INTEGRATING GENOMIC SELECTION AND GENOME EDITING STRATEGIES TO ACCELERATE POTATO BREEDING

By

Felix Eugenio Enciso Rodriguez

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Plant Breeding, Genetics and Biotechnology – Crop and Soil Sciences – Doctor of Philosophy

2019

ABSTRACT

INTEGRATING GENOMIC SELECTION AND GENOME EDITING STRATEGIES TO ACCELERATE POTATO BREEDING

By

Felix Eugenio Enciso Rodriguez

As a staple food, the potato (*Solanum tuberosum* L.) plays an important role in human nutrition and it is currently the third most important food crop after rice and wheat. However, the potato crop faces high production losses caused mainly by biotic factors. With the advent of cutting-edge technologies suitable for potatoes, there is an increasing possibility to accelerate genetic progress and variety generation. To contribute to the implementation of genomic strategies to accelerate potato breeding, three different approaches were used. First, whole genome regressions were conducted using additive and dominant allele dosage models for late blight and common scab resistance in tetraploid potatoes. Multiple Single Nucleotide Polymorphisms (SNPs), contribute to late blight resistance, uncovering the introgression history for this trait whereas an unreported locus with a sizable contribution to common scab resistance was detected. Prediction accuracy assessments demonstrated that 90% of the genetic variance could be captured with an additive model, demonstrating the applicability of genomic prediction for tetraploid potato breeding. Second, a genome editing approach was implemented to breakdown the *S-RNase*-based self incompatibly in diploid potatoes. New *S-RNase* allelic variants, with flower-restricted expression, were identified in two self-incompatible (SI) diploid potatoes and mapped to chromosome I in a low recombination region. A dual single-guide RNA strategy was used to generate *S-RNase* knock-out lines producing premature stop codons on each targeted *S-RNase* allele. Self-compatibility was achieved in T_0 knock-outs and stable transmitted to T_1 lines. Additionally, Cas-9 free plants were also obtained. Plasticity in the self-compatible response was

also observed in wild-type lines, presumably associated with non-stylar and environmental factors. Third, validation of the *IPI-O4*-mediated suppression of the *RB*-based late blight resistance was conducted using *in vivo* and *in vi*tro approaches. The hypersensitive response (HR) was confirmed when *IPI-O1* was co-infiltrated with the *RB* gene from *Solanum bulbocastanum* using a heterologous system. However, HR was observed when *IPI-O1* and *IPI-O4* were infiltrated in transgenic potato lines carrying a synthetic *RB* gene containing a Coiled-Coil (CC) domain from *S. pinnatisectum*. Further work should be conducted to confirm this un-reported interaction. Similarly, we could not validate CC-dimerization using yeast-two hybrid assays and therefore more extensive experiments should be conducted to confirm this result. Ultimately, these genomic approaches open a new window to accelerate the generation of new potato varieties. Genomic selection strategies along with targeted mutagenesis will expand the boundaries of both approaches, reducing the potato breeding cycle considerably while maintaining genetic diversity, and providing access to genomic regions with low or null recombination in potatoes.

Dedicated to my Grandfather

ACKNOWLEDGMENTS

I want to offer my special thanks to Dr. David Douches, for his mentoring and valuable constructive suggestions during the planning and development of this research.

I would also like to thank my graduate committee, Drs. David Douches, Gustavo de los Campos, Robin Buell and Noah Rosenzweig for accepting being part of my Ph.D. committee and their advice and assistance. Especially, I would like to thank Dr. Gustavo de los Campos for his valuable contribution and support for the development of the genomic selection project. Similarly, I would like to thank Dr. Robin Buell for her guidance and valuable contribution to the self-incompatibility project.

I want to extend my thanks to Drs. Swathi Nadakuduti, Daniel Zarka, Norma Manrique-Carpintero, and Mr. Joseph Coombs, for their valuable help and contribution to each aspect of my dissertation projects.

I wish to thank to various potato people from the MSU potato breeding program; Chen Zhang, Kate Shaw and Matt Zuehlke for their valuable support in the greenhouse and laboratory; Monica Hufnagel and Bernard Schroeter, for their assistance in tissue culture. I wish to thank Chen Zhang, Andrea Garcia, Swathi Nadakuduti, Norma Manrique-Carpintero, Maher Alsahlany, Kate Shaw, Natalie Kaiser, Donna Kells, Thylani Jayakody, Kelly Zarka, Juan Sandoval, Julio Valenzuela, Sandra Hernandez, Gina Garzón, Jaime Osorio and Yaneth Camargo for their friendship and support.

I would like to thanks to Colciencias Fulbright – Fulbright for sponsoring me, and the Colombian Corporation for Agricultural Research (AGROSAVIA) for providing me permission to pursue

my Ph.D., especially to Dr. Luz Stella Barrero for her support and encouraging me to continue my graduate program.

Finally, I would like to thank my family for all their valuable support to make this dream possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

ADVANCES IN GENOMIC STRATEGIES AND APPLICATIONS IN POTATO BREEDING

I. Potato origin and domestication

Potato (*Solanum tuberosum* L., Group Tuberosum) represents one of the most economically important species within the Solanaceae family along with tomato, eggplant, pepper, tobacco and petunia (Wu and Tanksley, 2010; Olmstead et al., 2008). The center of origin of the potato covers the Peruvian and Bolivian Andes where it evolved as a result of the hybridization of *Solanum stenotonum* complex with an unknown species (Grun, 1990). From there, the short-day sub-species *S. tuberosum* Group Andigena spread across the Andes, resulting later in the emergence of the long-day sub-species *S. tuberosum* Group Tuberosum (Hosaka, 2003; Hosaka and Hanneman, 1988).

Potato domestication started 3400 calendar years before the present (cal BP) in the western Titicaca basin, contributing to the transition from nomadism to sedentarism of the first settlements in this region (Rumold and Aldenderfer, 2016). After the Spanish colonization, the potato was introduced to Europe in the 16th century (Hawkes and Francisco-Ortega, 1993). Later, *S. tuberosum* Group Tuberosum was introduced, replacing old varieties to face the damages caused by late blight, one of the most devastating diseases affecting potato (Ames and Spooner, 2008; Hosaka and Hanneman, 1988). From there, the potato spread across Europe, Africa, and Asia and later were re-introduced to North America by the British colonists in the 17th century (Hawkes, 1992).

Currently, potatoes are the third most important food crop after rice and wheat (Devaux et al., 2014). In 2017, potatoes reached a global production of 388 million tons, positioning the United

States as the fifth largest producer with more than 20 million tons (FAOSTAT, 2019). As a staple food, potato plays an important role in human nutrition. Along with beans, potato is considered the cheapest source of fiber, minerals and vitamins necessary for child nutrition (Drewnowski and Rehm, 2013). It is also considered one of the main sources of carbohydrates (Kolasa, 1993). Besides its nutritional value, potato has high concentrations of phenolic compounds such as chlorogenic acid, opening a new market for potato's commercialization as a functional food due to its potential anti-oxidative activity (Friedman, 1997; Vinson et al., 2012).

II. Potato genetics and genomics-based resources

Among the angiosperms, the Solanaceae represents one of the major families, with more than 3000 species distributed worldwide (Wu and Tanksley, 2010). Within this family, *Solanum* is the biggest and most diverse genus, with more than 1250 species from shrubs to small trees (Weese and Bohs, 2007, Magoon et al., 1962). *Solanum* is classified into different sections, in which the *Petota* section is the most prominent group (Spooner et al., 1991). This section constitutes about 100 tuber-bearing species, ranging from wild relatives to cultivated potato varities distributed from United States to Chile (Ovchinnikova et al., 2011). Its members are predominantly diploid (2x=2n=24), although tetraploid (2n=2x=48) and hexaploid (2n=6x=72) species are also present (Huamán and Spooner, 2002).

Extensive genomic resources have been developed to decipher potato diversity and complexity. Genotyping platforms including 8K-20K Infinium Single Nucleotide Polymorphism (SNPs) arrays (Felcher et al., 2012b; Vos et al., 2015) and Genotyping-by-Sequencing (GBS) approaches have been widely used for potato germplasm characterization and allele variants discovery through Quantitative Trait Loci (QTL) mapping and Genome-Wide Association

(GWA) analysis (Bastien et al., 2018; Boudhrioua et al., 2017; Ellis et al., 2018; Endelman and Jansky, 2016; Manrique-Carpintero et al., 2015; Massa et al., 2015).

Cultivated potato *S. tuberosum* is a heterozygous tetraploid species. Therefore, to reduce the complexity The Potato Genome Sequencing Consortium used a homozygous doubled monoploid, *S. tuberosum* Gp. Phureja line 1-3 516 R44 (DM) to generate a genome assembly (The Potato Genome Sequencing Consortium, 2011). This reference genome sequence was further improved by integrating physical and genetic maps from the heterozygous diploid lines: RH 89-039-16 *S. tuberosum* Gp. Tuberosum (RH); D84 (an *S. tuberosum* × *S. chacoense* hybrid) and DRH (DM × RH hybrid). A total of 674 Mb from the 727 Mb of the DM assembly were integrated, including 37,482 of the 39,031 predicted genes (Felcher et al., 2012a; Sharma et al., 2013). To date, more than 200 potato accessions including wild and cultivated relatives from the *Petota* section have been sequenced (Hardigan et al., 2016; Li et al., 2018). In particular, sequencing data from species such as *S. americanum*, *S. chacoense*, *S. commersonii*, *S. verrucosum* among others (Aversano et al., 2015; Leisner et al., 2018; Lister et al., 2019; Paajanen et al., 2017), offer a new resource for mining disease-resistance genes and tuber-related traits.

Whole-genome sequencing projects have contributed to elucidating the genetic variation landscape in diploid and tetraploid potatoes. The high degree of structural variation in potato, revealed by genomic and cytological analysis, significantly contributes to the high genomic diversity of this species (Hardigan et al., 2016; Iovene et al., 2013). In particular, potato heterozygosity and genetic load are driven by multiallelism associated with multiple polymorphisms in coding regions and Copy Number Variants (CNV), respectively (Pham et al., 2017). Genomic analysis also demonstrated that potato domestication led to differential allele

expression and accumulation of deleterious mutations in tetraploid potatoes, this last compensated with the presence of heterozygous alleles, masking unfavorable mutations (Lian et al., 2019; Pham et al., 2017).

Comprehensive gene expression atlases have been generated to understand the dynamics of key regulatory genes associated with physiological responses under different biological and environmental treatments in potato. Transcriptomic profiles from DM using up to 32 vegetative and reproductive tissues under different growth and stress conditions detected more than 22,000 differential expressed genes (DEG) and co-expression gene modules associated to specific stress conditions (Massa et al., 2011, 2013). Additional transcriptomic profiling studies have uncovered key genes involved in metabolic pathways related to drought, salt tolerant and environmental-stress adaptation responses (*i.e*., auxin response factors and nitrogen supplementation) using *de novo* or available expression repositories (Charfeddine et al., 2015a, 2015b; Gálvez et al., 2016; Gong et al., 2015; Song et al., 2019; Sprenger et al., 2016).

Targeted and non-targeted metabolite profiles from various tissues and agronomic-related traits complement these transcriptomic atlases in potatoes. The chemical composition of potato leaves, tubers and resistance-related compounds have elucidated important metabolic pathways associated with the synthesis of carbohydrates, glycoalkaloid, vitamins and bioactive compounds such as hydroxycinnamic acid (Chaparro et al., 2018; Mariot et al., 2016; Piñeros-Niño et al., 2017; Rodríguez-Pérez et al., 2018; Tomita et al., 2017; Yogendra et al., 2015). Additional transcriptomic and metabolomic repositories are also available for autotetraploids derived from the wild relatives *S. commersonii* and *S. bulbocastanum,* representing a new resource for polyploidization and gene discovery studies (Fasano et al., 2016).

The aforementioned approaches have enabled researchers to identify, validate and reconstruct historical introgressions of adaptation-related traits such as plant maturity. For instance, association analyses and homologous searching of *S. tuberosum* Gp Tuberosum RH 89-039-16 (RH) genomic sequences revealed a major QTL controlled by the *S. tuberosum* Cycling Dof Factor (*StCDF* gene), located in Chromosome 5 (Collins et al., 1999; Kloosterman et al., 2013). Functional analysis confirmed the role of *StCDF* in plant maturity through complementation assays using the short-day *S. tuberosum* Group Andigenum (Kloosterman et al., 2013). Further investigations revealed the role of *StCDF* in the short-day to long-day transition through the introgression of this gene from diploid landraces into cultivated tetraploid potatoes (Hardigan et al., 2017). Together, these genomic resources offer a valuable tool for gene mining, evolutionary studies, germplasm characterization and potato breeding.

III. Vision for diploid potato breeding – Advantages and Challenges

Conventional breeding and genetic analyses are challenging in cultivated potato due to its tetraploidy, heterozygosity and vegetative mode of propagation. Re-inventing potato as a diploid inbred/F1 hybrid variety ($2n=2x=24$) would allow the application of efficient breeding methods (Jansky et al., 2016). For instance, the development of inbred diploid potatoes not only would accelerate the generation of new varieties with favorable allele combinations targeting yield, tuber quality, and disease resistance traits but will also aid in understanding the genetics underlying these traits. Diploid based approaches will enable increases in genetic gains per breeding cycle (Jansky et al., 2016). Compared with major staple crops such as maize and rice, in which heterosis has contributed significantly to their increase in global production (Fu et al., 2014), production increases have been non-significant in potato under the current breeding

system during the past decades (Figure 1.1). Therefore, diploid potato breeding could achieve comparable gains by exploiting heterosis supported with new breeding strategies.

Improvement of cultivated tetraploid potatoes also relies on the discovery and introgression of genes from wild species, especially for traits related to biotic and abiotic stresses. However, the introgression of critical dominant alleles in a triplex or quadruplex allelic configuration can take up to 15 years in a tetraploid breeding scheme (Mendoza et al., 1996). Moreover, vegetative propagation favors disease transmission (i.e., viruses), which requires the production of disease-free tubers increasing production costs (Simmonds, 1997).

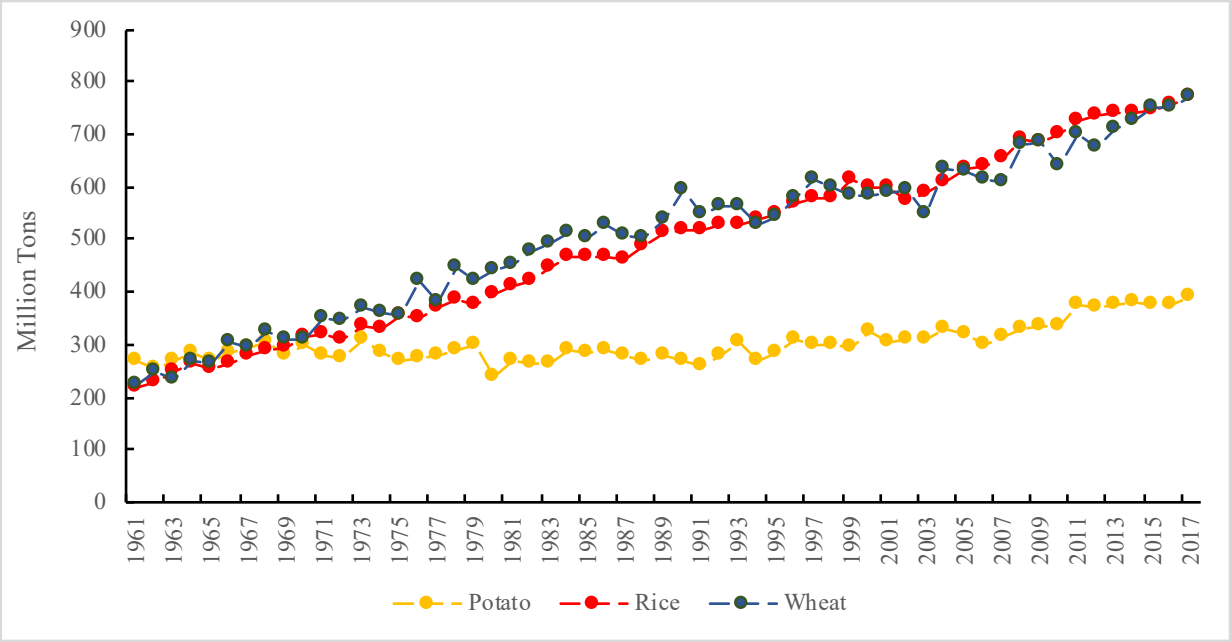

Figure 1.1. *Annual potato production from 1961 to 2017*. No significant increases in potato production has been observed when compared with the rice and wheat. Source: FAOSTAT (March 8, 2019).

Moreover, diploid potatoes will enable the commercialization of this crop through true seed instead of seed tubers for propagation, reducing transportation costs as well as disease transmission of bacterial and oomycete pathogens (Jansky et al., 2016). Additionally, genomics tools developed in other diploid species could now be accessible for potato. Allele dosage estimation, quantitative trait loci detection, breeding value estimation among others will be more feasible and accurate under a diploid system (Jansky et al., 2016).

One of the major challenges for future diploid breeding programs is male sterility. This phenomena occurs when *Solanum* relatives are used as males and crossed to *S. tuberosum*-derived dihaploids lines (Carroll, 1975; Jansky et al., 2016; Jansky and Peloquin, 2006), representing a limiting factor for diploid hybrid production. Linkage drag represents another major drawback for introgressive hybridization when using wild-potato relatives as donors (Gaiero et al., 2018). For instance, introgression of resistance-related genes into cultivated diploid germplasm could lead to fixation of undesirable traits such as high tuber glycoalkaloid content from donors such as *S. chacoense.* However, the gametophytic self-incompatibility (SI) represents one of the most significant barriers to diploid breeding, preventing the ability to generate diploid homozygous lines.

i. Self-incompatibility in diploid potatoes

In Solanaceae, a single multiallelic *S*-locus governs self-pollen rejection under the gametophytic SI system. This locus is composed of tightly linked genes, *SLF* (*S*-locus F-box) and *S-RNase* (*S*-locus RNase), expressed in the pollen and the style, respectively (McClure et al., 1989; Takayama and Isogai, 2005). The S-RNase protein produces cytotoxic effects that inhibit the elongation of self-pollen tubes via degradation of RNA from the pollen whereas *SLF* functions as

a component of a detoxification complex that mediates ubiquitination of non-self S-RNase proteins leading to degradation via the proteasome pathway (Kubo et al., 2015; Sijacic et al., 2004). Hence, when self-pollination occurs in self-incompatible individuals, *SLF* does not recognize its native *S-RNase*, and consequently, pollen tube growth in the style is inhibited due to the ribonuclease activity of the *S-RNase* (Hua et al., 2008).

Most species under the GSI system contain multiple *SLFs* to recognize a broader set of non-self *S-RNases* (Figure 1.2A) (Fujii et al., 2016; Kubo et al., 2010). This collaborative non-self-recognition system is under the control of the *S*-haplotype-specific F-box brothers (SFBB) as result of *SLF* inter-haplotype genetic exchanges (Kubo et al., 2015; Sassa et al., 2007). SFBB proteins provide an evolutionary advantage for outcrossing, increasing the number of potential mating partners, but also contributing to a self-compatibility transition recognizing self-pollen (Fujii et al., 2016; Kubo et al., 2010, 2015; Sassa et al., 2007).

Although GSI gene determinants confer specificity for pollen rejection, modifier genes are also required to modulate the SI response (Goldraij *et al.*, 2006; McClure *et al.*, 1999; O'Brien *et al.*, 2002). For instance, pollen determinant modifiers such as the *S-RNase* binding protein 1 (SBP1), a ring domain protein, form a complex with SLF, interacting with *S-RNase* conserved regions between hypervariable domains, leading to non-self *S-RNase* degradation (O'Brien et al., 2004; Sassa et al., 2007). Similarly, stylar factors such as the 120-KD arabinogalactan (120K) and Kunitz-type proteinase inhibitors (*NaStEP*) play an essential role in the pistil endomembrane protein trafficking and *S-RNase* toxicity (Figure 1.2B) (Goldraij et al., 2006; Jiménez-Durán et al., 2013; Lee et al., 2009).

Two different strategies have been used to develop diploid-self-compatible (SC) potato lines. The inbred line M6, generated from the wild tuber-bearing species, *S. chacoense* (Jansky et al.,

2014) harbors a dominant allele of the *S*-locus inhibitor (*Sli*) that inactivates the gametophytic SI system (Hosaka and Hanneman, 1998) leading to self-compatibility. However, introgression of *Sli* into other germplasm is time-consuming and could lead to linkage drag and fixation of undesirable traits such as high tuber glycoalkaloid content from the donor *S. chacoense*. Similarly, a targeted mutagenesis approach has been used to target the *S-RNase* gene, successfully generating self-compatible potato lines (Ye et al., 2018).

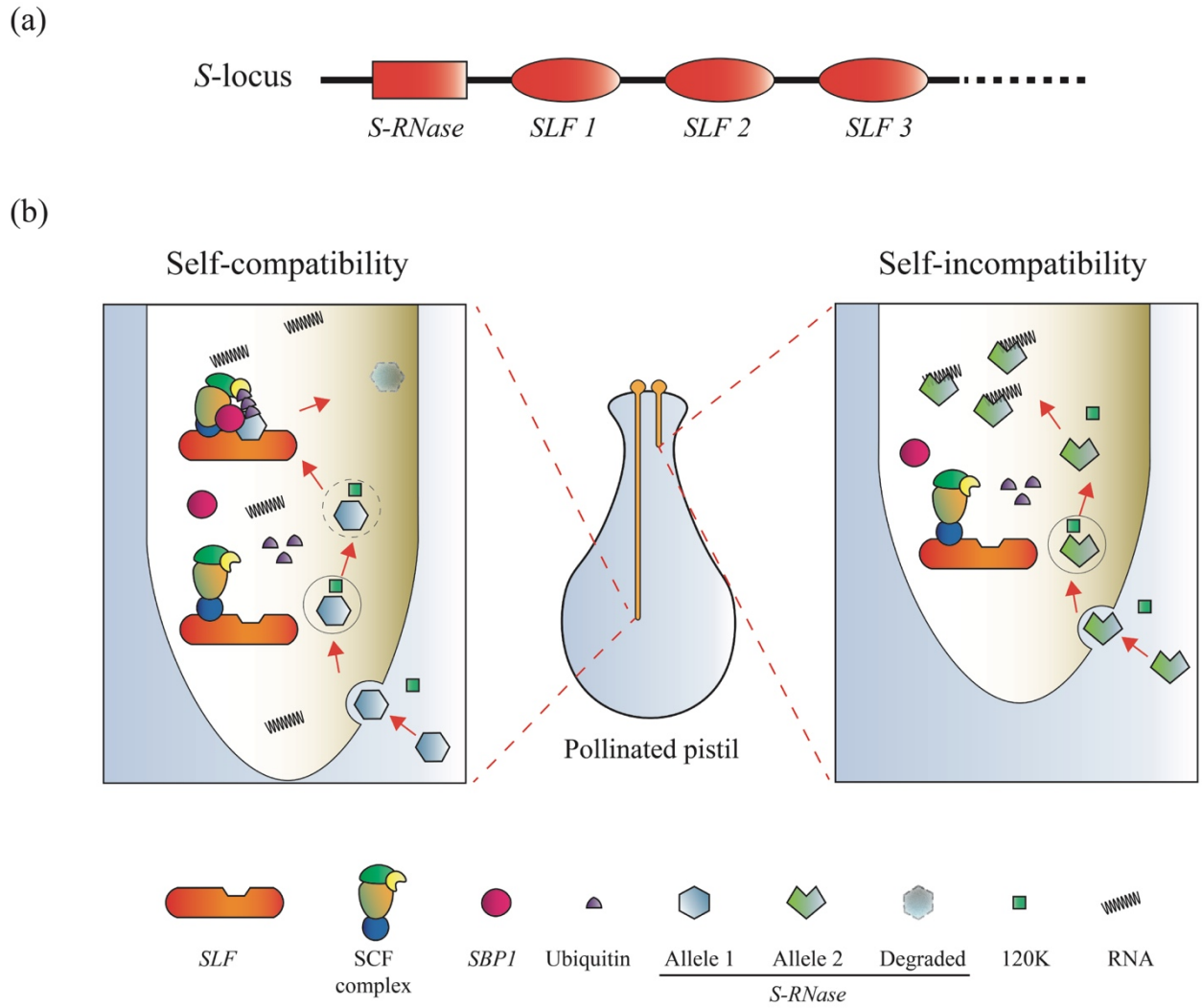

Figure 1.2. *Gametophytic self-incompatibility in Solanaceae.* (a) The *S*-locus is composed of tightly linked genes, *S-RNase* (*S*-locus RNase) and multiple *SLFs* (*S*-locus F-box proteins),

which are expressed in the style and pollen, respectively. (b) After pollination, *S-RNase* is imported into pollen tubes along with proteins such as 120K, which is required for their ability to reject incompatible pollen. In self-compatible reactions, the N terminal F-box domain from SLF recognize components of the SCF E3 ubiquitin ligase complexes. Likewise, SBP1 interact with the S-RNase N-terminal region and E3 ubiquitin ligases forming a complex with SLF-SCF, leading to the degradation of non-self *S-RNases*. In self-incompatible reactions, the SLF-SCF complex is not able to recognize its own *S-RNase* leading to the degradation of RNA involved in self-pollen tube development by the ribonuclease activity of *S-RNase*.

IV. Current production constraints in cultivated tetraploid potatoes

Despite its significant economic importance, potato faces production losses, with 1,331 tons in the United States alone (National Potato Council, 2016). While abiotic factors such as heat and drought can be one of the major limiting components in potato production, affecting the growing cycle and reducing tuber yield and quality, biotic factors have the most devastating impacts over this crop (George et al., 2017; Kromann et al., 2014). Pests and pathogens have a severe effect on sustainable potato production, generating production losses in field and storage (Kromann et al., 2014). The most recent data 2001-2003 (though it is now nearly 20 years old) shows that pathogens caused 14.5% of global potato production losses, followed by pests with 10.9 %, weeds with 8.3 and viruses with 6.6%, and it is estimated that without pest control total production losses could reach up to 75% (Oerke, 2006). Pathogens such as late blight and *Alternaria* are the most devastating and widely distributed constraints on potato, along with the viruses like potato leafroll luteovirus and potato potyvirus Y. Animal pests like potato cyst nematodes and the Colorado beetle are significant as well (Oerke, 2006). Among potato

pathogens, common scab and late blight are two primary production constraints faced by potato growers, causing up to 100% of losses, and reducing the quality and marketability of the tuber.

Although pathogen infection can be controlled by using protectants or systematic fungicides in potato, they can be ineffective if environmental conditions favor pathogen dispersion or because of the emergence of fungicide-resistant genotypes (Nowicki et al., 2011; Pomerantz et al., 2014). The most effective way to control the incidence of common scab and late blight in potato is through the generation of resistant varieties (Ahn and Park, 2013) and therefore this research seeks to contribute to the generation of resistant potato varieties using different genomic approaches, contributing to the reduction of the potato breeding cycle.

i. Common scab (Streptomyces scabies Thaxter)

Common scab is a soil-borne pathogen that reduces potato quality and marketability by causing superficial lesions on the tuber surface (Dees and Wanner, 2012). This pathogen, previously considered a fungus given its filamentous morphology, is a gram-positive bacteria which produces spores that allow it to survive on different substrates (Braun et al., 2017). Common scab penetrates young tubers through lenticels, spreading rapidly in growing plants cells, facilitated by the phytotoxin thaxtomin produced by this pathogen (Braun et al., 2017). This pathogen has a worldwide distribution, and due to its saprophytic 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). Susceptibility to common scab is dependent upon genotype, growing season and environmental conditions (Wanner, 2006; Wanner and Kirk, 2015), having a negative impact mainly in underground tissues in development, such as stolons and tubers.

ii. Late blight [Phytophthora infestans (Mont.) de Bary]

Late blight is known for being the causal agent of the Irish potato famine in the 19th century, leading to over a million human deaths (Grada, 2004). *infestans* a hemibiotrophic oomycete, originated in central Mexico and later spread all over the world (Grünwald and Flier, 2005). At the first stage of infection, late blight forms an asexual multinucleate sporangium or mononucleate zoospores produced by cleaved sporangia, entering into the host via stomata, wounds or appressorium formation in vegetative tissues (Leesutthiphonchai et al., 2018). Late blight also forms sexual spores known as oospores, that require two mating types (A1 and A2). Through sexual recombination, this mating system increases pathogen diversity when both mating types are present in a population, favoring late blight adaptation to extreme environmental conditions (Smart and Fry, 2001). Consequently, due to its ability to use both sexual and asexual reproduction, using wind and rain as a means of dispersal, *P. infestans* kills an entire plant within 7 to 10 days, and is responsible for annual losses of 16% to global potato production (Haverkort et al., 2009).

iii. Late blight-potato arms race

Plants lack an adaptive immune system, and therefore disease resistance relies on the recognition of molecular signatures associated with pathogens. Plant cells display a coordinated system to recognize threats through extramembrane receptors coupled with intramembrane kinases (Couto and Zipfel, 2016). These Pattern Recognition Receptors (PRRs) perceive pathogen-associated molecular patterns (PAMPs) or host-derived damage-associated molecular patterns (DAMPs), inducing a signaling cascade that activates the expression of defense-related genes, a process known as PRR-triggered immunity (PTI) (Couto and Zipfel, 2016; Jones and

Dangl, 2006).

However, fast-evolving pathogens promptly respond to PTI, translocating virulent effectors into the host cell, mostly via the Type Three Secretion System (TTSS), suppressing PTI. In turn, plants display an arsenal of intracellular receptors encoded by R genes, inducing a second layer of resistance response known as Effector-Triggered Immunity (ETI) upon sensing or recognition of pathogen effectors (Couto and Zipfel, 2016; Jones and Dangl, 2006). These multidomain protein receptors are composed of a Nucleotide-binding domain Leucine-rich Repeat (NLR) core with either a Coiled-Coil (CC) or a Toll Interleukin-1-like receptor (TIR) domain at the N terminal region (Sukarta et al., 2016). This host-pathogen arms race has shaped the genetic architecture of plant disease resistance. Previously, it was widely accepted that plant resistance responses were associated with a gene-by-gene or R-gene-mediated (Flor, 1971). However, current models suggest that plant pathogen resistance results from a dynamically co-evolving process involving resistance breakdown and recovery, known as the zig-zag model (Cook et al., 2015; Jones and Dangl, 2006). Understanding plant-pathogen interaction may provide a mechanism to counteract host resistance suppression in potato by modifying effector targeted sites, enhancing R gene-mediated resistance.

iv. Genetic architecture of late blight resistance in potatoes

The outcrossing nature of potato has enabled the transfer of genetic pools from wild species through hybridization, despite the partial sterility and different ploidy levels. Due to its closely phylogenetic distance, wild tuber bearing-potato have been the primary resistant genes source in cultivated potato (Hawkes, 1958). Thus, for late blight, resistance to *P. infestans* (*Rpi*) genes have been discovered using high-resolution maps and positional cloning in *Solanum* section

Petota species such as *S. venturii* (*Rpi-vnt1.1* and *Rpi-vnt1.3*), *S. pinnatisectum* (*Rpi1*) and *S. phureja* (*Rpi-phu1*). These particular genes are distributed across four potato chromosomes (5, 6,7, and 9) with a common CC-NB-LRR gene architecture (Kuhl et al., 2001; Park et al., 2005a; Pel et al., 2009; Vossen et al., 2005). Similar resistant genes/loci such as *R8-R9* and *Rpi-abpt* have been isolated from *S. bulbocastanum, S. demissum* and a quadruple hybrid (*S. acaule, S. bulbocastanum, S. phureja* and *S. tuberosum*), resulting in high resistance levels against different late blight strains (Park et al., 2005b). Additionally, different resistant genes/loci have been identified within *S. tuberosum* cultivars, including *R1, R3, R6* and *R7* (El-Kharbotly et al., 1996). *R1*, a NLR protein, is located in a hotspot region on Chromosome 5, clustered with genes that confer late blight resistance in potatoes (Ballvora et al., 2002).

