



# Global multi-environment resistance QTL for foliar late blight resistance in tetraploid potato with tropical adaptation

Hannele Lindqvist-Kreuze <sup>1,\*</sup>, Bert De Boeck,<sup>1</sup> Paula Unger,<sup>1</sup> Dorcus Gemenet,<sup>2</sup> Xianping Li,<sup>3</sup> Zhechao Pan <sup>3</sup>, Qinjun Sui,<sup>3</sup> Junhong Qin,<sup>4</sup> Gebremedhin Woldegjorgis,<sup>5</sup> Kassaye Negash,<sup>5</sup> Ibrahim Seid,<sup>5</sup> Betaw Hirut,<sup>6</sup> Manuel Gastelo,<sup>1</sup> Jose De Vega,<sup>7</sup> and Merideth Bonierbale<sup>1,†</sup>

<sup>1</sup>International Potato Center, CIP, Lima 15024, Peru,

<sup>2</sup>CIP Kenya, c/o ILRI, P.O. Box 30709, Nairobi, Kenya,

<sup>3</sup>Industrial Crops Research Institute, Yunnan Academy of Agricultural Sciences (YAAS), 2238 Beijing Road, Kunming, Yunnan 650205, P. R. China,

<sup>4</sup>CIP China, Beijing 100081, China,

<sup>5</sup>Ethiopian Institute of Agricultural Research, (EIAR), Holetta Agricultural research Center. P.O. Box 31, West Showa Zone, Oromia Region, Ethiopia,

<sup>6</sup>CIP Ethiopia, c/o ILRI Ethiopia P.O. Box 5689, Addis Ababa, Ethiopia, and

<sup>7</sup>Earlham Institute (EI), Norwich Research Park, Norwich NR4 7UZ, UK

\*Corresponding author: International Potato Center, CIP, Apartado 1558, Avenida La Molina 1895, La Molina, Lima, Peru. Email: h.lindqvist-kreuze@cgiar.org

†Present address: Duquesa Business Centre, P.O. Box 157, Malaga 29692, Spain.

## Abstract

The identification of environmentally stable and globally predictable resistance to potato late blight is challenged by the clonal and polyploid nature of the crop and the rapid evolution of the pathogen. A diversity panel of tetraploid potato germplasm bred for multiple resistance and quality traits was genotyped by genotyping by sequencing (GBS) and evaluated for late blight resistance in three countries where the International Potato Center (CIP) has established breeding work. Health-indexed, *in vitro* plants of 380 clones and varieties were distributed from CIP headquarters and tuber seed was produced centrally in Peru, China, and Ethiopia. Phenotypes were recorded following field exposure to local isolates of *Phytophthora infestans*. QTL explaining resistance in four experiments conducted across the three countries were identified in chromosome IX, and environment-specific QTL were found in chromosomes III, V, and X. Different genetic models were evaluated for prediction ability to identify best performing germplasm in each and all environments. The best prediction ability (0.868) was identified with the genomic best linear unbiased predictors (GBLUPs) when using the diploid marker data and QTL-linked markers as fixed effects. Genotypes with high levels of resistance in all environments were identified from the B3, LBHT, and B3-LTVR populations. The results show that many of the advanced clones bred in Peru for high levels of late blight resistance maintain their resistance in Ethiopia and China, suggesting that the centralized selection strategy has been largely successful.

**Keywords:** GWAS; polyploidy; breeding; genotyping by sequencing (GBS); *Phytophthora infestans*; GBLUP

## Introduction

Potato genetic resources comprise a polyploid series consisting of tremendously diverse germplasm of wild relatives and cultivated landraces (Ovchinnikova et al. 2011; Spooner et al. 2014). However, most commercially cultivated potato varieties are tetraploid ( $2n = 4x = 48$ ) with the genome consisting mostly of *Solanum tuberosum* Group Tuberosum with some introgressions from a few wild species and cultivated landraces (Bradshaw et al. 2006; reviewed by Bethke et al. 2017; reviewed by Gaiero et al. 2018). Tetraploid potato is a highly heterozygous, outcrossing autopolyploid crop, which complicates genetic analysis. Most early genetic mapping studies utilized bi-parental populations at the simpler, diploid level ( $2n = 2x = 24$ ) and several disease resistance loci have been identified this way (reviewed by Gebhardt and Valkonen 2001; Tiwari et al. 2013). However, this approach does

not permit the assessment of large gene pools or multi-allelic interactions that influence traits in polyploids. Significant progress has recently been made in the development of algorithms and software for genotype calling, linkage, and QTL analysis in polyploid species.

Some single nucleotide polymorphism (SNP) arrays have been developed for potato: 8K SolCAP (Hamilton et al. 2011; Felcher et al. 2012) and the 20K SolSTW arrays (Uitdewilligen et al. 2013; Vos et al. 2015). These were developed using North American and European potato germplasm, respectively, and are consequently not the best options for genotyping CIP germplasm, which contains more introgressions from the South American gene pool. According to our previous experience, less than 50% of the SNPs on the 8K SolCAP array were informative in a test sample of CIP germplasm (Lindqvist-Kreuze et al. 2014). Genotyping by sequencing (GBS) has been applied to tetraploid potato (Uitdewilligen

Received: August 12, 2020. Accepted: June 29, 2021

© The Author(s) 2021. Published by Oxford University Press on behalf of Genetics Society of America.

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

et al. 2013; Sverrisdóttir et al. 2017) and variant calling from short-read sequencing data considering allele dosage is now possible using several different tools, such as GATK, Freebayes, or SAMtools, to name a few (Clevenger et al. 2015). However, reliable dosage calling in the heterozygous individuals depends on the read depth in the SNP loci. It was recently demonstrated in autopolyploid forage grass *Urochloa* spp. that a read depth of 61 was adequate to reliably call allele dosage, while only 17 reads were needed to reliably classify simplex tetraploids as heterozygous with 95% accuracy (Matias et al. 2019). The identification of QTL in autopolyploids is facilitated by new tools, such as GWASpoly, that consider allele dosage effects (Rosyara et al. 2016). Together, these advances make genomic analysis of tetraploid potato more informative and applicable to evolutionary and breeding studies.

The goal of CIP's potato breeding program is to develop resilient, high-yielding, nutritious, and early-maturing varieties for smallholder farming systems in the developing world. We are targeting farming systems that must function with minimum input of pesticides; therefore, a high level of disease resistance is an indispensable trait. To this end, CIP's potato breeding program has developed breeding populations with high levels of resistance to late blight caused by the oomycete *Phytophthora infestans*, and resistance to Potato Virus Y (PVY), Potato Virus X (PVX), and Potato Leaf Roll Virus (PLRV). Previous studies have identified genomic regions in CIP-bred germplasm to explain resistance to late blight focusing on phenotypic data collected from field trials in Peru or using local pathogen strains in greenhouse conditions (Li et al. 2010; Lindqvist-Kreuzer et al. 2014; Jiang et al. 2018). Information on the late blight resistance of CIP-bred materials has been published sporadically in target regions where they have been distributed (Li et al. 2018; Sun et al. 2010; Muhinyuza et al. 2015; Hirut et al. 2017b); however, to our knowledge, no genetic analysis has been published that identifies QTL for resistance in CIP germplasm tested in environments outside Peru.

A major objective of this research was to evaluate the potential of genomic-based selection methods to support a global breeding strategy using late blight resistance as a case study. To systematically evaluate CIP's breeding materials in diverse environments, we established a trait observation network (TON) of collaborators and assembled a diversity panel that consists of representative advanced clones including elite materials from each of CIP's breeding populations. This so-called TON panel was then distributed from Peru to China and Ethiopia, where it was included in a series of trait evaluation experiments by national research institutions and CIP. The specific aims of this study were (i) to identify QTL underlying late blight resistance in and across environments; and (ii) to test prediction models to support the global deployment and use of promising resistance sources in local breeding and variety development programs.

We report the genotyping, estimation of linkage disequilibrium, and population structure of the TON panel and identify QTL for late blight resistance via genome-wide association (GWA). In addition, we present a case for genomics-assisted breeding for foliar late blight resistance and show how the use of genomics and pedigree information can be used to select "best bet" clones for breeding and variety development in diverse target environments.

## Materials and methods

### Germplasm

The TON panel consisted of 380 genotypes representing seven CIP breeding populations and a group of varieties with variable

origins (Table 1). "Population A" was developed at CIP between 1980 and 1990 with emphasis on late blight resistance. Sources of late blight resistance were improved materials with *S. demissum*-derived resistance from breeding programs around the world, including native Andean cultivars (*S. tuberosum* groups Andigena, Phureja, and Stenotomum) and wild species (*S. acaule* and *S. bulbocastanum*). "Population B3" genotypes were derived from "Population A" with emphasis on increasing frequencies and levels of quantitative resistance to late blight. The "B1 population" is derived from the *S. tuberosum* group, Andigena. The "LTVR population" is characterized primarily for its resistance to the most important virus diseases (PVY, PVX, and PLRV), short crop duration, and adaptability to warm environments. The "LB-HT" population combines late blight resistance from the "B3 population" and heat tolerance from North American and European-bred varieties and the LTVR population. The "B3-LTVR" population contains hybrid genotypes originating from crosses between "B3" and "LTVR populations." The "pre-Bred" population has genotypes with late blight resistance introduced from wild *Solanum* species into the tetraploid background of "B3" or "LTVR." The varieties group consists of several potato varieties or key breeding lines: "Desiree," "Atlantic," "Spunta," "Granola," "Yungay," "Tomas Condemayta," "DTO-33," "Kufri Yoti," and "Chucmarina." The CIP numbers and the parentage of the 380 genotypes are given in Supplementary Table S1.

