Integrating genomics assisted breeding and novel selection methods for enhancing application and uptake of potato technologies in South Asia

Final Report

A collaborative research project between
the International Potato Center (CIP)
and
Agriculture and Agri-Food Canada–Potato Research Center (AAFC–PRC)

Submitted to the CGIAR-Canada Linkage Fund of the
Canadian International Development Agency (CIDA)

11 May 2016

TABLE OF CONTENTS

ABBREVIATIONS

AAFC	Agriculture and Agri-Food Canada
AUDPC	Area under the disease progress curve
BW	Bacterial wilt
CCLF	CGIAR-Canada Linkage Fund
Chrs.	Chromosomes
CIDA	Canadian International Development Agency
CIP	International Potato Center
DAP	Days after planting
GEBV	Genomic estimated breeding value
GS	Genomic selection
GWAS	Genome wide association studies
HRM	High-resolution melting
LB	Late blight
LD	Linkage disequilibrium
LTVR	Lowland Tropic Virus Resistance
MAS	Marker-assisted selection
M&B	Mother and Baby
MLM	Mixed linear model
PVS	Participatory variety selection
QTL	Quantitative trait loci
RTB	CGIAR Research Program on Roots, Tubers and Bananas
SET	Standard evaluation trials
SNP	Single nucleotide polymorphism
SNC	Single node cutting
SolCAP	The Solanaceae Coordinated Agricultural Project (SolCAP)
STA	Sub-tropical Asia
TP	Training population

1. BACKGROUND AND JUSTIFICATION

Potato variety development in South Asia has relied on European and North American germplasm. Available seed is expensive, and most varieties are poorly adapted and thus unreliable under conditions of abiotic stress and pressure from disease. Early potato resistant to heat and viruses is a profitable and nutritious complement to low-income cereals in Bangladesh, India, North Vietnam, and the plains of Nepal and East Pakistan. Traditional potato breeding is time consuming and has had relatively low success rates. The International Potato Center (CIP) has taken a long-term approach to strategic use of genetic resources in the development of broad-based populations with resistance to potato late blight (LB) and adaptation to cool highland conditions of the tropics. CIP has also developed populations with resistance to virus diseases, relative earliness, and adaptation to warm, arid environments (Bonierbale et al. 2003). However, most germplasm with disease resistance is not adapted to the short-season environments typical in Canada or South Asia, and few progeny meet the required threshold for selection in these environments. The association of late maturity or the need for short crop duration and resistance to LB presents a distinct challenge to breeding resistant cultivars for intensive cropping systems of the subtropics. New genetic knowledge, alternate breeding strategies, and improved genomic tools have been used in this project to help bypass many of the inherent problems of conventional breeding approaches.

This report details the research performed as part of the Canadian International Development Agency (CIDA)-funded project "Integrating genomics assisted breeding and novel selection methods for enhancing application and uptake of potato technologies in South Asia." The project aimed to improve the efficiency of crop improvement through the incorporation of molecular breeding approaches and participatory selection. We also report on activities conducted to strengthen capacities on these approaches to help achieve project outputs.

2. PARTNERSHIPS AND ADDITIONAL RESOURCE MOBILIZATION BUIDING ON THE PROJECT

In 2012, an additional Canadian partner was included since Agriculture and Agri-Food Canada (AAFC) was not eligible to receive funds directly or indirectly from CIDA. Dr. Gefu Wang-Pruski of the Dalhousie University Faculty of Agriculture (Truro, Nova Scotia) agreed to participate in the project as the Canadian partner, and thus received funds designated for Canadian research under an agreement approved in March 2012. Dalhousie University generated molecular marker data for CIP genetic material using CGIAR-Canada Linkage Fund (CCLF) the first year of the project. It also provided academic training and co-supervision of M.Sc. thesis of a Peruvian postgraduate student in Dalhousie University from November 2012 onwards, together with the original Canadian partner, Dr. David De Koeyer from AAFC. In 2011, Canadian collaborator Dr. Koeyer contributed to the development of the inter-center proposal for the CGIAR Research Program on Roots, Tubers and Bananas (RTB), led by CIP, which is presently providing complementary funds to those from the CCLF collaboration aimed at implementation of the project's results.

Some major proposals for funding were submitted during the no-cost extension period of the project (2014–2015). This included a proposal "Integrating the Agile Potato into Cereal Based Systems in Central Asia" presented to BMZ/GIZ International Agricultural Research for Development call in 2015 that would benefit poor farmers in water stress-prone areas of Central Asia; it was unfortunately not funded. Outputs achieved from this CIDA project concerning genomic-assisted breeding based on genomic selection (GS) models would be applied in the

selection of new material for trait components associated with early response to warm long-day environments. A second proposal, "Trait Observation and Discovery Network for Agile Potato Technology," was also presented to BMZ/GIZ International Agricultural Research for Development and funded for 3 years (2015–2017 at €1,200,000). In this new project, CIP's panel of nearly 200 breeding lines used in the present CIDA project was extended to 300, and genotyping by sequencing (GBS) was performed to complement genotyping accomplished in the CIDA project using the Solanaceae Coordinated Agricultural project (SolCAP) single nucleotide polymorphism (SNP) array. A larger number of SNP markers would lead to enhanced resolution for genome wide association studies (GWAS) and improve accuracy of GS to predict phenotypic performance within segregating families. These families were generated from intercrossing breeding lines selected in the present CIDA project for disease resistance, heat tolerance, and response to photoperiod as a general measure to increase adaptation and yield in target environments and cropping systems.