Recently, the Resistance gene enrichment Sequencing (RenSeq) technology has enabled the identification of 331 unidentified R genes in potato, distributed across the potato genome, using NBL-targeting baits and Illumina sequencing (Jupe et al., 2013). This system has allowed identifying new late blight resistance genes such as *Rpi-amr3it* from *S. americanum* using single-molecule real-time (SMRT) sequencing, coupled with RenSeq (SMRT RenSeq) (Witek et al., 2016).

v. RB-mediated resistance in potato

The *RB* gene or *rpi-blb1* (resistant protein to *P. infestans* from *S. bulbocastanum*) is a member of a gene family with four genes located on Chromosome 8 of *S. bulbocastanum* (Song et al., 2003). This gene confers broad-spectrum disease resistance against late blight and has been successfully incorporated into different potato cultivars via *Agrobacterium*-mediated transformation (Halterman et al., 2008). Although it has been observed to result in partial

resistance in field, *RB*-carrying potatoes present a foliar resistant increase in growth chamber and greenhouse compared to wild type genotypes, without significant effects on tuber yield (Halterman et al., 2008).

The *RB* gene encodes a protein of 970 aa with typical R gene architecture, containing an N-terminal CC domain along with an NB (with a p-loop and two kinase motifs) and LRR (with 21 LRR motifs mostly imperfect) domains (Song et al., 2003). After infection, the avirulent protein IPI-OI, a class I effector from *P. infestans*, is translocated into the host cytoplasm, interacting with the CC domain from *RB* triggering a host defense responses (Chen et al., 2012).

Under normal conditions, *RB* remains in a resting state. After pathogen infection, RB proteins dimerize through CC domains mediated by IPI-O1 recognition. However, late blight strains carrying the class III variant effectors IPI-O4 suppress IPI-O1-mediated resistance by direct competition for the CC binding site, avoiding the CC dimerization (Chen et al., 2012; Chen and Halterman, 2017; Halterman et al., 2010). IPI-O4-carrying late blight strains cause more disease than those without it. However, this response is cultivar dependent indicating a possible presence of extra components acting against late blight infection, reducing the resistance response (Chen et al., 2012). The CC domain from the *S. pinnatisectum RB* homologous does not exhibit any interaction with IPI-O4, suggesting that this species may escape the IP-O4-mediated resistance suppression (Chen et al., 2012). Therefore, engineering the *RB*-protein may contribute to the generation of a durable late blight resistance response in *RB*-carrying potato. To achieve this goal, it is imperative to validate the IPI-O4 resistance suppression mediated in potato. Leveraging the molecular interactions between late blight and potato will enable the generation of durable late blight resistance, adopting strategies like CC domain swapping between the *RB* genes from *S. pinnatisectum* and *S. bulbocastanum*, as proposed by Chen et al. (2012).

V. Genomic approaches apply to potato breeding

Breeding for resistant varieties via phenotypic selection can take up to 15 years, which is time-consuming and ineffective against fast-evolving pathogens in potatoes (Lozano et al., 2012; The Potato Genome Sequencing Consortium, 2011). The use of statistical models in modern plant breeding has revealed new resistance gene sources, dissected the disease/resistance architecture and uncovered the mechanisms in which plants respond to pathogen attack (Boyd et al., 2013). Moreover, the capture and prediction of genetic variation for disease resistance using prediction models have contributed to maximizing the genetic gains while decreasing the breeding cycle in staple crops such as maize and wheat (Poland and Rutkoski, 2016). In potato, these strategies can support conventional breeding schemes, increasing the genetic gains and accelerating the generation of resistant potato varieties. The achievement of this goal must rely on the implementation of appropriate statistical models that consider the genetic complexity and allele dosage of this crop.

i. Whole genome association (GWA) analysis in potato

With the advent of high throughout next generation sequencing technologies (HT-NGS), it is now feasible to identify variants associated with disease resistance with relatively high resolution using GWA analysis. Unlike bi-parental QTL (quantitative trait locus) mapping which is restricted to the genetic variation coming from two parental lines, GWA is able to detect historical recombination events in germplasm collections, exploiting natural diversity, increasing mapping resolution and reducing research time (Yu and Buckler, 2006). GWA analysis relies on the linkage disequilibrium (LD) between a causative gene variant and candidate markers used for genotyping. Besides LD, GWA ability for QTL detection also depends on the population size,

marker density, relatedness and the control of non-genetic-related factors such as experimental design or field replicates (Lipka et al., 2015).

Confounding effects caused by population stratification can lead to false-positive discoveries decreasing the power of detection in association analysis (Hoffman, 2013). This problem was initially addressed through General Linear Models (GLM). Later, statistical methods such as principal component analysis (PCA), Bayesian model-based approaches and discriminant analysis of principal components (DAPC) were used to fit the hidden ancestry in a given population using Mixed Linear Models (MLM) (Corander and Tang, 2007; Hoffman, 2013; Lipka et al., 2015). These models also include random effects using variance-covariance matrixes of the individuals (kinship matrix) and have been widely used for QTLs detection in complex traits for most important crops (Zhang et al., 2010). Recently, Efficient Mixed-Model Association (EMMA) and Genome-wide Efficient Mixed-Model Association (GEMMA) algorithms have been used to address large marker and sample numbers, increasing the efficiency and computational speed in GWA analyses (Zhang et al., 2010; Zhou and Stephens, 2012). Complementary approaches use compression methods such as compress-MLM and P3D decreasing significantly computational time, reducing the sample size, by clustering similar individuals, and estimating the population structure (Zhang et al., 2010). To address population structure effect, multi-parental populations capturing historical and recent recombination events have been generated. Nested Association Mapping (NAM), Multi-parent Advance Generation Intercrosses (MAGIC) and Recombinant Inbred Advanced Intercross Lines (RIAILs) combine the advantages of GWAS and bi-parental populations, producing new chromosomal re-arrangements and increasing the genotypic diversity (Bandillo et al., 2013; Huang et al., 2015; Lipka et al., 2015; Rockman and Kruglyak, 2008).

In potato, association analyses have been conducted using linear and non-parametric (Mann–Whitney U test) models, for monogenic and complex traits such us pathogen/pest resistance, and tuber yield and quality, respectively (Rosyara et al., 2016; Schönhals et al., 2016). Though most of these studies have assessed allele dosage by using band intensities with AFLP (amplified fragment length polymorphism), overlapping peaks in electropherograms with SSRs (simple sequence repeats) and sequence variations with SNPs (single nucleotide polymorphisms) (Achenbach et al., 2009; D'hoop et al., 2014; D'Hoop et al., 2008; Gebhardt et al., 2004; Malosetti et al., 2007), considering both dominant and additive effects by the detection power of only an additive model, as a common practice in GWA (Bush and Moore, 2012), leading to biases in the final outcome. However, to date, no GWA analyses have been conducted to unravel the genetic architecture of disease resistance in potato, supported by high throughput genotyping technologies and appropriate statistical models representing the complex nature of potato.

ii. Genomic selection (GS) in potato

QTL mapping studies have contributed widely to the identification of genomic landscapes used for oligogenic trait selection through Marker Assisted Selection (MAS) to accelerate the generation of new cultivars. This molecular breeding approach depends on the reliability and tight linkage between the molecular marker and the causal gene of a particular trait (Collard and Mackill, 2008). GWA studies have reasonably good power to detect large effect-QTL in complex traits. However, GWA usually lacks power to detect associations for small effect variants. This can lead to high rates of missing heritability and low prediction accuracy (Desta and Ortiz, 2014).

To overcome this problem Meuwissen and Hayes (2001) proposed to use Whole-Genome Regression (WGR) methods for prediction of breeding values. This approach, also referred to as GS, has proved to be effective for prediction of breeding values in plant and animal breeding populations, such as maize and cattle, respectively (Beyene et al., 2015; Hayes et al., 2009; Zhang et al., 2014). A WGR model uses large number of variants (e.g., SNPs) distributed over the whole genome to predict breeding values for individuals with only genotypic information, using a trained model calculated from related individuals with both, genotypic and phenotypic information (Heffner et al., 2009). WGR can be extended by including SNPs detected in GWA analyses fitting them as fixed effects, increasing prediction accuracies (Spindel et al., 2016).

In potato, estimated breeding values (EBV) have been generated using best linear unbiased prediction (BLUP) methods and pedigree information. This method showed an improvement in progeny performance when used for traits with low heritability such as yield and plant maturity (Slater et al., 2014). Likewise, potato breeding programs have widely used MAS to select lines with desirable traits such as starch content and virus Y resistance, respectively (Freyre and Douches, 1994; Hämäläinen et al., 1997). However, these approaches are not able to capture total genetic variances since they are restricted only to the variance explained by major QTLs (Heffner et al., 2009). To address this problem, GS has been implemented in potato for yield, tuber quality and biotic/abiotic-related traits (Slater et al., 2016). Since GS requires a high marker density distributed across the genome, the Infinium 8303 Potato Array containing a set of SNP markers distributed across the 12 potato chromosomes (Felcher et al., 2012a) has been used for this purpose. More recently, a greater SNP number derived from Genotyping By Sequencing (GBS) and Diversity Array Technology (DArTseq) silicoDArT have been used to increase genetic gains using GS, obtaining high cross-validation accuracies (0.25-0.7) for complex traits

such a yield (Habyarimana et al., 2017; Sverrisdóttir et al., 2017). Overall, these studies demonstrate the feasibility of GS implementation in potato breeding programs as a contribution to the reduction of the breeding cycles in this crop (Figure 1.3). In particular, GS can contribute to improving pathogen resistance in potato cultivars, promptly responding toward fast-evolving pathogens such as late blight.

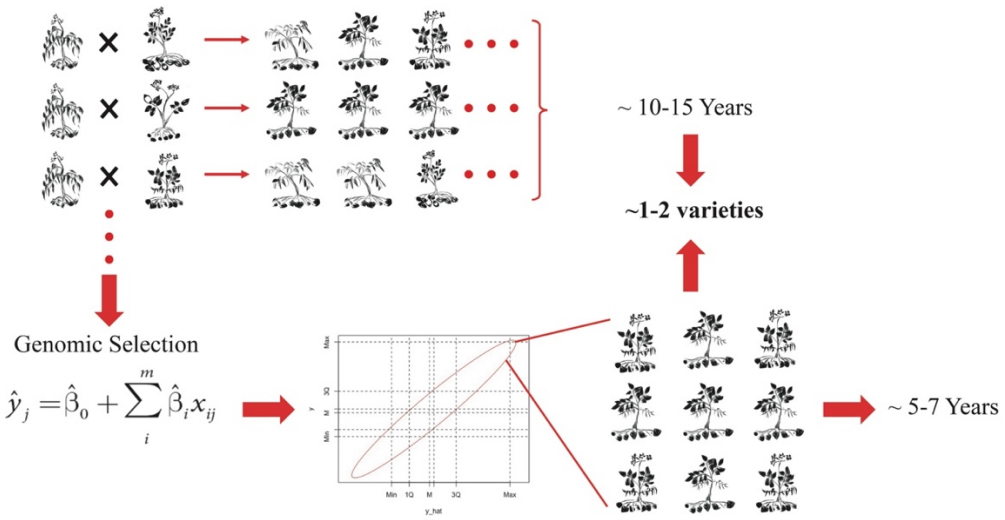

Figure 1.3. *Conventional vs. genomic selection-based breeding in potato.* Traditional breeding of potato relies on the generation of hundreds of crosses which progeny is evaluated and selected in the field up to 15 years to finally generated one or two new potato varieties. Genomic selection will reduce the breeding cycle up to seven years by selected superior lines based on their genetic merit using whole-genome regressions.

iii. Genome editing: An efficient method to target agronomic-related traits

GS has accelerated plant breeding, reducing the generation time of new cultivars by selecting individuals based on their genetic value. However, unlike GS, targeted mutagenesis (TM) represents a fast alternative to obtaining improved varieties by directly altering the expression of

genes controlling traits of interest. Genome editing technologies arise as an alternative tool for a precise and efficient targeted mutagenesis, since unlike conventional transformation, genome editing avoids the modification of regions other than the target sites (Bortesi and Fischer, 2015).

Using double strand breaks (DSB), engineered endonucleases cuts the DNA at the target site, triggering the response of endogenous cell repair mechanisms. Through homology directed repair (HDR) or non-homologous end joining (NHEJ) repairs, deletions (knockout-KO) or insertions (knockin-KI) are achieved, altering the expression of a particular gene (Bortesi and Fischer, 2015; Pellagatti et al., 2015).

a. Targeted mutagenesis technologies

Initially, TM was achieved through native engineered homing endonucleases (HEs) targeting genomic regions spanning 12-40 pb, and generating DSB or single strand-breaks that activate either NHEJ or HDR (Hafez and Hausner, 2012). Homing endonucleases have been used to create gene KO/KI on immortalized human cell lines and plant genes involved in disease and pest control (Hafez and Hausner, 2012). However, this technology was limited to naturally occurring enzymes, reducing the number of possible target sites (Belfort and Bonocora, 2014). To address this problem, customized sequence-specific endonucleases such as Zinc Finger (ZF) and Transcriptions Activator-Like Effector (TALE) nucleases emerged to disrupt or correct predeterminate genomic regions (Bogdanove and Voytas, 2011; Urnov et al., 2010). Zinc Finger nucleases couples the DNA-binding domain from zinc finger transcription factors with the nuclease domain from the restriction enzyme FokI (Urnov et al., 2010). This TM approach uses Cys2 His2 fingers in tandem arrays that recognizes 3 bp each, increasing target specificity when used as a heterodimer, binding up to 12 bp of the intended genomic regions on each direction

(Miller et al., 2007). In contrast, TALENs recognize single contiguous base-pairs using tandem polymorphic amino-acid repeats (Bogdanove and Voytas, 2011). This technology relies on the DNA-binding ability of the TALE protein derived from plant pathogen effectors coupled with FokI nucleases as ZFN. Transcriptions Activator-Like Effector nucleases use an amino-acid/DNA-type code to design effector proteins, whose specific amino-acid order recognizes an intended DNA sequence (Joung and Sander, 2013). This feature confers limitless targeting ranges unlike ZF nucleases which lack specific zinc finger combinations for some specific nucleotides triplets (Joung and Sander, 2013).

Recently, a more precise, amenable and economic technology has become the preferred approach within the scientific community for gene targeting and genome edition. This technology known as Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-associated Cas 9 protein (Cas9) system has been widely used to generate gene KO/KI of candidate genes related to agronomic traits in important crops (Jaganathan et al., 2018). The Cas9 target-DNA recognition is mediated by a single guide-RNA (sgRNA) bearing a 20 bp target-site complementary to a region adjacent to a protospacer-adjacent motif (PAM), 5′-NGG-3', resulting in the generation of a DSB (Doudna and Charpentier, 2014; Pellagatti *et al.*, 2015; Sternberg *et al.*, 2014) which can be leveraged to generate DSB of target genes.

CRISPR/Cas9 has extended its range of genomic targets and applicability by altering the Cas9 structure and creating new protein complexes, enabled it not only to address functional but structural analyses. For instance, catalytically inactive Cas9 (dCas9) proteins have been used to turn on/off genes to investigate gene function and localization when fused with proteins such as green fluorescent proteins (Doudna and Charpentier, 2014). Likewise, specific point mutations can be generated using cytidine or adenosine deaminases coupled with dCas9, converting C/G to

A/T and A/T to C/G bases, minimizing the generation of defective DBS-related products (i.e., translocations) (Rees and Liu, 2018). The ability of the Cas protein to recognize a wide range of target DNA and RNA has also been expanded. Cas variants such as Cas12a and Cas13b are capable of generating single-stranded DNA cleavage and gene knockdown, respectively, expanding the range of application of this technology to study essential genes and non-coding RNAs (Chen et al., 2018; Cox et al., 2017). Similarly, PAM sequence specificities have been broadened using pre-existing (SpCas9-NG) or evolved (xCas9) Cas versions (Hu et al., 2018; Nishimasu et al., 2018) expanding targetable genome loci. In contrast, anti-CRISPR associated (Aca) protein families, have been identified and studied for their ability to inhibit the CRISPR RNA (crRNA) binding, Cas9 dimerization, and conformational changes for cleavage activity (Pawluk et al., 2017). Anti-CRISPRs could be used as a safety measure for gene drive-derived technologies and contribute to the reduction of off-target effects by controlling Cas expression (Pawluk et al., 2017).

b. Genome editing in potato

Targeted mutagenesis has been successfully implemented in potato mainly targeting herbicide resistance and tuber quality-related traits. Transient and stable transfections assays have been conducted in potato protoplast and callus using TALENS, targeting the Acetolactate Synthase gene (ALS) and Vacuolar Invertase gene (Vlnv) (Butler et al., 2015; Clasen et al., 2016; Nicolia et al., 2015). Likewise, CRISPR-Cas9 technologies have been used in potato to eliminate the production of steroidal glycoalkaloids and Aux/IAA proteins by knocking-out the St16DOX and StIAA2 genes respectively (Nakayasu et al., 2018; Wang et al., 2015). CRISPR-Cas9 ribonucleoproteins (RNPs) have also been delivered into potato protoplasts, targeting the

Granule Bound Starch Synthase (GBSS) encoded gene to prevent Cas9 integration into the potato genome (Andersson et al., 2018).

Base editing has also been achieved in potato. For instance, specific point mutations targeting the GBSS and ALS genes have been generated using base editing in protoplasts, coupling the human A3A deaminase with a Cas9 nickase, increasing the deamination window to 17 nucleotides compared to the five nucleotides observed in other systems (Zong et al., 2018).

However, the complex nature of potato limits TM for targeting alleles on each potato homologous chromosomes. To address this limitation, an enhancer derived from the OsMac3 mRNA 5′-untranslated region (dMac3) has been introduced into a CRISPR/Cas9 cassette to increase chromosome-specific allele mutagenesis. dMac3 boosts Cas9 synthesis, increasing the ratio of GBSS KOs containing four mutant alleles (Kusano et al., 2018).

VI. Project objectives

This project aims to integrate genomic selection and genome editing strategies to accelerate potato breeding, contributing to reduce the breeding cycle. A strategy to understand the molecular interaction between potato and late blight is also presented. First, in Chapter 2 the generation of allele dosage models in tetraploid potato is presented under the title "*Genomic selection for late blight and common scab resistance in tetraploid potato (Solanum tuberosum).*" This research seeks primarily to develop an allele dosage model for polyploids that accounts for additive and dominance effects and to use that model to detect loci associated with late blight and common scab resistance in tetraploid potatoes. As a second specific aim, the expectation is to integrate the models developed into models for genomic prediction and to evaluate the ability of these models to predict resistance to late blight and Scab in a tetraploid potato population.

Second, in Chapter 3, a genome editing approach is implemented to breakdown the *S-RNase*-based self incompatibly in potato, under the title "Overcoming self-incompatibility in diploid potato using CRISPR-Cas9". The purpose of this chapter was first to identify *S-RNase* allelic variants in two self-incompatible diploid potatoes. As a second specific aim, it is expected to generate *S-RNase* KO lines using the CRISPR-Cas9 system in self-incompatible diploid potatoes. Finally, the third specific aim involves characterizing the gene-edited T_0 events and their phenotypic evaluation for self-compatibility.

Third, in Chapter 4, in vitro and in vivo approaches were used to confirm the interaction between RB CC domains and IPI-O effectors from late blight, under the title "Validation of the *RB* mediated-resistance suppression by the class III effector IPI-O4 from late blight". First, using RB and IPI-O co-infiltration assays on *N. benthamiana* and RB-transgenic potato lines, it is expected to confirm previously reported in vitro and in vivo observations for this pathosystem. Finally, a second specific aim seeks to validate the IP-O4 suppression effect over CC-self interaction using a yeast three-hybrid approach. Finally, the concluding Chapter 5 is presented, summarizing the main findings, drawbacks and future directions of each research chapter.

LITERATURE CITED

LITERATURE CITED

Achenbach, U., Paulo, J., Ilarionova, E., Lübeck, J., Strahwald, J., Tacke, E., et al. (2009). Using SNP markers to dissect linkage disequilibrium at a major quantitative trait locus for resistance to the potato cyst nematode *Globodera pallida* on potato chromosome v. Theoretical and Applied Genetics 118, 619–629.

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

Ames, M., and Spooner, D. M. (2008). DNA from herbarium specimens settles a controversy about origins of the European potato 1. 95, 252–257.

Andersson, M., Turesson, H., Olsson, N., Fält, A.-S., Ohlsson, P., Gonzalez, M. N., et al. (2018). Genome editing in potato via CRISPR-Cas9 ribonucleoprotein delivery. Physiologia Plantarum 164, 378–384.

Atul, T., and Ray, S. (2014). Certain medicinal plants of Solanaceae and their alkaloids screening. International Research Journal of Medical Sciences 2, 4–6.

Aversano, R., Contaldi, F., Ercolano, M. R., Grosso, V., Iorizzo, M., Tatino, F., et al. (2015). The Solanum commersonii Genome Sequence Provides Insights into Adaptation to Stress Conditions and Genome Evolution of Wild Potato Relatives. The Plant Cell 27, 954–968.

Ballvora, A., Ercolano, M. R., Weiß, J., Meksem, K., Bormann, C. A., Oberhagemann, P., et al. (2002). The *R1* gene for potato resistance to late blight (*Phytophthora infestans*) belongs to the leucine zipper/NBS/LRR class of plant resistance genes. Plant Journal 30, 361–371.

Bandillo, N., Raghavan, C., Muyco, P. A., Sevilla, M. A. L., Lobina, I. T., Dilla-Ermita, C. J., et al. (2013). Multi-parent advanced generation inter-cross (MAGIC) populations in rice: progress and potential for genetics research and breeding. Rice 6, 11.

Bastien, M., Boudhrioua, C., Fortin, G., and Belzile, F. (2018). Exploring the potential and limitations of genotyping-by-sequencing for SNP discovery and genotyping in tetraploid potato. Genome 61, 449–456.

Belfort, M., and Bonocora, R. P. (2014). Homing endonucleases: from genetic anomalies to programmable genomic clippers. Methods in molecular biology (Clifton, N.J.) 1123, 1–26.

Beyene, Y., Semagn, K., Mugo, S., Tarekegne, A., Babu, R., Meisel, B., et al. (2015). Genetic gains in grain yield through genomic selection in eight bi-parental maize populations under drought stress. Crop Science 55, 154–163.

Bogdanove, A. J., and Voytas, D. F. (2011). TAL Effectors: Customizable Proteins for DNA Targeting. Science 333, 1843 LP – 1846.

Bortesi, L., and Fischer, R. (2015). The CRISPR/Cas9 system for plant genome editing and beyond. Biotechnol. Adv. 33, 41–52.

Boudhrioua, C., Bastien, M., Légaré, G., Pomerleau, S., St-Cyr, J., Boyle, B., et al. (2017). Genotyping-by-Sequencing in Potato. in The Potato Genome, eds. S. Kumar Chakrabarti, C. Xie, and J. Kumar Tiwari (Cham: Springer International Publishing), 283–296.

Boyd, L. A., Ridout, C., O'Sullivan, D. M., Leach, J. E., and Leung, H. (2013). Plant–pathogen interactions: disease resistance in modern agriculture. Trends in Genetics 29, 233–240.

Braun, S., Gevens, A., Charkowski, A., Allen, C., and Jansky, S. (2017). Potato common scab: a review of the causal pathogens, management practices, varietal resistance screening methods, and host resistance. American Journal of Potato Research 94, 283–296.

Bush, W. S., and Moore, J. H. (2012). Chapter 11: Genome-wide association studies. PLoS Computational Biology 8, e1002822.

Butler, N. M., Atkins, P. A., Voytas, D. F., and Douches, D. S. (2015). Generation and inheritance of targeted mutations in potato (*Solanum tuberosum* L.) using the CRISPR/Cas system. PloS One 10, e0144591.

Carroll, C. P. (1975). The inheritance and expression of sterility in hybrids of dihaploid and cultivated diploid potatoes. Genetica 45, 149–162.

Chaparro, J. M., Holm, D. G., Broeckling, C. D., Prenni, J. E., and Heuberger, A. L. (2018). Metabolomics and ionomics of potato tuber reveals an influence of cultivar and market class on human nutrients and bioactive compounds. Frontiers in nutrition 5, 36.

Charfeddine, M., Saïdi, M. N., Charfeddine, S., Hammami, A., and Gargouri Bouzid, R. (2015a). Genome-wide analysis and expression profiling of the erf transcription factor family in potato (*Solanum tuberosum* L.). Molecular Biotechnology 57, 348–358.

Charfeddine, S., Saïdi, M. N., Charfeddine, M., and Gargouri-Bouzid, R. (2015b). Genome-wide identification and expression profiling of the late embryogenesis abundant genes in potato with emphasis on dehydrins. Molecular Biology Reports 42, 1163–1174.

Chen, J. S., Ma, E., Harrington, L. B., Da Costa, M., Tian, X., Palefsky, J. M., et al. (2018). CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. Science 360, 436 LP – 439.

Chen, Y., and Halterman, D. A. (2017). *Phytophthora infestans* effectors IPI-O1 and IPI-O4 each contribute to pathogen virulence. Phytopathology 107, 600–606.

Chen, Y., Liu, Z., and Halterman, D. a. (2012). Molecular determinants of resistance activation and suppression by *Phytophthora infestans* effector IPI-O. PLoS Pathogens 8, 1–10.

Clasen, B. M., Stoddard, T. J., Luo, S., Demorest, Z. L., Li, J., Cedrone, F., et al. (2016). Improving cold storage and processing traits in potato through targeted gene knockout. Plant Biotechnology Journal 14, 169–176.

Collard, B. C. Y., and Mackill, D. J. (2008). Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philos Trans R Soc Lond B Biol Sci 363, 557–572.

Collins, A., Milbourne, D., Ramsay, L., Meyer, R., Chatot-Balandras, C., Oberhagemann, P., et al. (1999). QTL for field resistance to late blight in potato are strongly correlated with maturity and vigor. Molecular Breeding 5, 387–398.

Cook, D. E., Mesarich, C. H., and Thomma, B. P. H. J. (2015). Understanding plant immunity as a surveillance system to detect invasion. Annual Review of Phytopathology 53, 541–563.

Corander, J., and Tang, J. (2007). Bayesian analysis of population structure based on linked molecular information. Mathematical Biosciences 205, 19–31.

Couto, D., and Zipfel, C. (2016). Regulation of pattern recognition receptor signaling in plants. Nature Reviews Immunology 16, 537.

Cox, D. B. T., Gootenberg, J. S., Abudayyeh, O. O., Franklin, B., Kellner, M. J., Joung, J., et al. (2017). RNA editing with CRISPR-Cas13. Science 358, 1019 LP – 1027.

D'hoop, B. B., Keizer, P. L. C., Paulo, M. J., Visser, R. G. F., van Eeuwijk, F. a, and van Eck, H. J. (2014). Identification of agronomically important QTL in tetraploid potato cultivars using a marker-trait association analysis. TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik 127, 731–48.

D'Hoop, B. B., Paulo, M. J., Mank, R. A., Van Eck, H. J., and Van Eeuwijk, F. A. (2008). Association mapping of quality traits in potato (Solanum tuberosum L.). Euphytica 161, 47–60.

Dees, M. W., and Wanner, L. A. (2012). In search of better management of potato common scab. Potato Research 55, 249–268.

Desta, Z. A., and Ortiz, R. (2014). Genomic selection: Genome-wide prediction in plant improvement. Trends in Plant Science 19, 592–601.

Devaux, A., Kromann, P., and Ortiz, O. (2014). Potatoes for sustainable global food security. Potato Research 57, 185–199.

Doudna, J. A., and Charpentier, E. (2014). The new frontier of genome engineering with CRISPR-Cas9. Science 346, 1258096.

Drewnowski, A., and Rehm, C. D. (2013). Vegetable cost metrics show that potatoes and beans provide most nutrients per penny. PloS One 8, e63277.

El-Kharbotly, A., Palomino-Sánchez, C., Salamini, F., Jacobsen, E., and Gebhardt, C. (1996). R6 and R7 alleles of potato conferring race-specific resistance to *Phytophthora infestans* (Mont.) de Bary identified genetic loci clustering with the *R3* locus on chromosome XI. Theor. Appl. Genet 92, 880–884.

Ellis, D., Chavez, O., Coombs, J., Soto, J., Gomez, R., Douches, D., et al. (2018). Genetic identity in genebanks: application of the SolCAP 12K SNP array in fingerprinting and diversity analysis in the global in trust potato collection. Genome 61, 523–537.

Endelman, J. B., and Jansky, S. H. (2016). Genetic mapping with an inbred line-derived F2 population in potato. Theoretical and Applied Genetics 129, 935–943.

FAOSTAT. (2019). FAOSTAT database collections. Food and Agriculture Organization of the United Nations. Available at: http://www.fao.org/faostat/ [Accessed January 27, 2019].

Fasano, C., Diretto, G., Aversano, R., D'Agostino, N., Di Matteo, A., Frusciante, L., et al. (2016). Transcriptome and metabolome of synthetic *Solanum* autotetraploids reveal key genomic stress events following polyploidization. New Phytologist 210, 1382–1394.

Felcher, K. J., Coombs, J. J., Massa, A. N., Hansey, C. N., Hamilton, J. P., Veilleux, R. E., et al. (2012a). Integration of two diploid potato linkage maps with the potato genome sequence. PloS One 7, e36347.

Felcher, K. J., Coombs, J. J., Massa, A. N., Hansey, C. N., Hamilton, J. P., Veilleux, R. E., et al. (2012b). Integration of two diploid potato linkage maps with the potato genome sequence. PloS One 7, e36347.

Flor, H. H. (1971). Current status of the gene-for-gene concept. Annual Review of Phytopathology 9, 275–296.

Freyre, R., and Douches, D. S. (1994). Development of a model for marker-assisted selection of specific gravity in diploid potato across environments. Crop Science 34, 1361–1368.

Friedman, M. (1997). Chemistry, biochemistry, and dietary role of potato polyphenols: A Review. Journal of Agricultural and Food Chemistry 45, 1523–1540.

Fu, D., Xiao, M., Hayward, A., Fu, Y., Liu, G., Jiang, G., et al. (2014). Utilization of crop heterosis: a review. Euphytica 197, 161–173.

Fujii, S., Kubo, K. I., and Takayama, S. (2016). Non-self- and self-recognition models in plant self-incompatibility. Nat. Plants 2, 16130.

Gaiero, P., Speranza, P., and de Jong, H. (2018). Introgressive hybridization in potato revealed by novel cytogenetic and genomic technologies. American Journal of Potato Research 95, 607–621.

Gálvez, J. H., Tai, H. H., Lagüe, M., Zebarth, B. J., and Strömvik, M. V (2016). The nitrogen responsive transcriptome in potato (*Solanum tuberosum* L.) reveals significant gene regulatory motifs. Scientific Reports 6, 26090.

Gebhardt, C., Ballvora, A., Walkemeier, B., Oberhagemann, P., and Schüler, K. (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. Molecular Breeding 13, 93–102.

George, T. S., Taylor, M. A., Dodd, I. C., and White, P. J. (2017). Climate change and consequences for potato production: a review of tolerance to emerging abiotic stress. Potato Research 60, 239–268.