### Environments

The field sites in Ethiopia and China are within important potato production areas. In the Peruvian field site, potato is not the main crop, but late blight is endemic with high pressure whenever the crop is grown (Table 2). The late blight pathogen populations have been described in each location. In Peru and Ethiopia, only the A1 mating type has been identified and different clonal lineages are present, frequently containing virulence to most of the known *S. demissum* R genes (Lindqvist-Kreuzer et al. 2020; Mihretu et al. 2020). Meanwhile, the A2 mating type has been found to dominate in southern China (Li et al. 2013). All locations are in tropical and subtropical regions with latitudes ranging between 9°N and nearly 25°S.

### Field trials

Standard protocols at CIP were followed for planning and conducting the field trials (Forbes et al. 2014). The statistical designs in each trial are shown in Table 2. Uniform tuber seed was produced centrally in each country following the introduction of *in vitro* plants or mini-tubers from CIP facilities in Peru or Kenya.

Late blight resistance was evaluated under endemic disease pressure. The disease level in the plots was recorded as "percent leaf area infected" at typically 7-day intervals until susceptible controls reached 100% infection. These values were used to calculate the area under the disease progress curve (AUDPC) and relative AUDPC (rAUDPC). AUDPC is expressed in infection percentage times days (Forbes et al. 2014), while the rAUDPC was calculated by dividing the AUDPC by the "maximum potential AUDPC," which is the AUDPC a genotype would have if it had 100% infection at all readings.

### Genotyping, variant calling, and filtering for association analysis

In total, 380 potato clones were genotyped. Library construction and GBS were outsourced to the Genomics Facility at Cornell University in 2015. The DNA was digested with *Eco*T221 restriction enzyme and the libraries were 48x multiplexed for sequencing.

**Table 1** CIP potato programs breeding populations utilized in the study, the main breeding objective of the populations, and the number of genotypes derived from each population and subpopulation

Breeding population (subpopulation)	Genotypes evaluated	Years of crosses made	Main breeding objective
A	13	1974, 1977, 1980, 1981, 1984, 1987, 1991	Late blight resistance, adaptation to high-lands
B1	11	1999	
B3 (B3C0, B3C1, B3C2)	100 (2, 51, 47)	1987, 1989, 1991, 1992, 1993, 1995, 1996	
LB-HT	37	1998	Late blight resistance, heat tolerance, adaptation to mid-elevation
B3-LTVR (intermediate-LT-LB, XY-LB)	25 (12, 13)	1996, 2001	Hybrid population combining late blight resistance, heat tolerance, virus resistance
LTVR (LTVR, BW)	186 (171, 15)	1979, 1980, 1985, 1988, 1989–2000, 2002–2004, 2009	Virus resistance, heat tolerance, drought tolerance, salinity tolerance, adaptation to lowlands, bacterial wilt resistance
PREBRED	2	1994	Late blight resistance
VARIETY	6	NA	Varied
Grand total	380		

**Table 2** Geographical location, year of implementation, and detailed description of the design of the field trials conducted to obtain estimates of late blight resistance of the potato genotypes in China, Ethiopia, and Peru

Country	Location	Year	Number of genotypes evaluated (statistical design)	Checks and shared varieties	H2 rAUDPC (Cullis et al. 2016)
Peru	Pasco, Oxapampa 10.5853°S, 75.4053°W	2014	240 (resolvable incomplete block design; 3 replicates)	Chucmarina Unica Tomas Desiree Spunta, Granola, Atlantic	0.99
China	Yunnan, Kunming 24.8801°N, 102.8329°E	2015	306 (RCBD; 3 replicates)	Chucmarina	0.95
		2016	336 (RCBD; 3 replicates)	C-88 Unica Desiree, Kexin, Yunshu, Zhongsu, Tomas	0.96
Ethiopia	Oromia, Holetta 9.0633°N, 38.4902°E	2017	60 (RCBD; 3 replicates)	Belete Gudene Jalene Unica	0.98
		2016	128 (CRD; 1.63a replicates)	Gudene Belete Jalene Unica Tomas	0.96

The estimate of the heritability of resistance based on the relative Area Under the Disease Progress Curve (rAUDPC) according to Cullis et al. (2016) for each trial is shown.

<sup>a</sup> The trial design does not have complete replicates. Therefore, the theoretical replication number is calculated as the harmonic mean of the number of replications across all genotypes.

RCBD, randomized complete block design; CRD, completely randomized design.

The diploid calling was performed using the Tassel pipeline (Bradbury et al. 2007). The resulting Variant Call Format (VCF) file was processed with Bcftools (<https://samtools.github.io/bcftools/>) to filter the variants for a minimum read depth (RD) of 17, a minimum genotype quality (GQ) of 30, and a minor allele frequency (MAF) of 0.03. The SNPs that did not pass these criteria were changed to missing calls. Only the SNP sites that contained less than 30% missing data were selected.

For polyploid calling, the raw FASTQ files were processed with Stacks (Catchen et al. 2013) to remove the barcodes and TrimGalore was used (<https://github.com/FelixKrueger/TrimGalore>) to trim the ends of the reads. The reads were

aligned to the reference genome version *S. tuberosum*\_448\_v4.03 (Sharma et al. 2013) using BWA (Li and Durbin 2009) and the resulting SAM files were converted to BAM files using Samtools (Li et al. 2009). The variants were called using the GATK HaplotypeCaller option (Poplin et al. 2018), disabling the duplicate read filter (as recommended for GBS data), and joint genotyped using the -ERC GVCF mode. From the VCF files, SNP calls were filtered using Bcftools for a minimum RD of 61, a minimum GQ of 30, and a MAF of 0.03. The samples that did not surpass these criteria were changed to missing calls. Only the SNP sites that contained less than 30% missing data were included in the analysis.

## Analysis of the population substructuring

Population structure was analyzed on a subset of 7597 diploid markers with no missing data. First, K-means clustering over K-values from 1 to 20 was repeated 100 times to identify potential clusters. The values K2-9 were further selected for discriminant analysis of principal components (DAPC) and for each K, the potato genotypes (clones) were assigned in the K groups based on the highest probability. To estimate the components of covariance among and within the clusters Analysis of molecular variance (AMOVA) was conducted using a farthest-neighbor algorithm. K-means clustering and AMOVA were conducted using Poppr v 2.9.0 (Kamvar et al. 2014).

## Linkage disequilibrium decay

Correlations between all pairs of markers were calculated using Pearson's correlation coefficient ( $r^2$ ) based on the SNP dosage scores (0–4) and plotted against the physical distance of the marker pairs. A spline was fitted on the 90th percentile using quantile regression (Koenker 2017) and the estimators for the LD decay were obtained from the spline at  $r^2$  of 0.1. In addition, another estimate for LD decay was obtained from the fitted spline: the distance at which half of the short-range LD had decayed based on the 90th percentile,  $LD_{1/2,90}$  as described by Vos et al. (2017).

## Genome-wide association

Marker trait associations were modeled for all trials independently with the GWASpoly package using both diploid and tetraploid marker sets (Rosyara et al. 2016). The BLUE values were used as phenotypic values for each experiment (Supplementary Table S2). For the tetraploid dataset, general, additive, simplex dominant (1-dom), and duplex dominant (2-dom) models were used while for the diploid dataset, diplo-general, diplo-additive, and the simplex dominant (1-dom) models were used. The parameters used for the GWAS modeling function (GWASpoly) in R were the following: no additional fixed effects; four principal components included as covariates to account for the population substructuring; a minimum MAF of 0.03; a maximum genotype frequency (after applying dominance relations) of 0.95; and P3D approximation. To detect statistical significance, the Bonferroni correction method was used, ensuring the genome-wide type I error would not be greater than 0.05. Manhattan plots were generated to display significant SNP in the different genetic models. In addition, Quantile–Quantile (Q–Q) plots of the observed vs expected  $-\log_{10}(p)$  values were used to evaluate that the population structure had been adequately controlled.

The genomic positions of the QTL, known genes, and the SNPs associated with plausible QTL for pathogen resistance in relation to other loci were determined using the genome browser available at <http://solanaceae.plantbiology.msu.edu/> and *S. tuberosum* Group Phureja DM1-3 516R44 v4.03 pseudomolecules. To approximate the physical location of markers for pathogen resistance present in literature, the position in cM was obtained from the GABI Primary Database (<https://www.gabipd.org/projects/Pomamo/>) and then translated to an approximate physical position in Mbp using the information provided in Sharma et al. (2013) that integrates the potato genome and physical and genetic maps.

## Statistical analysis of phenotypic data

From the weekly observations of the disease incidence in the plots, the AUDPC was calculated and then rescaled to the relative AUDPC (rAUDPC) to facilitate comparison among the different environments. Because the experimental design types varied

across environments, the raw data were analyzed using a stage-wise approach as elaborated in Piepho et al. (2012) and van Eeuwijk et al. (2016). In the first stage, genotypic best linear unbiased estimators (BLUEs) of rAUDPC for each environment separately were obtained by fitting a mixed model that considered the experimental design used in the specific environment (results presented in Supplementary Table S2). Single trial broad-sense heritabilities ( $H^2$ ) were estimated for rAUDPC using the methodology proposed by Cullis et al. (2006) (Table 2).