Funding from CIP, the United States Agency for International Development, and the Austrian Development Agency enabled collaboration with a number of partners in Asia for the introduction and participatory evaluation of candidate potato varieties. These partner organizations comprised the Tuber and Root Crops Research Center Bangladesh, Potato Development Program of the Ministry of Agriculture and Forests-(Buthan), National Potato Research Program of the (Nepal) Agricultural Research Council, and the Livelihoods and Biodiversity Initiatives for Research and Development, also in Nepal.

3. FULFILLMENT OF OBJECTIVES AND PRODUCTION OF OUTPUTS

The goal of the project was to help improve agricultural productivity, incomes, and dietary diversity for poor farmers and consumers, with reduced dependence on external inputs and risk from climate change. The project addressed variety needs of South Asia sub-region ecoregions where early heat-tolerant and virus-resistant potato varieties would enable farmers to profitably insert potato into mixed cropping systems in the extensive Indo Gangetic plains and in other regions. It is here that day-neutral varieties are desirable for multiplying seed under long-day length during the spring. For this purpose, broad genetic-base breeding germplasm developed from wild and South American native germplasm at CIP was used to develop day-neutral, heat-tolerant progenitors and progenies with virus and LB resistance. The goal of developing broadly adapted potato varieties in this project envisaged the implementation of innovative selection approaches based on newly emerging genomics and statistical models (i.e., GS). The poor and impoverished rural communities would thus benefit from technological advances that may reduce the timeframe for developing new varieties. The overall project objective was to develop capacities for GS in potato to select for resistance and component traits associated with adaptation to long photoperiods in warm areas. In addition, participatory approaches were incorporated for selection aimed at improving efficiency of potato breeding and variety uptake.

Section 3 summarizes the deliverables as outlined in the project.

3.1 BACKGROUND

The project developed communication systems to allow evaluation of gender-specific perception of potato, particularly on preferences in variety characteristics. Emphasis has been given to farmer participatory selection incorporating gender preferences into participatory evaluation schemes. Facilities in terms of database, protocol, software, and fieldbooks were developed, implemented,

and made available to the potato-breeding community. The feasibility and impact of adoption of the structured international testing approach that integrates researcher- and farmer-led experiments known as the "Mother and Baby" (M&B) trials (Snapp 1999) were assessed in several countries throughout the project period. This M&B scheme links replicated "mother" trials that can disaggregate confounding factors, such as management practices and varietal potential, with numerous simpler "baby" trials that can be designed and understood by farmers, in a robust analytical system that can sample multiple environments and conditions.

The project goal of combining resistance traits in early-maturing potato germplasm adapted to long-day environments was tackled by implementing novel breeding strategies based on emerging genomics and statistical methods to close the gap between genomic research in developed countries and traditional breeding research in developing countries. GWAS and GS were the two novel genomics-based methodologies applied in the project. GWAS aimed at identifying novel functional variation that can be deployed in cultivar development through marker-assisted selection (MAS) (Hamblin et al. 2011). GS (Meuwissen et al. 2001) is a breeding tool for improving the accuracy and frequency of parental and progeny selection.

Dense molecular marker coverage achieved using the SolCAP infinium array of 8,303 SNP markers enabled the integration of phenotypic and genotypic data so as to identify genomic regions and SNP markers influencing potato LB resistance, tuberization, and bulking-related traits affected by photoperiod and warm conditions by means of GWAS. Improved parental selection accuracy for these traits was achieved by GS, taking advantage of all available genome-wide marker information for directly predicting genetic values. The underlying concept in GS is to model the entire complement of quantitative trait loci (QTL) effects across the genome to produce a genomic estimated breeding value (GEBV) from which candidates (progenitors or clones) can be selected by genotyping. This project tested several GS models to assess the feasibility of applying this tool to improve selection of parental clones for LB resistance, early tuberization, and bulking under long- day length.

The success of GWAS efforts depends on the possibilities of separating linkage disequilibrium (LD) due to physical linkage from LD due to population structure and relatedness between genotypes. Highly significant LD between polymorphisms on different chromosomes because of population stratification may produce associations between a marker and a phenotype, even though the marker is not physically linked to the locus responsible for the phenotypic variation, causing spurious associations (Flint-Garcia et al. 2003). Relations between genotypes need to be adequately represented in statistical models for marker-trait associations to guarantee reliable inference. Another important issue in GWAS is the distance along the genome over which LD persists, as this determines the number and density of markers, and experimental design needed to perform an association analysis. Therefore, the resolution of associations between markers and traits—and thus of the efficiency of MAS—depends on LD extent. LD decay estimated in potato has been based on a limited amount of marker information, and was shown to vary from a rapid decay (<1 cM) (Gebhardt et al. 2004) to a slower decay (3 cM) (D'hoop et al. 2008) to a long-range decay (10 cM) (Simko et al. 2004). These results demonstrated that a higher marker density was needed to assess the LD decay in the potato genome. Therefore the potato SolCAP SNP array was useful for defining LD decay in the population used in this project.