Goldraij, A., Kondo, K., Lee, C. B., Hancock, C. N., Sivaguru, M., Vazquez-Santana, S., et al. (2006). Compartmentalization of S-RNase and HT-B degradation in self-incompatible *Nicotiana*. Nature 439, 805.

Gong, L., Zhang, H., Gan, X., Zhang, L., Chen, Y., Nie, F., et al. (2015). Transcriptome profiling of the potato (*Solanum tuberosum* L.) plant under drought stress and water-stimulus Conditions. PloS One 10, e0128041.

Grada, C. (2004). Ireland's great famine: An overview. Dublin. No.04/24

Grun, P. (1990). The evolution of cultivated potatoes. Economic Botany 44, 39–55.

Grünwald, N. J., and Flier, W. G. (2005). The biology of *Phytophthora infestans* at Its Center of Origin. Annual Review of Phytopathology 43, 171–190.

Habyarimana, E., Parisi, B., and Mandolino, G. (2017). Genomic prediction for yields, processing and nutritional quality traits in cultivated potato (*Solanum tuberosum* L.). Plant Breeding 136, 245–252.

Hafez, M., and Hausner, G. (2012). Homing endonucleases: DNA scissors on a mission. Genome 55, 553–569.

Halterman, D. A., Chen, Y., Sopee, J., Berduo-Sandoval, J., and Sánchez-Pérez, A. (2010). Competition between *Phytophthora infestans* effectors leads to increased aggressiveness on plants containing broad-spectrum late blight resistance. PloS One 5, e10536.

Halterman, D. a, Kramer, L. C., Wielgus, S., and Jiang, J. (2008). Performance of transgenic potato containing the late blight resistance gene RB. Plant Disease 92, 339–343.

Hämäläinen, J. H., Watanabe, K. N., Valkonen, J. P. T., Arihara, a., Plaisted, R. L., Pehu, E., et al. (1997). Mapping and marker-assisted selection for a gene for extreme resistance to potato virus Y. Theor Appl Genet, 192–197.

Hardigan, M. A., Crisovan, E., Hamiltion, J. P., Kim, J., Laimbeer, P., Leisner, C. P., et al. (2016). Genome reduction uncovers a large dispensable genome and adaptive role for copy number variation in asexually propagated *Solanum tuberosum*. Plant Cell 28, 388–405.

Hardigan, M. A., Laimbeer, F. P. E., Newton, L., Crisovan, E., Hamilton, J. P., Vaillancourt, B., et al. (2017). Genome diversity of tuber-bearing *Solanum* uncovers complex evolutionary history and targets of domestication in the cultivated potato. Proceedings of the National Academy of Sciences 114, E9999 LP-E10008.

Haverkort, A. J., Struik, P. C., Visser, R. G. F., and Jacobsen, E. (2009). Applied biotechnology to combat bate blight in potato caused by *Phytophthora Infestans*. Potato Research 52, 249–264.

Hawkes, J. G. (1958). Significance of wild species and primitive forms for potato breeding. Euphytica 7, 257–270.

Hawkes, J. G. (1992). "History of the potato," in The potato crop: the scientific basis for improvement, ed. P. Harris (Dordrecht: Springer Netherlands), 1–13.

Hawkes, J. G., and Francisco-Ortega, J. (1993). The early history of the potato in Europe. Euphytica 70, 1–7.

Hayes, B. J., Bowman, P. J., Chamberlain, A. J., and Goddard, M. E. (2009). Genomic selection in dairy cattle: progress and challenges. Journal of Dairy Science 92, 433–443.

Heffner, E. L., Sorrells, M. E., and Jannink, J.-L. (2009). Genomic selection for crop improvement. Crop Science 49, 1–12.

Hoffman, G. E. (2013). Correcting for population structure and kinship using the linear mixed model: Theory and extensions. PLoS One 8, e75707.

Hosaka, K. (2003). T-type chloroplast DNA in *Solanum tuberosum* L. ssp. tuberosum was conferred from some populations of *S. tarijense* Hawkes. American Journal of Potato Research 80, 21–32.

Hosaka, K., and Hanneman, R. E. (1988). The origin of the cultivated tetraploid potato based on chloroplast DNA. Theoretical and Applied Genetics 76, 172–176.

Hosaka, K., and Hanneman, R. E. (1998). Genetics of self-compatibility in a self-incompatible wild diploid potato species *Solanum chacoense*. 1. Detection of an S locus inhibitor (*Sli*) gene. Euphytica 99, 191–197.

Hu, J. H., Miller, S. M., Geurts, M. H., Tang, W., Chen, L., Sun, N., et al. (2018). Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. Nature 556, 57.

Hua, Z.-H., Fields, A., and Kao, T. (2008). Biochemical models for *S-RNase*-based self-incompatibility. Molecular Plant 1, 575–585.

Huamán, Z., and Spooner, D. M. (2002). Reclassification of landrace populations of cultivated potatoes (*Solanum* sect. Petota). American Journal of Botany 89, 947–965.

Huang, B. E., Verbyla, K. L., Verbyla, A. P., Raghavan, C., Singh, V. K., Gaur, P., et al. (2015). MAGIC populations in crops: current status and future prospects. Theoretical and Applied Genetics 128, 999–1017.

Iovene, M., Zhang, T., Lou, Q., Buell, C. R., and Jiang, J. (2013). Copy number variation in potato – an asexually propagated autotetraploid species. The Plant Journal 75, 80–89.

Jaganathan, D., Ramasamy, K., Sellamuthu, G., Jayabalan, S., and Venkataraman, G. (2018). CRISPR for crop improvement: An update review. Front. Plant. Sci. 9, 985.

Jansky, S. H., Charkowski, A. O., Douches, D. S., Gusmini, G., Richael, C., Bethke, P. C., et al. (2016). Reinventing potato as a diploid inbred line–based crop. Crop Science 56, 1412–1422.

Jansky, S. H., Chung, Y. S., and Kittipadukal, P. (2014). M6: A diploid potato inbred line for use in breeding and genetics research. J. Plant. Regist. 8, 195–199.

Jansky, S. H., and Peloquin, S. J. (2006). Advantages of wild diploid *Solanum* species over cultivated diploid relatives in potato breeding programs. Genetic Resources and Crop Evolution 53, 669–674.

Jiménez-Durán, K., McClure, B., García-Campusano, F., Rodríguez-Sotres, R., Cisneros, J., Busot, G., et al. (2013). NaStEP: A proteinase inhibitor essential to self-incompatibility and a positive regulator of HT-B stability in *Nicotiana alata* pollen tubes. Plant Physiology 161, 97 LP – 107.

Jones, J. D. G., and Dangl, J. L. (2006). The plant immune system. Nature 444, 323–329.
Joung, J. K., and Sander, J. D. (2013). TALENs: a widely applicable technology for targeted genome editing. Nature reviews. Molecular cell biology 14, 49–55.

Jupe, F., Witek, K., Verweij, W., Śliwka, J., Pritchard, L., Etherington, G. J., et al. (2013). Resistance gene enrichment sequencing (RenSeq) enables reannotation of the NB-LRR gene family from sequenced plant genomes and rapid mapping of resistance loci in segregating populations. Plant Journal 76, 530–544.

Kloosterman, B., Abelenda, J. A., Gomez, M. del M. C., Oortwijn, M., de Boer, J. M., Kowitwanich, K., et al. (2013). Naturally occurring allele diversity allows potato cultivation in northern latitudes. Nature 495, 246.

Kolasa, K. (1993). The potato and human nutrition. American Potato Journal 70, 375–384.

Kromann, P., Miethbauer, T., Ortiz, O., and Forbes, G. A. (2014). "Review of potato biotic constraints and experiences with integrated pest management interventions," in Integrated Pest Management. Pesticide problems, Vol.3, eds. D. Pimentel and R. Peshin (Dordrecht: Springer Netherlands), 245–268.

Kubo, K., Entani, T., Takara, A., Wang, N., Fields, A. M., Hua, Z., et al. (2010). Collaborative non-self recognition system in *S-RNase* based self-incompatibility. Science 330, 796–799.

Kubo, K., Paape, T., Hatakeyama, M., Entani, T., Takara, A., Kajihara, K., et al. (2015). Gene duplication and genetic exchange drive the evolution of *S-RNase*-based self-incompatibility in Petunia. Nat. Plants 1, 14005.

Kuhl, J. C., Hanneman Jr., R. E., and Havey, M. J. (2001). Characterization of and mapping of *Rpi1*, a late-blight resistance locus from diploid (1EBN) mexican *Solanum pinnatisectum*. Molecular Genetics and Genomics 265, 977–985.

Kusano, H., Ohnuma, M., Mutsuro-Aoki, H., Asahi, T., Ichinosawa, D., Onodera, H., et al. (2018). Establishment of a modified CRISPR/Cas9 system with increased mutagenesis frequency using the translational enhancer dMac3 and multiple guide RNAs in potato. Scientific Reports 8, 13753.

Lee, C. B., Kim, S., and McClure, B. (2009). A pollen protein, NaPCCP, that binds pistil arabinogalactan proteins also binds phosphatidylinositol 3-phosphate and associates with the pollen tube endomembrane system. Plant Physiology 149, 791 LP – 802.

Leesutthiphonchai, W., Vu, A. L., Ah-Fong, A. M. V, and Judelson, H. S. (2018). How does *Phytophthora infestans* evade control efforts? Modern insight into the late blight disease. Phytopathology 108, 916–924.

Leisner, C. P., Hamilton, J. P., Crisovan, E., Manrique-Carpintero, N. C., Marand, A. P., Newton, L., et al. (2018). Genome sequence of M6, a diploid inbred clone of the high-glycoalkaloid-producing tuber-bearing potato species *Solanum chacoense*, reveals residual heterozygosity.

Li, Y., Colleoni, C., Zhang, J., Liang, Q., Hu, Y., Ruess, H., et al. (2018). Genomic analyses yield markers for identifying agronomically important genes in potato. Molecular Plant 11, 473–484.

Lian, Q., Tang, D., Bai, Z., Qi, J., Lu, F., Huang, S., et al. (2019). Acquisition of deleterious mutations during potato polyploidization. Journal of Integrative Plant Biology 61, 7–11.

Lipka, A. E., Kandianis, C. B., Hudson, M. E., Yu, J., Drnevich, J., Bradbury, P. J., et al. (2015). From association to prediction: statistical methods for the dissection and selection of complex traits in plants. Current Opinion in Plant Biology 24, 110–118.

Lister, A., Heavens, D., Baker, D., Cugliandolo, F., Kettleborough, G., Macaulay, I., et al. (2019). A critical comparison of technologies for a plant genome sequencing project. GigaScience, giy163. doi:10.1093/gigascience/giy163.

Lozano, R., Ponce, O., Ramirez, M., Mostajo, N., and Orjeda, G. (2012). Genome-wide identification and mapping of NBS-Encoding Resistance Genes in *Solanum tuberosum* Group Phureja. PLoS One 7, e34775.

Malosetti, M., Van Der Linden, C. G., Vosman, B., and Van Eeuwijk, F. A. (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.

Manrique-Carpintero, N. C., Coombs, J. J., Cui, Y., Veilleux, R. E., Buell, C. R., and Douches, D. (2015). Genetic map and QTL analysis of agronomic traits in a diploid potato population using single nucleotide polymorphism markers. Crop Science 55, 2566–2579.

Mariot, R. F., de Oliveira, L. A., Voorhuijzen, M. M., Staats, M., Hutten, R. C. B., van Dijk, J. P., et al. (2016). Characterization and transcriptional profile of genes involved in glycoalkaloid biosynthesis in new varieties of *Solanum tuberosum* L. Journal of Agricultural and Food Chemistry 64, 988–996.

Massa, A. N., Childs, K. L., and Buell, C. R. (2013). Abiotic and biotic stress responses in *Solanum tuberosum* Group Phureja DM1-3 516 R44 as measured through whole transcriptome Sequencing. The Plant Genome 6. doi:10.3835/plantgenome2013.05.0014.

Massa, A. N., Childs, K. L., Lin, H., Bryan, G. J., Giuliano, G., and Buell, C. R. (2011). The transcriptome of the reference Potato Genome *Solanum tuberosum* Group Phureja clone DM1-3 516R44. PloS One 6, e26801.

Massa, A. N., Manrique-Carpintero, N. C., Coombs, J. J., Zarka, D. G., Boone, A. E., Kirk, W. W., et al. (2015). Genetic linkage mapping of economically important traits in cultivated tetraploid potato (*Solanum tuberosum* L.). G3: Genes, Genomes, Genetics 5, 2357–2364.

McClure, B. A., Haring, V., Ebert, P. R., Anderson, M. A., Simpson, R. J., Sakiyama, F., et al. (1989). Style self-incompatibility gene products of *Nicotlana alata* are ribonucleases. Nature 342, 955–957.

McClure, B., Mou, B., Canevascini, S., and Bernatzky, R. (1999). A small asparagine-rich protein required for *S*-allele-specific pollen rejection in *Nicotiana*. Proc. Natl. Acad. Sci. U S A. 96, 13548–13553.

Mendoza, H. A., Mihovilovich, E. J., and Saguma, F. (1996). Identification of triplex (YYYy) Potato Virus Y (PVY) immune progenitors derived from *Solanum tuberosum* ssp. *andigena*. Am. Potato J. 73, 13–19.

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

Miller, J. C., Holmes, M. C., Wang, J., Guschin, D. Y., Lee, Y.-L., Rupniewski, I., et al. (2007). An improved zinc-finger nuclease architecture for highly specific genome editing. Nature Biotechnology 25, 778.

Nakayasu, M., Akiyama, R., Lee, H. J., Osakabe, K., Osakabe, Y., Watanabe, B., et al. (2018). Generation of α-solanine-free hairy roots of potato by CRISPR/Cas9 mediated genome editing of the *St16DOX* gene. Plant Physiology and Biochemistry 131, 70–77.

National Potato Council (2016). Potato statistical year book. Washington D.C.

Nicolia, A., Proux-Wéra, E., Åhman, I., Onkokesung, N., Andersson, M., Andreasson, E., et al. (2015). Targeted gene mutation in tetraploid potato through transient TALEN expression in protoplasts. Journal of Biotechnology 204, 17–24.

Nishimasu, H., Shi, X., Ishiguro, S., Gao, L., Hirano, S., Okazaki, S., et al. (2018). Engineered CRISPR-Cas9 nuclease with expanded targeting space. Science 361, 1259 LP – 1262.

Nowicki, M., Foolad, M. R., Nowakowska, M., and Kozik, E. U. (2011). Potato and tomato late blight caused by *Phytophthora infestans*: An overview of pathology and resistance breeding. Plant Disease 96, 4–17.

O'Brien, M., Kapfer, C., Major, G., Laurin, M., Bertrand, C., Kondo, K., et al. (2002). Molecular analysis of the stylar-expressed *Solanum chacoense* small asparagine-rich protein family related to the HT modifier of gametophytic self-incompatibility in *Nicotiana*. Plant J 32, 985–996.

O'Brien, M., Major, G., Chantha, S.-C., and Matton, D. P. (2004). Isolation of *S-RNase* binding proteins from *Solanum chacoense*: identification of an SBP1 (RING finger protein) orthologue. Sexual Plant Reproduction 17, 81–87.

Oerke, E. C. (2006). Crop losses to pests. The Journal of Agricultural Science 144, 31–43.

Olmstead, R. G., Bohs, L., Migid, H. A., Santiago-valentin, E., Garcia, V. F., and Collier, S. M. (2008). A molecular phylogeny of the Solanaceae. Molecular phylogenetics 57, 1159–1181.

Ovchinnikova, A., Krylova, E., Gavrilenko, T., Smekalova, T., Zhuk, M., Knapp, S., et al. (2011). Taxonomy of cultivated potatoes (*Solanum* section Petota: Solanaceae). Botanical Journal of the Linnean Society 165, 107–155.

Paajanen, P., Kettleborough, G., Lopez-Girona, E., Giolai, M., Heavens, D., Baker, D., et al. (2017). A critical comparison of technologies for a plant genome sequencing project. bioRxiv, 201830. doi:10.1101/201830.

Park, T.-H., Gros, J., Sikkema, A., Vleeshouwers, V. G., Muskens, M., Allefs, S., et al. (2005a). The late blight resistance locus *Rpi-bib3* from *Solanum bulbocastanum* belongs to a major late blight R gene cluster on chromosome 4 of potato. Molecular plant-microbe interactions: MPMI 18, 722–729.

Park, T.-H., Vleeshouwers, V. G. A. A., Hutten, R. C. B., van Eck, H. J., van der Vossen, E., Jacobsen, E., et al. (2005b). High-resolution mapping and analysis of the resistance locus *Rpi-abpt* against *Phytophthora infestans* in Potato. Molecular Breeding 16, 33–43. doi:10.1007/s11032-005-1925-z.

Pawluk, A., Davidson, A. R., and Maxwell, K. L. (2017). Anti-CRISPR: discovery, mechanism and function. Nature Reviews Microbiology 16, 12.

Pel, M. a, Foster, S. J., Park, T.-H., Rietman, H., van Arkel, G., Jones, J. D. G., et al. (2009). Mapping and cloning of late blight resistance genes from *Solanum venturii* using an interspecific candidate gene approach. Molecular plant-microbe interactions: MPMI 22, 601–615.

Pellagatti, A., Dolatshad, H., Valletta, S., and Boultwood, J. (2015). Application of CRISPR/Cas9 genome editing to the study and treatment of disease. Arch. Toxicol. 89, 1023–1034.

Pham, G. M., Newton, L., Wiegert-Rininger, K., Vaillancourt, B., Douches, D. S., and Buell, C. R. (2017). Extensive genome heterogeneity leads to preferential allele expression and copy number-dependent expression in cultivated potato. The Plant Journal 92, 624–637.

Piñeros-Niño, C., Narváez-Cuenca, C.-E., Kushalappa, A. C., and Mosquera, T. (2017). Hydroxycinnamic acids in cooked potato tubers from *Solanum tuberosum* group Phureja. Food Science & Nutrition 5, 380–389.

Poland, J., and Rutkoski, J. (2016). Advances and Challenges in Genomic Selection for Disease Resistance. Annual Review of Phytopathology 54, 79–98.

Pomerantz, A., Cohen, Y., Shufan, E., Ben-Naim, Y., Mordechai, S., Salman, A., et al. (2014). Characterization of *Phytophthora infestans* resistance to mefenoxam using FTIR spectroscopy. Journal of Photochemistry and Photobiology B: Biology 141, 308–314.

Rees, H. A., and Liu, D. R. (2018). Base editing: precision chemistry on the genome and transcriptome of living cells. Nature Reviews Genetics 19, 770–788.

Rockman, M. V, and Kruglyak, L. (2008). Breeding designs for recombinant inbred advanced intercross lines. Genetics 179, 1069–1078.

Rodríguez-Pérez, C., Gómez-Caravaca, A. M., Guerra-Hernández, E., Cerretani, L., García-Villanova, B., and Verardo, V. (2018). Comprehensive metabolite profiling of *Solanum tuberosum* L. (potato) leaves by HPLC-ESI-QTOF-MS. Food Research International 112, 390–399.

Rosyara, U. R., De Jong, W. S., Douches, D. S., and Endelman, J. B. (2016). Software for genome-wide association studies in autopolyploids and its application to potato. The Plant Genome 9, 1–10.

Rumold, C. U., and Aldenderfer, M. S. (2016). Late Archaic-Early Formative period microbotanical evidence for potato at Jiskairumoko in the Titicaca Basin of southern Peru. Proceedings of the National Academy of Sciences 113, 13672–13677.

Sassa, H., Kakui, H., Miyamoto, M., Suzuki, Y., Hanada, T., Ushijima, K., et al. (2007). S locus F-box brothers: multiple and pollen-specific F-box genes with S haplotype-specific polymorphisms in apple and Japanese pear. Genetics 175, 1869–1881.

Schönhals, E. M., Ortega, F., Barandalla, L., Aragones, A., Ruiz de Galarreta, J. I., Liao, J.-C., et al. (2016). Identification and reproducibility of diagnostic DNA markers for tuber starch and yield optimization in a novel association mapping population of potato (*Solanum tuberosum* L.). Theoretical and Applied Genetics 129, 767–785.

Sharma, S. K., Bolser, D., de Boer, J., Sønderkær, M., Amoros, W., Carboni, M. F., et al. (2013). Construction of Reference Chromosome-Scale Pseudomolecules for Potato: Integrating the Potato Genome with Genetic and Physical Maps. G3: Genes|Genomes|Genetics 3, 2031 LP – 2047.

Sijacic, P., Wang, X., Skirpan, A. L., Wang, Y., Dowd, P. E., McCubbin, A. G., et al. (2004). Identification of the pollen determinant of *S-RNase*-mediated self-incompatibility. Nature 429, 302–305.

Simmonds, N. W. (1997). A review of potato propagation by means of seed, as distinct from clonal propagation by tubers. Potato Research 40, 191–214.

Slater, A. T., Cogan, N. O. I., Forster, J. W., Hayes, B. J., and Daetwyler, H. D. (2016). Improving genetic gain with genomic selection in autotetraploid potato. The Plant Genome 9, 1–15.

Slater, A. T., Wilson, G. M., Cogan, N. O. I., Forster, J. W., and Hayes, B. J. (2014). Improving the analysis of low heritability complex traits for enhanced genetic gain in potato. TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik 127, 809–20.

Smart, C. D., and Fry, W. E. (2001). Invasions by the late blight pathogen: Renewed sex and enhanced fitness. Biological Invasions 3, 235–243.

Song, J., Bradeen, J. M., Naess, S. K., Raasch, J. a, Wielgus, S. M., Haberlach, G. T., et al. (2003). Gene RB cloned from *Solanum bulbocastanum* confers broad spectrum resistance to potato late blight. Proceedings of the National Academy of Sciences of the United States of America 100, 9128–9133.

Song, S., Hao, L., Zhao, P., Xu, Y., Zhong, N., Zhang, H., et al. (2019). Genome-wide Identification, Expression Profiling and Evolutionary Analysis of Auxin Response Factor Gene Family in Potato (*Solanum tuberosum* Group Phureja). Scientific Reports 9, 1755.

Spindel, J. E., Begum, H., Akdemir, D., Collard, B., Redoña, E., Jannink, J.-L., et al. (2016). Genome-wide prediction models that incorporate de novo GWAS are a powerful new tool for tropical rice improvement. Heredity 116, 395.

Spooner, D. M., Sytsma, K. J., and Conti, E. (1991). Chloroplast dna evidence for genome differentiation in wild potatoes (Solanum sect. petota: Solanaceae). American Journal of Botany 78, 1354–1366.

Sprenger, H., Kurowsky, C., Horn, R., Erban, A., Seddig, S., Rudack, K., et al. (2016). The drought response of potato reference cultivars with contrasting tolerance. Plant, Cell & Environment 39, 2370–2389.

Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C., and Doudna, J. A. (2014). DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature 507, 62–7.

Sukarta, O. C. A., Slootweg, E. J., and Goverse, A. (2016). Structure-informed insights for NLR functioning in plant immunity. Seminars in Cell & Developmental Biology 56, 134–149.

Sverrisdóttir, E., Byrne, S., Sundmark, E. H. R., Johnsen, H. Ø., Kirk, H. G., Asp, T., 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.

Takayama, S., and Isogai, A. (2005). Self-Incompatibility in Plants. Annu. Rev. Plant. Biol. 56, 467–489.

The Potato Genome Sequencing Consortium (2011). Genome sequence and analysis of the tuber crop potato. Nature 475, 189–195.

Tomita, S., Ikeda, S., Tsuda, S., Someya, N., Asano, K., Kikuchi, J., et al. (2017). A survey of metabolic changes in potato leaves by NMR-based metabolic profiling in relation to resistance to late blight disease under field conditions. Magnetic Resonance in Chemistry 55, 120–127.

Urnov, F. D., Rebar, E. J., Holmes, M. C., Zhang, H. S., and Gregory, P. D. (2010). Genome editing with engineered zinc finger nucleases. Nature Reviews Genetics 11, 636.

Vinson, J. a., Demkosky, C. a., Navarre, D. a., and Smyda, M. a. (2012). High-antioxidant potatoes: Acute in vivo antioxidant source and hypotensive agent in humans after supplementation to hypertensive subjects. Journal of Agricultural and Food Chemistry 60, 6749–6754.

Vos, P. G., Uitdewilligen, J. G. A. M. L., Voorrips, R. E., Visser, R. G. F., and van Eck, H. J. (2015). Development and analysis of a 20K SNP array for potato (*Solanum tuberosum*): an insight into the breeding history. TAG Theoretical and applied genetics. 128, 2387–2401.

Vossen, E. A. G., Gros, J., Sikkema, A., Muskens, M., Wouters, D., Wolters, P., et al. (2005). The *Rpi-blb2* gene from *Solanum bulbocastanum* is an Mi-1 gene homolog conferring broad-spectrum late blight resistance in potato. The Plant Journal 44, 208–222.

Wang, S., Zhang, S., Wang, W., Xiong, X., Meng, F., and Cui, X. (2015). Efficient targeted mutagenesis in potato by the CRISPR/Cas9 system. Plant Cell Reports.

Wanner, L. A. (2006). A survey of genetic variation in Streptomyces isolates causing potato common scab in the United States. Phytopathology 96, 1363–1371.

Wanner, L. A., and Kirk, W. W. (2015). *Streptomyces* – from basic microbiology to role as a plant pathogen. American Journal of Potato Research 92, 236–242.

Weese, T. L., and Bohs, L. (2007). A Three-Gene Phylogeny of the Genus *Solanum* (Solanaceae). Systematic Botany 32, 445–463.

Witek, K., Jupe, F., Witek, A. I., Baker, D., Clark, M. D., and Jones, J. D. G. (2016). Accelerated cloning of a potato late blight–resistance gene using RenSeq and SMRT sequencing. Nature Biotechnology 34, 656–660.

Wu, F., and Tanksley, S. D. (2010). Chromosomal evolution in the plant family Solanaceae. BMC genomics 11, 182.

Ye, M., Peng, Z., Tang, D., Yang, Z., Li, D., Xu, Y., et al. (2018). Generation of self-compatible diploid potato by knockout of *S-RNase*. Nat. Plants 4, 651–654.

Yogendra, K. N., Kushalappa, A. C., Sarmiento, F., Rodriguez, E., and Mosquera, T. (2015). Metabolomics deciphers quantitative resistance mechanisms in diploid potato clones against late blight. Functional Plant Biology 42, 284–298.

Yu, J., and Buckler, E. S. (2006). Genetic association mapping and genome organization of maize. Current Opinion in Biotechnology 17, 155–160.

Zhang, X., Pérez-Rodríguez, P., Semagn, K., Beyene, Y., Babu, R., López-Cruz, M. a, et al. (2014). Genomic prediction in biparental tropical maize populations in water-stressed and well-watered environments using low-density and GBS SNPs. Heredity, 1–9.

Zhang, Z., Ersoz, E., Lai, C.-Q., Todhunter, R. J., Tiwari, H. K., Gore, M. A., et al. (2010). Mixed linear model approach adapted for genome-wide association studies. Nature Genetics 42, 355.

Zhou, X., and Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. Nature Genetics 44, 821.

Zong, Y., Song, Q., Li, C., Jin, S., Zhang, D., Wang, Y., et al. (2018). Efficient C-to-T base editing in plants using a fusion of nCas9 and human APOBEC3A. Nature Biotechnology 36, 950.

Chapter 2

GENOMIC SELECTION FOR LATE BLIGHT AND COMMON SCAB RESISTANCE IN TETRAPLOID POTATO (*Solanum tuberosum*)

[Published in *G3: Genes, Genomes, Genetics.* 8(7):2471-2481]

Felix Enciso-Rodriguez[1], David Douches[1], Marco Lopez-Cruz[1], Joseph Coombs[1] and Gustavo de Los Campos.[2*]

[1] Department of Plant, Soil, and Microbial Sciences, Michigan State University. East Lansing, Michigan, 48824. [2] Department of Epidemiology & Biostatistics, Department of Statistics & Probability and Institute for Quantitative Health Science and Engineering, Michigan State University. East Lansing, Michigan, 48824.

*Corresponding author (deloscampos@epi.msu.edu)

I. 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) genetic markers and field evaluations for 1,763 late blight phenotypes collected in seven years and 3,885 common scab phenotypes in nine years. We report moderately high genomic heritability estimates (0.46±0.04 and 0.45±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.

II. Introduction

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 total global potato production (Haverkort et al., 2009). Under increasingly variable 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).

Although pathogen infection can be controlled by using protectants or systematic fungicides, they 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, which is time-consuming and 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, Endelman, *et al.* 2017) 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.

III. Materials and Methods

i. 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 even late blight distribution in the field. After approximately 60 days, all plots were inoculated with a zoospore suspension of late blight at $3x10^6$ 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* genotype (clonal lineage) detected in each year in the trial can be found in Table S2.1. 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 for 35-50 days post inoculation (DPI). Finally, the percent defoliation data was 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 heterozygous alleles (AAAB, AABB, ABBB, respectively) and 4, homozygous allele (BBBB). Finally, 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 systematic 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): $\boldsymbol{GRM} = \frac{\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{X}}'}{ncol(\widetilde{\boldsymbol{X}})}$. Here, $\boldsymbol{GRM}$ is a matrix describing genomic relationships between genotypes, $\widetilde{\boldsymbol{X}} = \{[X_{im} - mean(X_{im})]/sd(X_{im})\}$ 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, $ncol(\widetilde{\boldsymbol{X}})$, makes the average diagonal value of $\boldsymbol{GRM}$ 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 $\boldsymbol{GRM}$.

ii. 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:

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

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, μ 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 ε_{ijk} are error terms, which were treated as normal, independently and identically distributed (*iid*) with year-specific variances, that is $\varepsilon_{ijk} \underset{\sim}{^{iid}} N(0, \sigma_j^2)$.

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 $g_i \underset{\sim}{^{iid}} N(0, \sigma_g^2)$. 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 α_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 $g_i = \sum_{m=1}^{4110} \tilde{X}_{im}\alpha_m + \sum_{m=1}^{4110} \widetilde{D}_{im}d_m$ where $\widetilde{D}_{im} = [D_{im} - mean(D_{im})]/sd(D_{im})$ 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, $g_i = \sum_{m=1}^{4110} \sum_{n=0}^{4} \widetilde{W}_{imn}\gamma_{mn}$, where $\widetilde{W}_{imn}$ are (centered and standardized) dummy variables for genotypes carrying *n* copies of the reference allele and γ_{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.

a. 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, $ge_{ij} \overset{iid}{\sim} N(0, \sigma^2_{ge})$.

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 2.1.

Table 2.1. *Sequence of models.*

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

[a] M1-M7 are model numbers. [b] Random effect of the genotype (no SNPs used, no assumption about gene action are made). [c] Principal components, [d] linear regression on allele content (0/1/2/3/4), [e] Simple dominance (1 degree of freedom per locus representing heterozygous) and [f] General model for additive + dominance (with up to 4 degrees of freedom per locus). [g] Genotype-by-year interaction. An '×' indicates that the effects was included in the model.
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.

b. 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). 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.

c. 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 to unit variance; therefore, $Var(X_{im}\alpha_m) = \alpha_m^2$. We used

samples from the posterior distribution of SNP effects from the A model to assess the contribution of individual loci to additive variance.

d. Prediction accuracy evaluation.

We implemented two cross-validation schemes. First, we used a 5-fold 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 5-fold 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^{th} fold, data from the j^{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 i^{th} genotype on the j^{th} year all the data from the i^{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.

e. 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.

IV. Results

The distribution of late blight and common scab infection varied substantially between years (Figure 2.1). In general, RAUDPC median values decreased from 2010 to 2012, with US-22 as the prevalent late blight genotype on infected plants. In subsequent years, a differential response for late blight resistance was observed when US-23 was the prevalent genotype. 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 2.1).

