Genomic Prediction of Breeding Values when Modeling Genotype × Environment Interaction using Pedigree and Dense Molecular Markers. *Crop Sci.* 52: 707–719. <https://doi.org/10.2135/cropsci2011.06.0299>

Buscaill, P., and S. Rivas, 2014 Transcriptional control of plant defence responses. *Curr. Opin. Plant Biol.* 20: 35–46. <https://doi.org/10.1016/j.pbi.2014.04.004>

Crossa, J., G. de los Campos, P. Pérez, D. Gianola, J. Burgueño *et al.*, 2010 Prediction of Genetic Values of Quantitative Traits in Plant Breeding Using Pedigree and Molecular Markers. *Genetics* 186: 713 LP-724. <https://doi.org/10.1534/genetics.110.118521>

Daetwyler, H. D., M. P. L. Calus, R. Pong-Wong, G. de los Campos, and J. M. Hickey, 2013 Genomic Prediction in Animals and Plants: Simulation of Data, Validation, Reporting, and Benchmarking. *Genetics* 193: 347 LP-365.

Dees, M. W., and L. A. Wanner, 2012 In Search of Better Management of Potato Common Scab. *Potato Res.* 55: 249–268. <https://doi.org/10.1007/s11540-012-9206-9>

de los Campos, G., J. M. Hickey, R. Pong-Wong, H. D. Daetwyler, and M. P. L. Calus, 2013 Whole-Genome Regression and Prediction Methods Applied to Plant and Animal Breeding. *Genetics* 193: 327 LP-345. <https://doi.org/10.1534/genetics.112.143313>

de los Campos, G., D. Sorensen, and D. Gianola, 2015 Genomic Heritability: What Is It? *PLoS Genet.* 11: e1005048. <https://doi.org/10.1371/journal.pgen.1005048>

Drewnowski, A., and C. D. Rehm, 2013 Vegetable Cost Metrics Show That Potatoes and Beans Provide Most Nutrients Per Penny. *PLoS One* 8: e63277. <https://doi.org/10.1371/journal.pone.0063277>

Driscoll, J., J. Coombs, R. Hammerschmidt, W. Kirk, L. Wanner *et al.*, 2009 Greenhouse and field nursery evaluation for potato common scab tolerance in a tetraploid population. *Am. J. Potato Res.* 86: 96–101. <https://doi.org/10.1007/s12230-008-9065-8>

Falconer, D., and T. Mackay, 1996 *Introduction to Quantitative Genetics* (4e). Pearson, London.

FAOSTAT, 2016 *FAOSTAT Database Collections* Food and Agriculture Organization of the United Nations, Rome.

Fry, W. E., 1978 Quantification of General Resistance of Potato Cultivars and Fungicide Effects for Integrated Control of Potato Late Blight. *Phytopathology* 68: 1650–1655. <https://doi.org/10.1094/Phyto-68-1650>

Gebhardt, C., A. Ballvora, B. Walkemeier, P. Oberhagemann, and K. Schüller, 2004 Assessing genetic potential in germplasm collections of crop plants by marker-trait association: A case study for potatoes with quantitative variation of resistance to late blight and maturity type. *Mol. Breed.* 13: 93–102. <https://doi.org/10.1023/B:MOLB.0000012878.89855.df>

Habyarimana, E., B. Parisi, and G. Mandolino, 2017 Genomic prediction for yields, processing and nutritional quality traits in cultivated potato (*Solanum tuberosum* L.). *Plant Breed.* 136: 245–252. <https://doi.org/10.1111/pbr.12461>

Haverkort, A. J., P. C. Struik, R. G. F. Visser, and E. Jacobsen, 2009 Applied Biotechnology to Combat Late Blight in Potato Caused by *Phytophthora Infestans*. *Potato Res.* 52: 249–264. <https://doi.org/10.1007/s11540-009-9136-3>

Haynes, K. G., and B. J. Christ, 1999 Heritability of resistance to foliar late blight in a diploid hybrid potato population of *Solanum phureja* × *Solanum stenotomum*. *Plant Breed.* 118: 431–434. <https://doi.org/10.1046/j.1439-0523.1999.00394.x>

Haynes, K. G., B. J. Christ, C. R. Burkhart, and B. T. Vinyard, 2009 Heritability of Resistance to Common Scab in Diploid Potatoes. *Am. J. Potato Res.* 86: 165–170. <https://doi.org/10.1007/s12230-009-9068-0>