The efficiency of GWAS and GS depends on the phenotyping efficiency. The genetic material used in the project consisted of a panel of 162 advanced breeding clones generated by CIP's breeding program. The panel was represented by 109 clones from the Lowland Tropics Virus Resistance

(LTVR) population (Bonierbale et al. 2003), adapted to the subtropical lowlands with virus resistance, medium-maturity, and heat tolerance; 28 clones from a population derived from a wide range of LB-resistance sources known as B3 population (Mihovilovich et al. 2015); 11 from crosses between LTVR and LB-resistance cultivars from *S. tuberosum* L. Andigena group; and 11 and 3 clones, respectively, from past breeding activities on BW resistance and pre-breeding for LB resistance. Since breeding at CIP has emphasized selection for adaptation to the short-day length of the lowland subtropics, few clones might meet the required threshold for selection in temperate environments. Despite this, variability for adaptation to long photoperiod was expected in this panel from long-day length–adapted germplasm of *S. tuberosum* Group Tuberosum bred under the summer conditions of the Northern Hemisphere (Cubillos and Plaisted 1976), Neotuberosum clones (Plaisted 1987), and hybrids of Tuberosum with *S. demissum* present in most pedigrees. This panel was built based on their variation of tuber characteristics (i.e., tuber shape, skin and flesh color); agronomical attributes (i.e., yield and early bulking, cooking, and processing quality); and virus and LB-resistance information retrieved from CIP's corporate database of "Standard Evaluation Trials" https://research.cip.cgiar.org/confluence/display/SET/Standard+Evaluation+Trials. CIP's series of variety assessment experiments, called standard evaluation trials (SET), complements the breeding and selection process. It provides uniform data for agronomic and quality characteristics across all populations of advanced clones using standardized protocols. SET comprise a series of intentional exposure trials in which advanced clones are challenged with each biotic or abiotic stress separately, and are evaluated for yield and quality under recommended production conditions. For example, the SET for LB resistance is conducted in a disease-prone area and season, such as the rainy season of the mid-elevation subtropics of Peru in Oxapampa (Pasco, 1,814 masl), where LB is endemic. Historical data from replicated LB-resistance trials from 2005 to 2011 were retrieved and used in the project. In addition, new protocols were developed for traits of importance for variety acceptance by farmers and consumers, including appropriate fieldbooks. All were implemented in Data Collector, a software developed by CIP for field trial planning, documentation, analysis, and reporting.

3.2 MAJOR PROJECT ACHIEVEMENTS BY DELIVERABLE

Deliverable 1. Enhanced understanding of local and gender-specific preferences for intrinsic characteristics of new potato varieties for integration into breeding schemes.

- CIP's protocol for participatory variety selection (PVS) under the M&B trial design was improved, including definitions, guidelines for procedures, and electronic fieldbooks. All were linked to Data Collector software and the BioMart facility of CIP for data management, analysis, and exchange.
- PVS trials in Bhutan successfully engaged women (11) and men (9) in each stage of the evaluation process (during flowering, at harvest, and during storage). Disease resistance, medium plant height, thick and strong stems, large round leaves, early bulking, high yield, mid-sized tubers, red skin color, late sprouting, strong and short sprouts, and resistance to rotting were recorded as user-preferred characteristics.
- In Bangladesh, 10 women and 20 men participated in M&B trials. The following preference characteristics were recorded: LB resistance, strong stem, drought tolerance, tolerance to waterlogging, high yield, good behavior in storage, good flavor when consumed fresh, red skin color, resistance to rotting, no change to taste after storage, late sprouting, and tight skin.
- In Nepal, an average of 13 women and 15–37 men participated in M&B trials at various stages of evaluation. Farmers preferred abundant flowers, high stem number, thick stems, high yield,

medium flowering, disease resistance, insect resistance, red skin color, and uniform mid-sized tubers.

Deliverable 2. Researchers' capacity to apply analytical procedures will be enhanced and efficiency of international potato evaluation trials improved.

- New protocols were developed for SET on early tuber bulking and tuber dormancy. Fieldbooks, analytical procedures, and report outputs were implemented in CIP's Data Collector software (https://research.cip.cgiar.org/confluence/display/GDET4RT/ Protocols)
- A new user-friendly interface, "HYDAP," was developed to facilitate use of Data Collector for planning and conduct of experiments. It allowed for integrating aspects of statistical analysis and visualization of results via simple reports drawn on database contents.

Deliverable 3. Availability (via a web portal) of new genetic markers statistical tools, practices (GS), and information (trial data and GEBV) for selection of early-maturing clones and progenitors for breeding programs and variety testing.