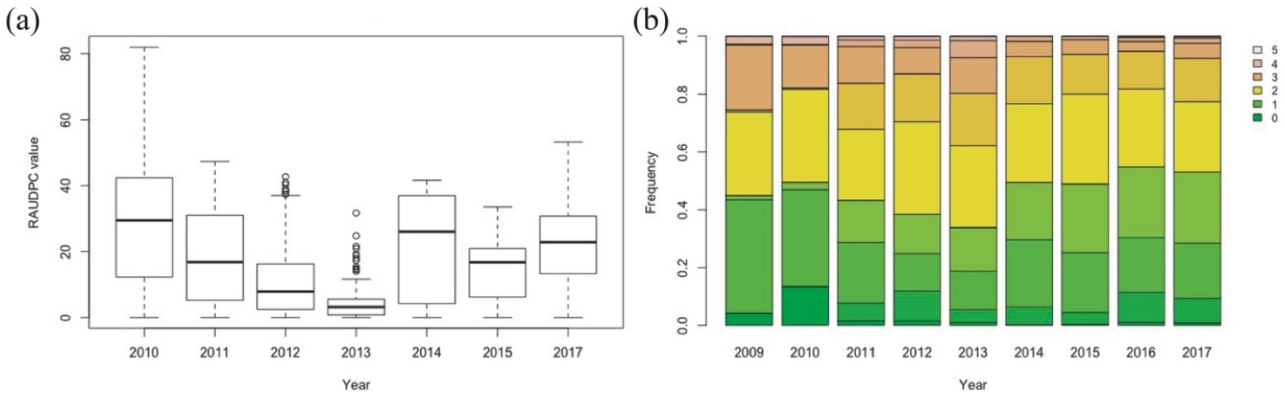

Figure 2.1. *(a) Boxplot of late blight scores (relative area under the disease progress curve-RAUDPC), and (b) bar plot for common scab scores (0-5 rating scale).*

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 2.2). The eigenvalues associated to the first two PCs explained about 8% of the total genotype variance (Figure 2.2).

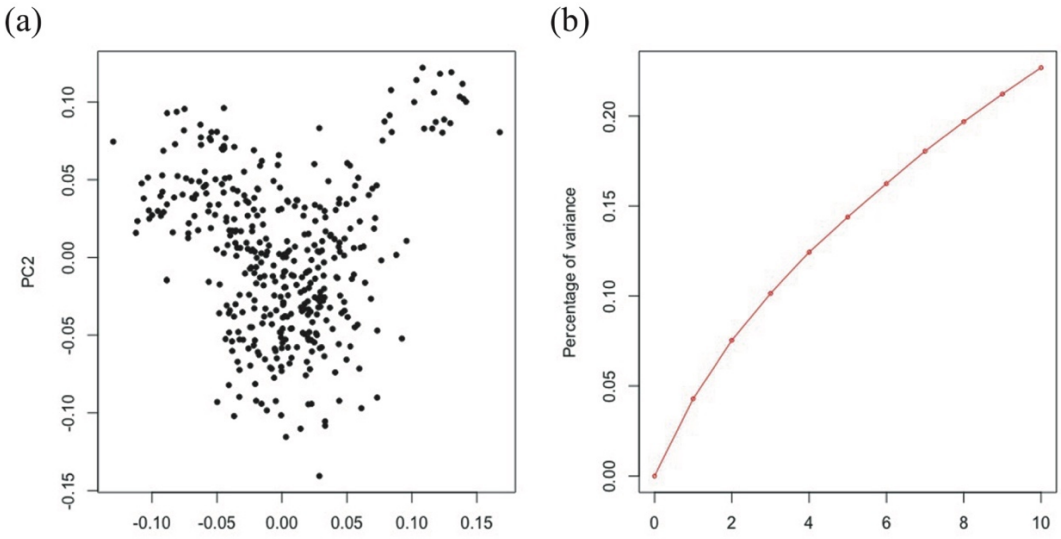

Figure 2.2. *Principal component analysis of the Michigan State University's potato breeding genotypes derived from 4,110 SNPs*: loadings on the first two marker-derived principal components (a) and proportion of variance explained by the top 10 principal components (b).

A cluster analysis using a correlation matrix derived from SNP markers supports the PC-analysis results (Figure 2.3). 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 2.3).

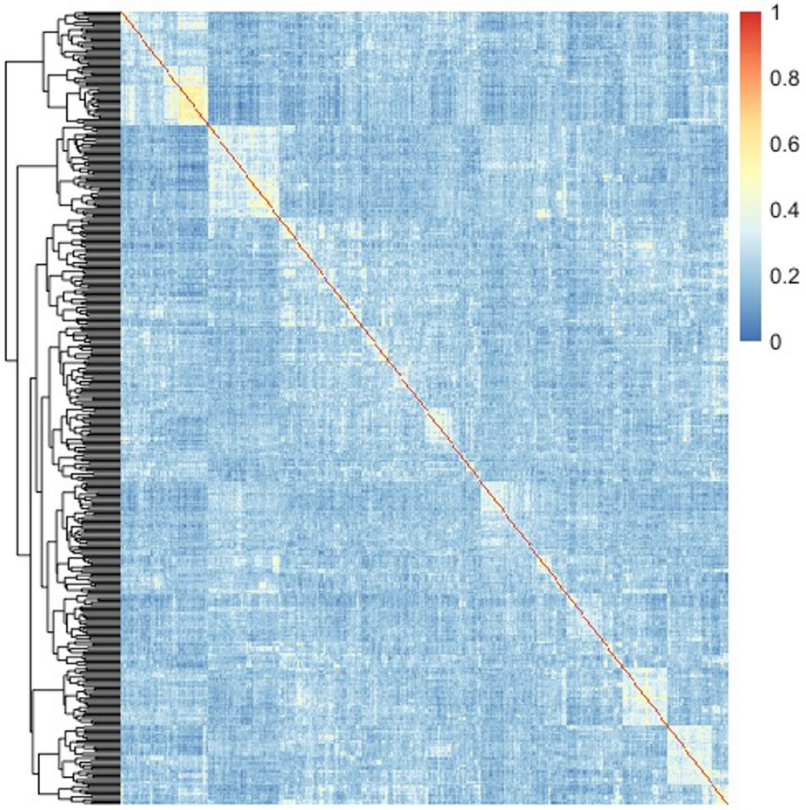

Figure 2.3. *Heatmap of the genomic relationship matrix (GRM) from the Michigan State University's potato breeding genotypes.*

i. Variance Components Estimates

The variance components analyses for ***late blight*** resistance (Tables 2.2 and S2.2) 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).

Table 2.2. *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 var variance explained by each component. Results obtained with the fully Gaussian model (BRR) are presented in Table S2.2.

Model # (label)[a]	*Year*	*Genetic*							
		Genotype	*Marker-derived*				Total genetic[f]	Genotype-by-year[g]	Error
			PC[b]	Additive[c]	Dominance[d]	General[e]			
Late blight									
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)

Table 2.2. (cont'd)

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 -0.06		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)

[a] M1-M7 are model numbers (label). The effects included in each of them are described in the columns. [b] Principal components, [c] linear regression on allele content (0/1/2/3/4), [d] Simple dominance (1 degree of freedom per locus representing heterozygous) and [e] General model for additive + dominance (with up to 4 degrees of freedom per locus). [f] Total genetic variance, [g] Genotype-by-year interaction.

For ***common scab*** (Tables 2.2 and S2.2) 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. The proportion of the total genetic variance that could be attributed to the first-5 PCs was substantial for late blight (~30%, computed as 0.107/0.352) and low for common scab (~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 2.4a and 2.4b). Additionally, we report in Figures S2.1 and S2.2, 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 2.4a and Table S2.3) 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 2.4b and Table S2.4), that

stands out for its contribution to variance and a few SNPs with a moderate contribution to phenotypic variance.

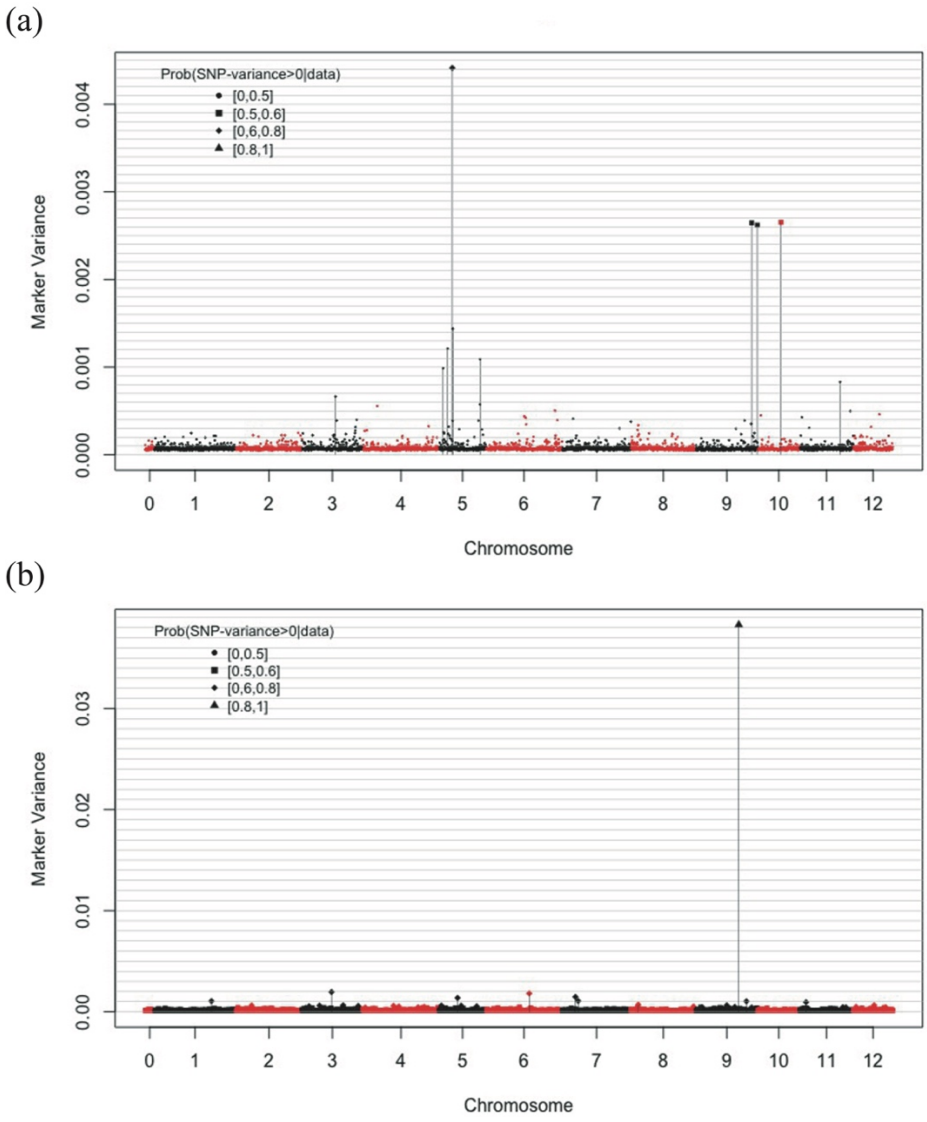

Figure 2.4. *Estimated SNP-variances derived from BayesB model using the additive model for late blight (a) and common scab (b).* (In both cases, phenotypes where diseases scores standardize to a variance equal to one. Vertical lines indicate the positions of the top-10, according to estimated SNP-variance markers).

ii. Cross-validation analysis

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 2.3). In the case of ***common scab***, the A model (with a prediction correlation of ~0.27) outperformed the A+D (correlation ~0.26) and G (correlation ~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.

Table 2.3. *Cross-validation correlations obtained with BRR and BayesB models by trait and model.*

Prior[a]	**Model # (label)** [b]	**CV-Correlation**		**Proportion of times that the model in row gave a higher correlation than the model in columns**					
				BRR[a]			**BayesB**[a]		
		Average[c]	SD[d]	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	0.33	0.91	0
	M6 (A+D)	0.241	0.023	0.04		0	0.04	0.57	0
	M7 (G)	**0.312**	0.017	1	1		1	1	0.5
BayesB	M5 (A)	0.26	0.024	0.67	0.96	0		0.94	0
	M6 (A+D)	0.24	0.024	0.09	0.43	0	0.06		0
	M7 (G)	**0.313**	0.017	1	1	0.5	1	1	
Common scab									
BRR	M5 (A)	**0.268**	0.025		0.81	0.99	0.1	0.55	1
	M6 (A+D)	0.259	0.023	0.19		0.99	0.07	0.2	0.99
	M7 (G)	0.218	0.022	0.01	0.01		0	0.02	0.63
BayesB	M5 (A)	**0.278**	0.026	0.9	0.93	1		0.91	1

Table 2.3. (cont'd)

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	0.02	

[a] BRR uses a Gaussian prior for effects, BayesB uses a prior that has a point of mass at zero and a scaled-t slab. [b] A: Additive model, A+D: additive+dominance; G: general model (with up to 4 degrees of freedom per locus).[c] Average from 100 cross-validations. [d] Standard deviation.

Note that in Table 2.3 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 (Tables 2.4 and S2.5). 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).

Table 2.4. *Year cross-validation correlations obtained with BayesB model by trait and model.*

Year	**Model # (label)** [a]				
	M2	M3	M5 (A)	M6 (A+D)	M7 (G)
Late blight					
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	0.64	0.655
2015	0.719	0.73	0.678	0.696	0.745
2017	0.508	0.504	0.471	0.491	0.506

Table 2.4. (cont'd)

Average	0.579	0.578	0.56	0.568	**0.597**
SD	0.098	0.104	0.087	0.089	0.095
			Common scab		
2009	0.459	0.46	0.472	0.471	0.466
2010	0.52	0.522	0.535	0.533	0.517
2011	0.61	0.611	0.622	0.625	0.618
2012	0.625	0.628	0.628	0.626	0.634
2013	0.75	0.759	0.731	0.737	0.75
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	0.51	0.519	0.52	0.515
Average	0.597	0.599	0.606	**0.607**	0.605
SD	0.088	0.09	0.082	0.083	0.089

[a]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).

V. 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 polypoid 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±0.04 for late blight and 0.45±0.02 for common scab) can be explained using 4,110 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 versus tetraploid, hybrids versus genotypes) and of the environmental conditions (natural versus induced infection) makes the comparisons across studies difficult (Nelson 1978; Braun, Gevens, *et al.* 2017). 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, Pajerowska-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. stenotonum*) 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 was 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 2.2 and Table S2.2). 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 late blight case. 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 2.4). 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 SNP-variances that compared with most of them were large. For late blight, these analyses lead to the finding of many relatively large-variance SNPs located mainly in chromosomes V and IX (Figure 2.4a and Table S2.3). For late blight resistance, multiple quantitative trait loci (QTLs) have been reported across the 12 potato chromosomes in tetraploid and diploid potato populations (Tiwari *et al.* 2013). Most of these major QTLs 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; Pajerowska-Mukhtar *et al.* 2009; Álvarez *et al.* 2017). For instance, genes involved in carbohydrate metabolism such as sucrose synthase (Table S2.3) 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) or non-host pathogens and priming the synthesis of metabolites associated with the production of defense-related compounds (Stamler

et al. 2015). 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; 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 (Figure 2.4 and Table S2.4). 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 2.4). 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 S2.2).

For instance, we found one SNP located in chromosome III (Table S2.4) 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 QTLs 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, Endelman, *et al.* 2017). 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.

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 (~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 2.3 with those for M2 in Table 2.4). 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 2.4) 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.

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, 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.

APPENDIX

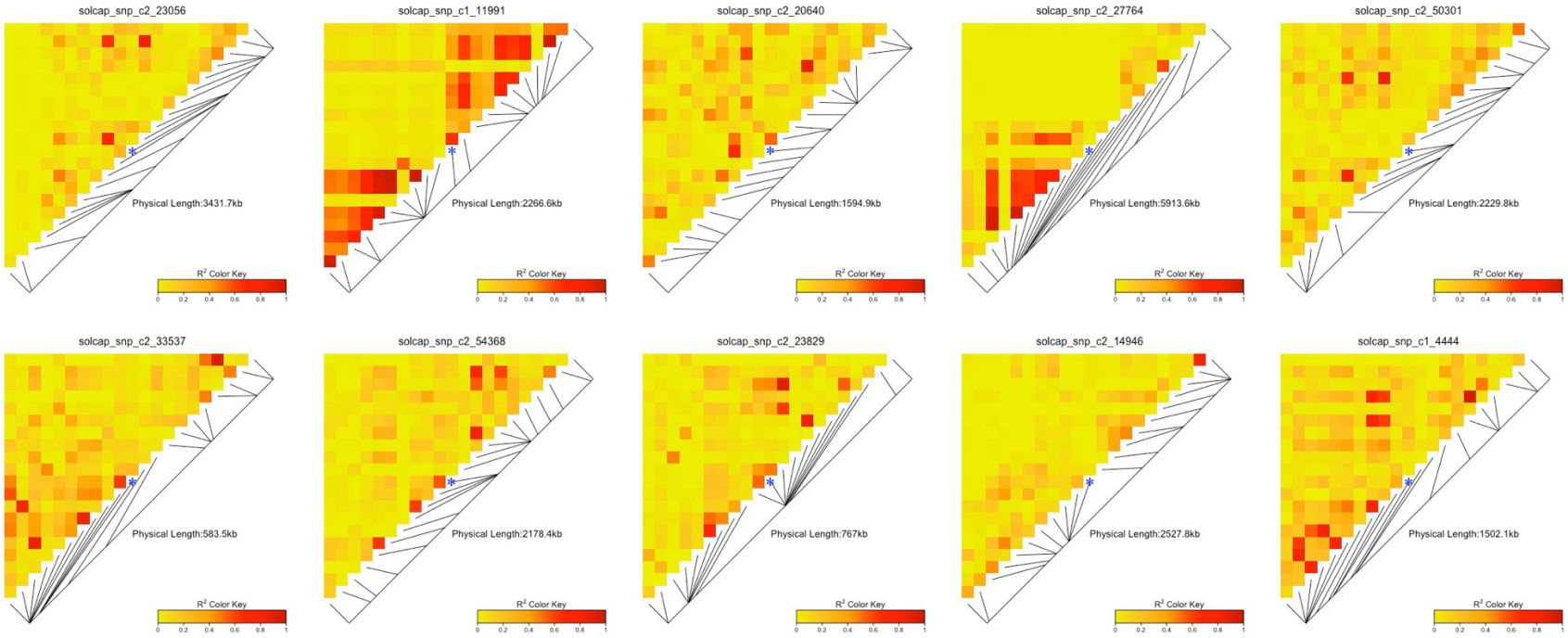

Figure S2.1. *Linkage disequilibrium plots for the top ten ranked loci according to their contribution to the late blight genotypic variance.* The blue asterisk shows the SNP position on each plot.

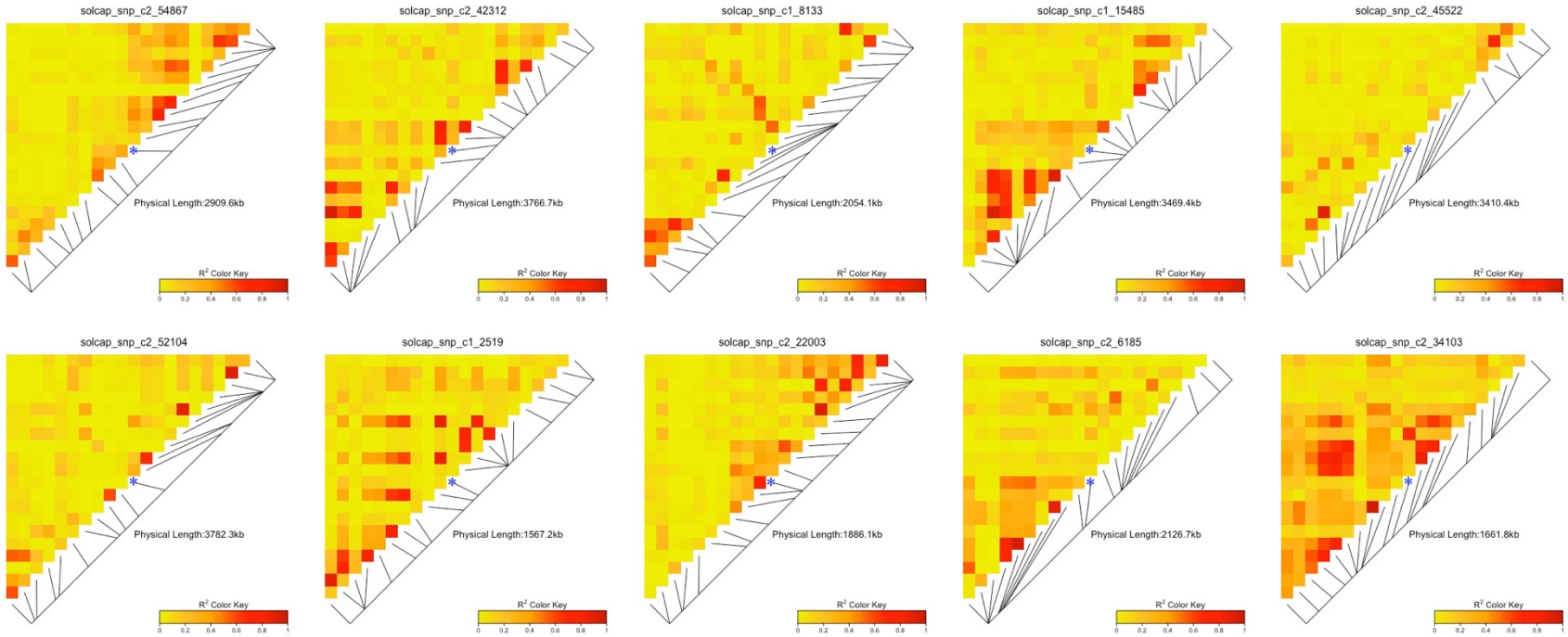

Figure S2.2. *Linkage disequilibrium plots for the top ten ranked loci according to their contribution to the common scab genotypic variance.* The blue asterisk shows the SNP position on each plot.

Table S2.1. *Late blight clonal lineages detected throughout seven years at the Michigan State University Clarksville Research Center (Clarksville, MI).*

Year	Clonal lineage
2010	US-22
2011	US-22
2012	US-22
2013	US-23
2014	US-23
2015	US-23
2017	US-23

Table S2.2. *Variance components estimates (posterior standard deviation) derived from BRR 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.*

Model # (label)[a]	Year	Genetic							
		Genotype	Marker-derived				Total genetic [f]	Genotype-by-year [g]	Error
			PC [b]	Additive[c]	Dominance[d]	General[e]			
Late blight									
M1	0.266 (0.021)								0.735 (0.025)
M2	0.257 (0.014)	0.434 (0.018)					0.434 (0.018)		0.303 (0.011)
M3	0.249 (0.028)	0.340 (0.031)					0.340 (0.031)	0.253 (0.019)	0.144 (0.006)
M4	0.247 (0.026)	0.266 (0.029)	0.097 (0.025)				0.353 (0.033)	0.250 (0.018)	0.144 (0.006)
M5 (A)	0.238 (0.028)		0.139 (0.069)	0.270 (0.046)			0.314 (0.036)	0.285 (0.022)	0.144 (0.006)
M6 (A+D)	0.242 (0.027)		0.132 (0.060)	0.162 (0.061)	0.13 (0.045)		0.329 (0.035)	0.272 (0.020)	0.144 (0.006)
M7 (G)	0.245 (0.027)		0.105 (0.038)			0.271 (0.033)	0.346 (0.033)	0.253 (0.019)	0.144 (0.006)
Common Scab									
M1	0.033 (0.006)								0.971 (0.022)
M2	0.030 (0.004)	0.456 (0.017)					0.456 (0.017)		0.523 (0.012)
M3	0.029 (0.005)	0.440 (0.020)					0.440 (0.020)	0.060 (0.008)	0.483 (0.013)
M4	0.029 (0.005)	0.419 (0.021)	0.030 (0.012)				0.447 (0.023)	0.056 (0.009)	0.485 (0.012)
M5 (A)	0.032 (0.006)		0.117 (0.076)	0.499 (0.070)			0.443 (0.025)	0.061 (0.009)	0.485 (0.013)
M6 (A+D)	0.031 (0.006)		0.079 (0.052)	0.278 (0.085)	0.205 (0.073)		0.445 (0.025)	0.057 (0.009)	0.486 (0.013)
M7 (G)	0.031 (0.006)		0.054 (0.031)			0.441 (0.031)	0.448 (0.023)	0.056 (0.009)	0.485 (0.013)

[a] M1-M7 are model numbers (label). The effects included in each of them are described in the columns. [b] Principal components, [c] linear regression on allele content (0/1/2/3/4), [d] Simple dominance (1 degree of freedom per locus representing heterozygous) and [e] General model for additive + dominance (with up to 4 degrees of freedom per locus). [f] Total genetic variance, [g] Genotype-by-year interaction.

Table S2.3. *Top ten ranked loci from the whole genome regression derived from BayesB model, according to their additive variance estimates for late blight.*

Locus	Chromosome	Proportion of variance explained (Prob.)*	Putative function
solcap_snp_c2_23056	V	0.00441 (0.669)	60S ribosomal protein L34
solcap_snp_c1_11991	X	0.00227 (0.596)	Pre-mRNA-splicing factor syf2
solcap_snp_c2_20640	IX	0.00264 (0.590)	Ammonium transporter 1 member 1
solcap_snp_c2_27764	IX	0.00262 (0.578)	Sucrose synthase
solcap_snp_c2_50301	V	0.00143 (0.490)	Gene of unknown function
solcap_snp_c2_33537	V	0.00121 (0.475)	6-phosphogluconate dehydrogenase, decarboxylating
solcap_snp_c2_54368	V	0.00109 (0.467)	Chalcone synthase
solcap_snp_c2_23829	V	0.00098 (0.427)	Dihydropterin pyrophosphokinase-dihydropteroate synthase
solcap_snp_c2_14946	XI	0.00083 (0.449)	Xanthine dehydrogenase
solcap_snp_c1_4444	III	0.00066 (0.409)	HCF106; proton motive force dependent protein transmembrane transporter

* Estimated posterior probability of having a SNP variance greater than zero given the data.

Table S2.4. *Top ten ranked loci from the whole genome regression derived from BayesB model, according to their additive variance estimates for common scab.*

Locus	Chromosome	Proportion of variance explained (Prob.)*	Putative function
solcap_snp_c2_54867	IX	0.0306 (0.977)	WRKY transcription factor
solcap_snp_c2_42312	III	0.0019 (0.695)	Fructokinase
solcap_snp_c1_8133	VI	0.0018 (0.766)	TAF5
solcap_snp_c1_15485	VII	0.0015 (0.690)	Conserved gene of unknown function
solcap_snp_c2_45522	V	0.0014 (0.698)	DEAD-box ATP-dependent RNA helicase 22
solcap_snp_c2_52104	VII	0.0011 (0.678)	Auxin response factor
solcap_snp_c1_2519	I	0.001 (0.699)	Leucine-rich repeat-containing protein
solcap_snp_c2_22003	IX	0.001 (0.683)	Homocysteine s-methyltransferase
solcap_snp_c2_6185	XI	0.0009 (0.643)	Dead box ATP-dependent RNA helicase
solcap_snp_c2_34103	VIII	0.0007 (0.409)	Conserved gene of unknown function

***** Estimated posterior probability of having a SNP variance greater than zero given the data.

Table S2.5. *Year cross-validation correlations obtained with BRR model by trait and model.*

Year	Model # (label) [a]				
	M2	M3	M5 (A)	M6 (A+D)	M7 (G)
Late blight					
2010	0.551	0.531	0.456	0.463	0.518
2011	0.651	0.661	0.605	0.612	0.643
2012	0.583	0.611	0.574	0.581	0.614
2013	0.424	0.419	0.482	0.492	0.49
2014	0.621	0.597	0.634	0.654	0.647
2015	0.72	0.724	0.683	0.692	0.735
2017	0.509	0.504	0.467	0.491	0.499
Average	0.58	0.578	0.557	0.569	0.592
SD	0.097	0.102	0.09	0.089	0.092
Common scab					
2009	0.458	0.461	0.469	0.471	0.466
2010	0.52	0.521	0.534	0.534	0.519
2011	0.61	0.613	0.627	0.628	0.618
2012	0.625	0.629	0.626	0.626	0.634
2013	0.751	0.758	0.734	0.736	0.75
2014	0.615	0.612	0.63	0.634	0.636
2015	0.649	0.654	0.67	0.672	0.66
2016	0.639	0.639	0.65	0.646	0.647
2017	0.507	0.51	0.517	0.518	0.514
Average	0.597	0.6	0.606	**0.607**	0.605
SD	0.088	0.09	0.083	0.083	0.089

[a]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 dominance modeled using up to 4 degrees of freedom per locus ('General' model).

LITERATURE CITED

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 Science 63: 595–603.

Álvarez, M. F., M. Angarita, M. C. Delgado, C. García, J. Jiménez-Gomez et al. (2017). Identification of novel associations of candidate genes with resistance to late blight in *Solanum tuberosum* Group Phureja. Frontiers in Plant Science 8: 1040.

Barone, A. (2004). Molecular marker-assisted selection for potato breeding. American Journal of Potato Research 81: 111–117.

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). Theoretical and Applied Genetics 116: 193–211.

Braun, S. R., J. B. Endelman, K. G. Haynes, and S. H. Jansky. (2017). Quantitative trait loci for resistance to common scab and cold-induced sweetening in diploid potato. The Plant Genome 10: 1–9.

Braun, S., A. Gevens, A. Charkowski, C. Allen, and S. Jansky. (2017). potato common scab: a review of the causal pathogens, management practices, varietal resistance screening methods, and host resistance. American Journal of Potato Research. 94(4): 283–296.

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 Science 52: 707–719.

Buscaill, P., and S. Rivas. (2014). Transcriptional control of plant defense responses. Current Opinion in Plant Biology 20: 35–46.

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.

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 Research 55: 249–268.

Drewnowski, A., and C. D. Rehm. (2013). Vegetable cost metrics show that potatoes and beans provide most nutrients per penny. PLoS One 8(5): e63277.

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. American Journal of Potato Research 86: 96–101.

Falconer, D., and T. Mackay. (1996). Introduction to quantitative genetics (4e). Pearson United Kingdom.

FAOSTAT. (2018). FAOSTAT database collections. Food and Agriculture Organization of the United Nations.

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.

Gebhardt, C., A. Ballvora, B. Walkemeier, P. Oberhagemann, and K. Schüler. (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. Molecular Breeding 13: 93–102.

Habyarimana, E., B. Parisi, and G. Mandolino. (2017). Genomic prediction for yields, processing and nutritional quality traits in cultivated potato (*Solanum tuberosum* L.). Plant Breeding 136: 245–252.

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 Research 52: 249–264.

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 Breeding 118: 431–434.

Haynes, K. G., B. J. Christ, C. R. Burkhart, and B. T. Vinyard. (2009). Heritability of resistance to common scab in diploid potatoes. American Journal of Potato Research 86: 165–170.

Haynes, K. G., R. W. Goth, and R. J. Young. (1997). Genotype × environment interactions for resistance to common scab in tetraploid potato. Crop Science 37: 1163–1167.

Heffner, E. L., M. E. Sorrells, and J.-L. Jannink. (2009). Genomic selection for crop improvement. Crop Science 49: 1–12.

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]. Physiological Plant Pathology 3: 121–131.