In the second stage, the genotype-by-environment table of BLUEs and their standard errors were used as starting point for the multi-environment trial (MET) analysis. The BLUEs were weighted according to Method 2, as described in Möhring and Piepho (2009), to fit a mixed model that considers possible genotype-by-environment interaction (GEI). In the presence of GEI, the more realistic models often allow for heterogeneity of genetic variances and covariances across environments (Malosetti et al. 2013). The best-fitting model was chosen using Akaike's information criterion (AIC):  $\underline{\mu}_{ij} = \mu + G_{ij} + E_j + \epsilon_{ij}$  where  $\mu$  is the overall mean (intercept),  $E_j$  the fixed effect of the  $j^{\text{th}}$  environment ( $j = 1, \dots, 5$ ),  $G_{ij}$  the random environment-specific genetic effect of the  $i^{\text{th}}$  genotype in environment  $j$  and  $\epsilon_{ij} \sim N(0, 1)$  standard normally distributed and independent random error effects (for all  $i = 1, \dots, I$  and  $j = 1, \dots, 5$ ). The underlined terms in the model are random effects and all other terms are fixed effects. Note that observations  $\underline{\mu}_{ij}$  are the reweighted BLUEs. In contrast to the residual errors, the random genetic effects are not necessarily independent for all  $i$  and  $j$ . The 5-dimensional random genetic effect  $\underline{G}_i = (G_{i1}, \dots, G_{i5})^T$  of the  $i^{\text{th}}$  genotype follows a multivariate normal distribution  $\underline{G}_i \sim N(0, \Sigma_E)$  with variance-covariance matrix  $\Sigma_E$  allowing for flexible GEI modelling. The variance-covariance matrix  $\Sigma_E$  of the random genetic effect  $\underline{G}_i$  was parametrized using the second-order factor analytic model that accommodates heterogeneity of genetic variances and genetic covariances across environments in a parsimonious manner (Piepho 1998). The concatenated random effects vector  $\underline{G} = (\underline{G}_1, \dots, \underline{G}_I)^T$  was assumed to follow a multivariate normal distribution with mean  $\mathbf{0}$  and variance-covariance matrix  $\Sigma$  where  $\Sigma = \Sigma_G \otimes \Sigma_E$ . The symbol  $\otimes$  denotes the Kronecker product and  $\Sigma_G$  denotes the kinship matrix defining the genetic relationships between the different genotypes. The matrix  $\Sigma_E$  was estimated when fitting the model, but the  $I \times I$  matrix  $\Sigma_G$  had to be stated as prior knowledge before fitting the model. Choosing a different approach to determine this matrix  $\Sigma_G$ , therefore, led to different kinds of final models. The considered options are described below.

In a first simplified approach, it was assumed that  $\Sigma_G = I_I$ , where  $I_I$  is the  $I$ -dimensional identity matrix. The corresponding model therefore assumes that there are no genetic correlations between genotypes, but the final variance-covariance matrix  $\Sigma$  of the random genetic effect  $\underline{G}$  allows for heterogeneity of genetic variances and covariances across environments through  $\Sigma_E$ . The genotypic best linear unbiased predictors (BLUPs), predicting rAUDPC per genotype and possibly per environment, were obtained from the selected mixed model that describes the genetic effect and the GEI in terms of heterogeneity of variances and covariances across environments.

Alternative—and more realistic—approaches recognize that genetic relationships between the genotypes exist. In the following models, genetic relationships were incorporated, first by using genetic theory (pedigree-based relationship or kinship matrix) and then by using the available molecular information (molecular-based relationship or kinship matrix). In the standard



approach described above, the random genetic effects  $\mathbf{G}_i$  are assumed independent for  $i = 1, \dots, I$ . It is, however, more realistic to consider a variance-covariance structure for the genetic effects that integrates the expected or observed relationships between genotypes. In the first case, the relationship matrix  $\Sigma_G$  was calculated using pedigree of up to four generations deep that was not complete for all genotypes. The resulting mixed model is called the PBLUP model. Four GBLUP models were fitted calculating the relationship matrix  $\Sigma_G$  based on the same 2x and 4x marker sets as used in the GWA analysis. For both ploidy levels, a GBLUP model was considered without including the diagnostic SNP markers as fixed effects in the model, and a separate GBLUP model was considered by including the diagnostic SNP markers listed in Tables 4 and 5 as fixed effects in the model and testing for the significance of these QTL and QTL-by-environment (QTLx E) interaction effects. Using the tetraploid SNP marker data, this approach led to a final GBLUP model with fixed effects for the dosage of 0\_36073482, 3\_3319097 and 9\_60067335 SNP markers. Using the diploid SNP marker data, this approach led to a final GBLUP model with fixed effects for the dosage of 0\_36073482, 5\_5572873, 9\_59967523 and 9\_60067335 SNP markers. The pedigree BLUPs (PBLUPs) and the genomic BLUPs (GBLUPs), predicting rAUDPC per genotype and possibly per environment, were obtained from the selected mixed model that describes the genetic effect and the GEI in terms of heterogeneity of variances and covariances across environments, but also accounts for the correlations between genotypes within environments.

A fivefold cross-validation was performed to compare the different prediction models for genetic rAUDPC predictions per environment. For this validation procedure, the set of available adjusted phenotypic values (the environment-specific genetic BLUEs of all five environments) was randomly subdivided into

five disjoint subsets. The environment-specific genetic rAUDPC values of each of the five subsets were predicted while the other four subsets were used as training data set to fit the prediction model. This cross-validation procedure was repeated 100 times to obtain an empirical sampling distribution of the prediction ability (i.e., the Pearson correlation between the adjusted phenotypic values and the validation predictions of the genetic values). The mean and standard deviation of this empirical distribution were calculated to characterize the expected value and variability of the prediction ability, respectively.

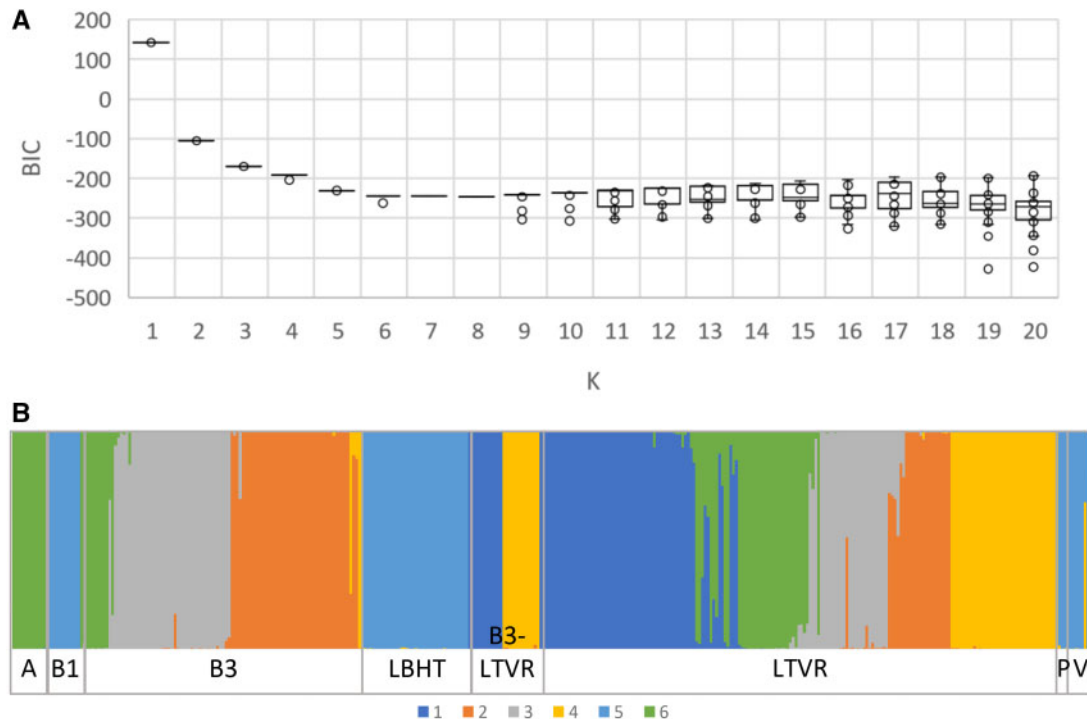
All calculations, fitting the mixed models and calculating BLUEs and BLUPs, were done by using the ASREML-R package in R.

## Results and discussion

### SNP markers

The tetraploid dataset included a total of 305,345 SNPs after GATK variant calling, while the diploidized dataset after Tassel pipeline had an SNP count of 312,727. After applying filtering parameters, including the read depth thresholds recommended based on different ploidy levels (Matias et al. 2019), the number of SNPs were reduced to 3239 tetraploid SNPs and 55,748 diploid SNPs.