- Data information for practicing GWAS and GS available on the project's website: https://research.cip.cgiar.org/confluence/display/cclf/Home
  - A panel of 162 advanced breeding clones and nine known varieties was SNP genotyped using the Potato SolCAP SNP array. A table (Table S2. Matrix of genotypic data) comprising genotypic data for 4,738 SolCAP SNPs is available on the project's website
  - Phenotypic data for LB resistance and component traits associated with tuberization and bulking are also available on the project's website. A table (Table S1. Adjusted mean values) with the adjusted mean values for each trait is also available there.
  - The panel's population membership matrix and SNP-based kinship relationship matrix, both required to perform GWAS or GS, have been uploaded to the project website.
  - User manuals of R packages used for GWAS and GS analysis (i.e., GAPIT and BGLR) are also available.
- This project has contributed information on SolCAP SNP markers associated with component traits, including LB resistance, early tuberization, and bulking that can be used for MAS development. Details are available below in "Achievements by Outputs" (the appendix of this report) and in the thesis research developed under the project "Identification of Chromosomal Regions in Potato Associated with Tuberization Traits Affected by Photoperiod Using SNP Markers" available in Dalhousie University link https://dalspace.library.dal.ca/handle/10222/71386 and on the project's website following this link https://research.cip.cgiar.org/confluence/display/cclf/Home:
  - Three SolCAP SNP markers annotated for candidate genes for tuberization photoperiodic response in potato on Chrs. 4, 8, and 9, respectively, were found associated with tuber initiation in the present study.
  - Twenty genes described in the plant model *Arabidopsis thaliana* controlling photoperiod response were mapped in the proximities or within the same regions of clusters of SNP markers associated with traits related to early tuberization and bulking in potato.
  - Two SNP markers were identified on Chrs. 4 and 9 with alleles associated with decreased susceptibility to LB.
  - Two candidate genes recognized to play a role in pathogen recognition and response were found in the vicinities of a SNP marker associated with LB resistance on Chr. 9.

Deliverable 4. New progenitors and resistant germplasm with improved adaptation to short-season environments identified for evaluation in South Asian national programs.

- By applying GS, predictive models for LB resistance, and component traits related with tuberization and bulking recorded under long-day length and warm conditions were developed for a sample of 130 CIP breeding clones regarded as a "training population" (TP). The accuracy of the prediction was verified by cross validation correlation between predicted breeding values and phenotypic data in 41 additional advanced clones, nominated as "test population" in which GEBV were estimated based on the models developed in the TP. Prediction accuracies were high enough (r^2>0.3) for most traits, indicating the GS can be applied in breeding to improved accuracy of parental selection for resistance to LB, early tuberization, and bulking in warm and long-day length conditions.
- Thirty top breeding clones with virus resistance were selected for early tuberization and good bulking under warm and long-day length conditions. Availability of most of these clones was confirmed in the breeding germplasm distributed from CIP to sub-tropical Asian (STA) countries.
- True seed from intercrossing 15 out of the 30 breeding lines were generated. Some 200 true seed from each of 24 progenies were distributed to Vietnam and Uzbekistan as part of the deliverables expected from this project and with complementary support from RTB. Validation of GS can be performed in these countries by genotyping and phenotyping selections of first clonal generation. An improvement of predictive accuracies will be expected as the training population will become more representative of these families.

4. CONTRIBUTIONS TO RESEARCH EXCHANGE AND KNOWLEDGE DISSEMINATION

Scientific presentations

Mihovilovich, E., D. De Koeyer, and M. Bonierbale. 2013. Response of short day-length adapted potato breeding lines to long photoperiod. 97th Annual meeting of the Potato Association of America. July 28–August 1, Quebec, QC.

Caraza, M., D. De Koeyer, E. Mihovilovich, M. Bonierbale, and G. Wang-Pruski. 2013. Identification of candidate genes associated with photoperiod for improving adaptation of potatoes to diverse environments. 9th Canadian Plant Genomics Workshop. August 12–15, Halifax, NS.

Caraza, M. 2015. Identification of chromosomal locations in potatoes associated with tuberization Related traits Affected by Photoperiod using SNP Markers. Master's thesis Retrieved from https://dalspace.library.dal.ca/handle/10222/71386

Research exchange

Maria Caraza Salas (formerly at CIP) was a B.Sc. in biology from the Biology Faculty at the University Jorge Basadre Grohmann (Tacna, Peru) developing her undergraduate thesis research and travelled to Canada to pursue a master's degree at Dalhousie University. She successfully completed her thesis, "Identification of chromosomal locations in potatoes associated with tuberization related traits affected by photoperiod using SNP Markers," in 2015 under the supervision of project partners Dr. Wang-Pruski, professor at the Department of Plant and Animal Sciences of Dalhousie University, and Dr. De Koeyer, research scientist at the Potato Research Centre, AAFC. During her three-year stay in Canada, Ms. Caraza obtained skills on statistical analysis and management of statistical software required for GWAS and GS.

CIP scientist Elisa Mihovilovich travelled to Canada for 15 days in October 2014, to visit AAFC and Dalhousie University. She was there to increase her knowledge of how to apply new molecular tools and analytical procedures used in the parental selection of the genetic material used in this research project.