Kolasa, K. (1993). The potato and human nutrition. American Potato Journal 70: 375–384.

Lehermeier, C., G. de los Campos, V. Wimmer, and C.-C. Schön. (2017). Genomic variance estimates: With or without disequilibrium covariances? Journal of Animal Breeding and Genetics 134: 232–241.

Li, D., H. Liu, H. Zhang, X. Wang, and F. Song. (2008). *OsBIRH1*, a DEAD-box RNA helicase with functions in modulating defense responses against pathogen infection and oxidative stress. Journal of Experimental Botany 59: 2133–2146.

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.

de los Campos, G., D. Sorensen, and D. Gianola. (2015). Genomic heritability: What is it?. PLOS Genetics 11(5): e1005048.

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(4): e34775.

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.

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(6): 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. Frontiers in Genetics 6: 1–19.

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.

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. American Journal of Potato Research 92: 326–338.

Nelson, R. R. (1978). Genetics of horizontal resistance to plant diseases. Annual Review of Phytopathology 16: 359–378.

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 Disease 96: 4–17.

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 Physiology 150: 1648–1655.

Pérez, P., and G. de los Campos. (2014). Genome-wide regression & prediction with the BGLR Statistical Package. Genetics 198(2): 483–495

Pomerantz, A., Y. Cohen, E. Shufan, Y. Ben-Naim, S. Mordechai et al. (2014). Characterization of *Phytophthora infestans* resistance to mefenoxam using FTIR spectroscopy. Journal of Photochemistry and Photobiology B: Biology 141: 308–314.

Solano, J., I. Acuña, F. Esnault, and P. Brabant. (2014). Resistance to *Phytophthora infestans* in *Solanum tuberosum* landraces in Southern Chile. Tropical Plant Pathology 39: 307–315.

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. Global Change Biology 20: 3621–3631.

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: 1–19.

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. Theoretical and Applied Genetics 130: 2091–2108.

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.

The Potato Genome Sequencing Consortium. (2011). Genome sequence and analysis of the tuber crop potato. Nature 475: 189–195.

The R Development Core Team (2010). R : A Language and Environment for Statistical Computing. Retrieved from 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 Breeding 132: 237–245.

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. Journal of Dairy Science 91: 4414–4423.

Wanner, L. A. (2006). A survey of genetic variation in *Streptomyces* isolates causing potato common scab in the United States. Phytopathology 96: 1363–1371.

Wanner, L. A., and W. W. Kirk. (2015). *Streptomyces* – from basic microbiology to role as a plant pathogen. American Journal of Potato Research 92: 236–242.

Zimaro, T., N. Gottig, B. S. Garavaglia, C. Gehring, and J. Ottado. (2011). Unraveling plant responses to bacterial pathogens through proteomics. Journal of Biomedicine and Biotechnology 2011: 354801.

Chapter 3

OVERCOMING SELF-INCOMPATIBILITY IN DIPLOID POTATO USING CRISPR-CAS9

[*Front. Plant Sci. doi: 10.3389/fpls.2019.00376*]

Felix Enciso-Rodriguez[1], Norma C. Manrique-Carpintero[1], Satya Swathi Nadakuduti[1], C. Robin Buell[2,3,4], Daniel Zarka[1], David Douches.[1,4*]

[1] Department of Plant, Soil and Microbial Sciences, Michigan State University, East Lansing, Michigan, 48824. [2] Department of Plant Biology, Michigan State University, East Lansing, Michigan, 48824. [3] Plant Resilience Institute, Michigan State University, East Lansing, Michigan. [4] Michigan State University AgBioResearch, East Lansing, Michigan.

* Corresponding author (douchesd@msu.edu)

I. Abstract

Potato breeding can be redirected to a diploid inbred/F1 hybrid variety breeding strategy if self-compatibility can be introduced into diploid germplasm. However, the majority of diploid potato clones (Solanum spp.) possess gametophytic self-incompatibility (SI) that is primarily controlled by a single multiallelic locus called the *S*-locus which is composed of tightly linked genes, *S-RNase* (*S*-locus RNase) and multiple *SLFs* (*S*-locus F-box proteins), which are expressed in the style and pollen, respectively. Using *S-RNase* genes known to function in the Solanaceae gametophytic SI mechanism, we identified *S-RNase* alleles with flower-specific expression in two diploid self-incompatible potato lines using genome resequencing data. Consistent with the location of the *SLF* gene in potato, we genetically mapped the *S-RNase* gene using a segregating population to a region of low recombination within the pericentromere of chromosome I. To generate self-compatible diploid potato lines, a dual single-guide RNA (sgRNA) strategy was used to target conserved exonic regions of the *S-RNase* gene and generate targeted knock-outs using a Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR-associated approach. Self-compatibility was achieved in nine *S-RNase* knock-out (KO) T_0 lines which contained bi-allelic and homozygous deletions/insertions in both genotypes, transmitting SC to T_1 progeny. This study demonstrates an efficient approach to achieve stable, consistent self-compatibility through *S-RNase* KO for use in diploid potato breeding approaches.

II. Introduction

Cultivated potato (*Solanum tuberosum* L.) is the third most important food crop after rice and wheat (Devaux et al. 2014) and plays an essential role in human nutrition as a primary source of carbohydrates. Although global production of potato totaled 388 million tons in 2017

(FAOSTAT 2019), potatoes face high production losses due to biotic and abiotic stresses that will increase with global warming (Raymundo et al. 2017). While improvement of cultivated potatoes (2n=4x=48) relies on the discovery and introgression of genes from wild species for traits such as disease resistance, the polyploid nature of cultivated tetraploid potato hampers the fixation of desirable alleles in new cultivars. For example, the introgression of critical dominant alleles such as the potato virus Y (PVY) disease-resistance gene in a triplex or quadruplex allelic configuration can take up to 15 years (Mendoza et al. 1996). Re-inventing potato as a diploid inbred/F1 hybrid variety (2n=2x=24) would allow the application of efficient breeding methods (Jansky et al. 2016) as inbred potatoes would accelerate the generation of new varieties with favorable allelic combinations targeting yield, tuber quality, and resistance traits. A significant barrier to this approach is the occurrence of gametophytic self-incompatibility (SI) in a majority of the diploid potato germplasm, thereby preventing the ability to generate diploid homozygous lines.

In diploid potato, the gametophytic SI system is controlled by a single multiallelic locus called the S-locus (Porcher and Lande 2005). This locus is composed of tightly linked genes, *S-RNase* (*S*-locus RNase) and *SLFs* (*S*-locus F-box) genes known also as *S*-haplotype-specific F-box brothers (SFBB), expressed in the style and pollen, respectively (Takayama and Isogai 2005; Sassa et al. 2007; Kubo et al. 2010; Bush and Moore 2012). The S-RNase protein produces cytotoxic effects that inhibit the elongation of self-pollen tubes via degradation of RNA from the pollen whereas *SLF* function as a component of a detoxification complex that mediates ubiquitination of non-self S-RNase proteins leading to degradation via the proteasome pathway (Sijacic et al. 2004; Kubo et al. 2015). Hence, when self-pollination occurs in self-incompatible

individuals, the *SLFs* genes do not recognize their native *S-RNase* and consequently, pollen tube growth in the style is inhibited due to the ribonuclease activity of the *S-RNase* (Hua et al. 2008).

In an effort to develop diploid-self-compatible (SC) potato lines, the inbred line M6 was generated from the wild tuber-bearing species, *Solanum chacoense* (Jansky et al. 2014). In M6, a dominant allele of the *S*-locus inhibitor (*Sli*) inactivates the gametophytic SI system (Hosaka and Hanneman 1998) leading to self-compatibility. However, introgression of *Sli* into other germplasm is time-consuming and could lead to linkage drag and fixation of undesirable traits such as high tuber glycoalkaloid content from the donor *S. chacoense*. An alternative strategy to *Sli* introgression is the use of genome editing to accelerate the generation of SC diploid lines by targeting genes involved in SI.

Genome editing by Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-associated protein 9 (Cas9) system has been widely used to generate gene knock-outs (KO) of candidate genes related to agronomic traits in important crops (Jaganathan et al. 2018). Cas9 induces double-strand breaks (DSB) in the DNA at the target site, triggering the response of endogenous cell repair mechanisms. One of the cellular mechanisms to repair DSBs is non-homologous end joining (NHEJ), which can generate insertions and deletions in the coding region resulting in a KO of gene function (Bortesi and Fischer 2015; Pellagatti et al. 2015). The target-DNA recognition is mediated by a single guide-RNA (sgRNA) bearing a 20 bp target-site complementary to region adjacent to a protospacer-adjacent motif (PAM), 5′-NGG-3’, resulting in the generation of a DSB (Doudna and Charpentier, 2014; Pellagatti et al., 2015; Sternberg et al., 2014) which can be leveraged to generate DSB of target genes.

Previous studies in tomato wild relatives demonstrated that missense mutations and gene loss prevent S-RNase ribonuclease activity in *S. peruvianum* and *S. pennellii*, leading to self-

compatibility (Royo et al. 1994; Kowyama et al. 1994; Covey et al. 2010; Li and Chetelat 2015). Considering that *S-RNase* is the gametophytic SI component directly implicated in degradation of RNA in self-pollen tubes, inhibiting the *S-RNase* function is a straightforward strategy to confer SC in potato. In an effort to contribute to the development of diploid inbred potato lines, we generated SC diploid lines by targeted mutagenesis of *S-RNase* using CRISPR-Cas9, obtaining stable self-compatibility in T_0 and T_1 generations. Contemporaneous with the writing of this manuscript, Ye et al. (2018) published their findings using a similar approach. However, this study provides further insight into SI in diploid potatoes, reporting three new *S-RNase* alleles, their localization within a low recombination pericentromeric region consistent with the location of the *S*-locus, generation of stable SC in KO lines, and documentation of plasticity in the phenotype of SI in two diploid lines.

III. Material and Methods

i. Plant material

After an initial test of self-compatibility with more than 50 self-pollinations, the SI diploid potato lines (2n=2x=24) DRH-195 and DRH-310 F1 lines were generated from a cross between S. tuberosum Gp. Phureja DM 1-3 516 R44 (DM) and S. tuberosum Gp Tuberosum RH 89-039-16 (RH) at Virginia Tech and used in this study. Plants were maintained in vitro, propagated on Murashige and Skoog (MS) medium (MS basal salts plus vitamins, 3% sucrose, 0.7% plant agar, pH 5.8) (Murashige and Skoog, 1962) and cultured in growth chambers with 16-h-light/8-h-dark photoperiod at 22°C and average light intensity of 200 µmoles m-2s-1.

ii. Allelic identification, annotation and phylogenetic analysis of S-RNase

TBLASTN (BLAST- basic local alignment search tool) searches were performed using reported S-RNase protein sequences (Table 3.1) from the Solanaceae family against the DM v4.04 assembly (Hardigan et al. 2016) using BLAST v2.2.31 (Altschul et al. 1990) with default parameters. A candidate *S-RNase* gene was selected using the top blast hits. Expression abundances across a range of developmental stages, tissues, and organs were determined using available gene expression atlases for DM and RH (The Potato Genome Sequencing Consortium, 2011). To identify *S-RNase* allelic variants in the diploid potato clones, genomic and complementary DNA sequence data from DRH-195 and DRH-310 leaf and tuber tissues were retrieved from the Sequence Read Archive (SRA) of the National Center for Biotechnology Information (Table S3.1) and aligned to the DM v4.04 assembly using BWA-MEM (Li 2013). Duplicate reads were removed using Picard Tools v1.113 (http://broadinstitute.github.io/picard) and consensus sequences were obtained using the mpileup utility from Samtools v1.2 (Li et al. 2009) with the consensus option from bcftools v1.2 (Li 2011).

A primer set was designed to amplify the predicted ORF of the *S-RNase* gene in DRH-195 and DRH-310 using the detected *S-RNase* variants (Table S3.2). S-RNase amino acid sequences reported in Table 3.1, including the alleles reported by Ye et al. (2018), along with the deduced amino acid sequences from the S-RNase variants identified in this study were aligned using Clustal Omega (Sievers et al. 2011). A phylogenetic tree was constructed using the Neighbor joining method with 1000 bootstrap replicates in MEGA version 7.0 (Kumar et al. 2016). Amino acid similarities percentages were calculated using BioEdit (Hall 1999).

Table 3.1. *S-RNase sequences from seven Solanaceae species used in this study.*

Gene/protein[†],	**Accession**[*]	**Species**
Ribonuclease S-2	Q01796	*Solanum tuberosum*
RNase	CAA05306	*Nicotiana sylvestris*
S-RNase	BAC00940	*Solanum neorickii*
S1-RNase	BAC00934	*Solanum chilense*
S11	AAA50306	*Solanum chacoense*
S2 self-incompatibility ribonuclease precursor	AAG21384	*Petunia integrifolia subsp. inflata*
Sx-protein	AAA33729	*Petunia* x *hybrida*

[*] Gene or protein name, [‡]National Center for Biotechnology Information (NCBI) accession identifier.

iii. S-RNase linkage mapping

The previously reported diploid DRH mapping population was used to genetically map the *S-RNase* locus (Manrique-Carpintero et al., 2015). DNA was isolated from DRH-195 and DRH-310 young leaves using the DNeasy Plant Mini Kit (Qiagen, Hilden, Germany), and used for PCR with a Q5® High-Fidelity DNA Polymerase (New England Biolabs, Ipswich, MA) with the following thermocycler conditions: one cycle of initial denaturation for 4 min at 94 °C, followed by 34 cycles for 15 s at 94 °C, 45 s at 56 °C and 1 min at 72 °C and a final extension of 5 min at 72 °C. *S-RNase* amplicons were gel-purified using the QIAquick PCR Purification Kit (Qiagen, Hilden, Germany) and cloned into the Zero Blunt TOPO PCR Cloning vector (ThermoFisher, Carlsbad, CA). Ten colonies for each line were sequenced by the Sanger method and aligned using Clustal Omega (Sievers et al. 2011). DM and RH allelic sequences were confirmed and used to design S-RNase RH-allele specific primers (Table S3.2). These primers were screened across 80 individuals of the DRH mapping population. The genotype from the presence/absence of an RH allele was coded as nnxnp and used for mapping in JoinMap4.1 with the same parameters as previously reported by Manrique-Carpintero et al. (2015).

iv. sgRNA identification, assembly and validation

A double sgRNA construct targeting predicted conserved regions from the first (sgRNA 1) and second (sgRNA 2) *S-RNase* exons were designed using CRISPR RGEN tools (Table S3.2, Park et al., 2015). A gene KO construct containing the sgRNA combination (sgRNA 1-2) was assembled in the pHSE40 vector containing the CRISPR-Cas9 cassette as described by Xing et al. (2014) and transferred into *Agrobacterium tumefaciens* strain GV3101 pMP90 (Koncz et al. 1994) by electroporation.

v. Agrobacterium-mediated transformation

Agrobacterium-mediated transformation was performed using leaf segments from four-week-old tissue culture plants of DRH 195 and DRH 310 as described by Li et al. (1999). Briefly, explants were pre-cultured on a step I media (MS salts, 3% sucrose, 5 g/l phytoagar, 1 mg/l thiamine-HCl, 0. 8 mg/l zeatin-riboside and 2 mg/l 2,4-D) for four days and inoculated with *Agrobacterium*. After three days, explants were rinsed with sterile distilled water containing 250 mg/l cefotaxime and 200 mg/l carbenicillin and placed onto step II media (MS salts, 3% sucrose, 5 g/l phytoagar, 1 mg/l thiamine-HCl, 0.8 mg/l zeatin- riboside, 2 mg/l gibberellic acid, 20 mg/l hygromycin and 150 mg/l ticarcillin disodium and clavulanate potassium). Explants were transferred to fresh step II media every week. After approximately 30 days, transformation events (T_0 lines) were selected from step II media and transferred to root induction media containing MS medium supplemented with antibiotics for selection as described above.

vi. Molecular characterization of KO lines

DNA from T_0 and T_1 plants was isolated as described above. PCR was carried out using the GoTaq DNA polymerase (Promega, Fitchburg, WI) with the following thermocycler conditions: one cycle of initial denaturation for 4 min at 94 °C, followed by 34 cycles for 15 s at 94 °C, 45 s at 56 °C and 1 min at 72 °C and a final extension of 5 min at 72 °C. Amplicons were visualized on 1% (w/v) agarose gels. Allelic mutations of positive transformation events were identified by insertion/deletion presence. Selected transformation events were amplified with the Q5 High-Fidelity DNA Polymerase (New England Biolabs, Ipswich, MA). Then, purified PCR products were cloned into the Zero Blunt TOPO PCR Cloning vector (ThermoFisher, Carlsbad, CA), and transformed into DH5α competent cells (ThermoFisher, Carlsbad, CA). Colonies carrying the alleles from each event were Sanger sequenced.

vii. Assessment of self-compatibility

One month old *in-vitro* plants were planted in one gallon plastic pots with a peat and perlite growth medium mixture (Bacto professional planting mix) and placed into a greenhouse with a light intensity of 250 µmoles m−2 s−1, 16/8-h light/dark photoperiod and a temperature of 25 °C. Plants were fertilized with Peters Professional® 20: 20: 20 fertilizer (The Scotts Co., Marysville, OH) at a rate of 500 mg/l twice a week. Around 50 flowers per plants were hand self-pollinated to test for SC. Pollen staining with acetocarmine-glycerol (Ordoñez 2014a) and cross-pollination were also done to test male and female viability, respectively. T_0 fruits were harvested three-four weeks after self-pollination and kept at room temperature for two weeks. Extracted T_1 seeds were sterilized and subjected to overnight treatment with 1500 ppm of gibberellic acid then allowed to germinate. T_1 seedlings were transferred to greenhouse and self-

pollinated as described above. Additionally, chloroplast counting of guard cells was performed according to Ordoñez (2014a) to discard possible chromosome doubling in each selected *S-RNase* KO line.

viii. S-RNase expression analysis

Twenty-five flowers from wild-type (WT) and DRH-195/310-derived T_0 KO lines (DRH-195.158 and DRH-310.21) were self-pollinated at anthesis. Pollinated pistils were excised 24-hour post pollination (hpp) and preserved in -80 ºC until use. Total RNA was isolated using the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany) and DNase treated using the TURBO DNA-free kit (ThermoFisher, Carlsbad, CA) following manufacturer's instructions. RNA was quantified using a NanoDrop spectrophotometer (ThermoFisher Scientific, Grand Island, NY) and a reverse-transcription polymerase chain reaction (RT-PCR) was carried-out with 1 µg of total RNA using the Super-Script One-Step RT-PCR kit (Thermo Fisher Scientific, Carlsbad, CA). Primers designed to amplify the *S-RNase* ORF and the elongation-factor one alpha (*EF1α*) housekeeping internal control were used for the RT-PCR reaction (Table S3.2).

IV. Results

i. Identification of the S-RNase gene in potato

For this study, we used available genomic and gene expression data from the sequenced doubled monoploid DM and the SC heterozygous breeding line RH (The Potato Genome Sequencing Consortium 2011). To identify the *S-RNase* gene, the DM genome sequence was selected as the reference genome and utilized in sequence similarity searches using seven Solanaceae *S-RNase* genes (Table 3.1). Candidate genomic regions encoding the *S-RNase* gene were identified and

located within 3,948,850 - 3,949,581 bp of the unanchored scaffold PGSC0003DMB000000091 and the annotated DM *S-RNase* allele, PGSC0003DMG400026738, which encodes a 738 bp open reading frame and a 216 amino acids (aa) predicted protein composed of five conserved and two hypervariable regions, characteristic of SI *S-RNases* (Figure 3.1A; Ioerger et al., 1991). The detected DM and RH *S-RNase* alleles resembles class III *S-RNases* (Figure 3.1B) and is comprised of two exons and one small intron, which is located at position five of the 11 recognized intron positions for this gene family (Igic and Kohn 2001; Ramanauskas and Igić 2017).

We found that the DM *S-RNase* (referred hereafter as *S. tub*_S_{p5} from *S. tuberosum S-RNase* allele five of *S. tuberosum* Group Phureja) is highly expressed in mature flowers (245.5 Fragments Per Kilobase of exon model per Million mapped reads – FPKM) compared with no expression in leaves or tubers in DM. High *S-RNase* expression levels were also detected in carpels (4342.7 FKPM) consistent with its role in preventing SC (Kao and Tsukamoto 2004). More limited gene expression data is available for RH (The Potato Genome Sequencing Consortium 2011) and consistent with the expression of *S-RNase* in potato, it is expressed in flowers (167.04 FPKM) and not tubers or leaves. Together with the functional annotation, these results suggest we have identified the *S-RNase* gene involved with SI.

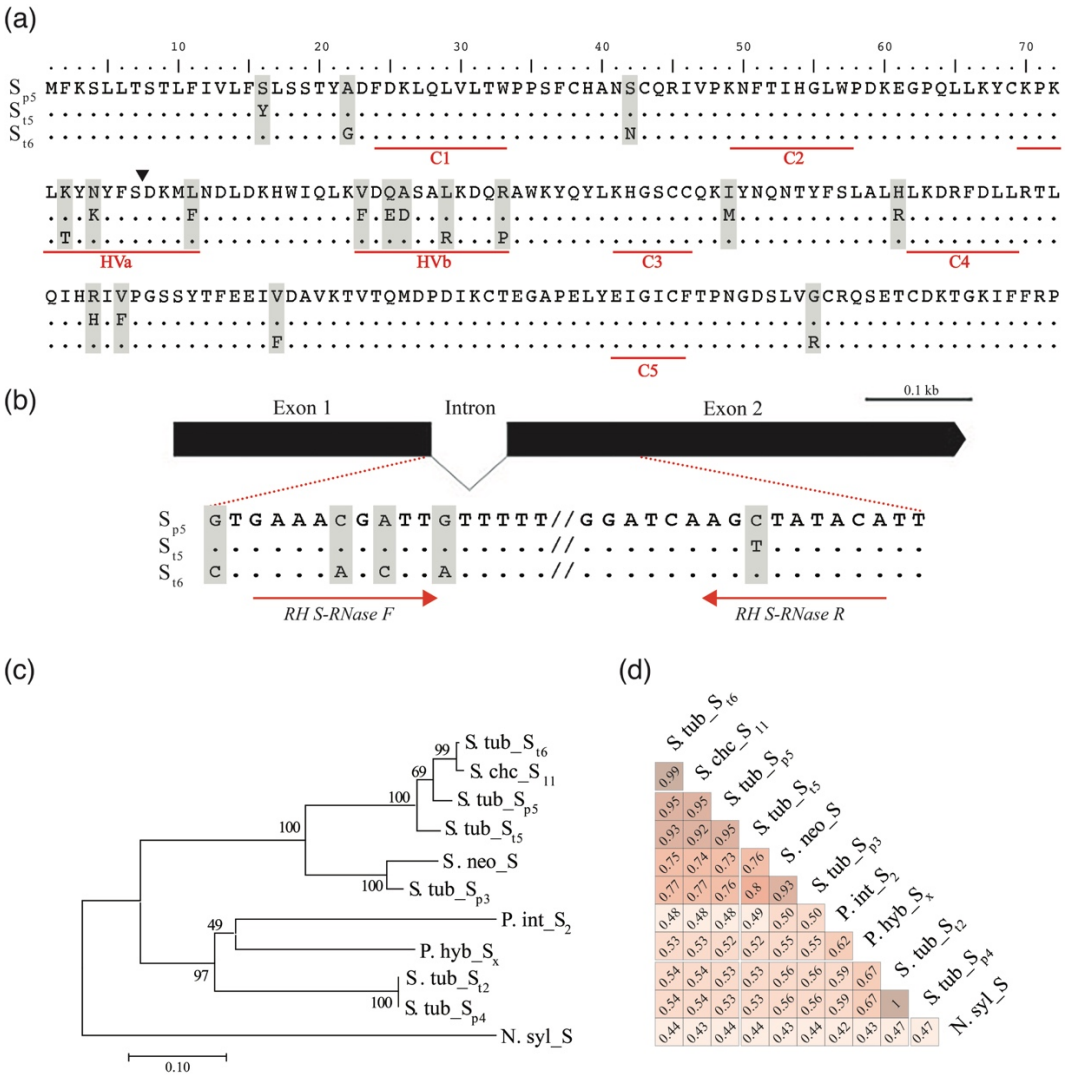

Figure 3.1. *S-RNase gene structure and allelic variants in diploid potato and related species.* (a) S-RNase predicted amino-acid sequence alignment of the DM (Sp5) and RH (S_{t5} and S_{t6}) alleles. Underlined regions in red represent the typical five conserved regions (C1 to C5) and two hypervariable regions (HVa and HVb) of the *S-RNase* gene family. Exon/intron boundary is indicated with a filled triangle within the HVa region. (b) *S-RNase* gene structure. The *S-RNase* open reading frame is composed of two exons separated by one small intron. Zoomed-in regions are shown within dotted lines indicating the intronic and exonic regions used for RH-specific primer design within the reported *S-RNase* alleles. (c) Phylogenetic tree constructed using the

Neighbor Joining method based in the proportion of S-RNase amino acid differences. S-RNase from *N. sylvestris* was used as out-group. Numbers above each branch represent bootstrapping percentages from1000 replications. (d) Pairwise amino acid similarity of S-RNase in *Solanum* species and detected *S-RNase* alleles. *S-RNase* alleles are represented first by a species name abbreviation followed by S and the allele number for Tuberosum (t) or Phureja (p) group in *S. tuberosum*. For other species, a similar pattern is used, and allele numbers or letters (i.e. Sx for *P. hybrida*) are added if reported. S. tub: *S. tuberosum*, S. chc: *S. chacoense*, S. neo: *S. neorickii*, P. int: *P. integrifolia*, P. hyb: *P. hybrida*, N. syl: N. sylvestris.

ii. Identification of allelic variants of S-RNase in diploid potato lines

The DRH-195 and DRH-310 F1 diploid self-incompatible lines derived from a cross between DM and RH were used to identify *S-RNase* allelic variants. Using whole genome resequencing data for these two lines, the DM and RH *S-RNase* alleles were identified in DRH-195 and DRH-310 and validated using Sanger sequencing. As previously observed for DM, the *S. tub*_S_{p5} predicted 216 amino acid sequence was detected in both lines, beside one of the two RH *S-RNase* alleles in each line, hereafter referred as *S. tub*_S_{t5} and $_{S.\ tub}$_S_{t6} (*S. tuberosum S-RNase* alleles five and six of *S. tuberosum* Group Tuberosum group respectively, Figure 3.1A).

A phylogenetic tree was constructed using available S-RNase amino acid sequences from Solanum species and the allelic variants identified in this study (Figure 3.1C). Two main clades with high confidence bootstrap values were observed. The *S. tub*_S_{p5}, *S. tub*_S_{t5} and *S. tub*_S_{t6} alleles and the *S. chacoense* (*S. chc*_S_{11}) allele exhibited highest percentage of similarity relative to the other species (Table S3.3, Figure 3.1D), whereas the *Petunia* (*P. int*_S_2 and *P. hyb*_S_x) and *S. tuberosum* (*S. tub*_S_{t2}) *S-RNases* clustered in two separate subgroups. Two *S. tuberosum*

Group Phureja alleles, *S. tub*_S_{p3} and *S. tub*_S_{p4}, were located in separate clades (Figure 3.1C). Overall, the amino acid sequence identity between S-RNase alleles ranged from 42 to 100%, showing 92% similarity between the DM and RH *S-RNase* alleles and 93% between the two RH alleles (Figure 3.1D). Notably, the *S. tub*_S_{p4} allele had 100% similarity with a previously reported *S-RNase* from *S. tub*_S_{t2} (NCBI accession: Q01796), presumably representing the same allele. Moreover, *S. tub*_S_{p5}, *S. tub*_S_{t5}, and *S. tub*_S_{t6} alleles had similar amino acid identity (53-54%) to the reported *S. tub*_S_{p3} allele when compared with *S. tub*_S_{p4} (Ye et al. 2018).

iii. S-RNase is located within a pericentromeric region of chromosome I in potato

Using the segregating DRH population, linkage mapping indicated that the *S-RNase S. tub*_S_{t6} allele mapped to the pericentromeric region of chromosome I spanning a region between 13.7 and 17.8 cM (solcap_snp_c2_27882 and solcap_snp_c1_16425 markers, respectively), corresponding to 6.1 Mb and 18.9 Mb of chromosome I in the physical map (Figure 3.2). These results are consistent with the region corresponding to the map location of the S-locus in potato (Gebhardt et al. 1991).

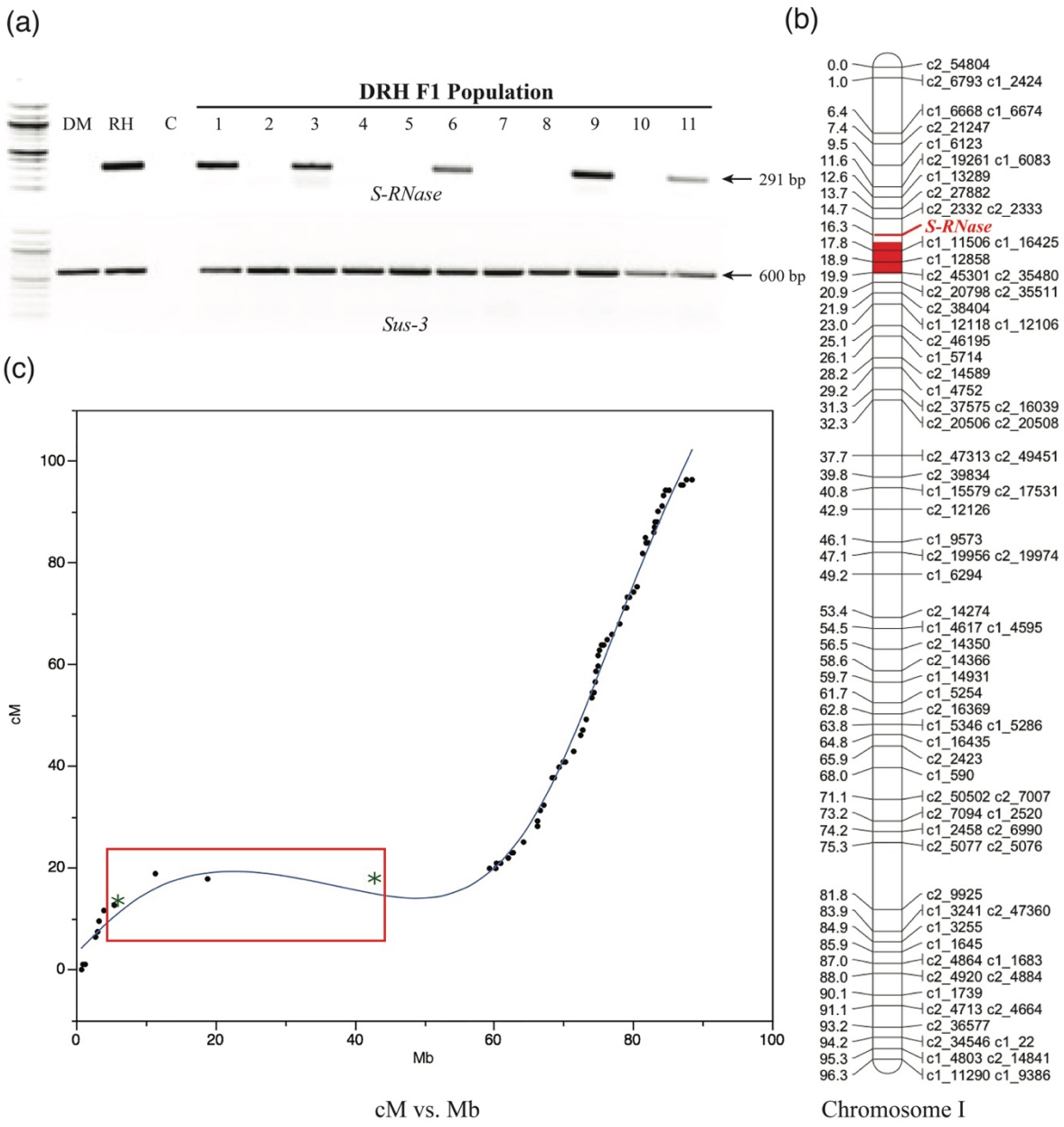

Figure 3.2. *S-RNase gene mapping in diploid potato.* (a) RH *S-RNase* (*S. tub*_S_{t6}) allelic screening on the DRH F1 population using the *S-RNase* and housekeeping *Sucrose synthase 3* gene (*Sus-3*) primers. DM and RH parental lines are shown in the first two lanes followed by the negative control [C], and the RH-*S-RNase* segregation pattern of 11 F1-derived lines. (b) The *S-RNase* gene mapped to 16.3 cM on the short arm near the centromeric region of chromosome I (red). (c) Marey map of physical (Mb) versus genetic (cM) distances from chromosome I

showing the *S-RNase* gene within a low-recombination region (red box). Asterisks within the red box represent SNPs spanning the region between to 6.1Mb and 18.9 Mb in the potato physical map (solcap_snp_c2_27882 and solcap_snp_c1_16425 markers, respectively).

iv. Targeted mutagenesis of S-RNase in diploid potatoes using CRISPR/Cas9 results in self compatibility

A dual gRNA strategy (sgRNA 1 and sgRNA 2) was used to efficiently generate *S-RNase* KOs and disrupt the *S-RNase* function in DRH-195 and DRH-310 (Figure 3.3A). Multiple T_0 plants were recovered for each line due to a 98% regeneration and transformation efficiency for DRH-195 and 93% for DRH-310 (Table 3.2). Based upon PCR analysis using primers to the *S-RNase* and gel detection of insertion/deletion polymorphisms, biallelic *S-RNase* mutations were recovered in both the DRH-195 and DRH-310-derived T_0 lines (Figure 3.3B, 3.3C). Specifically, seven *S-RNase* KOs exhibiting polymorphic deletions with up to 580 bp were detected for DRH-195-derived T_0 lines. In contrast, for DRH-310, only three *S-RNase* KOs were detected with up to 524 bp monomorphic deletions. To further characterize the CRISPR-targeted regions in selected T_0 lines, both T_0 KOs and WT-like *S-RNase* amplicons were sequenced (Figure 3.3D, 3.3E). A distinct nucleotide deletion was detected in each KO line, ranging from small bi-allelic deletions (1 bp) to large homozygous deletions (527 bp) in both *S-RNase* alleles of each DRH-derived T_0 lines. Insertions (1 to 18 bp) and inversions (486 bp) were also observed in a bi-allelic configuration. Similarly, besides the described mutation types, chimeric mutations were detected in T_0 lines, which has been reported in other species subjected to CRISPR-mediated mutagenesis, potentially due to late embryogenesis editing (Gomez et al. 2018).