- Haynes, K. G., R. W. Goth, and R. J. Young, 1997 Genotype  $\times$  Environment Interactions for Resistance to Common Scab in Tetraploid Potato. *Crop Sci.* 37: 1163–1167. <https://doi.org/10.2135/cropsci1997.0011183X003700040023x>
- Heffner, E. L., M. E. Sorrells, and J.-L. Jannink, 2009 Genomic Selection for Crop Improvement. *Crop Sci.* 49: 1–12. <https://doi.org/10.2135/cropsci2008.08.0512>
- Hirsch, C. N., C. D. Hirsch, K. Felcher, J. Coombs, D. Zarka *et al.*, 2013 Retrospective View of North American Potato (*Solanum tuberosum* L.) Breeding in the 20th and 21st Centuries. *G3: Genes, Genomes. Genetics* 3: 1003–1013.
- Killick, R. J., and J. F. Malcolmson, 1973 Inheritance in potatoes of field resistance to late blight [*Phytophthora infestans* (Mont.) de Bary]. *Physiol. Plant Pathol.* 3: 121–131. [https://doi.org/10.1016/0048-4059\(73\)90028-3](https://doi.org/10.1016/0048-4059(73)90028-3)
- Kolasa, K., 1993 The potato and human nutrition. *Am. Potato J.* 70: 375–384. <https://doi.org/10.1007/BF02849118>
- Lehermeier, C., G. de los Campos, V. Wimmer, and C.-C. Schön, 2017 Genomic variance estimates: With or without disequilibrium covariances? *J. Anim. Breed. Genet.* 134: 232–241. <https://doi.org/10.1111/jbg.12268>
- Li, D., H. Liu, H. Zhang, X. Wang, and F. Song, 2008 OsBIRH1, a DEAD-box RNA helicase with functions in modulating defence responses against pathogen infection and oxidative stress. *J. Exp. Bot.* 59: 2133–2146. <https://doi.org/10.1093/jxb/ern072>
- Lozano, R., O. Ponce, M. Ramirez, N. Mostajo, and G. Orjeda, 2012 Genome-Wide Identification and Mapping of NBS-Encoding Resistance Genes in *Solanum tuberosum* Group Phureja. *PLoS One* 7: e34775. <https://doi.org/10.1371/journal.pone.0034775>
- Malosetti, M., C. G. Van Der Linden, B. Vosman, and F. A. Van Eeuwijk, 2007 A mixed-model approach to association mapping using pedigree information with an illustration of resistance to *Phytophthora infestans* in potato. *Genetics* 175: 879–889. <https://doi.org/10.1534/genetics.105.054932>
- Massa, A. N., N. C. Manrique-Carpintero, J. J. Coombs, D. G. Zarka, A. E. Boone *et al.*, 2015 Genetic Linkage Mapping of Economically Important Traits in Cultivated Tetraploid Potato (*Solanum tuberosum* L.). *G3: Genes, Genomes. Genetics* 5: 2357–2364.
- Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard, 2001 Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157: 1819–1829.
- Mosquera, T., M. F. Alvarez, J. M. Jiménez-gómez, A. Draffehn, A. Hofmann *et al.*, 2016 Targeted and Untargeted Approaches Unravel Novel Candidate Genes and Diagnostic SNPs for Quantitative Resistance of the Potato (*Solanum tuberosum* L.) to *Phytophthora infestans* Causing the Late Blight Disease. *PLoS One* 11: e0156254.
- Muktar, M. S., J. Lübeck, J. Strahwald, and C. Gebhardt, 2015 Selection and validation of potato candidate genes for maturity corrected resistance to *Phytophthora infestans* based on differential expression combined with SNP association and linkage mapping. *Front. Genet.* 6: 1–19. <https://doi.org/10.3389/fgene.2015.00294>
- Murphy, A. M., H. De Jong, and G. C. C. Tai, 1995 Transmission of resistance to common scab from the diploid to the tetraploid level via 4x-2x crosses in potatoes. *Euphytica* 82: 227–233. <https://doi.org/10.1007/BF00029565>
- Navarro, F. M., K. T. Rak, E. Banks, B. D. Bowen, C. Higgins *et al.*, 2015 Strategies for Selecting Stable Common Scab Resistant Clones in a Potato Breeding Program. *Am. J. Potato Res.* 92: 326–338. <https://doi.org/10.1007/s12230-015-9435-y>
- Nelson, R. R., 1978 Genetics of horizontal resistance to plant diseases. *Annu. Rev. Phytopathol.* 16: 359–378. <https://doi.org/10.1146/annurev.py.16.090178.002043>