Canadian partners Drs. De Koeyer and Wang-Pruski visited CIP in December 2014, to discuss advances of activities of the project, possible collaborative projects, and meet with CIP scientists working on integrating high-throughput genotyping using SOLCAP chip and phenotyping for GS.

Ms. Caraza and two Colombian scientists interns at AAFC attended via video a workshop on "Methods and Models for Plant Genomic Prediction and Selection," held at CIP on 13–17 October.

These reciprocal scientific exchanges, knowledge transfer, and a co-supervised student strengthened collaborations between CIP and AAFC.

5. ACHIEVEMENTS BY OUTPUTS

5.1 OUTPUT 1: UNDERSTANDING LOCAL AND GENDER-SPECIFIC PREFERENCES FOR INTRINSIC CHARACTERISTICS OF NEW POTATO VARIETIES AND THEIR INTEGRATION INTO CROP IMPROVEMENT SCHEMES

CIP's protocol for PVS under the M&B trial design was improved during the project. The protocol comprises definitions, guidelines for procedures, and electronic fieldbooks, all linked to HIDAP-Data Collector software and the BioMart facility of CIP for data management, analysis, and exchange. The protocol compiles and extends CIP's and partners' experience in adaptation of the M&B trial design (Snapp 1999) to potato variety evaluation involving multiple stakeholders, with attention to gender. CIP's gender specialists from its Social, Health and Innovation Science Division reviewed the protocol developed by its crop improvement teams, with attention to engaging women in the PVS process and obtaining gender-disaggregated data on users' preferences.

CIP presented the procedures for M&B trials in five countries (Peru, Uzbekistan, Bangladesh, Nepal, and Bhutan). In all but Uzbekistan, where the lack of healthy seed stock set the progress back, the design has been incorporated into projects engaging multiple stakeholders. Data sets from five consortia in Peru (4 years, >35 trials) and from four trials conducted over two years in Bangladesh were compiled into a database for analysis and to develop implications for the variety improvement process.

Conclusions and Lessons Learned

Interest levels and willingness to apply the M&B trial design varied greatly across countries. The process is knowledge-intensive and time-consuming. It has been more successful in countries with less-regimented variety testing practices and where a dedicated researcher has been present for more than three complete selection cycles to offer guidance and fully demonstrate the procedures by training of trainers with local language skills. For example, following three years of intensive collaboration and support from CIP's dissemination expert and breeders, Peru's national potato program has taken up M&B trials as a primary means of multi-locational variety testing. A range of institutions, from nongovernmental organizations (NGOs) to farmers' groups and schools, has been engaged, such that five PVS consortia are now sustained with minimal backstopping from CIP. In Bangladesh, on the other hand, national program breeders are not encouraged to engage farmers in the varietal selection process, and the method was only demonstrated and led by CIP during the project period. Thus it has not been taken up at the national program level. Bangladesh considers

potato a "notified crop," which entails formal procedures, including the release of varieties before they can be shared with farmers. And although the "Baby" trials may be tended by farmers, they are more akin to secondary trial locations and do not provide for management according to farmers' practices. Nepal presented an intermediate situation: an NGO and the national program appreciated the value of PVS and practiced the M&B trial design for two seasons with successful results but require more time to assemble additional institutions and systematize the learning and data collection and interpretation. In Peru, this method of PVS has already shortened the variety release time by attention to the cyclic nature of selection, release, and acquisition of new candidate varieties. In Nepal, researchers' understanding of PVS procedures has been enhanced, and the methods will likely be implemented after the project's completion. Likewise in Bhutan, researchers were receptive to the method, training of trainers proceeded well, and the program will continue to use PVS. They adapted CIP's protocol to their needs by developing a short and practical PVS protocol (http://176.34.251.32/potatoknowledge/pvs.html). In Bangladesh, additional training of trainers as well as policy action will likely be required if PVS is to be adopted by researchers and extensionists for incorporation into statutory variety testing and release procedures. Farmers' preference traits were recorded and will be considered when recommending candidate varieties for testing in each target country (see Achievements by deliverables).

Individual consults via Skype with 34 researchers from Bangladesh and Bhutan (2012–2014) encouraged use of standard evaluation protocols and electronic PVS fieldbooks in Asian countries. To help realize synergies in research information from different sources, the nomenclatures of trials and farmers' preferences from PVS were standardized following the ontologies procedure and structures built in BioMart database (http://germplasmdb.cip.cgiar.org/index.jsp). Thirty-eight PVS fieldbooks (31 from Latin America, 7 from South Asia) developed from 2011 to 2014 were submitted to Biomart. Also, 4 fieldbooks from Buthan and 1 from Nepal received during 2014 and 2015 are currently being processed to be included in the database. Farmers' preference traits were recorded and will be considered when recommending candidate varieties for testing in each target country.