The number of SNP obtained through GBS depends on the complexity of the genome, restriction enzyme(s) used, and the sequencing depth. Although GBS methodology was initially developed using a single restriction enzyme (Elshire et al. 2011), it was soon followed by a method that uses two enzymes (Poland et al. 2012). In tetraploid potato, ApeKI or MspI/PstI combination has been reported (Sverrisdóttir et al. 2017; Bastien et al. 2018). It was shown that by using a single enzyme, more markers can be obtained, albeit with lower read coverage per marker as



**Figure 1** Population substructuring of the potato genotypes based on K-means clustering and discriminant analysis of principal components (DAPC). (A) K-means clustering over 20 K values. (B) The posterior probabilities of the group assignments of each potato genotype for  $K = 6$ . The potato genotypes were derived from pre-defined CIP potato breeding populations [A, B1, B3, LBHT, B3\_LTVR, PREBRED (=P)] and from a group of old varieties (V). See Table 1 for details.

**Table 3** Analysis of molecular variance (AMOVA) to test for population differentiation among the potato genotypes based on K-means clustering of  $n = 2-9$ 

Variance components	K = 2	K = 3	K = 4	K = 5	K = 6	K = 7	K = 8	K = 9
$\sigma_p^2$	0.00231	0.003557	0.10983	0.10818	0.10580	0.06733	0.06439	0.05695
$\sigma_{g(p)}^2$	0.04507	0.04385	0.01477	0.00958	0.00721	0.00723	0.00725	0.00728
$\sigma_p^2 / (\sigma_p^2 + \sigma_{g(p)}^2)$	0.04619	0.07110	0.86318	0.89853	0.91492	0.87225	0.86982	0.84989

**Table 4** SNP markers that were significantly associated with the late blight resistance phenotype in the field trials in China (Kun2015, Kun2016), Ethiopia (Hol2016), and Peru (Oxa2014)

Marker (chromosome followed by position)	Ref	Alt	Trial	Model	Threshold	Score	Effect			
0_36073482	G	A	Oxa2014	General	4.71	9.13	NA			
				Additive	4.78	9.13	-0.19			
				1-dom-alt	4.71	9.13	-0.19			
			Kun2015	General	4.74	19.23	NA			
				Additive	4.78	19.23	-0.29			
				1-dom-alt	4.71	19.23	-0.29			
			Kun2016	General	4.76	22.22	NA			
				Additive	4.78	22.22	-0.27			
				1-dom-alt	4.71	22.22	-0.27			
3_3319097	A	T	Hol2017	General	4.49	5.83	NA			
9_58779951	G	A	Oxa2014	Additive	4.78	4.95	-0.11			
				1-dom-alt	4.71	5.62	-0.13			
			Kun2015	1-dom-alt	4.71	5.53	-0.14			
				Kun2016	1-dom-alt	4.71	5.01	-0.12		
			9_59967523	A	T	Kun2015	General	4.74	7.5	NA
Additive	4.78	6.92					-0.14			
1-dom-alt	4.71	8.39					-0.18			
Kun2016	General	4.76				8.59	NA			
	Additive	4.78				9.18	-0.14			
	1-dom-alt	4.71				9.15	-0.16			
9_60067335	A	G				Oxa2014	General	4.71	11.94	NA
							Additive	4.78	12.54	-0.21
							1-dom-alt	4.71	12.94	-0.21
			Kun2015	General	4.74	18.71	NA			
				Additive	4.78	18.29	-0.26			
				1-dom-alt	4.71	19.61	-0.27			
			Kun2016	General	4.76	19.13	NA			
				Additive	4.78	19.94	-0.24			
				1-dom-alt	4.71	20.18	-0.25			

The associations were modeled using tetraploid SNP with GWASpoly software and general, additive, and 1-dominance models.

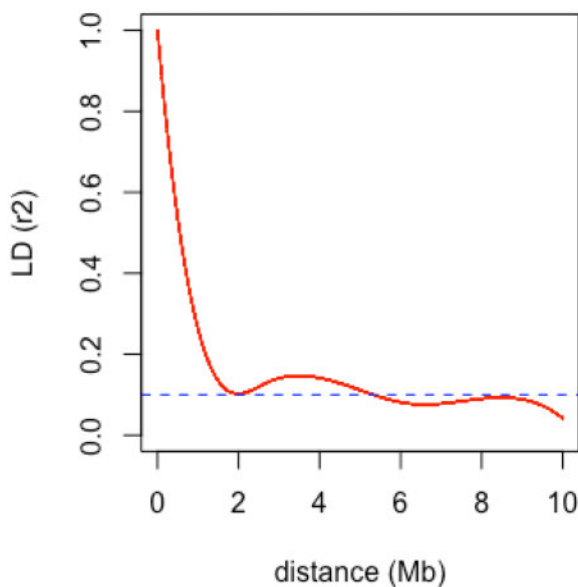
compared to the two-enzyme combo (Bastien et al. 2018). Thus, the single enzyme choice in our study may have contributed to low read coverage and thereby loss of markers after the subsequent filtering steps.

## Population structure

Population structure that is not considered in GWA may result in the discovery of false-positive associations. Previous studies in tetraploid potato report either weak (D'hoop et al. 2010; Uitdewilligen et al. 2013; Vos et al. 2017) or absence of population structure (Li et al. 2010; Fischer et al. 2013; Stich et al. 2013). Based on K-means clustering and the lowest BIC value, the optimal number of subgroups in the CIP TON panel was  $K = 6$  (Figure 1A). The result was corroborated by AMOVA since with  $K = 6$ , 91.5% of the variance was found between the groups (Table 3). However, these six subgroups do not strictly follow the pre-assigned breeding populations listed in Table 1 (Figure 1B). For instance, the LTVR breeding population splits into five groups, while the B3 population consists of four groups, which are also found in the LTVR population. As any

long-term breeding program, CIP's potato breeding efforts have addressed a wide and dynamic range of improvement goals throughout its approximately 40-year history. The germplasm characterized in this paper is a cross-section of advanced clones from crosses made between 1974 and 2009. During this time priorities have shifted several times and new trait sources have been incorporated that can be expected to influence the composition of breeding populations. The main drivers of the trait prioritization include growing awareness and increased understanding of (i) pathogen dynamics and host-pathogen interactions driving the need for new sources of resistance to late blight; coincident with donor support for pre-breeding; (ii) global warming trends and climate extremes requiring greater tolerance for heat and drought, and (iii) political influences that change development agendas to new world regions and thus growing conditions.

The substructure encountered was included as a co-factor in the GWA to avoid false-positive associations. This was done directly in the GWASpoly package by incorporating the first four principal components in the mixed model.



**Figure 2** Linkage disequilibrium (LD) decay estimated in the potato genotypes. Pearson correlation coefficient ( $r^2$ ) was plotted against the physical map distance (Mb) between pairs of SNP. The red line is a regression spline fitted on the 90th percentile and the blue dotted line indicates the  $r^2$  value of 0.1.

### Linkage disequilibrium

The extent to which LD persists when the physical distance between the SNP markers increases is used to measure the rate of LD decay. The rate of LD decay is directly linked to the feasibility to assess marker-trait associations, since crops with low LD decay rates require lower marker density as compared to those with a faster decay. To be able to compare the LD in our potato germplasm with other recently published studies in tetraploid potato (Vos et al. 2017; Sharma et al. 2018) we estimated the LD decay on a short distance of 10 Mb. By plotting the pairwise  $r^2$  values with the physical distance of the markers and fitting a regression spline on the 90th percentile of the observations, we obtained an estimate for the LD decay with the  $r^2$  value of 0.1 at 2 Mb (Figure 2). This physical distance is equivalent to a genetic distance of 5cM, similar to that observed by D'hoop et al. (2010) in a collection of tetraploid potato cultivars representative of the cultivated gene pool in Europe and North America. As suggested by Vos et al. (2017), more accurate estimates may be obtained when using the LD-decay estimate value for  $r^2_{1/2\max, 90}$ . This estimate in CIP germplasm was 0.55 Mb, which corroborates values reported in recent European potato varieties (0.6 Mb in Vos et al. 2017; 0.91 Mb in Sharma et al. 2018). The average  $r^2$  for the short distance in our dataset was 0.091, which is a bit lower than the average  $r^2$  (0.19–0.22) reported for European varieties (Vos et al. 2017), indicating that there were probably more founder haplotypes in our diversity panel than in the European pooled varieties. In summary, the LD decay estimated in CIP germplasm was moderate, and comparable to the LD decay found in the European and North American potato germplasm (D'hoop et al. 2010; Vos et al. 2017; Sharma et al. 2018).

### QTL for late blight resistance

Several SNPs were significantly associated with late blight resistance in the field trials with a total of 16 markers tagging possible QTL (Tables 4 and 5). In the tetraploid data, 5 markers for late blight resistance were found in chromosomes 0, III, and IX, while

in the diploid dataset 14 markers on chromosomes 0, III, V, VI, IX, and X were associated with the resistance phenotype. Populations of *P. infestans* are diverse and there is a trend of increasing diversity in potato-growing regions worldwide (Cooke and Lees 2004). Taking this into account, the markers on chromosomes IX and X could indicate a QTL for broad resistance not specific to regional late blight strains, while markers associated to the QTL on chromosomes III and V could be environment-specific (Tables 4 and 5). The highest number of SNPs associated to late blight resistance in the GWAS were mapped between 59 and 61.2 Mbp of chromosome IX in a region that had been previously associated with late blight resistance in Peru (Lindqvist-Kreuzer et al. 2014). In addition, the markers were within or surrounding the segment that forms a large cluster of putative resistance genes. For example, the locus PGSC0003DMG400020587 encodes a homolog of *Rpi-vnt1* (Mosquera et al. 2016), a major gene for resistance to *P. infestans* that has been previously cloned and is characterized in the wild potato species *Solanum venturii* (Foster et al. 2009; Pel et al. 2009).