- Nowicki, M., M. R. Foolad, M. Nowakowska, and E. U. Kozik, 2011 Potato and Tomato Late Blight Caused by *Phytophthora infestans*: An Overview of Pathology and Resistance Breeding. *Plant Dis.* 96: 4–17. <https://doi.org/10.1094/PDIS-05-11-0458>
- Pajerowska-Mukhtar, K., B. Stich, U. Achenbach, A. Ballvora, J. Lübeck *et al.*, 2009 Single Nucleotide Polymorphisms in the Allene Oxide Synthase 2 Gene Are Associated With Field Resistance to Late Blight in Populations of Tetraploid Potato Cultivars. *Genetics* 181: 1115 LP-1127.
- Pandey, S. P., and I. E. Somssich, 2009 The Role of WRKY Transcription Factors in Plant Immunity. *Plant Physiol.* 150: 1648–1655. <https://doi.org/10.1104/pp.109.138990>
- Pérez, P., and G. de los Campos, 2014 Genome-Wide Regression & Prediction with the BGLR Statistical Package. *Genetics* 198: 483–495. <https://doi.org/10.1534/genetics.114.164442>
- Pomerantz, A., Y. Cohen, E. Shufan, Y. Ben-Naim, S. Mordechai *et al.*, 2014 Characterization of *Phytophthora infestans* resistance to mefenoxam using FTIR spectroscopy. *J. Photochem. Photobiol. B* 141: 308–314. <https://doi.org/10.1016/j.jphotobiol.2014.10.005>
- Solano, J., I. Acuña, F. Esnault, and P. Brabant, 2014 Resistance to *Phytophthora infestans* in *Solanum tuberosum* landraces in Southern Chile. *Trop. Plant Pathol.* 39: 307–315. <https://doi.org/10.1590/S1982-56762014000400005>
- Sparks, A. H., G. A. Forbes, R. J. Hijmans, and K. A. Garrett, 2014 Climate change may have limited effect on global risk of potato late blight. *Glob. Change Biol.* 20: 3621–3631. <https://doi.org/10.1111/gcb.12587>
- Stamler, R. A., O. Holguin, B. Dungan, T. Schaub, S. Sanogo *et al.*, 2015 BABA and *Phytophthora nicotianae* Induce Resistance to *Phytophthora capsici* in Chile Pepper (*Capsicum annuum*). *PLoS One* 10: e0128327. <https://doi.org/10.1371/journal.pone.0128327>
- Sverrisdóttir, E., S. Byrne, E. H. R. Sundmark, H. Ø. Johnsen, H. G. Kirk *et al.*, 2017 Genomic prediction of starch content and chipping quality in tetraploid potato using genotyping-by-sequencing. *Theor. Appl. Genet.* 130: 2091–2108. <https://doi.org/10.1007/s00122-017-2944-y>
- Tai, G. C. C., A. M. Murphy, and X. Xiong, 2009 Investigation of long-term field experiments on response of breeding lines to common scab in a potato breeding program. *Euphytica* 167: 69–76. <https://doi.org/10.1007/s10681-008-9862-7>
- The Potato Genome Sequencing Consortium, 2011 Genome sequence and analysis of the tuber crop potato. *Nature* 475: 189–195. <https://doi.org/10.1038/nature10158>
- The R Development Core Team, 2010 R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Available at: <http://www.R-project.org>.
- Tiwari, J. K., S. Siddappa, B. P. Singh, S. K. Kaushik, S. K. Chakrabarti *et al.*, 2013 Molecular markers for late blight resistance breeding of potato: an update. *Plant Breed.* 132: 237–245. <https://doi.org/10.1111/pbr.12053>
- VanRaden, P. M., 2008 Efficient Methods to Compute Genomic Predictions. *J. Dairy Sci.* 91: 4414–4423. <https://doi.org/10.3168/jds.2007-0980>
- Wanner, L. A., 2006 A survey of genetic variation in *Streptomyces* isolates causing potato common scab in the United States. *Phytopathology* 96: 1363–1371. <https://doi.org/10.1094/PHYTO-96-1363>
- Wanner, L. A., and W. W. Kirk, 2015 *Streptomyces* – from basic microbiology to role as a plant pathogen. *Am. J. Potato Res.* 92: 236–242. <https://doi.org/10.1007/s12230-015-9449-5>
- Zimaro, T., N. Gottig, B. S. Garavaglia, C. Gehring, and J. Ottado, 2011 Unraveling Plant Responses to Bacterial Pathogens through Proteomics. *J. Biomed. Biotechnol.* 2011: 354801. <https://doi.org/10.1155/2011/354801>

Communicating editor: T. Mackay