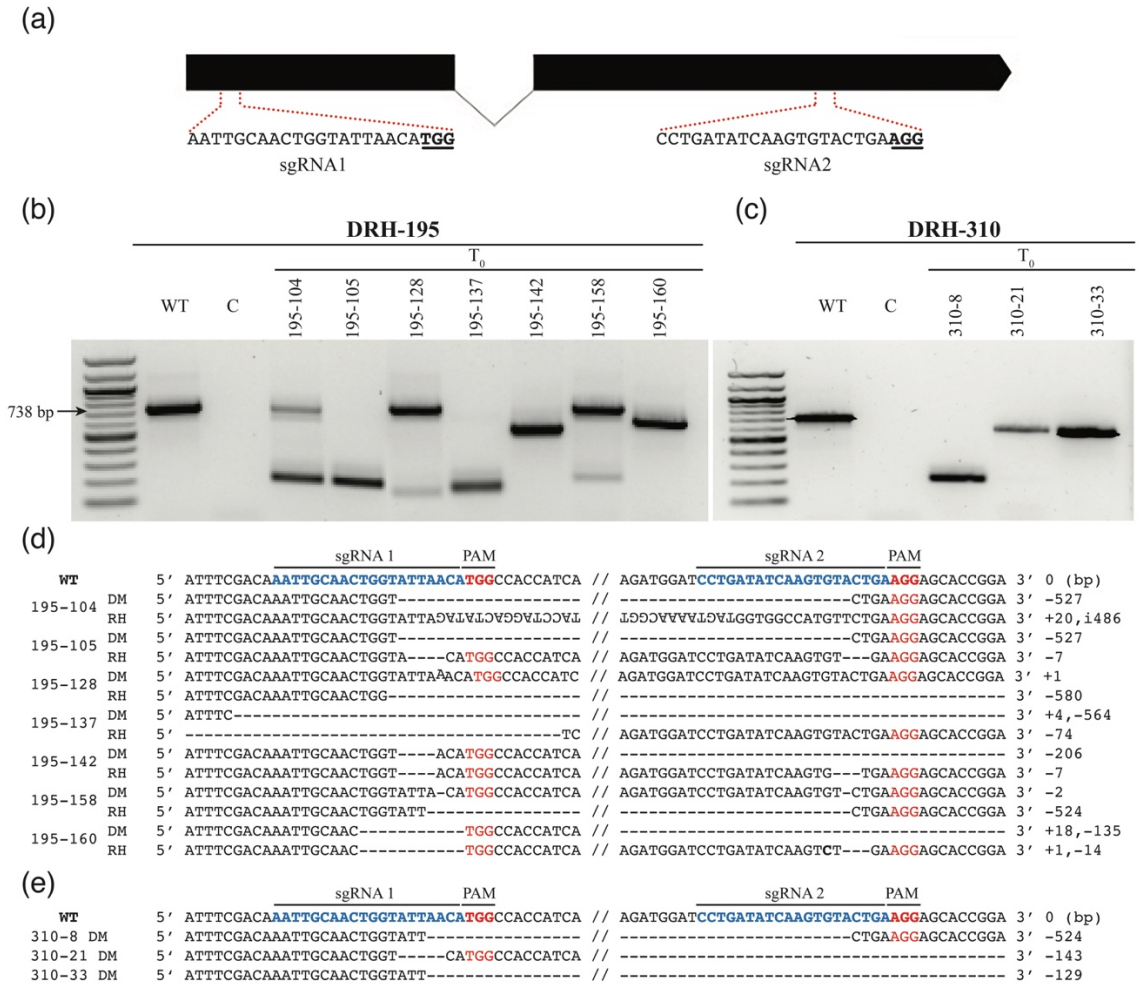

Figure 3.3. *CRISPR/Cas9 mediated mutagenesis in two SI diploid potato lines*. (a) Single-guide RNAs (sgRNAs) designed to target the *S-RNase* exon1 (sgRNA 1) and exon 2 (sgRNA 2). Zoomed-in regions within the dotted lines show sgRNA and PAM sequences. (b) Seven and (c) three bi-allelic *S-RNase* knock-outs (KOs) DRH-195 and DRH-310 T_0 lines, respectively, detected by insertion/deletion polymorphisms compared to wild type (WT) and negative controls [C]. All DRH-195 KOs exhibit polymorphic deletions and DRH-310 monomorphic deletions (lines 195-105,195-137,195-142 and 195-160 presented a faint mutated or WT-like bands that is not observed in the figure). (d) Different mutation types detected by amplicon sequencing in selected KOs for DRH-195, and (e) DRH-310 T_0 lines respectively. Wild-type (WT) *S-RNase*

exhibiting selected sgRNAs (sgRNA1 and sgRNA2) and PAM sequences are shown at the top of each alignment. Different types of mutations including deletions (-), insertions (+), and inversions (i) detected in DM (DM) and RH (RH) *S-RNase* alleles of each T_0 KO DRH-derived lines shown as '195-' or '310-'.

To test whether the selected T_0 KOs lines underwent spontaneous chromosome doubling, chloroplasts were counted in stomatal guard cells. One out of ten *S-RNase* KO lines (DRH-195.104) revealed chromosome doubling which has also been observed in a related *S-RNase* KO approach (Ye et al., 2018). This phenomenon, known as endopolyploidization, is frequently observed in potatoes subjected to regeneration processes, in which structural cell and chromosome rearrangements at mitosis results in increased chromosome numbers (Karp et al. 1984; Owen et al. 1988). The tetraploid KO line was not considered for further analysis.

Table 3.2. *DRH-195 and DRH-310-derived T_0 and indel-based selected S-RNase KO lines with bi-allelic mutations.*

Line	Num. Explants	T_0 lines	Transformation Efficiency (%)*	Mutant deletion polymorphism type	
				Single	***Double***
DRH-195	186	162	98		7
DRH-310	276	78	93	3	

*Calculated as the percentage of T_0 lines with Cas9 integration

To confirm the *S-RNase* mutant phenotype, T_0 KO lines were self-pollinated in two separate replications under greenhouse conditions. In both replications, all T_0 KO lines set fruit (Figure 3.4A, Tables S3.4 and S3.5). In one of the replications, both wild-type non-transformed lines (DRH-195 and DRH-310) also exhibited a limited number of specific self-pollination events with fruit set that either had complete development (DRH-195, Table S3.4) or arrest of fruit set

two weeks after self-pollination (DRH-310, Figure S3.5A). However, after one-week, new self-pollinations of wild-type lines did not set fruit (Figure 3.4B) suggesting plasticity of self-compatibility, a phenomenon observed previously in *Solanum* (Saba-El-Leil et al. 1994; Mena-Ali and Stephenson 2007).

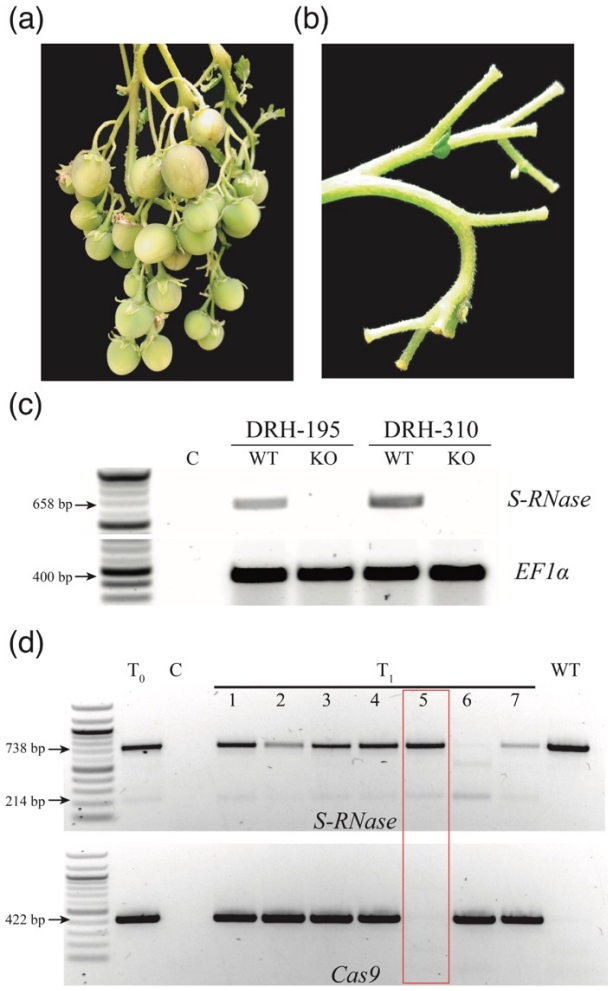

Figure 3.4. *S-RNase expression and knock-out phenotype in SI diploid potato lines.* (a) Fruits obtained after five weeks of self-pollination in an *S-RNase* DRH-195-derived T_0 mutant line. (b) Dropped flowers after self-pollination in the wild type DRH-195. (c) Semi-quantitative reverse transcription PCR (RT-PCR) in DRH-195 and DRH-310 self-pollinated wild-type (WT) and knock-out (KO) T_0 lines (DRH-195.158 and DRH-310-21, respectively). RNA was isolated from

pistils 24 hours after self-pollination revealing *S-RNase* expression in WT but not in KO lines as compared with the housekeeping gene control (EF1α). (d) T_1 plants derived from the DRH-195.158 T_0 line were screened with the *S-RNase* primers. WT-like bands with 1 bp deletion on each target site (causing frameshift leading to a premature stop codon, Figure S3.1) are observed in T_0 and T_1 lines. Cas9 gene did not transmit to T_1 line 5 (lane 5). A previously undetected band observed in lane 6 is potentially the result of transgenerational CRISPR/Cas9 activity. T_0: DRH-195.158, C: Negative control. The red box is showing a T_1 line segregating out Cas9 while maintaining the *S-RNase* KO.

This plasticity however, represents an unreliable source of self-compatibility as was evident in the ratio of fruit set per pollination observed, with self-pollination success in *S-RNase* KO lines being an order of magnitude higher than the WT in both DRH-195 and DRH-310 KO lines (Tables S3.4 and S3.5). To further investigate if this result was associated with the suppression of *S-RNase* expression, a semi-quantitative RT-PCR was performed. As shown in Figure 3.4C, *S-RNase* transcripts were detected in both wild-type lines 24 hpp yet no expression was detected in T_0 KOs, confirming *S-RNase* expression in WT but not mutant lines.

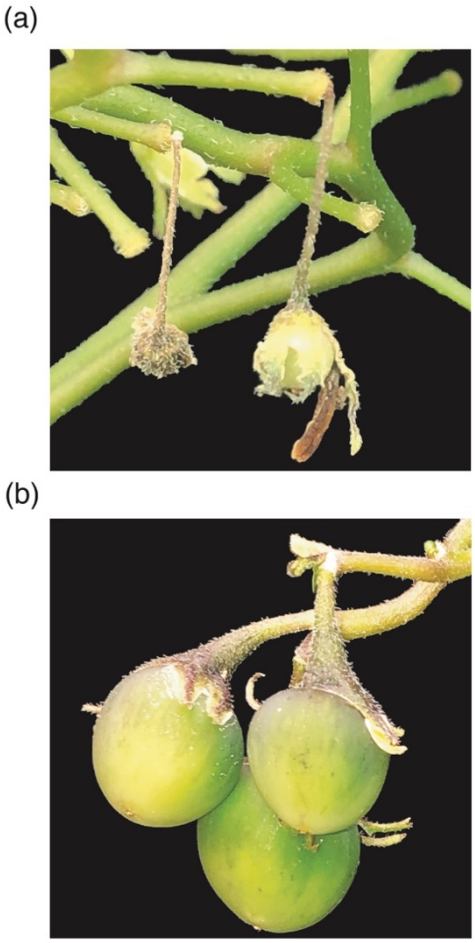

Figure 3.5. *Fruit formation in WT and T_1 S-RNase KO diploid potatoes lines.* (a). Fruit setting arrest two weeks after self-pollination in the wild type DRH-310. (b). Fruits obtained after four weeks of self-pollination in an *S-RNase* DRH-195-derived T_1 mutant line.

Viable T_1 seeds were obtained for each *S-RNase* T_0 KO line. Self-compatibility was confirmed in T_1 lines after self-pollination so far, demonstrating the inheritance and stability of the *S-RNase* KO phenotype (Figure 3.5B). Cas9 inheritance in the T_1 lines exhibited a segregation ratio associated with a hemizygous multi-copy integration of Cas9 (4 out of 135), in addition to the segregation of the mutated *S-RNase* alleles (Figure 3.4D). Likewise, because of the activity of

integrated Cas9, a potential transgenerational deletion was observed in a T_1 line (DRH-195.158.6). These results demonstrate the advantage of using CRISPR/Cas9 to generate Cas9-free edited plants and the potential to transmit stable gene mutations through different generations.

V. Discussion

Self-incompatibility has been a limiting factor for inbred/F1 hybrid cultivar development in diploid potatoes because efforts involving crossing with wild SC relatives result in many undesirable traits segregating in the progeny. To redirect potato breeding toward an efficient inbred/F1 hybrid generation strategy, we exploited the *S-RNase*-based SI system in diploid potatoes and generated KO lines using CRISPR-based genome editing to achieve self-compatibility.

Amino acid sequence variation within S-RNase was observed among *S. tub*_S_{p5}, *S. tub*_S_{t5}, and *S. tub*_S_{t6} alleles (Figure 3.1A). Nearly half of these variants were within the hypervariable domains (HVa and HVb) and not in conserved domains (C1-5) consistent with data that show the *S-RNase* variable regions are the determinants for allele specificity in different *Solanum* species (Matton et al. 1997, 1999; Brisolara-Corrêa et al. 2015). Specifically, four amino acids within these variable regions (T74, N76, Y77, and R101) have been reported as the sole factors for allele conversion of the pollen rejection phenotype in *S. chacoense* (Matton et al. 1997). Three of these amino acid changes were present within the *S. tub*_S_{p5}, *S. tub*_S_{t5}, and *S. tub*_S_{t6} alleles indicating that these variations could be sufficient to confer allele specificity while preserving their catalytic activity which is associated with two of the five conserved domains (Kao and Tsukamoto, 2004).

Ioerger et al. (1990) observed that S-RNase inter-specific similarities were higher than intra-specific similarities in Solanaceae, concluding that S-RNase divergence pre-dates speciation in this clade. The results observed in our study further confirm this previous observation. A high degree of inter-specific S-RNase amino acid sequence similarity was observed in *Solanum S-RNases* (*S. tuberosum* and *S. chacoense*). Conversely, a clear intra-specific separation within the *S. tuberosum S-RNase* alleles (Figure 3.1C) was also observed, consistent with the hypothesis of a single ancestral origin of *S-RNase* and conservation of specific polymorphisms throughout evolution governing allelic diversity (Ioerger et al. 1991; Dzidzienyo et al. 2016).

The *S-RNase* gene mapped to chromosome I within a region of low recombination consistent with the hypothesis to promote outbreeding due to a reduction in recombination events between the *S-RNase* and *SLF* genes (Kubo et al. 2015; Fujii et al. 2016). This chromosome position has also been reported in other Solanaceae members. For instance, *S-RNase* is located on chromosome I in *S. lycopersicum* and *S. peruvianum* within highly complex and repetitive genomic regions (Kubo et al. 2015; Fujii et al. 2016). Furthermore, Kubo et al. (2015) mapped an SLF, the other component of the S-locus, also to chromosome I in potato genome within a repeat-rich sub-centromeric region, suggesting that the *S-RNase* location was at the same position since these genes are reported to be closely linked (Sijacic et al., 2004).

All edited T_0 lines had a frameshift in the coding region close to sgRNA 1, leading to a premature stop codon. The resulting truncated sequence prevented the amplification of the *S-RNase* gene by removing the primer annealing site at the 3' end (Figure S3.1). Similarly, the consistent mutations generated by the two sgRNAs allowed detection of *S-RNase* size polymorphisms. It should be noted that this strategy was selected for the potential to use PCR for a quick and facile screen for large deletions in T_0 lines. However, undetected insertions/deletions

or inversions could be present in T_0 lines. For instance, sequencing data revealed a single bp insertion and deletion in the DM *S-RNase* allele of DRH-195.128 and DRH-195.158, respectively, showing a similar amplicon size in agarose gels as WT (Figure 3.3B, 3.3D). These observations indicate that a large number of allelic KOs can be generated given the high transformation efficiency observed in both diploid lines.

The DRH-195.158 T_0 KO line, which exhibited a single bp deletion at each sgRNA targeting site in the RH *S. tub*_S_{p5} allele (Figure 3.3D), showed a new *S-RNase* deletion in a T_1-derived line (Figure 3.4D). Given the Cas9 mismatching tolerance, this allele possibly underwent a new mutagenesis event, displaying a different mutation pattern in the T_1 generation. In different plant species, it has been found that editing occurs at a higher frequency across generations, therefore new mutations segregate from WT alleles in heterozygous T_0 as a result of constitutive expression of Cas9 (Feng et al. 2014; Xu et al. 2015; Wang et al. 2018).

Two independent self-pollination assays were conducted in DRH-195 and DRH-310 wild-type lines in 2015 and 2018 with a minimal of 50 flowers, demonstrating their SI nature. However, a third biological replicate in 2018 resulted in fruit formation suggesting plasticity in the strength of SI. Environmental effects along with plant phenology have been associated with unstable SI in angiosperms. For instance, temperature fluctuations, photoperiod, glucose starvation and humidity significantly reduced SI in *S. peruvianum* after selfing (Webb and Williams 1988). This process, known as pseudo-self-incompatibility, has also been reported in grasses in which artificial self-pollination techniques can contribute to SI breakdown (Do Canto et al. 2016). Similarly, sporadic fruit set has been observed across Solanaceae species such as *Witheringia solanacea*, *S. carolinense*, *S. peruvianum* and *N. alata* in which floral age, flowering stage and delayed floral abscission has been associated with fruit set in SI populations (Stone et

al. 2006; Mena-Ali and Stephenson 2007; Miller and Kostyun 2011; Liao et al. 2016). This phenomenon has also been observed in species under sporophytic SI, in which floral age reduces the expression of the *S*-locus associated genes in *Brassica oleracea* resulting in SI breakdown (Hadj-Arab et al. 2010). In natural populations of *Campanula rapunculoides*, strong SI has been observed in young flowers. However, self-fruit formation is also evident in old flowers as a consequence of pollen scarcity and low fruit production from prior inflorescences (Stephenson et al. 2000). Therefore, environmental conditions favoring SI breakdown (plant age, plant health, and greenhouse conditions) could lead to fruit set in one of the wild-type biological replicates in this study.

Unlike pseudo-self-incompatibility, the *S-RNase* KO proved to be both stable and consistent across different replications and generations, presenting a higher ratio of fruit set per pollination when compared with self-fruit WT lines (Tables S3.4 and S3.5). Although the SC phenotype appears to be line dependent, distinctive *S-RNase* KO lines exhibited either high fruit set or seed formation. These results also indicated that genes other than *S-RNase* could be contributing to the strength of the SI response in both WT and *S-RNase* KOs. In fact, besides the *S-RNase* gene, other SI modifier loci can modulate the pollen rejection response in several *Solanum* species (Goldraij et al., 2006; McClure et al., 1999; O'Brien et al., 2002). For instance, *S-RNase*-independent stylar factors such as eEF1A or High Top-Band (HT-B) proteins, can directly or indirectly interact with *S-RNase* contributing to the SI response (Goldraij et al. 2006; Soulard et al. 2014). Similarly, unintended somaclonal variation and chromosomal rearrangements associated with the potato regeneration processes and Cas9 activity respectively, could also contribute to variations in the observed ratio of fruit set per pollination within the *S-RNase* KOs.

This hypothesis is further supported by Peterson et al. (2016) which identified several genomic regions associated with self-fertility in a DRH F1 population, located on chromosomes IV, IX, XI and XII. They also found that a specific SNP associated to the RH allele, fixed in selfed populations, is likely the primary factor for self-fertility in the DRH F1 progeny. Overall, this study demonstrates that *S-RNase* is the primary component for self-pollen rejection in DRH-195 and DRH-310. However, external evidence suggests that besides RH self-fertility mechanisms, *S-RNase*-independent stylar factors and environmental conditions could play a role in spontaneous self-compatibility observed in the WT lines in this study.

In this study, we generated self-compatible potato diploid lines by targeting the *S-RNase* gene using the CRISPR-cas9 system. We first computationally identified three new *S-RNase* alleles in SI diploid lines (a DM and two RH alleles, each inherited to DRH-195 and DRH-310, respectively) and mapped this gene to chromosome I within the peri-centromeric region consistent with the localization of the *S*-locus to a low recombination region on chromosome I. *S-RNase* KO lines were obtained using a dual sgRNA strategy in which premature stop codons were generated. After self-pollination, fruits were set in selected KO lines in T_0 and T_1 lines. Cas9-free KO lines were also identified in T_1 lines. Our results demonstrated the inheritance and stablity of the *S-RNase* KO phenotype, which can contribute to utilization of SC as a first step for the generation of commercial diploid cultivars.

APPENDIX

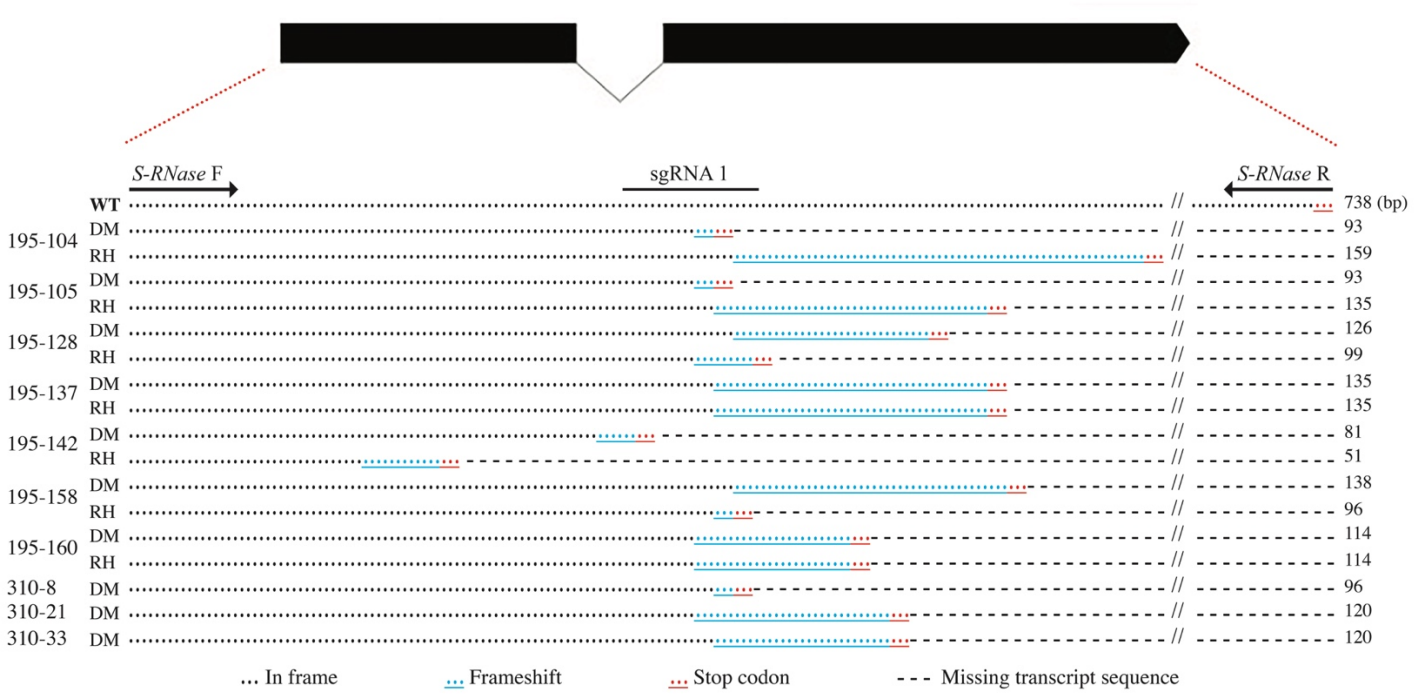

Figure S3.1. *S-RNase open reading-frames in T_0 knock-out (KO) lines derived in two SI diploid potato lines.* Sequences shown within the dotted lines contain 5' and 3' mRNA borders of this gene. Primer sequences designed to amplify the *S-RNase* ORF are shown as well as the sgRNA 1 located in exon 1. Individual dots represent nucleotides within the *S-RNase* gene. All T_0 KO lines had a frameshift near the sgRNA target region (blue dotted line) creating a premature stop codon (red dotted line). Black dotted line represents in frame *S-RNase* regions. nucleotides. Black dotted dash lines represent missing transcript sequence. *S-RNase* alleles from DM and RH are shown for each KO line.

Table S3.1. *Retrieved Illumina reads for genomic assembly of the DRH potato lines and assembly statistics.*

Tissue	**DRH-195**		**DRH-310**	
	SRA accession	**Pair-end Reads**	**SRA accession**	**Pair-end Reads**
Mature leaflet	SRR4018191	30,799,234	SRR4018197	24,891,913
Leaf	SRR4018147	30,423,428	SRR4018153	29,113,282
Tuber	SRR4018170	31,800,294	SRR4018174	31,283,414
Total:		93,022,956		85,288,609

Table S3.2. *Primers and sgRNAs used in this study.*

Name	Sequence	Description
S-RNase F	ATGTTTAAATCACTGCTTACATCAAC	*S-RNase* forward primer
S-RNase R	TCAGGGACGGAAAAATATTTTCCCTG	*S-RNase* reverse primer
S-RNaseRH F	GTTTTGTTTAATTTACTGAAAAGCTTA	RH-specific *S-RNase* forward primer
S-RNaseRH R	AAAGATTTCTTCAAATGTATAACT	RH-specific *S-RNase* reverse primer
EF1α F	GGTGGTTTTGAAGCTGGTATCTCT	Elongation factor one alpha forward primer
EF1α R	CCAGTAGGGCCAAAGGTCACA	Elongation factor one alpha reverse primer
sgRNA1	AATTGCAACTGGTATTAACA**TGG***	Single-guide RNA 1 targeting exon 1
sgRNA2	CCTGATATCAAGTGTACTGA**AGG**	Single-guide RNA 2 targeting exon 2

* In bold, protospacer adjacent motif (PAM) sequence

Table S3.3. *Top TBLASTN hits of reported S-RNase proteins in the DM genome assembly.*

Species	Scaffold	HSP* Number	Hit Score	E value	Length	Percent Id
Nicotiana sylvestris (CAA05306.1)	chr00	1	192	8.00E-55	244	41.8
Petunia integrifolia (AAG21384.1)	chr00	1	159	3.00E-43	222	44.14
Petunia x hybrida (AAA33729.1)	chr00	1	226	2.00E-66	251	48.21
Solanum chacoense (AAA50306.1)	chr00	1	415	3.00E-132	245	84.08
Solanum chilense (BAC00934.1)	chr00	1	313	1.00E-96	244	63.93
Solanum neorickii (BAC00940.1)	chr00	1	321	2.00E-99	244	64.75
Solanum tuberosum (Q01796.1)	chr00	1	204	1.00E-58	230	49.57

* High-scoring Segment Pair

Table S3.4. *Fruit set and seed count upon self-pollination in DRH-195 wild type and S-RNase-derived KO lines.*

Line	Self-pollinated flowers**	Total fruit set	Ratio of fruit set per pollination	Number of cluster and seeds per cluster
DRH-195 (WT)*	130	4	0.03	50
DRH-195.105	38	29	0.76	200 60 240
DRH-195.128	80	9	0.11	19 7 21 6 10 11
DRH-195.137	29	14	0.48	80 24 34 50
DRH-195.142	125	36	0.29	20 30 40 50 250 30
DRH-195.158	121	46	0.38	50 100 40 50 38 19 51 14 38 45

Table S3.4. (cont'd)

				39
				45
				226
				50
				50
				200
				29
				8
DRH-195.160	71	23	0.32	33
				11
				15
				200

*Wild type. ** Total of flowers in two replicates.

Table S3.5. *Fruit set and seed count upon self-pollination in DRH-310 wild type and S-RNase-derived KO lines.*

Line	Self-pollinated flowers**	Total fruit set	Ratio of fruit set per pollination	Number of cluster and seeds per cluster
DRH-310 (WT)*	123	0	0	0
DRH-310.8	38	5	0.13	2 3 15
DRH-310.21	52	16	0.31	50 50 20 28 50 63
DRH-310.33	53	15	0.28	10 19 20 20 18 10

* Wild type. ** Total of flowers in two replicates

LITERATURE CITED

LITERATURE CITED

Altschul S. F., Gish W., Miller W., Myers E. W., Lipman D. J. (1990). Basic Local Alignment Search Tool. J. Mol. Biol. 215: 403–410.