The 48x multiplexing of samples during the sequencing and stringent filtering for minimum read depth in all samples yielded relatively few SNP (less than 4 K). Therefore, to increase the chances of finding significant associations, the GWA was done using both tetraploid and the diploidized data. Three markers (9\_58779951, 9\_59967523, 9\_60067335) map in the same region of chromosome IX previously found associated with late blight resistance in Peru (Li et al. 2010; Lindqvist-Kreuzer et al. 2014). These markers are physically separated in the DM genome by 1.3 Mb, which fits the estimate for the LD decay in our diversity panel, suggesting that this could be a single QTL. The R8 gene from *S. demissum* was recently identified within the QTL dPI09c reported by Li et al. (2010; Jiang et al. 2018). The QTL dPI09c interval begins at 60615044 bp in potato DM1-3 516 R44 (Potato Genome Sequencing Consortium 2011), which is over 0.6 Mbp from the nearest GBS marker (9\_60067335) and identified in the current research in the tetraploid marker set. None of the SNP in our tetraploid GBS marker set is located in the dPI09c interval (Jiang et al. 2018), and hence the markers in the tetraploid set could be indicative of another QTL. In the diploid marker set, the QTL in chromosome IX extends beyond that discovered in the tetraploid marker set, reaching also to the QTL dPI09c interval.

### Best linear unbiased models to predict late blight resistance

Phenotypic evaluation of potato germplasm in multiple environments is constrained by the difficulty of sharing vegetatively propagated tuber seed. Distribution of clonal material is restricted due to phytosanitary risks and strictly regulated. Furthermore, clonal seed multiplication rates are low from one generation to the next, and tuber seed is bulky and perishable. Therefore, it would be advantageous to be able to predict the performance of potato clones in a given environment based on their performance in another environment. This may be feasible for traits that exhibit low levels of genotype by environment interaction. Late blight resistance in CIP germplasm has been shown largely to be controlled by a large effect QTL (Li et al. 2012) and the late blight resistance gene R8 has been identified in the QTL (Jiang et al. 2018). R8 confers a broad-spectrum resistance with predicted global effect because virulence toward R8 is only rarely encountered (Swiezyński et al. 2000; Haynes et al. 2002; Zhang and Kim 2007; Vossen et al. 2016; Lindqvist-Kreuzer et al. 2020).

**Table 5** GBS markers that were significantly associated with the late blight resistance phenotype in the field trials in China (Kun2015, Kun2016), Ethiopia (Hol2017), and Peru (Oxa2014)

Marker (chromosome followed by position)	Ref	Alt	Trial	Model	Threshold	Score	Effect
0_36073482	G	A	Oxa2014	Diplo-general	6	10.66	NA
				Diplo-additive	6.03	10.66	-0.20
				1-dom-alt	6.02	10.66	-0.20
			Kun2015	Diplo-general	6.01	25.19	NA
				Diplo-additive	6.03	25.19	-0.31
				1-dom-alt	6.02	25.19	-0.31
			Kun2016	Diplo-general	6.02	29.27	NA
				Diplo-additive	6.03	29.27	-0.29
				1-dom-alt	6.02	29.27	-0.29
3_45458723	A	C	Hol2017	Diplo-general	5.94	6.22	NA
				Diplo-additive	6.02	6.22	0.47
				1-dom-alt	6.02	6.22	0.47
3_45458753	G	A	Hol2017	Diplo-additive	6.02	6.22	0.47
				1-dom-alt	6.02	6.22	0.47
3_45458754	A	T	Hol2017	Diplo-additive	6.02	6.22	0.47
				1-dom-alt	6.02	6.22	0.47
5_5572873	G	A	Oxa2014	Diplo-additive	6.03	6.87	0.16
				1-dom-alt	6.02	6.87	0.16
6_45694949	G	A	Kun2015	Diplo-general	6.01	7.09	NA
				1-dom-alt	6.02	7.64	-0.15
			Kun2016	Diplo-general	6.02	7.33	NA
				1-dom-alt	6.02	8.13	-0.13
			Oxa2014	Diplo-general	6	8.61	NA
				Diplo-additive	6.03	8.61	-0.17
9_58779951	G	A	Kun2015	1-dom-alt	6.02	8.61	-0.17
				diplo-general	6.01	7.78	NA
				diplo-additive	6.03	7.78	-0.17
			Kun2016	1-dom-alt	6.02	7.78	-0.17
				Diplo-general	6.02	8.05	NA
				Diplo-additive	6.03	8.05	-0.16
			Oxa2014	1-dom-alt	6.02	8.05	-0.16
				Diplo-general	6	6.01	NA
				Diplo-general	6.01	13.09	NA
9_59967523	A	T	Kun2015	Diplo-additive	6.03	13.09	-0.21
				1-dom-alt	6.02	13.09	-0.21
			Kun2016	Diplo-general	6.02	16.79	NA
				Diplo-additive	6.03	16.79	-0.21
				1-dom-alt	6.02	16.79	-0.21
			Oxa2014	Diplo-general	6	8.84	NA
9_59997331	T	C	Kun2015	Diplo-additive	6.03	8.84	-0.17
				1-dom-alt	6.02	8.84	-0.17
			Kun2016	Diplo-general	6.01	18.25	NA
				Diplo-additive	6.03	18.25	-0.25
				1-dom-alt	6.02	18.25	-0.25
			Oxa2014	Diplo-general	6.02	19.58	NA
				Diplo-additive	6.03	19.58	-0.23
				1-dom-alt	6.02	19.58	-0.23
			Oxa2014	Diplo-general	6	15.15	NA
9_60067335	A	G	Kun2015	Diplo-additive	6.03	15.15	-0.23
				1-dom-alt	6.02	15.15	-0.23
			Kun2016	Diplo-general	6.01	25.64	NA
				Diplo-additive	6.03	25.64	-0.31
				1-dom-alt	6.02	25.64	-0.31
			Oxa2014	Diplo-general	6.02	27.56	NA
				Diplo-additive	6.03	27.56	-0.28
				1-dom-alt	6.02	27.56	-0.28
			Kun2015	Diplo-general	6.01	6.89	NA
9_61106174	C	T	Kun2015	Diplo-additive	6.03	7.54	-0.14
				1-dom-alt	6.02	7.05	-0.14
			Kun2016	Diplo-general	6.02	11.4	NA
				Diplo-additive	6.03	12.31	-0.15
				1-dom-alt	6.02	11.83	-0.15
			Oxa2014	Diplo-additive	6.03	6.71	-0.12
9_61108928	T	G	Kun2015	1-dom-alt	6.02	6.66	-0.12
				Diplo-general	6.01	8.94	NA
			Kun2016	Diplo-additive	6.03	9.87	-0.16
				1-dom-alt	6.02	9.56	-0.17
				Diplo-general	6.02	10.73	NA
			Oxa2014	Diplo-additive	6.03	11.69	-0.15

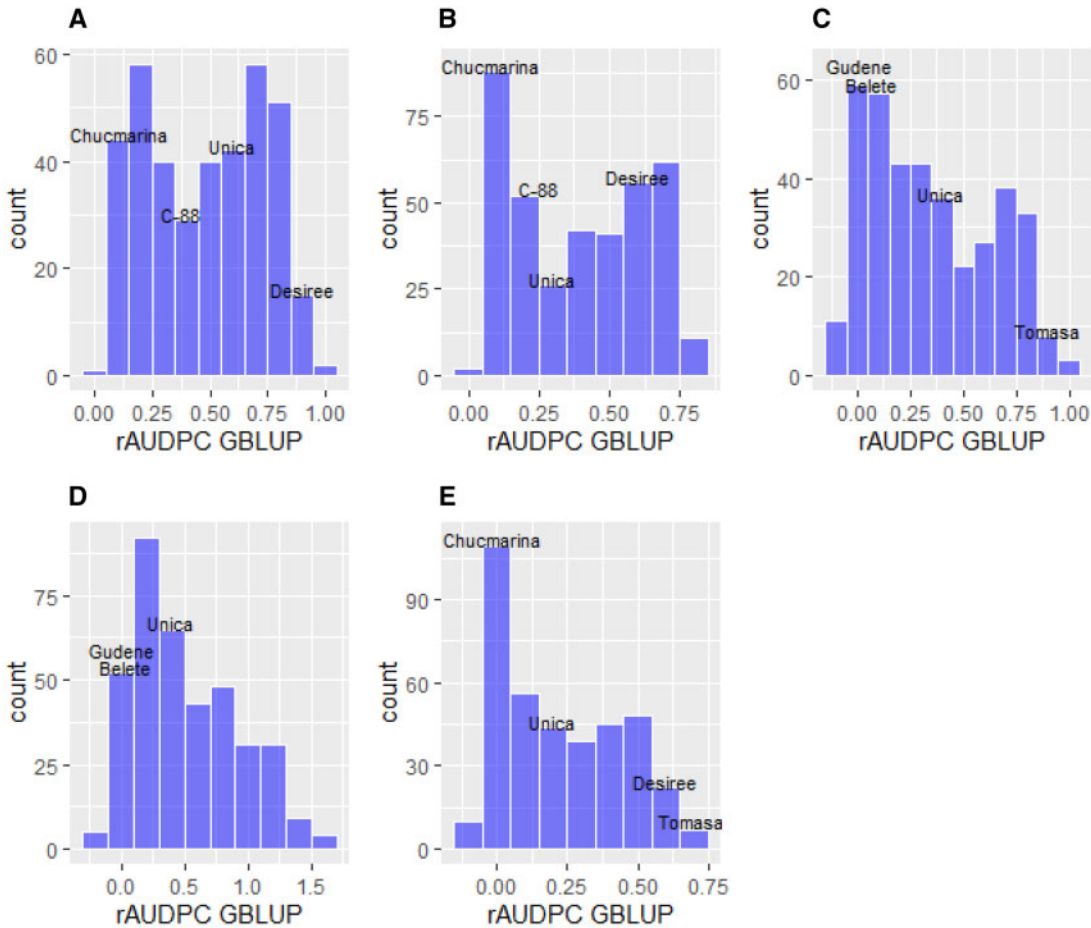
(continued)