5.2 Output 2. Improved Statistical Procedures and Researchers' Capacities for Analyzing Variety Trials and Other Breeding Data

The SET portfolio was augmented with an improved protocol for evaluating earliness of tuber bulking, quality, and dormancy of potato tubers. During the project, a new user-friendly interface (HYDAP) was developed to facilitate planning, conduct of experiments, data collection, and integrating aspects of statistical analysis and visualization of results via simple reports drawn on database content. Six modules of CIP's International Cooperator's Guide for evaluation of advanced potato clones have been implemented in HYDAP. HYDAP was compared with the CGIAR-supported tool BMS (Breeding Management System) such that complementary features of these options could be identified and improvements to both systems planned.

All data collected in the project was updated in BioMart database. BioMart is a web query system on phenotypic and genotypic information about potatoes. BioMart, as a central repository database, maintains databases from genetic material distributed by CIP. Additionally, data integration combines heterogeneous data from different sources, providing a unified view for researchers.

To date, BioMart has four modules: (1) Healthy Tuber Yield (60 variables), (2) Late Blight Exposure Trials (15 variables), (3) Phytochemical Evaluation (46 variables for minerals, vitamin C, carotenoids, and phenolics), and (4) Participatory Variety Selection). This database provides access to 450 biological datasets from 10 countries and 146 fieldbooks: (1) biochemical trials—minerals

(54), vitamins (54), phenolics (30), and carotenoids (8); (2) yield performance trials (155); (3) LB evaluation trials (146); (4) physiological trait assessments (4); and (5) PVS (40).

Additionally, BioMart provides direct access to information on 325 clones evaluated in SET as catalogues available for international distribution (www.cipotato.org/catalogue). This catalogue showed information of clones on disease resistances, yield, mineral and vitamin C content, earliness, processing and boiling quality, dormancy period, and sprouting patterns.

Conclusions

HYDAP database and structures are intended to integrate independent components of CIP's Global Trial Data Management System. Besides recovering information of performance of clones distributed by CIP to stakeholders, BioMart performs data integration by combining heterogeneous data from different sources, thus providing a better overview of set of activities and connections between them, as well as the facility for retrieving and communicating results across projects and regions.

The long process required for adapting tools and the lack of timely feedback have made breeding progress slow and somewhat inefficient for the identification and diffusion of promising clones in terms of time and research demands. Improving policies on data management at institutional level and increasing capacities and availability of training materials in regional breeding programs will improve feedback and accelerate the adoption of new varieties.

5.3 Output 3. Molecular Genetic Information and Tools to Help Breeders Select Potato Genotypes and Parental Lines with LB Resistance that Tuberize Early under South Asian Conditions

See Appendix for report of activities concerning this output and for figures or tables denoted with "S."

Project activities of this output were implemented as part of an M.Sc. student's research program. Ms. Caraza, from Peru, travelled to Nova Scotia to follow postgraduate studies in Dalhousie University. She developed a significant part of this objective under thesis research on the identification of chromosomal locations in potatoes associated with tuberization-related traits affected by photoperiod using SNP markers. The research was co-supervised by the project partners Dr. Wang-Pruski, a professor at Research Centre, AAFC.

A cluster of SNPs was found associated with different tuberization and bulking component traits on different chromosomes, including SNPs for:

- Stolon length, small and total tuber number, and tuber induction under short-day length within a region of 1 Mbp between 58 and 59 Mbp on Chr. 3
- Stolon length and total tuber number under long-day length within a region of ~1.5 Mbp between 65.7 and 67.1 Mbp on Chr. 4
- Bulking ratio and number of small tubers under short-day length within a region of ~1.5 Mbp between 51.2 and 52.6 Mbp on Chr. 6
- Tuber induction and small tuber number under short-day length, and number of stolons in long-day length within a region of ~1.5 Mbp between 54.3 and 55.7 Mbp on Chr. 7
- Tuber induction and marketable tuber number under short-day length, and small and total tuber number under long-day length within a region of ~ 2.5 Mbp between 0.4 and 2.7 Mbp on Chr. 11 (Fig. S4).

Notably, SNPs significantly associated with multiple tuberization- and bulking-related traits (tuber induction, bulking ratio, number of marketable tubers, and stolon length under long-day length) were identified within a region of 6 Mbp between 0.7 and 6.4 Mbp on Chr. 5 (Fig. S4). This finding agrees with previous studies reporting a major QTL for maturity on Chr. 5 spanning this region (Visker et al. 2003).

Twenty genes responsible for photoperiod responses described in the model plant *Arabidopsis thaliana* (Table 1) were downloaded from https://www.arabidopsis.org/servlets/blast. Blast alignment was performed on the potato genome using the basic local alignment search tool (BLAST) option of the SPUD DB (Hirsch et al. 2014) (http://solanaceae.plantbiology.msu.edu/blast.shtml) in order to locate their positions in the potato genome. From this procedure, we found genes controlling photoperiod response in the proximities or within the same regions where clusters of SNPs aforementioned were found associated with different component traits related to tuber initiation (tuber induction) and bulking (Fig. 1). Some of these genes were the following: a Cycling DOF factor (CDF3) found ~2 Mb from the cluster of associated SNPs on Chr. 3; a CONSTANS gene located within the same region of associated SNPs on Chr. 4; a Cycling DOF factor (CDF5) gene located at 0.4 Mbp position close to the cluster of associated SNPs on Chr. 5; a Cycling DOF factor (CDF) gene located at 51.6 Mbp position within the same region of associated SNPs on Chr. 6; and Flowering locus T (FT) at 3.2 Mbp position, close to the cluster of associated SNPs on Chr. 11.