Bortesi L., Fischer R. (2015). The CRISPR/Cas9 system for plant genome editing and beyond. Biotechnol. Adv. 33: 41–52.

Brisolara-Corrêa L., Thompson C. E., Fernandes C. L., Freitas L. B. de. (2015). Diversification and distinctive structural features of *S-RNase* alleles in the genus Solanum. Mol. Genet. Genomics. 290: 987–1002.

Bush W. S., Moore J. H. (2012). Chapter 11: Genome-Wide Association Studies (F Lewitter and M Kann, Eds.). PLoS Computational Biology 8: e1002822.

Canto J. Do, Studer B., Lubberstedt T. (2016). Overcoming self-incompatibility in grasses: a pathway to hybrid breeding. Theor. Appl. Genet. 129: 1815–1829.

Covey P. A., Kondo K., Welch L., Frank E., Sianta S., Kumar A., Nuñez R., Lopez-Casado G., Knaap E. Van Der, Rose J. K. C., McClure B. A., Bedinger P. A. (2010). Multiple features that distinguish unilateral incongruity and self-incompatibility in the tomato clade. The Plant Journal 64: 367–378.

Devaux A., Kromann P., Ortiz O. (2014). Potatoes for sustainable global food security. Potato Research 57: 185–199.

Doudna J. A., Charpentier E. (2014). The new frontier of genome engineering with CRISPR-Cas9. Science 346: 1258096.

Dzidzienyo D. K., Bryan G. J., Wilde G., Robbins T. P. (2016). Allelic diversity of *S-RNase* alleles in diploid potato species. Theor. Appl. Genet. 129: 1985–2001.

FAOSTAT. (2019). FAOSTAT database collections. Food and Agriculture Organization of the United Nations.

Feng Z., Mao Y., Xu N., Zhang B., Wei P., Yang D.-L., Wang Z., Zhang Z., Zheng R., Yang L., Zeng L., Liu X., Zhu J.-K. (2014). Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in Arabidopsis. Proc. Natl. Acad. Sci. U S A. 111: 4632–4637.

Fujii S., Kubo K. I., Takayama S. (2016). Non-self- and self-recognition models in plant self-incompatibility. Nat. Plants 2: 16130.

Gebhardt C., Ritter E., Barone A., Debener T., Walkemeier B., Schachtschabel U., Kaufmann H., Thompson R. D., Bonierbale M. W., Ganal M. W., Tanksley S. D., Salamini F. (1991). RFLP maps of potato and their alignment with the homoeologous tomato genome. Theor. Appl. Genet. 83: 49–57.

Goldraij A., Kondo K., Lee C. B., Hancock C. N., Sivaguru M., Vazquez-Santana S., Kim S., Phillips T. E., Cruz-Garcia F., McClure B. (2006). Compartmentalization of *S-RNase* and *HT-B* degradation in self-incompatible *Nicotiana*. Nature 439: 805.

Gomez M. A., Lin Z. D., Moll T., Chauhan R. D., Hayden L., Renninger K., Beyene G., Taylor N. J., Carrington J. C., Staskawicz B. J., Bart R. S. (2018). Simultaneous CRISPR/Cas9-mediated editing of cassava eIF4E isoforms nCBP-1 and nCBP-2 reduces cassava brown streak disease symptom severity and incidence. Plant. Biotechnol. J.: 1–14.

Hadj-Arab H., Chèvre A.-M., Gaude T., Chable V. (2010). Variability of the self-incompatibility reaction in *Brassica oleracea* L. with S15 haplotype. Sex. Plant. Reprod. 23: 141–151.

Hall T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series 41: 95–98.

Hardigan M. A., Crisovan E., Hamiltion J. P., Kim J., Laimbeer P., Leisner C. P., Manrique-Carpintero N. C., Newton L., Pham G. M., Vaillancourt B., Yang X., Zeng Z., Douches D., Jiang J., Veilleux R. E., Buell C. R. (2016). Genome reduction uncovers a large dispensable genome and adaptive role for copy number variation in asexually propagated *Solanum tuberosum*. Plant Cell 28: 388–405.

Hosaka K., Hanneman R. E. (1998). Genetics of self-compatibility in a self-incompatible wild diploid potato species *Solanum chacoense*. 1. Detection of an *S* locus inhibitor (*Sli*) gene. Euphytica 99: 191–197.

Hua Z.-H., Fields A., Kao T. (2008). Biochemical Models for *S-RNase*-Based Self-Incompatibility. Molecular Plant 1: 575–585.

Igic B., Kohn J. R. (2001). Evolutionary relationships among self-incompatibility *RNases*. Proc. Natl. Acad. Sci. U S A. 98: 13167–13171.

Ioerger T. R., Clark A. G., Kao T. H. (1990). Polymorphism at the self-incompatibility locus in Solanaceae predates speciation. Proc. Natl. Acad. Sci U S A. 87: 9732 LP – 9735.

Ioerger T. R., Gohlke J. R., Xu B., Kao T. I. H. (1991). Sexual Plant lleproduction Primary structural features of the self-incompatibility protein in Solanaceae. Sex. Plant. Reprod 4: 81–87.

Jaganathan D., Ramasamy K., Sellamuthu G., Jayabalan S., Venkataraman G. (2018). CRISPR for crop improvement: An update review. Front. Plant. Sci. 9: 985.

Jansky S. H., Chung Y. S., Kittipadukal P. (2014). M6: A diploid potato inbred line for use in breeding and genetics research. J. Plant. Regist. 8: 195–199.

Jansky S. H., Charkowski A. O., Douches D. S., Gusmini G., Richael C., Bethke P. C., Spooner D. M., Novy R. G., Jong H. De, Jong W. S. De, Bamberg J. B., Thompson A. L., Bizimungu B., Holm D. G., Brown C. R., Haynes K. G., Sathuvalli V. R., Veilleux R. E., Miller J. C., Bradeen J. M., Jiang J. (2016). Reinventing potato as a diploid inbred line–based crop. Crop Science 56: 1412–1422.

Kao T., Tsukamoto T. (2004). The molecular and genetic bases of *S-RNase*-based self-incompatibility. Plant Cell 16: S72–S83.

Karp A., Risiott R., Jones M. G. K., Bright S. W. J. (1984). Chromosome doubling in monohaploid and dihaploid potatoes by regeneration from cultured leaf explants. Plant Cell Tissue Organ Cult. 3: 363–373.

Koncz C., Martini N., Szabados L., Hrouda M., Bachmair A., Schell J., 1994 Specialized vectors for gene tagging and expression studies. In: Gelvin SB, Schilperoort RA (Eds.), Plant Molecular Biology Manual, Springer Netherlands, Dordrecht, pp. 53–74.

Kowyama Y., Kunz C., Lewis I., Newbigin E., Clarke A. E., Anderson M. A. (1994). Self-compatibility in a *Lycopersicon peruvianum* variant (LA2157) is associated with a lack of style S-RNase activity. Theoretical and Applied Genetics 88: 859–864.

Kubo K., Entani T., Takara A., Wang N., Fields A. M., Hua Z., Toyoda M., Kawashima S., Ando T., Isogai A., Kao T., Takayama S. (2010). Collaborative non-self recognition system in *S-RNase* based self-incompatibility. Science 330: 796–799.

Kubo K., Paape T., Hatakeyama M., Entani T., Takara A., Kajihara K., Tsukahara M., Shimizu-Inatsugi R., Shimizu K. K., Takayama S. (2015). Gene duplication and genetic exchange drive the evolution of *S-RNase*-based self-incompatibility in *Petunia*. Nat. Plants 1: 14005.

Kumar S., Stecher G., Tamura K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33: 1870–1874.

Li W., Zarka K. A., Douches D. S., Coombs J. J., Pett W. L., Grafius E. J. (1999). Coexpression of potato PVYo coat protein and cryV-Bt genes in potato. J. Amer. Soc. Hort. Sci. 124: 218–223.

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R., Subgroup 1000 genome project data processing. (2009). The sequence alignment/map format and SAMtools. Bioinformatics 25: 2078–2079.

Li H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27: 2987–2993.

Li H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. eprint arXiv:1303.3997: arXiv:1303.3997.

Li W., Chetelat R. T. (2015). Unilateral incompatibility gene ui1.1 encodes an *S*-locus F-box protein expressed in pollen of *Solanum* species. Proc. Natl. Acad. Sci. U S A. 112: 4417–22.

Liao J., Dai J., Kang H., Liao K., Ma W., Wang J., Chen S. (2016). Plasticity in the self-incompatibility system of cultivated *Nicotiana alata*. Euphytica 208: 129–141.

Manrique-Carpintero N. C., Coombs J. J., Cui Y., Veilleux R. E., Buell C. R., Douches D. (2015). Genetic map and QTL analysis of agronomic traits in a diploid potato population using single nucleotide polymorphism markers. Crop Science 55: 2566–2579.

Matton D. P., Maes O., Laublin G., Xike Q., Bertrand C., Morse D., Cappadocia M. (1997). Hypervariable domains of self-incompatibility RNases mediate allele-specific pollen recognition. Plant Cell 9: 1757–1766.

Matton D. P., Luu D. T., Xike Q., Laublin G., O'Brien M., Maes O., Morse D., Cappadocia M. (1999). Production of an *S RNase* with dual specificity suggests a novel hypothesis for the generation of new *S* alleles. Plant Cell 11: 2087–2097.

McClure B., Mou B., Canevascini S., Bernatzky R. (1999). A small asparagine-rich protein required for *S*-allele-specific pollen rejection in *Nicotiana*. Proc. Natl. Acad. Sci. U S A. 96: 13548–13553.

Mena-Ali J. I., Stephenson A. G. (2007). Segregation analyses of partial self-incompatibility in self and cross progeny of *Solanum carolinense* reveal a leaky *S*-allele. Genetics 177: 501–510.

Mendoza H. A., Mihovilovich E. J., Saguma F. (1996). Identification of triplex (YYYy) Potato Virus Y (PVY) immune progenitors derived from *Solanum tuberosum* ssp. andigena. Am. Potato J. 73: 13–19.

Miller J. S., Kostyun J. L. (2011). Functional gametophytic self-incompatibility in a peripheral population of *Solanum peruvianum* (Solanaceae). Heredity 107: 30–39.

Murashige T., Skoog F. (1962). A revised medium for rapid growth and bio assays with tobacco tissue cultures. Physiol. Plant 15: 473–497.

O'Brien M., Kapfer C., Major G., Laurin M., Bertrand C., Kondo K., Kowyama Y., Matton D. P. (2002). Molecular analysis of the stylar-expressed *Solanum chacoense* small asparagine-rich protein family related to the HT modifier of gametophytic self-incompatibility in *Nicotiana*. Plant J 32: 985–996.

Ordoñez B. (2014a). Pollen Viability Assessment. International Potato Center (CIP). 8p, Lima, Peru.

Ordoñez B. (2014b). Assessment of ploidy by chloroplast count in stomatal guard cells. : 4p.

Owen H. R., Veilleux R. E., Levy D., Ochs D. L. (1988). Environmental, genotypic, and ploidy effects on endopolyploidization within a genotype of *Solanum phureja* and its derivatives. Genome 30: 506–510.

Park J., Bae S., Kim J. (2015). Cas-Designer: a web-based tool for choice of CRISPR-Cas9 target sites. Bioinformatics 31: 4014–4016.

Pellagatti A., Dolatshad H., Valletta S., Boultwood J. (2015). Application of CRISPR/Cas9 genome editing to the study and treatment of disease. Arch. Toxicol. 89: 1023–1034.

Peterson B. A., Holt S. H., Laimbeer F. P. E., Doulis A. G., Coombs J., Douches D. S., Hardigan M. A., Buell C. R., Veilleux R. E. (2016). Self-fertility in a cultivated diploid potato population examined with the Infinium 8303 Potato Single-Nucleotide Polymorphism Array. The Plant Genome 9: 1–13.

Porcher E., Lande R. (2005). Loss of gametophytic self-incompatibility with evolution of inbreeding depression. Evolution 59: 46–60.

Ramanauskas K., Igić B. (2017). The evolutionary history of plant T2/S-type ribonucleases. PeerJ 5: e3790.

Raymundo R., Asseng S., Robertson R., Petsakos A., Hoogenboom G., Quiroz R., Hareau G., Wolf J. (2017). Climate change impact on global potato production. Eur. J. Agron.: j.eja.2017.11.008.

Royo J., Kunz C., Kowyama Y., Anderson M., Clarke A. E., Newbigin E. (1994). Loss of a histidine residue at the active site of *S*-locus ribonuclease is associated with self-compatibility in *Lycopersicon peruvianum*. Proc Natl Acad Sci U S A 91: 6511–6514.

Saba-El-Leil M. K., Rivard S., Morse D., Cappadocia M. (1994). The S11 and S13 self incompatibility alleles in *Solanum chacoense* Bitt. are remarkably similar. Plant. Mol. Biol. 24: 571–583.

Sassa H., Kakui H., Miyamoto M., Suzuki Y., Hanada T., Ushijima K., Kusaba M., Hirano H., Koba T. (2007). *S* locus F-box brothers: multiple and pollen-specific F-box genes with *S* haplotype-specific polymorphisms in apple and Japanese pear. Genetics 175: 1869–1881.

Sievers F., Wilm A., Dineen D., Gibson T. J., Karplus K., Li W., Lopez R., McWilliam H., Remmert M., Söding J., Thompson J. D., Higgins D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7: 539.

Sijacic P., Wang X., Skirpan A. L., Wang Y., Dowd P. E., McCubbin A. G., Huang S., Kao T. H. (2004). Identification of the pollen determinant of *S-RNase*-mediated self-incompatibility. Nature 429: 302–305.

Soulard J., Boivin N., Morse D., Cappadocia M., 2014 *eEF1A* is an *S-RNase* binding factor in self-incompatible *Solanum chacoense*. PloS One 9: e90206.

Stephenson A. G., Good S. V, Vogler D. W. (2000). Interrelationships among inbreeding depression, plasticity in the self-incompatibility system, and the breeding system of *Campanula rapunculoides* L. (Campanulaceae). Ann. Bot. 85: 211–219.

Sternberg S. H., Redding S., Jinek M., Greene E. C., Doudna J. A. (2014). DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature 507: 62–7.

Stone J. L., Sasuclark M. A., Blomberg C. P. (2006). Variation in the self-incompatibility response within and among populations of the tropical shrub *Witheringia solanacea* (Solanaceae). Am. J. Bot. 93: 592–598.

Takayama S., Isogai A. (2005). Self-Incompatibility in Plants. Annu. Rev. Plant. Biol. 56: 467–489.

The Potato Genome Sequencing Consortium. (2011). Genome sequence and analysis of the tuber crop potato. Nature 475: 189–195.

Wang W., Pan Q., He F., Akhunova A., Chao S., Trick H., Akhunov E. (2018). Transgenerational CRISPR-Cas9 activity facilitates multiplex gene editing in allopolyploid wheat. The CRISPR Journal 1: 65–74.

Webb M. C., Williams E. G. (1988). Effects of temperature, light, nutrients and carbon dioxide on the strength of the self-incompatibility response in detached flowers of *Lycopersicon peruvianum*. Ann. Bot. 61: 395–404.

Xing H.-L., Dong L., Wang Z.-P., Zhang H.-Y., Han C.-Y., Liu B., Wang X.-C., Chen Q.-J. (2014). A CRISPR/Cas9 toolkit for multiplex genome editing in plants. BMC Plant. Biol. 14: 327.

Xu R.-F., Li H., Qin R.-Y., Li J., Qiu C.-H., Yang Y.-C., Ma H., Li L., Wei P.-C., Yang J.-B. (2015). Generation of inheritable and “transgene clean” targeted genome-modified rice in later generations using the CRISPR/Cas9 system. Sci. Rep. 5: 11491.

Ye M., Peng Z., Tang D., Yang Z., Li D., Xu Y., Zhang C., Huang S. (2018). Generation of self-compatible diploid potato by knockout of *S-RNase*. Nat. Plants 4: 651–654.

Chapter 4

VALIDATION OF THE RB-MEDIATED RESISTANCE SUPPRESSION BY THE IPI-O4 EFFECTOR FROM *P. infestans* IN POTATO

I. Abstract

Potato (*Solanum tuberosum* L.) is considered the third most important food crop after rice and wheat. However, pathogens such as late blight (*Phytophthora infestans* (Mont.) de Bary), can generate annual losses over 16% of the total global potato production. The most effective way to control this pathogen is through the introgression of resistant genes from landraces or wild-relative species. In particular, resistance to the late blight gene (*RB*) from *S. bulbocastanum*, with a Coiled-Coil (CC)-Nucleotide-binding Leucine-rich Repeat (NLR) gene architecture, confers a wide-spectrum resistance to late blight in transgenic potato lines. Under pathogen attack, the late blight IPI-O1 effector elicits the RB dimerization through the CC domain, triggering a disease resistance response, while IPI-O4 suppress the IPI-O1-mediated resistance to late blight. In this study, we used in vivo and in vitro approaches to confirm the molecular interactions leading to the activation or suppression of late blight resistance responses through CC/IPI-O protein interactions. We observed a hypersensitive response (HR) when *RB* was co-infiltrated with *IPI-O1* via *N. benthamiana* agroinfiltration. However, no HR was observed when *IPI-O1* or *IPI-O4* were co-infiltrated using a synthetic *RB* gene carrying a CC domain from *S. pinnatisectum* (*RB-SP*). In contrast, HR was elicited when *IPI-O1* and *IPI-O4* were infiltrated in a *RB-SP* transgenic line. Moreover, this study was unable to demonstrate CC-self association using a yeast-two-hybrid system. These results suggest that in addition to *RB*, other proteins or *RB* domains may be involved in resistance response in vivo. Further analysis should be conducted to confirm non self-CC dimerization and its relationship with observed HR mediated by unreported CC/IPI-O interactions.

II. Introduction

Cultivated potato (*Solanum tuberosum* L.) is the third most important food crop after rice and wheat (Devaux et al., 2014) and plays an essential role in human nutrition as a primary source of carbohydrates. This staple food is also known for its high content of phenolic compounds, opening a new window for the potato's commercialization as a functional food, due to its potential anti-oxidative activity (Friedman, 1997; Vinson et al., 2012). Although global production of potato totaled 388 million tons in 2017 (FAOSTAT, 2019), potatoes face major production losses due to biotic and abiotic stresses that will increase with global warming (Raymundo et al., 2017).

In particular, pathogens such as late blight (*Phytophthora infestans* (Mont.) de bary), represent one of the most devastating diseases for this crop, infecting vegetative tissues and killing the entire plant within 7 to 10 days after infection. Late blight accounts for annual losses of 16% of total global potato production (Haverkort et al., 2009). Additionally, under the current fluctuating weather conditions, it is expected that late blight incidences will increase worldwide, affecting mainly highlands in developing countries in upcoming decades (Sparks et al., 2014).

Late blight is a hemibiotrophic oomycete that infects vegetative plant tissues through multinucleate sporangium. The asexual stage uses mononucleate zoospores for germination, produced from cleaved sporangia, entering into the host via stomata, wounds or appressorium formation (Leesutthiphonchai et al., 2018). However, sexual spores, known as oospores, are also produced and requires two mating types; A1 and A2. This system increases late blight diversity through sexual recombination when both mating types are present in a population, favoring its adaptation to extreme environmental conditions (Smart and Fry, 2001).

Although pathogen infection in late blight of potato can be controlled using protectants or systemic fungicides, they are ineffective if environmental conditions favor pathogen dispersion (Nowicki et al., 2011) and also promote the emergence of fungicide-resistant strains (Pomerantz et al., 2014). The most effective way to control the incidence of late blight in potatoes is through the generation of resistant varieties (Ahn and Park, 2013). However, this process can take up to 15 years, which is time-consuming and ineffective against fast-evolving pathogens such as late blight (Lozano et al., 2012; The Potato Genome Sequencing Consortium, 2011).

Due to their close phylogenetic relationship, wild tuber bearing-species have been the primary source of resistance (R) genes for late blight in cultivated potato (Rodewald and Trognitz, 2013). Multiple R genes have been discovered in several potato wild relatives. In particular, the late blight resistance gene (*RB*) from *S. bulbocastanum* (*blb*), confers broad-spectrum disease resistance against this pathogen (Song et al., 2003). Although partial resistance in has been observed field, *RB*-transgenic potatoes present an increase in foliar resistance in growth chamber and greenhouse experiments when compared with wild type genotypes, without significant effects on tuber yield (Halterman et al., 2008).

The *RB* gene encodes a modular resistance protein comprising an N-terminal Coiled-Coil (CC) domain and a Nucleotide-binding Leucine-rich Repeat (NLR) domain core (Song et al., 2003). RB remains in a resting state prior to late blight infection. After pathogen attack, a conformational change elicited by the recognition of the IPI-O1 effector from the late blight pathogen, results in RB dimerization through the CC domain, leading to a hypersensitive response (HR) (Chen et al., 2012). However, when the IPI-O4 effector is present, *RB*-mediated resistance is suppressed by direct competition with IPI-O1, preventing the interaction of the CC domains or other resistance-related signaling components (Chen et al., 2012; Chen and

Halterman, 2017).

The CC domain from the *Solanum pinnatisectum* (*pnt*) homologous *RB* gene does not exhibit any interaction with either *IPI-O1* or *IPI-O4*, suggesting that this species may escape *IP-O4*-mediated resistance suppression (Chen et al., 2012). In an attempt to generate a durable late blight resistance response in *RB*-transgenic potato, we validated the IPI-O4 mediated suppression in vitro and in vivo. *Agrobacterium*-infiltration (agroinfiltration) assays in *Nicotiana benthamiana* using a synthetic *RB* gene (*RB-SP*) carrying a CC domain from *S. pinnatisectum* that led to a lack of HR when challenged with IPI-O effectors, confirming previously reported results. However, HR was observed when *IPI-O1* and *IPI-O4* were infiltrated in a *RB-SP* transgenic potato line, suggesting that additional mechanisms may be involved in a resistance response in vivo. Moreover, CC dimerization was not observed in yeast-two-hybrid (Y2H) assays. Further work is required to confirm protein expression in yeast and validate the role of the IPI-O effectors on the activation/suppression of late blight resistance in potatoes.

III. Material and Methods

i. Plant material

The diploid potato DRH S5 28-5 line (referred hereafter as DRH-S5) was used in this study. This line was generated from five rounds of self-pollination using a F1 founder from a cross between the doubled monoploid, *S. tuberosum* Group Phureja DM1-3 516 R44, and a heterozygous diploid breeding line, *S. tuberosum* Group Tuberosum RH89-039-16. Plants were maintained in vitro, propagated on Murashige and Skoog (MS), medium (MS basal salts plus vitamins, 3% sucrose, 0.7% plant agar, pH 5.8) (Murashige and Skoog, 1962) and cultured in a growth

chamber with 16-h-light/8-h-night photoperiod at 22°C and average light intensity of 200 µmoles $m^{-2}s^{-1}$.

ii. RB and IPI-O constructs and Agrobacterium-mediated transformation

A plasmid containing a synthetic *RB* gene with a CC domain from *pnt* and a NLR domain core from *blb* (referred here after as *RB-SP*) along with *Agrobacterium tumefaciens* cultures containing the late blight IPI-O1 and IPI-O4 effectors (Chen et al., 2012) were kindly donated by Dr. Halterman (University of Wisconsin), and used in this study. The pSPUD69 construct, carrying the *RB* gene from *S. bulbocastanum* under control of its native promoter and terminator (Kuhl et al., 2007), was used as a control. *RB-SP* and *RB* constructs with the *NPTII* gene as a selectable marker, were transferred into *A. tumefaciens* strain GV3101 pMP90 (Koncz et al., 1994) by electroporation.

Agrobacterium-mediated transformations were performed using leaf segments from four-week-old tissue culture plants of DRH-S5 as described by Li *et al.* (1999). Briefly, explants were pre-cultured on a step I media (MS salts, 3% sucrose, 5 g/l phytoagar, 1 mg/l thiamine-HCl, 0. 8 mg/l zeatin-riboside and 2 mg/l 2,4-D) for four days and inoculated with *A. tumefaciens* carrying the *RB* genes. After three days, explants were rinsed with sterile distilled water containing 250 mg/l cefotaxime and 200 mg/l carbenicillin and placed onto step II media (MS salts, 3% sucrose, 5 g/L phytoagar, 1 mg/L thiamine-HCl, 0.8 mg/L zeatin-riboside, 2 mg/L gibberellic acid, 50 mg/L kanamycin and 150 mg/L ticarcillin disodium and clavulanate potassium). Explants were transferred to fresh step II media every week. After approximately 30 days, transformation events were selected from step II media and transferred to root induction media containing MS medium supplemented with antibiotics for selection as described above.

iii. Leaf Agroinfiltration on N. benthamiana and transgenic potatoes

Four week-old *N. benthamiana* leaves were co-infiltrated with the *IPI-O* effectors and *RB-SP*/*RB* genes as described by Ma et al. (2012). Briefly, *A. tumefaciens* containing the *RB/RB-SP* and the *IPI-O1/IPI-O4* constructs were inoculated in LBmani media (10 g/L Bacto-tryptone, 5 g/L yeast extract, 2.5 g/L sodium chloride, 10 g/L mannitol) supplemented with 20 mM acetosyringone, 10 mM 2-(n-morpholino)ethanesulfonic acid (MES) and 50 mg/L kanamycin. Cells were harvested when cultures reached an optical density at 600 nm (OD600) of 0.8 and suspended in MMAi medium (5 g/L MS salts, 20 g/L sucrose, 200 µM acetosyringone and 10 mM MES). For HR induction, *Agrobacterium* suspensions containing either *RB-SP* or *RB* were syringe-infiltrated at the abaxial side of each leaf. After drying, suspensions containing *IPI-O* effectors were syringe-infiltrated overlapping the *RB-SP* and *RB* infiltrations respectively, for a total of four R/Avr gene interactions: *RB/IPI-O1, RB/IPI-O4, RB-SP/IPI-O1*, and *RB-SP/IPI-O4.* The *RB*/*IPI-O1* co-infiltration and an empty *A. tumefaciens* inoculum were used as a positive and negative control, respectively.

One month-old in-vitro DRH transgenic plants were planted in 10 cm square plastic pots with a peat and perlite growth medium mixture (Bacto professional planting mix) and placed into a grow chamber with a light intensity of 250 µmoles m−2 s−1, 16/8-h light/dark photoperiod and a temperature of 25 °C. Plants were fertilized with Peters Professional® 20: 20: 20 fertilizer (The Scotts Co., Marysville, OH) at a rate of 500 mg/l twice a week. Three weeks later, plants were agroinfiltrated as described above with the *IPI-O4* effector using the *IPI-O1* and an empty *A. tumefaciens* inoculum as a positive and negative control, respectively. For both assays, at least three independent leaves were agroinfiltrated. Plants were kept under growth chamber conditions and the HR response was assessed one week after infiltration.

iv. Molecular characterization of RB-SP and RB transgenic lines.

DNA was isolated from young leaves using the DNeasy Plant Mini Kit (Qiagen, Hilden, Germany). *NPTII* and *RB*-derived primers (Colton et al., 2006) were used to select positive transformation events (Table 4.1). PCR was carried out using GoTaq DNA polymerase (Promega, Fitchburg, WI) with the following thermocycler conditions: one cycle of initial denaturation for 4 min at 94 °C, followed by 34 cycles for 15 s at 94 °C, 45 s at 55 °C and 1 min at 72 °C and a final extension of 5 min at 72 °C. Amplicons were visualized on 1% (w/v) agarose gels.

Table 4.1. Primers used in this study

Name	Sequence	Description
RB-F	CACGAGTGCCCTTTTCTGAC	*RB* forward primer
RB-R	ACAATTGAATTTTTAGACTT	*RB* reverse primer
NPTII-F	ATGATTGAACAAGATGGATTGCAC	*NPTII* (kanamycin) forward primer
NPTII-R	CCAAGCTCTTCAGCAATATCACGG	*NPTII* (kanamycin) reverse primer
SUS3-F	CTGCAAGCTAAGCCTGATCTTATTAT	Sucrose synthase gene forward primer
SUS3-R	TTCGGAGTATGGAAAATAGAGATTCA	Sucrose synthase gene reverse primer

Expression analyses were performed on transformation events carrying either the *RB-SP* or *RB* transgenes. Total RNA was isolated from six-week-old plant leaves using the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany) and DNase treated using the TURBO DNA-free kit (ThermoFisher, Carlsbad, CA) following the manufacturer's instructions. RNA was quantified using a NanoDrop spectrophotometer (ThermoFisher Scientific, Grand Island, NY) and reverse-transcription polymerase chain reaction (RT-PCR) was carried-out using 1 μg of total RNA using the Super-Script One-Step RT-PCR kit (Thermo Fisher Scientific, Carlsbad, CA). The *RB* primers, designed to amplify a 218 bp of the NLS core domain, and the Sucrose Synthase 3 Gene

(*SUS3*) housekeeping internal control, were used for the RT-PCR reaction (Table 4.1). DRH-S5 (WT) were used as a negative control.

v. Protein interaction assays

The Matchmaker® Gold Y2H System (Takara Bio USA) was used for protein interaction screening. Full-length coding regions of the *blb* and *pnt* CC domains (495 bp each) were individually cloned into the multiple cloning site (MCS) of the pBridge and pGADT7 vectors, encoding a bait (Gal4-BD) and prey (Gal4-AD) fusion proteins, respectively. Likewise, the coding region of the *IPI-O4* effector (396 bp) was cloned into a second MCS within the pBridge vector. Both vectors were co-transformed using a lithium acetate procedure into the Y2HGold haploid yeast strain according to the Matchmaker user manual. Co-transformants were selected by culturing in double dropout minimum (SD) medium (DDO) lacking leucine (-Leu) and tryptophan (-Trp) at 30°C for 4-5 days. Selected yeast colonies were sub-cultured in DDO supplemented with X-α-Gal and Aureobasidin A (DDO/X/A). Healthy 2 mm blue colonies were patched onto the higher stringency quadruple dropout minimum medium (QDO) lacking adenine (-Ade), Histidine (-His), leucine (-Leu) and tryptophan (-Trp), supplemented with X-α-Gal and Aureobasidin A (QDO/X/A). For IPI-O4 expression, healthy 2 mm colonies were patched from DDO onto DDO/X/A and QDO/X/A media lacking methionine (-Met). The pGBKT7-53/pGAT7-T and pGBKT7-53/ pGBKT7-Lam vectors were co-transformed and used as positive and negative controls, respectively. Autoactivation tests were conducted using an empty pBridge vector co-transformed with pGADT7+CC.

IV. Results

i. HR-induced assessment in N. benthamiana

N. benthamiana leaves were co-infiltrated to test whether the synthetic *RB* gene interacts with *IPI-O4* in vivo. Co-infiltrated areas with *RB* and *IPI-O1* triggered HR as expected. However, no HR was observed when *RB-SP* and *IPI-O4* were co-infiltrated along with *RB-SP/IPI-O1, RB/IPI-O4* and empty control vector (Figure 4.1). These results are consistent with those of Chen et al. (2012) who reported a lack of interaction between the CC domain from *pnt* and the IPI-O effectors using in vitro approaches. These findings demonstrate that RB-SP is not able to recognize either IPI-O1 or IPI-O4 effectors via the CC domain when co-expressing in a heterologous system.