Table 5. (continued)

Marker (chromosome followed by position)	Ref	Alt	Trial	Model	Threshold	Score	Effect
9_61261167	A	C	Oxa2014	1-dom-alt	6.02	11.38	−0.15
				Diplo-general	6	14.23	NA
				Diplo-additive	6.03	13.88	−0.21
			Kun2015	1-dom-alt	6.02	15.15	−0.23
				Diplo-general	6.01	24.53	NA
				Diplo-additive	6.03	24.33	−0.30
			Kun2016	1-dom-alt	6.02	25.64	−0.31
				Diplo-general	6.02	26.41	NA
				Diplo-additive	6.03	26.33	−0.27
10_51544544	A	G	Oxa2014	1-dom-alt	6.02	27.56	−0.28
				Diplo-general	6	13.39	NA
				Diplo-additive	6.03	13.39	−0.22
			Kun2015	1-dom-alt	6.02	13.39	−0.22
				Diplo-general	6.01	23.14	NA
				Diplo-additive	6.03	23.14	−0.30
			Kun2016	1-dom-alt	6.02	23.14	−0.30
				Diplo-general	6.02	29.56	NA
				Diplo-additive	6.03	29.56	−0.29
				1-dom-alt	6.02	29.56	−0.29

The associations were modeled using diploid SNP with GWASpoly software and the diplo-general, diplo-additive, and 1-dominance models.



**Figure 3** Histograms of genetic rAUDPC GBLUP values (diploid model with QTLxE fixed effects) in Kunming, China in 2015(A) and in 2016 (B), Holeta, Ethiopia in 2016 (C) and 2017 (D), and Oxapampa, Peru at 2014 (E). The shared control genotypes (checks) or well-known varieties in each trial are indicated in the plots based on their adjusted phenotypic rAUDPC value.

**Table 6** Estimated prediction ability of the different models for genetic and environment-specific predictions

Model	Prediction ability
BLUP	0.815 ± 0.009
PBLUP	0.855 ± 0.006
GBLUP (2x)	0.858 ± 0.006
GBLUP (4x)	0.855 ± 0.007
GBLUP (2x) with fixed QTL(xE) effects	0.868 ± 0.005
GBLUP (4x) with fixed QTL(xE) effects	0.864 ± 0.006

Fivefold cross-validation was obtained for the first two models.

In this research project, more than 300 advanced tetraploid clones from CIP were shared with partners, but for various reasons not all were evaluated in all environments and the total overlapping set of genotypes that were evaluated in all environments was less than 60 (Table 2). Varieties with known levels of late blight resistance were included as controls in each environment and the high level of infection in the susceptible control varieties suggests that the environmental conditions were conducive for the disease development (Figure 3). The heritability of the late blight resistance phenotype was high in each trial (Table 2), indicating that the trait is most likely controlled by a major disease resistance gene or a QTL with a large effect. We set out to test the prediction ability of six different mixed models to estimate the genetic correlations across environments and to predict the late blight resistance of all clones in all environments. Relatively high prediction ability ranging from 0.815 to 0.864 was obtained with all the tested models (Table 6). The prediction ability of the mixed model improved after incorporating pedigree and marker information. Indeed, the fivefold cross-validation results demonstrated that both the pedigree PBLUP model and the diploid marker GBLUP model have superior performance compared to the BLUP model (Table 6). The highest prediction ability was obtained for the GBLUP model that utilized the diploidized marker set and included QTL(xE) fixed effects. The top 10 ranked genotypes in each environment, according to this model, are listed in Table 7. However, if molecular marker information is not available, a significant improvement for the prediction ability can be obtained from the pedigree information in the model. Another advantage of the PBLUP model is that predictions can also be obtained for the pedigree parents that were not phenotyped or genotyped (Supplementary Table S3). In summary, our results show that genomic prediction for late blight resistance in potato is feasible and has added value, confirming the results from previous studies (Stich and Van Inghelandt 2018; Enciso-Rodriguez et al.

2018). The complete prediction results with the best-performing models are listed in Supplementary Table S4.

The CIP potato germplasm evaluated included many highly resistant genotypes based on the predicted rAUDPC values (GBLUPs) (Figure 3). The resistance of these genotypes is comparable to the resistant control genotype, which is released as a variety called “Chucmarina” in Peru and as “Belete” in Ethiopia (Figure 3). Notably, most of the genotypes tested in China were more resistant than the local variety “Cooperation-88” which has been popular for many years due to its good late blight resistance.

A biplot for predicted GBLUP values using the diploid marker-based kinship matrix shows the performance of the genotyped clones in all environments (Figure 4). In this figure, the genotypes that locate on the left side quadrants of the graph, and further away from the plot center, display the highest level of resistance. Genotypes near the horizontal mid-line on the left quadrants display stable resistance across environments, while genotypes toward the top and bottom extremes on the left quadrants display environment-specific resistance. Thus, some genotypes’ resistance to late blight is environment-specific, however, several genotypes show stable resistance across environments. The most resistant genotypes belong to the B3 and LB-HT populations, while only a few from the LTVR and B3-LTVR populations display high levels of resistance (Figure 4). The correlations among environments were high (Figure 5 and Table 8). The environment in Peru, especially, correlates highly with all the other environments, suggesting that resistant clones selected in Peru will also likely have good resistance in these other environments.

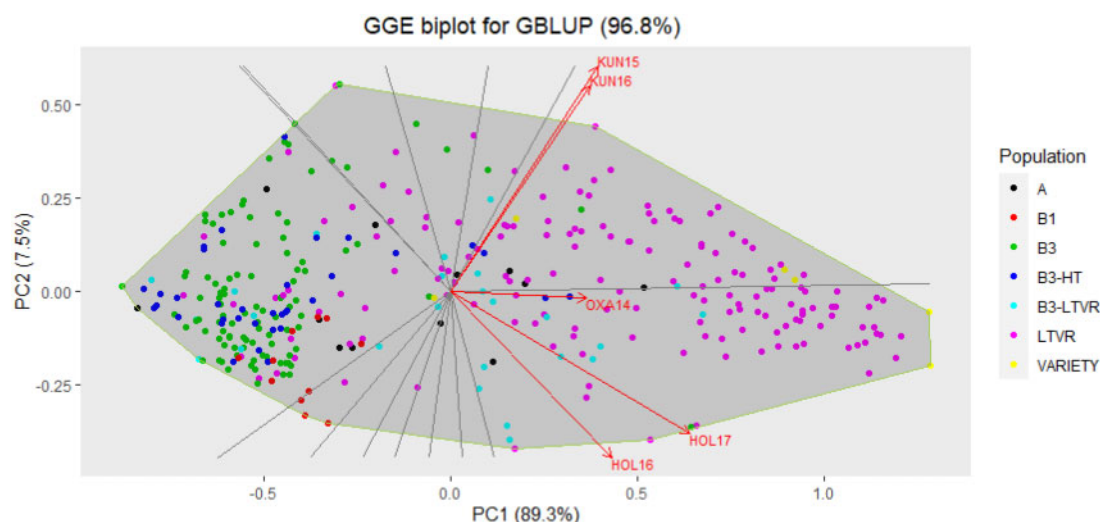
In the 1990s, CIP’s breeding strategy focused on improving the quantitative resistance in the “B3” population by phenotypic recurrent selection under endemic pressure from the *P. infestans* in the Peruvian Andes, supplemented by progeny tests to identify parents with good general combining ability and to eliminate those resulting in segregation for hypersensitive response against test isolates. The pathogen population in this area is currently dominated by the A1 mating type and EC-1 clonal lineage, which is highly aggressive and complex in its virulence (Lindqvist-Kreuz et al. 2020). Despite differences in the mating type and clonal lineages of the pathogen populations among the countries, it seems that phenotypic selection for late blight resistance in Peru was largely successful and the results are transferable across the environments tested here.