Other photoperiod-related genes found in the vicinities of associated SNP were CONSTANS (CO); CONSTANS 3 (CDF3); and Zinc finger protein found ~1.6 Mbp apart from SolCAP_c1_11459 located at 47.6 Mbp position on Chr. 2 and identified associated with tuber induction at 59 days after planting (DAP) under short-day length; a Flowering locus T at 51 Mbp position on Chr. 5 found 2.5 Mbp from SolCAP_c2_10358 associated with bulking ratio and small tuber number at 90 DAP under long-day length; CONSTANS at 2 Mbp position on Chr. 8 found on the same genomic region of SolCAP_c2_27452 associated with tuber induction at 40 DAP under long-day length; ~2 Mbp from SolCAP_c2_34103 associated with stolon number at 75 DAP under short-day length; and GIGANTEA at 54.7 Mbp position on Chr. 12 found ~3 Mbp from SolCAP_c2_18855 associated with tuber induction at 59 and 74 DAP under short-day length.

In addition, there were chromosomal locations annotated for potato genes which may control tuberization photoperiodic response such as SolCAP_c1_10762, associated with tuber induction under short-day length, is annotated for a photoperiod responsive protein in Chr. 4; SolCAP_c2_27452, associated with tuber induction under long-day length is annotated for a FLAVIN-BINDING KELCH protein in Chr. 8, known to down-regulate CDF genes; and most importantly, SolCAP_c2_40879, also associated with tuber induction under short-day length is annotated for a CONSTANS (CO) gene in Chr. 9.

As for LB resistance, susceptibility of the panel in the highland tropical location of Oxapampa (Pasco, Peru) was associated with two markers located within 9 Kb at 60.2 Mbp position on Chr. 4 and one marker located at 61.1 Mbp position on Chr. 9 (Table 2). For all three markers, genotypes with minor alleles showed a decreasing level of susceptibility (allele effect). The effect of different genotype combinations at associated marker loci was evaluated (Table 3). Levels of susceptibility to LB decreased significantly when minor allele (allele B, Table 3) at associated SNP on Chr. 9 was in multiplex dosage, or when minor alleles of associated SNPs on Chrs. 4 and 9 were present in the same genotype (Table 3). Genes annotated for associated SNPs (Table 2) did not have any role in disease resistance. For this reason, narrow areas around the markers were investigated to look for

candidate genes. We found two SolCAP SNP markers, solcap_snp_c2_27719 and solcap_snp_c2_27716, annotated for a receptor protein kinase and a nucleotide binding protein, respectively, localized 60 and 82 Kbp downstream marker solcap_snp_c1_8560. These were associated with LB resistance on Chr. 9 in this study. Receptor protein kinases and nucleotide binding proteins have been recognized as playing a role in pathogen recognition and response.

Conclusions

High-throughput genotyping with SolCAP SNPs provided high genome coverage and therefore high power in mapping. Meanwhile, the decay of LD to r^2=0.2 over 0.3 Mbp (Fig. S2) provided a high resolution in terms of the physical proximity of the marker-trait association, which may allow a further dissection of the genomic region of interest. GWAS showed the complexity of component traits for tuberization and bulking of potato. Nevertheless, this study provides information on genomic hotspot regions of SNP associated with different component traits, and pointed out candidate genes for tuberization photoperiodic response located in the vicinities or within these genomic regions. The same was true for LB resistance, for which candidate genes for pathogen recognition and response were found less than 85 kbp from associated SNPs on Chr 9. The two SNPs identified associated with LB resistance might be used in the future as markers for assisted selection, pyramiding best alleles to further enhance resistance.

This project has contributed information on SolCAP SNP markers associated with component traits for LB resistance, early tuberization, and bulking that, due to their close physical linkage to the trait, can be used for MAS development—even if not all genetically associated SNPs might have a functional relationship with the variation in the trait. On the other hand, three associated SNP markers found annotated for candidate genes for tuberization photoperiodic response in potato, as well as several genes described in the plant model *A. thaliana* controlling photoperiod response that were mapped in the proximities or within the same regions of clusters of associated SNPs, deserve further confirmation and validation as tools for assisting selection. It would be interesting to investigate these and candidate genes identified close to LB resistance-associated SNPs by means of a target resequencing approach in a group of clones from the panel and develop allele-specific assays as has been initiated for the CDF1 candidate gene for early maturity (Klosterman et al. 2013).

Since GWAS-associated SNP jointly explained up to 60% of the phenotypic variation of the traits, MAS developed for these traits could be efficiently used for pre-screening of seedlings before further field evaluation. This would contribute to shifting the mean of the selection population rather than reducing the length of the selection cycle, as could be expected by GS in which all markers are simultaneously fitted for capturing most of the trait heritability.