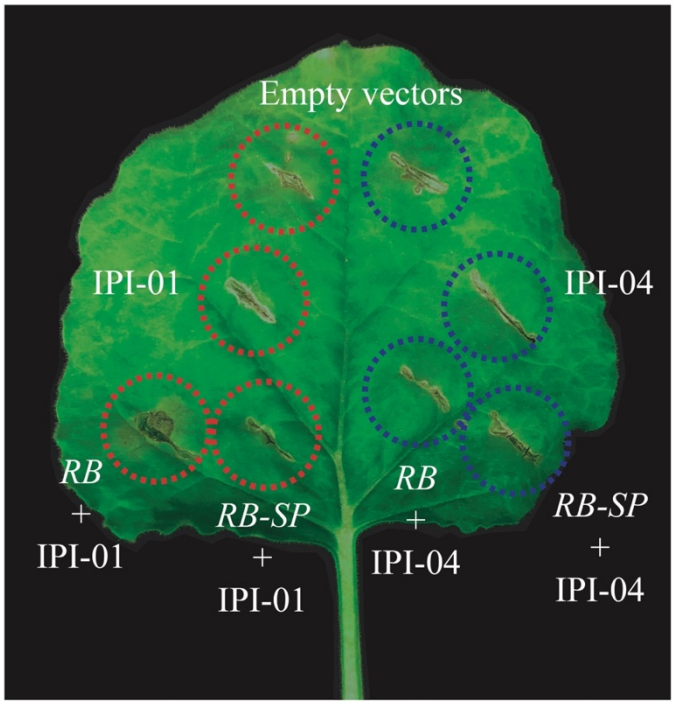

Figure 4.1. *Induction of hypersensitive response (HR) in Nicotiana benthamiana. Agrobacterium tumefaciens* expressing *RB, RB-PS*, *IPI-O1* and *IPI-O4* were co-infiltrated into *N. benthamiana* leaves. HR was observed when *RB* was co-infiltrated with *IPI-O1* (used as a positive control),

six days after infiltration. Empty *A. tumefaciens* was used as negative control. Red and blue dashed circles represent infiltrations/co-infiltration for *IPI-O1* and *IPI-04*, respectively.

ii. RB-SP/RB-transgenic potato lines

To further confirm the previous results, *RB-SP* and *RB* transgenic potato lines were generated to test whether or not these genes interact with the *IPI-O* effectors in vivo. Several transformation events were recovered for each *RB* gene construct (Table 4.2). All transformation events revealed the integration of the *NPTII* gene based upon PCR analysis and agarose gel detection, with a transformation efficiency of 26.6% for *RB-SP* and 15.5% for *RB* (Table 4.2). Nine and seven transformation events for *RB-SP* and *RB* transgenic lines respectively, were discarded due to phenotypic abnormalities.

Table 4.2. *Transformation efficiency for RB-transgenic lines.*

Construct	Num. Explants	N. Trans. Events**	Transformation Efficiency (%)***	Selected transgenic plants
*RB-SP**	110	26	26.6	8
RB	97	17	15.5	10

*Synthetic *RB* gene with a CC domain from *S. pinnatisectum* and NLR domain core from *S. bulbocastanum.*
** Number of transformation events.
*** Calculated as the percentage of transformation events with *NPTII* integration.

RT-PCR was conducted to confirm the expression of *RB-SP* and *RB* in selected transgenic lines. One transgenic line expressing *RB-SP* was detected (PNT-22, Figure 4.2A) but no expression was observed in the remaining lines despite having *RB-SP* or *RB* integrated (Figure 4.2A). An agroinfiltration assessment was conducted to confirm whether HR is elicited after challenging with IPI-O1 and IPI-O4 on PNT-22. HR was triggered using IPI-O1 as expected (Figure 4.2B). However, a weak HR was observed when IPI-O4 was infiltrated.

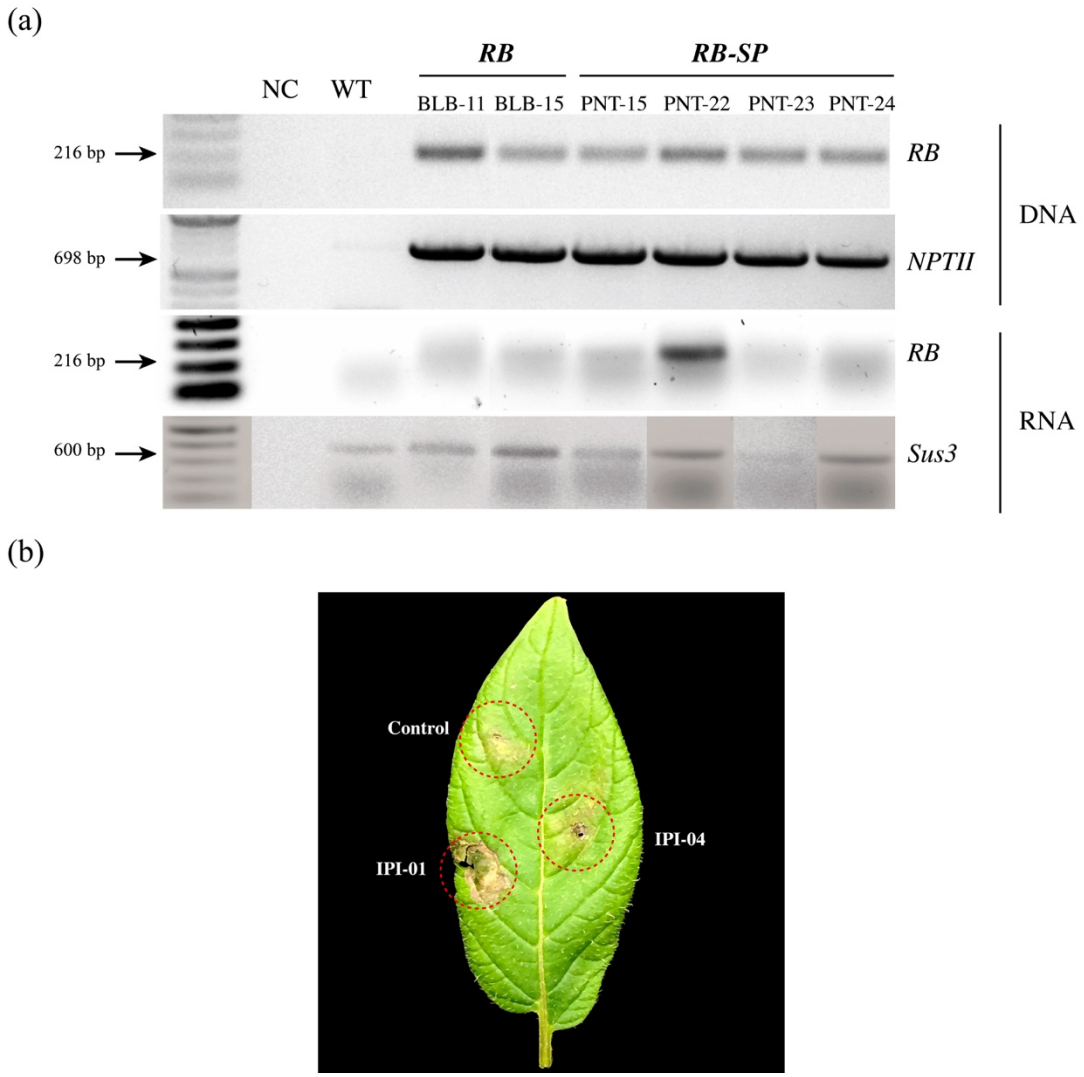

Figure 4.2. *Agroinfiltration of IPI-O effectors and RB-SP/RB insertion and expression in DRH-S5 transgenic potato lines.* (a). *RB-SP* detection and expression within selected transgenic lines. *NPTII* and *SUS3* were used to confirm *RB-SP* integration and as an internal control for RT-PCR, respectively. NC: Negative control. WT: Wild type. (b) A DRH-S5 transgenic line (PNT-22) carrying the *RB-SP* synthetic gene revealed HR after challenging with both IPI-O effectors. A weak HR-type of response is observed on negative control caused by mechanical damage.

iii. Coiled-Coil homodimerization disruption mediated by IPI-O4

A Y2H assay was conducted to validate the CC dimerization suppression when IPI-O4 is present. A weak to null self-association was observed between CC domains from *blb* and *pnt* respectively (Figure 4.3). These results were consistent across each technical replicate and only the positive control presented a strong interaction in both DDO/X/A and QDO/X/A dropout media. Due to this lack of interaction, we were unable to determine the role IPI-O4 plays in CC dimerization.

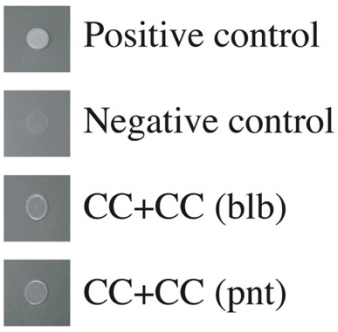

Figure 4.3. *Coiled-Coil (CC) domain dimerization assay using yeast-two-hybrid.* Representative CC-domain self -associations are shown for either *S. bulbocastanum* (blb) and *S. pinnatisectum* (pnt) CC domains. Three independent yeast transformations presented weak to null growth when compared with a positive control.

V. Discussion

Plant disease resistance relies on the ability to recognize pathogen-associated signatures through specialized plant receptors. Plant breeding programs have leveraged this detection system to introduce multi-pathogen resistance into new cultivars, significantly reducing crop production losses caused by plant diseases (Gupta et al., 2015). However, much of the research conducted to

date has been restricted to investigations of host-pathogen interactions in vitro, limiting the recognition of additional factors contributing to disease resistance responses (Gupta et al., 2015). In this study, in vitro and in vivo approaches were used to elucidate the interaction of potato CC domains from a synthetic *RB* gene and the late blight *IPI-O* effectors.

The impact of *IPI-O1* and *IPI-O4* effectors over the *RB*-mediated resistance in vivo was accessed using transient and stable *RB* expression. *RB* and *IPI-O1* co-infiltration in *N. benthamiana* leaves elicited HR (Figure 4.1), supporting in vitro and in vivo observations reported by Liu and Halterman (2009), Halterman *et al*. (2010), and Chen *et al*. (2012). However, HR was observed when *IPI-O1* and *IPI-O4* effectors were infiltrated in a transgenic line expressing the synthetic *RB-SP* gene (Figure 4.2). Although some technical replicates did not elicit a strong HR when *IPI-O4* was infiltrated, these results suggest that additional mechanisms may contribute to HR when *RB-SP* is present in potato. Moreover, these findings, also indicate that *IPI-O1* may not be the sole factor for triggering HR. Further analysis should be conducted using more transgenic lines expressing *RB-SP* to confirm these results.

Interestingly, all transformation events subjected to PCR amplified an *RB-SP* or *RB* band. However, most of them did not express these genes as revealed by RT-PCR (Figure 4.2A). Previous studies have shown a similar finding, associated mainly to transgene silencing mechanisms in plants subjected to transgenesis (Fagard and Vaucheret, 2000; Finnegan and McElroy, 1994). In particular, transgene co-suppression, correlated with a high gene copy number, but also DNA methylation, have been reported as the principal causes of transcriptional repression (Fagard and Vaucheret, 2000; Finnegan and McElroy, 1994). Thus, *RB* expression/translation may be suppressed by internal silencing cell machinery in DRH-S5.

Chen *et al*. (2012) reported no physical interactions between the CC domains from either *blb* or *pnt*, and the IPI-O4 effector using Y2H. In the present study, no HR was observed when *RB-SP* and *RB* were independently co-infiltrated with IPI-O4, confirming that IPI-O4 does not interact with the CC domain in *N. benthamiana*. Interestingly, unlike most *RB-SP* and *RB* transgenic lines, the PNT-22 line expressed *RB-SP* and triggered an HR after challenging with both IPI-O effectors (Figure 4.2B). These results suggest that other proteins in addition to *RB-SP* may be involved in HR in potato. Although IPI-O1 and IPI-O4 do not exhibit physical interactions to the CC domain from *pnt* in vitro (Chen et al., 2012), it is possible, that undetected specific posttranscriptional and posttranslational changes, such as protein folding, must be required to interact with *RB-SP* in vivo. However, with nearly null *RB* and *RB-SP* expression in transgenic lines, caution must be applied, and further analysis should be conducted to elucidate these interactions in vivo. For instance, a detached leaf bioassay using the US-23 *P. infestans* strain could be conducted to confirm these results.

These findings raise intriguing questions regarding the nature and extent of the *RB*-mediated resistance through the recognition of IPI-O1 and consequent CC dimerization. The present study could not validate the interaction with CC domains from either *blb* or *pnt in vitro*. Despite presenting weak yeast growth on DDO/X/A and QDO/X/A dropout media (comparable with negative control), these results are not conclusive.

Further research should be undertaken to investigate the *RB*-mediated resistance suppression by the *IPI-O4* effector from late blight in transgenic potato lines. Although in vitro and in vivo observations agree with previous studies using transient expression in *N. benthamiana*, only one line carrying a synthetic *RB* gene was able to trigger an HR upon *IPI-O* effectors infiltration. A previously undetected interaction was also observed (*RB-SP*/*IPI-O4*) suggesting that additional

proteins or recognition mechanism may be involved in an HR in transgenic potato. Moreover, we were unable to confirm CC-self associations and therefore further analysis should be conducted to confirm these findings.

LITERATURE CITED

LITERATURE CITED

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

Chen, Y., and Halterman, D. A. (2017). *Phytophthora infestans* effectors IPI-O1 and IPI-O4 each contribute to pathogen virulence. Phytopathology 107:600–606.

Chen, Y., Liu, Z., and Halterman, D. a. (2012). Molecular determinants of resistance activation and suppression by *Phytophthora infestans* effector IPI-O. PLoS Pathogens 8:1–10.

Colton, L. M., Groza, H. I., Wielgus, S. M., and Jìang, J. (2006). Marker-assisted selection for the broad-spectrum potato late blight resistance conferred by gene *RB* derived from a wild potato species. Crop Science 46: 589–594.

Devaux, A., Kromann, P., and Ortiz, O. (2014). Potatoes for sustainable global food security. Potato Research 57: 185–199.

Fagard, M., and Vaucheret, H. (2000). (Trans)gene silencing in plants: How many mechanisms? Annual Review of Plant Physiology and Plant Molecular Biology 51: 167–194.

FAOSTAT (2019). FAOSTAT database collections. Food and Agriculture Organization of the United Nations. Available at: http://www.fao.org/faostat/ [Accessed January 27, 2019].

Finnegan, J., and McElroy, D. (1994). Transgene inactivation: plants fight back!. Bio/Technology 12: 883–888.

Friedman, M. (1997). Chemistry, biochemistry, and dietary role of potato polyphenols: A Review. Journal of Agricultural and Food Chemistry 45: 1523–1540.

Gupta, R., Lee, S. E., Agrawal, G. K., Rakwal, R., Park, S., Wang, Y., et al. (2015). Understanding the plant-pathogen interactions in the context of proteomics-generated apoplastic proteins inventory. Frontiers in plant science 6, 352.

Halterman, D. A., Chen, Y., Sopee, J., Berduo-Sandoval, J., and Sánchez-Pérez, A. (2010). Competition between *Phytophthora infestans* effectors leads to increased aggressiveness on plants containing broad-spectrum late blight resistance. PloS One 5(5):e10536.

Halterman, D. a, Kramer, L. C., Wielgus, S., and Jiang, J. (2008). Performance of transgenic potato containing the late blight resistance gene *RB*. Plant Disease 92: 339–343.

Haverkort, A. J., Struik, P. C., Visser, R. G. F., and Jacobsen, E. (2009). Applied biotechnology to bombat bate blight in potato aused by *Phytophthora Infestans*. Potato Research 52: 249–264.

Koncz, C., Martini, N., Szabados, L., Hrouda, M., Bachmair, A., and Schell, J. (1994). "Specialized vectors for gene tagging and expression studies," in Plant Molecular Biology Manual, eds. S. B. Gelvin and R. A. Schilperoort (Dordrecht: Springer Netherlands), 53–74.

Kuhl, J. C., Zarka, K., Coombs, J., Kirk, W. W., and Douches, D. S. (2007). Late blight resistance of *RB* transgenic potato lines. Journal of the American Society for Horticultural Science 132: 783–789.

Leesutthiphonchai, W., Vu, A. L., Ah-Fong, A. M. V, and Judelson, H. S. (2018). How does *Phytophthora infestans* evade control efforts?. Modern insight into the late blight disease. Phytopathology 108: 916–924.

Li, W., Zarka, K. A., Douches, D. S., Coombs, J. J., Pett, W. L., and Grafius, E. J. (1999). Coexpression of potato PVYo coat protein and cryV-Bt genes in potato. J. Amer. Soc. Hort. Sci. 124: 218–223.

Liu, Z., and Halterman, D. (2009). Analysis of proteins differentially accumulated during potato late blight resistance mediated by the *RB* resistance gene. Physiological and Molecular Plant Pathology 74: 151–160.

Lozano, R., Ponce, O., Ramirez, M., Mostajo, N., and Orjeda, G. (2012). Genome-Wide Identification and Mapping of NBS-Encoding Resistance Genes in *Solanum tuberosum* Group Phureja. PloS One 7(4):e34775.

Ma, L., Lukasik, E., Gawehns, F., and Takken, F. L. W. (2012). "The use of agroinfiltration for transient expression of plant resistance and fungal effector proteins in *Nicotiana benthamiana* leaves," in Plant Fungal Pathogens: Methods and Protocols, eds. M. D. Bolton and B. P. H. J. Thomma (Totowa, NJ: Humana Press), 61–74.

Murashige, T., and Skoog, F. (1962). A revised medium for rapid growth and bio assays with tobacco tissue cultures. Physiol. Plant 15: 473–497.

Nowicki, M., Foolad, M. R., Nowakowska, M., and Kozik, E. U. (2011). Potato and tomato late blight caused by *Phytophthora infestans*: An overview of pathology and resistance breeding. Plant Disease 96: 4–17.

Pomerantz, A., Cohen, Y., Shufan, E., Ben-Naim, Y., Mordechai, S., Salman, A., et al. (2014). Characterization of *Phytophthora infestans* resistance to mefenoxam using FTIR spectroscopy. Journal of Photochemistry and Photobiology B: Biology 141: 308–314.

Raymundo, R., Asseng, S., Robertson, R., Petsakos, A., Hoogenboom, G., Quiroz, R., et al. (2017). Climate change impact on global potato production. Eur. J. Agron. 100: 87-98.

Rodewald, J., and Trognitz, B. (2013). *Solanum* resistance genes against *Phytophthora infestans* and their corresponding avirulence genes. Molecular Plant Pathology 14: 740–757.

Smart, C. D., and Fry, W. E. (2001). Invasions by the late blight pathogen: Renewed sex and enhanced fitness. Biological Invasions 3: 235–243.

Song, J., Bradeen, J. M., Naess, S. K., Raasch, J. a, Wielgus, S. M., Haberlach, G. T., et al. (2003). Gene *RB* cloned from *Solanum bulbocastanum* confers broad spectrum resistance to potato late blight. Proceedings of the National Academy of Sciences of the United States of America 100: 9128–9133.

Sparks, A. H., Forbes, G. A., Hijmans, R. J., and Garrett, K. A. (2014). Climate change may have limited effect on global risk of potato late blight. Global Change Biology 20: 3621–3631.

The Potato Genome Sequencing Consortium (2011). Genome sequence and analysis of the tuber crop potato. Nature 475: 189–195.

Vinson, J. a., Demkosky, C. a., Navarre, D. a., and Smyda, M. a. (2012). High-antioxidant potatoes: Acute in vivo antioxidant source and hypotensive agent in humans after supplementation to hypertensive subjects. Journal of Agricultural and Food Chemistry 60: 6749–6754.

Xing, S., Wallmeroth, N., Berendzen, K. W., and Grefen, C. (2016). Techniques for the analysis of protein-protein interactions *in vivo*. Plant physiology 171: 727–758.

Chapter 5

GENERAL CONCLUSIONS

I. Improving pathogen-resistance in potato using genomic selection

The genomic selection study described here uses whole genome-regressions to access the genetic architecture and prediction accuracy for late blight and common scab resistance in a breeding population from the Michigan State University (MSU) potato breeding program. In this study, high-quality Single Nucleotide Polymorphisms (SNPs) and field data collected during seven and nine years for late blight and common scab were used. These data reveal a moderately high genomic heritability value for late blight (0.46±0.04) and common scab (0.45±0.017).

The extent of genotype-by-year interaction was high for late blight and low for common scab, reflecting high disease pressures in the field caused by changes of pathogen aggressiveness and late blight strain prevalence across years. In particular, for late blight, we revealed that multiple SNPs contribute to late blight resistance on previously reported resistance hotspot regions. These results uncover the disease resistance introgression history of our germplasm collection coming from species such as *S. demissum* and *S. berthaultii* (Massa et al. 2015; Manrique-Carpintero, personal communication). Unlike late blight, a new locus located in a WRKY transcription factor, with a sizable contribution to inter-individual differences, was detected for common scab. This SNP along with those with larger-SNP variance detected in this study can be used to support marker-assisted selection programs using genotyping technologies such as KASP, representing the first attempt to date for early selection of common scab resistance lines.

Prediction accuracy assessments demonstrated the applicability of genomic prediction for tetraploid potato breeding. For both traits, more than 90% of the genetic variance could be

captured with an additive model. For late blight, small but statistically significant gains in prediction accuracy were obtained using a model that accounted for both additive and dominance effects. For common scab, the highest prediction accuracy was achieved using an additive model.

Further efforts to increase prediction accuracies for late blight and common scab resistance should consider expanding sample size and marker density and also integrate this breeding strategy into the MSU potato breeding scheme. In particular, the development of a 35K SNP array will not only allow increased marker density but also reduce the ascertainment bias by including new and rare genotypic variants. Additionally, the models evaluated in this work can also be extended and integrated into current diploid breeding strategies.

II. Contributing to diploid potato breeding using targeted mutagenesis

The targeted mutagenesis study described here uses the CRISPR/Cas9 system to overcome self-incompatibility (SI) in diploid potato. We first computationally identified three new *S-RNase* alleles, with flower-specific expression, in two diploid SI potato lines (DRH-195 and DRH-310). These alleles presented high inter-restricted amino-acid similarities to reported Solanaceae *S-RNase* sequences, supporting the theory that *S-RNase* divergence precedes speciation (Ioerger et al., 1990).

An RH-specific *S-RNase* allele was mapped to chromosome I within the peri-centromeric region, consistent with the localization of the *S*-locus in a low recombination region, favoring outcrossing. *S-RNase* knock-out (KO) lines were obtained using a dual single-guide RNA (sgRNA) strategy, generating premature stop codons by targeting conserved regions on each *S-RNase* exon. After self-pollination, fruits were set in selected KO lines in T_0 and T_1, demonstrating a stable *S-RNase* KO transmission. Moreover, Cas9-free KO lines were also

detected in T_1 lines, representing a significant advance towards the generation of transgene-free self-compatible (SC) potato cultivars that may not be regulated by the United States Department of Agriculture, Animal and Plant Health Inspection Service (USDA APHIS).

Plasticity in SI was also observed in wild-type (WT) plants after self-pollination. These results demonstrated that besides the *S-RNase*-based SI system, other modifier loci can modulate pollen rejection in gametophytic SI, as reported in several *Solanum* species (Goldraij *et al.*, 2006; McClure *et al.*, 1999; O'Brien *et al.*, 2002). In particular, proteins such as eEF1A or 120K, can directly interact with *S-RNase*, contributing to the strength of the SI response (Goldraij et al., 2006; Soulard et al., 2014). For instance, the suppression of pollen tube inhibition has been achieved targeting a gene encoding the 120K protein via RNA interference (Liu et al., 2009). Therefore, knocking-out these modifier loci along with *S-RNase* may introduce a consistent and stable self-compatibility in diploid potato.

Environmental factors also play an essential role in SI plasticity. For instance, temperature fluctuations, photoperiod, glucose starvation, and humidity significantly reduced SI in species such as *S. peruvianum* after self-pollination (Webb and Williams, 1988). To further explore whether conditions such as temperature contribute to the strength of the SI response in WT and *S-RNAse* KO lines, in vivo and in vitro analysis could be performed. As reported in citrus (Distefano et al., 2012), a species under gametophytic SI, self-pollinated plants and excised flowers can be placed in a growth chamber under different temperature conditions (i.e., 20, 25 and 30 ˚C). The effect of temperature on pollen germination and pollen tube growth can be observed under microscope after style staining. This approach could provide insight on the impact of specific environmental conditions over SI plasticity and may be considered as an alternative strategy for the generation of SC diploid potato. Moreover, heritability estimates can

be accessed to investigate the genetic and environmental component explaining self-compatibility in WT and KO lines.

Overall, additional work must be conducted to increase the strength and stability of the SC phenotype. Targeting primary fertility-related genes including *S-RNase* and non-stylar factors will contribute to incorporate self-compatibility into diploid potato inbred lines.

III. Understanding late blight resistance suppression

The validation of the *RB*-mediated resistance suppression in potato assessment described here uses in vivo and in vitro approaches to confirm this suppression. Using a heterologous system, we confirmed the interaction between the Coiled-Coil (CC) domain from the resistance to the late blight (RB) protein and the late blight IPI-O1 effector leading to a hypersensitive responses (HR).

Diploid transgenic potato lines were generated using a synthetic *RB* gene carrying a CC domain from *S. pinnatisectum* and an NLR domain core from *S. bulbocastanum* (*RB-SP*), using an *RB* gene as a control. The PNT-22 transgenic line, expressing the *RB-SP* gene was selected for agroinfiltration assays using the *IPI-O1* and *IPI-O4* effectors. The HR was elicited for both late blight effectors, suggesting that additional mechanisms may be involved in the resistance response to late blight in transgenic diploid potato. Further analyses should be conducted to explore whether proteins other than *RB-SP* or unreported interactions between the *IPI-O* effectors and specific domains from *RB*-SP, contribute to HR. Moreover, additional transgenic lines expressing *RB-SP* should be generated to confirm the observed phenotype and determine a possible relationship between HR strength and *RB-SP* copy number. Likewise, a detached-leaf

bioassay using the US-23 late blight genotype will also provide a broader comprehension of the *RB-SP*-mediated resistance in diploid potato.

This study was unable to demonstrate CC-self association using a yeast-two-hybrid system (Y2H), and therefore we could not validate the *IPI-O4* suppression role over the *RB*-mediated resistance. Despite observing a weak yeast growth in selection media for using CC domains for both *RB* genes, the results were not conclusive. To confirm whether these results were a product of bait toxicity, further analysis should be conducted including yeast co-transformations using empty bait plasmids along with CC. Similarly, to determine whether the lack of self-association was due to a lack of CC expression, western blot assays should be conducted to confirm this hypothesis. Similarly, additional work should be conducted to test if the lack of CC-interaction was a result of a loss of yeast glycosylation signatures related to posttranscriptional modification as suggested by Xing et al. (2016) for Y2H systems.

IV. Final considerations

Advances in crop-based breeding have enabled the identification of thousands of genomic variants that can be harnessed for selection of superior genotypes without multiple phenotypic cycles through genomic selection (GS) (Hu et al., 2018). However, multi-trait GS is restricted to highly correlated traits (Jia and Jannink, 2012), which underscores the need for integrating genomic technologies towards the improvement of multiple related or unrelated traits to respond rapidly to increasing food demand. In particular, genome editing technologies could contribute to the generation of superior breeding materials by targeting genomic regions associated with deleterious genes or enhancing the expression of important agronomic-related traits.

Recently, a targeted recombination strategy, combining GS and genome editing tools, was proposed to increase genetic gains for several crops. This approach involves the use of homologous recombination or crossing-over induction on genomic regions where marker effects have been estimated (Bernardo, 2017). An increase of genetic gains could be achieved for those lines in which chromosome segments with desirable allele combinations are present (Bernardo, 2017). Despite the inability to target all regions with high estimated markers effects, simulation analysis on major crops such as maize, soybean, wheat, barley, and pea (Bernardo, 2017; Ru and Bernardo, 2018), have demonstrated the feasibility of targeted recombination, directed to reduce the breeding cycle in crops like potato.

A constant improvement in potato genotyping platforms, including sequencing initiatives such as the potato pan-genome, will enable the integration of genomics tools for potato improvement. The recognition of new causative genetic variants for agronomic-related traits using genome wide-association and GS analysis will become a primary input for targeted recombination in potato. This new breeding method may allow the generation of novel allele combinations in potato breeding, translating genomic and data modeling into genetic gains in the field.

LITERATURE CITED

LITERATURE CITED

Bernardo, R. (2017). Prospective targeted recombination and genetic gains for quantitative traits in maize. The plant genome 10: 1–9.

Distefano, G., Hedhly, A., Las Casas, G., La Malfa, S., Herrero, M., and Gentile, A. (2012). Male–female interaction and temperature variation affect pollen performance in Citrus. Scientia Horticulturae 140: 1–7.

Goldraij, A., Kondo, K., Lee, C. B., Hancock, C. N., Sivaguru, M., Vazquez-Santana, S., et al. (2006). Compartmentalization of *S-RNase* and *HT-B* degradation in self-incompatible *Nicotiana*. Nature 439(7078): 805-10.

Hu, H., Scheben, A., and Edwards, D. (2018). Advances in integrating genomics and bioinformatics in the plant breeding pipeline. Agriculture 8, 75.

Ioerger, T. R., Clark, A. G., and Kao, T. H. (1990). Polymorphism at the self-incompatibility locus in Solanaceae predates speciation. Proceedings of the National Academy of Sciences 87: 9732 LP – 9735.

Jansky, S. H., Charkowski, A. O., Douches, D. S., Gusmini, G., Richael, C., Bethke, P. C., et al. (2016). Reinventing potato as a diploid inbred line–based crop. Crop Science 56: 1412–1422.

Jia, Y., and Jannink, J.-L. (2012). Multiple-trait genomic selection methods increase genetic value prediction accuracy. Genetics 192: 1513–1522.

Liu, B., Morse, D., and Cappadocia, M. (2009). Compatible pollinations in *Solanum chacoense* decrease both *S-RNase* and *S-RNase* mRNA. PloS One 3;4(6):e5774.

Massa, A. N., Manrique-Carpintero, N. C., Coombs, J. J., Zarka, D. G., Boone, A. E., Kirk, W. W., et al. (2015). Genetic linkage mapping of economically important traits in cultivated tetraploid potato (*Solanum tuberosum* L.). G3: Genes, Genomes, Genetics 5: 2357–2364.

McClure, B., Mou, B., Canevascini, S., and Bernatzky, R. (1999). A small asparagine-rich protein required for *S*-allele-specific pollen rejection in *Nicotiana*. Proceedings of the National Academy of Sciences. 96: 13548–13553.

O'Brien, M., Kapfer, C., Major, G., Laurin, M., Bertrand, C., Kondo, K., et al. (2002). Molecular analysis of the stylar-expressed *Solanum chacoense* small asparagine-rich protein family related to the HT modifier of gametophytic self-incompatibility in *Nicotiana*. Plant J 32: 985–996.

Ru, S., and Bernardo, R. (2018). Targeted recombination to increase genetic gain in self-pollinated species. Theoretical and Applied Genetics, 132(2):289-300.

Soulard, J., Boivin, N., Morse, D., and Cappadocia, M. (2014). eEF1A is an *S-RNase* binding factor in self-incompatible *Solanum chacoense*. PLOS ONE 9(2): e90206.

Webb, M. C., and Williams, E. G. (1988). Effects of temperature, light, nutrients and carbon dioxide on the strength of the self-incompatibility response in detached flowers of *Lycopersicon peruvianum*. Ann. Bot. 61: 395–404.

Xing, S., Wallmeroth, N., Berendzen, K. W., and Grefen, C. (2016). Techniques for the analysis of protein-protein interactions *in vivo*. Plant physiology 171: 727–758.