## Conclusions

METs for foliar late blight resistance with a set of advanced tetraploid potato clones bred at International Potato Center in Peru

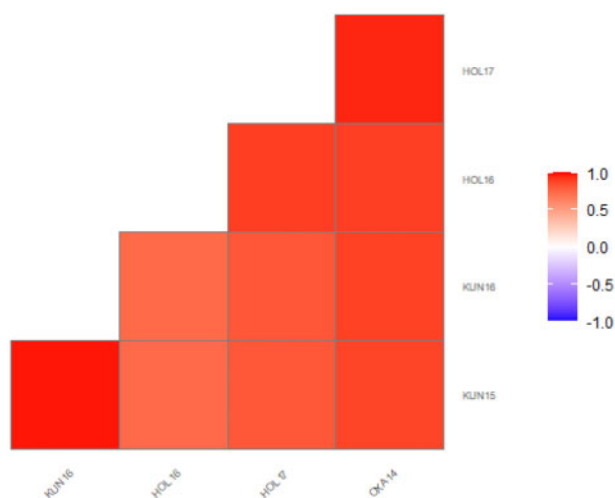
**Table 7** List of the ten most late blight resistant genotypes per environment based on the diploid GBLUP model with fixed QTL(xE) effects

Rank	Hol2016	Hol2017	Kun2015	Kun2016	Oxa2014
1	CIP396012.266	CIP398098.205	CIP301029.18	CIP395017.242	CIP398098.205
2	CIP393280.57	CIP398098.231	CIP399085.30	CIP396240.23	CIP398190.615
3	CIP396018.241	CIP398190.200	CIP393371.164	CIP395017.229	CIP396012.266
4	CIP301024.14	CIP396033.102	CIP399085.23	CIP393280.82	CIP398098.231
5	CIP396037.215	CIP393084.31	CIP395123.6	CIP396012.266	CIP394898.13
6	CIP392637.10	CIP398098.119	CIP395017.242	CIP393227.66	CIP384321.3
7	CIP396004.263	CIP393073.179	CIP399085.17	CIP301029.18	CIP398098.119
8	CIP384321.3	CIP393077.159	CIP384321.3	CIP384321.3	CIP301024.14
9	CIP393248.55	CIP398190.735	CIP391919.3	CIP392650.12	CIP398180.612
10	CIP304347.6	CIP398190.605	CIP399075.32	CIP395123.6	CIP398098.570



**Figure 4** GGE biplot for predicted late blight resistance [based on the diploid GBLUP model with fixed QTL(xE) effects] of the potato genotypes in China, Kunming 2015 (KUN15) and 2016 (KUN16); Peru, Oxapampa 2014 (OXA14), and Ethiopia, Holetta 2016 (HOL16) and 2017 (HOL17).

**Correlations of trials for GBLUP**



**Figure 5** Heat-plot on the GBLUP correlations among the field trials in China (KUN15, KUN16), Ethiopia (HOL16, HOL17), and Peru (OXA14) based on the diploid marker GBLUP model with fixed QTL(xE) effects.

identified QTL that were shared among environments and those that were environment-specific. QTL with the largest effect was identified in chromosome IX. Good predictive ability for the foliar LB resistance across environments was obtained using the BLUP model, and this was further improved by incorporating the pedigree and genotypic information in the mixed model. The best predictive ability was obtained with the GBLUP model for the diploid marker set using fixed QTL(xE) effects. The results show that many of the advanced clones bred in Peru for high levels of late blight resistance maintain their resistance in Ethiopia and China, suggesting that the centralized selection strategy has been largely successful.

## Data availability

The field books containing the phenotypic data used in the analysis are available in the CIP Dataverse (<https://data.cipotato.org/dataset.xhtml?persistentId=doi:10.21223/P3/JJJQV0>) and (<https://>

**Table 8** Estimated genetic correlations between the five environments obtained by fitting the diploid marker GBLUP model

	Hol2016	Hol2017	Kun2015	Kun2016	Oxa2014
Hol2016	1.00	0.67	0.52	0.51	0.80
Hol2017	0.67	1.00	0.53	0.51	0.82
Kun2015	0.52	0.53	1.00	0.94	0.76
Kun2016	0.51	0.51	0.94	1.00	0.76
Oxa2014	0.80	0.82	0.76	0.76	1.00

[data.cipotato.org/dataset.xhtml?persistentId=doi:10.21223/6TRC9T](https://data.cipotato.org/dataset.xhtml?persistentId=doi:10.21223/6TRC9T)). Population denominations and parentage of the potato genotypes are detailed in Supplementary Table S1. The genotypic best linear unbiased estimators (BLUEs) of rAUDPC for each environment are in Supplementary Table S2; the pedigree best linear unbiased predictors (PBLUPs) are in Supplementary Table S3; and the genomic best linear unbiased predictors (GBLUPs) using the model with the best prediction ability are in Supplementary Table S4. The genotypic data are available in variant calling format (vcf), in diploid format (10.6084/m9.figshare.12786398), and in tetraploid format (10.6084/m9.figshare.12789383).

## Acknowledgments

Thanks to Ms. Janet Higgins at Earlham Institute for helpful conversations on data analysis and assistance in bioinformatics. Designed experiments and obtained funding: H.L.K. and M.B. Implemented field trials and collected data: M.G., X.L., J.Q., G.W., B.H., Z.P., Q.S., K.N., and I.S. Analyzed and interpreted data: H.L.K., D.G., B.D.B., and P.U. Bioinformatics: H.L.K., J.D.V., and D.G. Statistics: B.D.B. Wrote the manuscript: H.L.K., P.U., M.B., and B.D.B. All authors accepted the final version of the manuscript.

## Funding

The work has received funding from GIZ through the project “Accelerating the Development of Early-Maturing-Agile Potato for Food Security through a Trait Observation and Discovery

Network” between 2015 and 2018. This work has also received funding from the CGIAR Research Program and Roots, Tubers and Bananas (RTB), and USAID. Global Challenges Research Fund—part of the UK AID budget, supported the secondment of H. Lindqvist-Kreuzer to Earlham Institute (EI) within the CABANA project. Bert De Boeck was co-funded by the Centre for International Migration and Development (CIM), a joint operation of GIZ and the German Federal Employment Agency on behalf of the German Federal Government.

## Conflicts of interest

None declared.

## Literature cited

- Bastien M, Boudhrioua C, Fortin G, Belzile F. 2018. Exploring the potential and limitations of genotyping-by-sequencing for SNP discovery and genotyping in tetraploid potato. *Genome*. 61: 449–456.
- Bethke PC, Halterman DA, Jansky S. 2017. Are we getting better at using wild potato species in light of new tools? *Crop Sci*. 57: 1241–1258.
- Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, et al. 2007. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*. 23:2633–2635.
- Bradshaw JE, Bryan GJ, Ramsay G. 2006. Genetic resources (including wild and cultivated *Solanum* species) and progress in their utilisation in potato breeding. *Potato Res*. 49:49–65.
- Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. 2013. Stacks: an analysis tool set for population genomics. *Mol Ecol*. 22: 3124–3140.
- Clevenger J, Chavarro C, Pearl SA, Ozias-Akins P, Jackson SA. 2015. Single nucleotide polymorphism identification in polyploids: a review, example, and recommendations. *Mol Plant*. 8:831–846.
- Cooke DEL, Lees AK. 2004. Markers, old and new, for examining *Phytophthora infestans* diversity. *Plant Pathol*. 53:692–704.
- Cullis BR, Smith AB, Coombes NE. 2006. On the design of early generation variety trials with correlated data. *J Agric Biol Environ Sci*. 11:381–393.
- D’hoop BB, Paulo MJ, Kowitwanich K, Sengers M, Visser RGF, et al. 2010. Population structure and linkage disequilibrium unravelled in tetraploid potato. *Theor Appl Genet*. 121:1151–1170.
- Elshire RJ, Glaubitz JC, Sun Q, Poland J, Kawamoto K, et al. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*. 6:e19379.
- Enciso-Rodriguez F, Douches D, Lopez-Cruz M, Coombs J, de Los Campos G. 2018. Genomic selection for late blight and common scab resistance in tetraploid potato *Solanum tuberosum*. G3 (Bethesda). 8:2471–2481.
- Felcher KJ, Coombs JJ, Massa AN, Hansey CN, Hamilton JP, et al. 2012. Integration of two diploid potato linkage maps with the potato genome sequence. *PLoS One*. 7:e36347.
- Fischer M, Schreiber L, Colby T, Kuckenberg M, Tacke E, et al. 2013. Novel candidate genes influencing natural variation in potato tuber cold sweetening identified by comparative proteomics and association mapping. *BMC Plant Biol*. 13:113–115.
- Forbes G, Perez W, Andrade Piedra J. 2014. Field Assessment of Resistance in Potato to *Phytophthora infestans*: International Cooperators Guide. Lima, Peru: International Potato Center.
- Foster SJ, Park TH, Pel M, Brigneti G, Śliwka J, et al. 2009. Rpi-vnt1.1, a Tm-22 homolog from *Solanum venturii*, confers resistance to potato late blight. *Mol Plant Microbe Interact*. 22:589–600.
- Gaiero P, Speranza P, de Jong H. 2018. Introgressive hybridization in potato revealed by novel cytogenetic and genomic technologies. *Am J Potato Res*. 95:607–621.
- Gebhardt C, Valkonen JP. 2001. Organization of genes controlling disease resistance in the potato genome. *Annu Rev Phytopathol*. 39: 79–102.
- Hamilton JP, Hansey CN, Whitty BR, Stoffel K, Massa AN, et al. 2011. Single nucleotide polymorphism discovery in elite North American potato germplasm. *BMC Genomics*. 12:1–12.
- Haynes KG, Christ BJ, Weingartner DP, Douches DS, Thill CA, et al. 2002. Foliar resistance to late blight in potato clones evaluated in national trials in 1997. *Am J Pot Res*. 79:451–457.
- Hirut B, Shimelis H, Fentahun M, Bonierbale M, Gastelo M, et al. 2017. Combining ability of highland tropic adapted potato for tuber yield and yield components under drought. *PLoS One*. 12: e0181541.
- Jiang R, Li J, Tian Z, Du J, Armstrong M, et al. 2018. Potato late blight field resistance from QTL dPI09c is conferred by the NB-LRR gene R8. *J Exp Bot*. 69:1545–1555.
- Kamvar ZN, Tabima JF, Grünwald NJ. 2014. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*. 2:e281.
- Koenker R, Portnoy S, Ng PT, Zeileis A, Grosjean P, et al. 2018. Package ‘quantreg’. Reference Manual. R-CRAN: <https://cran.rproject.org/web/packages/quantreg/quantreg.pdf>.
- Koenker R., Portnoy S., Ng P. T., Zeileis A., Grosjean P., et al. 2018. Package ‘quantreg’. Reference manual available at R-CRAN: <https://cran.rproject.org/web/packages/quantreg/quantreg.pdf>.
- Li Y., Van der Lee T., Zhu J. H., Jin G. H., Lan C. Z., et al. 2013. Population structure of *Phytophthora infestans* in China—geographic clusters and presence of the EU genotype Blue\_13. *Plant Pathology*. 62:932–942.
- Li XP, Qin J, Jiang W, Tiang Y, Bao L, et al. 2018. Evaluation and screening of international potato center imported resources in Yunnan. In: Chinese Potato Essays. Harbin Cartographic Publishing House. p. 174–177.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*. 25:1754–1760.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics*. 25: 2078–2079.
- Li J, Lindqvist-Kreuzer H, Tian Z, Liu J, Song B, et al. 2012. Conditional QTL underlying resistance to late blight in a diploid potato population. *Theor Appl Genet*. 124:1339–1350.
- Li L, Paulo MJ, van Eeuwijk F, Gebhardt C. 2010. Statistical epistasis between candidate gene alleles for complex tuber traits in an association mapping population of tetraploid potato. *Theor Appl Genet*. 121:1303–1310.
- Lindqvist-Kreuzer H, Gastelo M, Perez W, Forbes GA, de Koeijer D, et al. 2014. Phenotypic stability and genome-wide association study of late blight resistance in potato genotypes adapted to the tropical highlands. *Phytopathology*. 104:624–633.
- Lindqvist-Kreuzer H, Gamboa S, Izarra M, Pérez W, Correa MY, et al. 2020. Population structure and host range of the potato late blight pathogen *Phytophthora infestans* in Peru spanning two decades. *Plant Pathol*. 69:334–346.
- Malosetti M, Ribaut JM, van Eeuwijk FA. 2013. The statistical analysis of multi-environment data: modeling genotype-by-environment interaction and its genetic basis. *Front Physiol*. 4:44.



- Matias FI, Xavier Meireles KG, Nagamatsu ST, Lima Barrios SC, Borges do Valle C, et al. 2019. Expected genotype quality and diploidized marker data from genotyping-by-sequencing of *Urochloa* spp. Tetraploids. *Plant Genome*. 12:1–9.
- Mihretu E, Mohammad W, Kassa B, Lindqvist-Kreuze H. 2020. Virulence Spectrum of *Phytophthora infestans* and Spatial Distribution of Physiological Races in Northwestern Ethiopia. *Ethiopian J Agric Sci*. 30:69–85.
- Mosquera T, Alvarez MF, Jiménez-Gómez JM, Muktar MS, Paulo MJ, 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.
- Möhring J, Piepho HP. 2009. Comparison of weighting in two-stage analysis of plant breeding trials. *Crop Sci*. 49:1977–1988.
- Muhinyuza JB, Shimelis H, Melis R, Sibiya J, Gahakwa D, et al. 2015. Yield response and late blight reaction of potato genotypes in Rwanda. *Am J Potato Res*. 92:10–22.
- Ovchinnikova A, Krylova E, Gavrilenko T, Smekalova T, Zhuk M, et al. 2011. Taxonomy of cultivated potatoes (*Solanum* section *Petota*: *Solanaceae*). *Bot J Linn Soc*. 165:107–155.
- Pel MA, Foster SJ, Park TH, Rietman H, van Arkel G, et al. 2009. Mapping and cloning of late blight resistance genes from *Solanum venturii* using an interspecific candidate gene approach. *Mol Plant Microbe Interact*. 22:601–615.
- Piepho HP. 1998. Empirical best linear unbiased prediction in cultivar trials using factor-analytic variance-covariance structures. *Theor Appl Genet*. 97:195–201.
- Piepho HP, Möhring J, Schulz-Streeck T, Ogutu JO. 2012. A stage-wise approach for the analysis of multi-environment trials. *Biom J*. 54: 844–860.
- Poland JA, Brown PJ, Sorrells ME, Jannink JL. 2012. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS One*. 7: e32253.
- Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, et al. 2018. Scaling accurate genetic variant discovery to tens of thousands of samples. *BioRxiv*. 201178.
- Potato Genome Sequencing Consortium 2011. Genome sequence and analysis of the tuber crop potato. *Nature*. 475:189.
- Rosyara UR, De Jong WS, Douches DS, Endelman JB. 2016. Software for genome-wide association studies in autopolyploids and its application to potato. *Plant Genome*. 9: doi: 10.3835/plantgenome2015.08.0073.
- Sharma SK, MacKenzie K, McLean K, Dale F, Daniels S, et al. 2018. Linkage disequilibrium and evaluation of genome-wide association mapping models in tetraploid potato. *G3 (Bethesda)*. 8: 3185–3202.
- Sharma SK, Bolser D, de Boer J, Sønderkær M, Amoros W, et al. 2013. Construction of reference chromosome-scale pseudomolecules for potato: integrating the potato genome with genetic and physical maps. *G3 (Bethesda)*. 3:2031–2047.
- Spooner DM, Ghislain M, Simon R, Jansky SH, Gavrilenko T. 2014. Systematics, diversity, genetics, and evolution of wild and cultivated potatoes. *Bot Rev*. 80:283–383.
- Stich B, Van Inghelandt D. 2018. Prospects and potential uses of genomic prediction of key performance traits in tetraploid potato. *Front Plant Sci*. 9:159.
- Stich B, Urbany C, Hoffmann P, Gebhardt C. 2013. Population structure and linkage disequilibrium in diploid and tetraploid potato revealed by genome-wide high-density genotyping using the SolCAP SNP array. *Plant Breed*. 132:718–724.
- Sun HH, Ye G, Wang F, Zhou Y, Wang J. 2010. Field identification the resistance of potato germplasm from CIP for late blight in Qinghai. *Acta Agriculturae Boreali-Occidentalis Sinica*. 19:68–70.
- Sverrisdóttir E., Byrne S., Sundmark E. H. R., Johnsen H. Ø., Kirk H. G., et al. 2017. Genomic prediction of starch content and chipping quality in tetraploid potato using genotyping-by-sequencing. *Theoretical and Applied Genetics*, 130:2091–2108.
- Swiezynski KM, Domański L, Zarzycka H, Zimnoch-Guzowska E. 2000. The reaction of potato differentials to *Phytophthora infestans* isolates collected in nature. *Plant Breeding*. 119:119–126.
- Tiwari J.K., Siddappa S., Singh B.P., Kaushik S.K., Chakrabarti S.K., et al. 2013. Molecular markers for late blight resistance breeding of potato: an update. *Plant Breeding*, 132:pp.237–245.
- Uitdewilligen JG, Wolters AMA, Bjorn B, Borm TJ, Visser RG, et al. 2013. A next-generation sequencing method for genotyping-by-sequencing of highly heterozygous autotetraploid potato. *PLoS One*. 8:e62355.
- Van Eeuwijk FA, Bustos-Korts DV, Malosetti M. 2016. What should students in plant breeding know about the statistical aspects of genotype × environment interactions? *Crop Sci*. 56:2119–2140.
- Vos PG, Paulo MJ, Voorrips RE, Visser RG, van Eck HJ, et al. 2017. Evaluation of LD decay and various LD-decay estimators in simulated and SNP-array data of tetraploid potato. *Theor Appl Genet*. 130:123–135.
- Vos PG, Uitdewilligen JG, Voorrips RE, Visser RG, van Eck HJ. 2015. Development and analysis of a 20K SNP array for potato (*Solanum tuberosum*): an insight into the breeding history. *Theor Appl Genet*. 128:2387–2401.
- Vossen JH, van Arkel G, Bergervoet M, Jo KR, Jacobsen E, et al. 2016. The *Solanum demissum* R8 late blight resistance gene is an *Sw-5* homologue that has been deployed worldwide in late blight resistant varieties. *Theor Appl Genet*. 129:1785–1796.
- Zhang X-Z, Kim B-S. 2007. Physiological races of *Phytophthora infestans* in Korea. *Plant Pathol J*. 23:219–222.

Communicating editor: J-L. Jannink