The thesis "Identification of Chromosomal Regions in Potato Associated with Tuberization Traits Affected by Photoperiod Using SNP Markers" is available under Dalhousie University link https://dalspace.library.dal.ca/handle/10222/71386 and at the project's website https://research.cip.cgiar.org/confluence/display/cclf/Home. The Appendix contains tables of SolCAP SNP identified associated with component traits for early tuberization and bulking under long- and short-day length conditions.

5.4 Output 4. Distribution of Parental Clones for LB Resistance and Fast Tuber Bulking under Subtropical Conditions to Bangladesh and Additional South Asian Countries for Incorporation into Breeding Programs

See Appendix for report of activities concerning this output and for figures or tables denoted with "S."

Prediction accuracies of ≥0.3 (Table S3) can be regarded as sufficiently high to merit implementation of GS (Heffner 2010). These values were attained for tuberization induction, stolon length, bulking ratio at 75 and 90 DAP, and marketable tuber number at 75 DAP. Therefore, they can be predicted by GS and used in breeding for improved earliness and adaptation to warm and long-day length environments. Prediction accuracies can be improved by increasing the size of the training population. But we should not expect a significant improvement due to the low heritability of traits associated with yield components as those evaluated in this study.

As for LB resistance, the high prediction accuracy should be considered important for the endemic areas of Andean countries, where isolates have been shown to belong to the EC-1 clonal lineage of the A1 mating type. However, genotype by environment studies should be taken into account as resistance stability is also associated with pathogen diversity and aggressiveness. This points to the necessity of evaluating the panel or representative samples for resistance in target countries, not only in Peru.

Conclusions

Prediction accuracies for traits associated with early tuberization and bulking under long-day length were within the range needed to apply GS. Considering that GS models capture the entire complement of QTL effects across the genome to produce GEBV, selection accuracy is expected to be better than that provided by traditional breeding values. However, selection accuracy will depend on the phenotyping efficiency. For instance, tuber induction measured by the tuberization response of cuttings taken from plants growing in the field provided one of the highest prediction accuracies. The same was true for tuber bulking measured as the number of medium-sized and large tubers over the total number of tubers (tiny + small + medium + large) multiplied by the ratio of tuberized plants.

As for LB resistance, GS took advantage of QTL with minor effect as well as major genes present in the panel, thanks to good coverage of the genome. The high accuracy of the prediction might be accounted for by the presence of major genes matching avirulence genes present in the local isolate. Therefore, testing the panel in other disease-endemic environments may contribute to develop predictions based on multiple environments and achieve stability of the resistance.

Clones selected for early tuberization and good bulking under warm and long-day length conditions, as well as derived true seed progenies, are already available in one STA (Vietnam) and one Central Asian country (Uzbekistan). A table of the 30 top clones selected for their early tuberization and bulking based on GEBV and their derived true seed progenies is available in the Excel file “Selected early bulking clones and derived progenies” on the project’s website (https://research.cip.cgiar.org/confluence/display/cclf/Home).

Table 1. Positions on the potato genome of 20 genes controlling photoperiod responses in the model plant *Arabidopsis thaliana* after blast alignment using the basic local alignment search tool (BLAST) option of the SPUD DB (Hirsch et al. 2014)

Gene code	Chr. no.	Position start	Position ending	Annotation	Abbr.
PGSC0003DMG400019971	1	531783	536380	FLAVIN-BINDING KELCH	FKF1
PGSC0003DMG400026690	1	44702627	44704151	CONSTANS	CO
PGSC0003DMG400025129	2	25587999	25591776	Zinc finger protein	ZFP
PGSC0003DMG402010056	2	45088022	45092647	CONSTANS	CO
PGSC0003DMG401010056	2	45098374	45101577	CONSTANS 3	CO3
PGSC0003DMG400001330	2	46143997	46147444	Zinc finger protein	ZFP
PGSC0003DMG400001110	3	14265389	14266279	GIGANTEA	GI
Sotub03g010860.1.1	3	15893963	15899148	Flowering locus T	FT
PGSC0003DMG400019528	3	55882563	55885296	Cycling DOF Factor	CDF3
PGSC0003DMG400006387	4	65533306	65538392	CONSTANS	CO
PGSC0003DMG400018408	5	4538879	4541736	Cycling DOF Factor	CDF5
PGSC0003DMG400025414	5	16834678	16838305	CONSTANS-like 15	COL15
PGSC0003DMT400013625	5	35575037	35577321	CONSTANS-like 9	COL9
PGSC0003DMG400001263	5	42736398	42738599	CONSTANS-like	COL
PGSC0003DMG400023365	5	51319127	51320774	Flowering locus T	FT
PGSC0003DMG400033046	6	51598496	51601151	Cycling DOF Factor	CDF
PGSC0003DMG400027475	7	2275709	2277328	CONSTANS	CO
PGSC0003DMG400026311	8	2052882	2054826	CONSTANS	CO
PGSC0003DMG400016180	11	3249173	3251562	Flowering locus T	FT
PGSC0003DMG400042340	12	18944286	18945393	CONSTANS	CO
PGSC0003DMG400018791	12	54722763	54731694	GIGANTEA	GI
PGSC0003DMG400029365	12	58153724	58155118	CONSTANS	CO

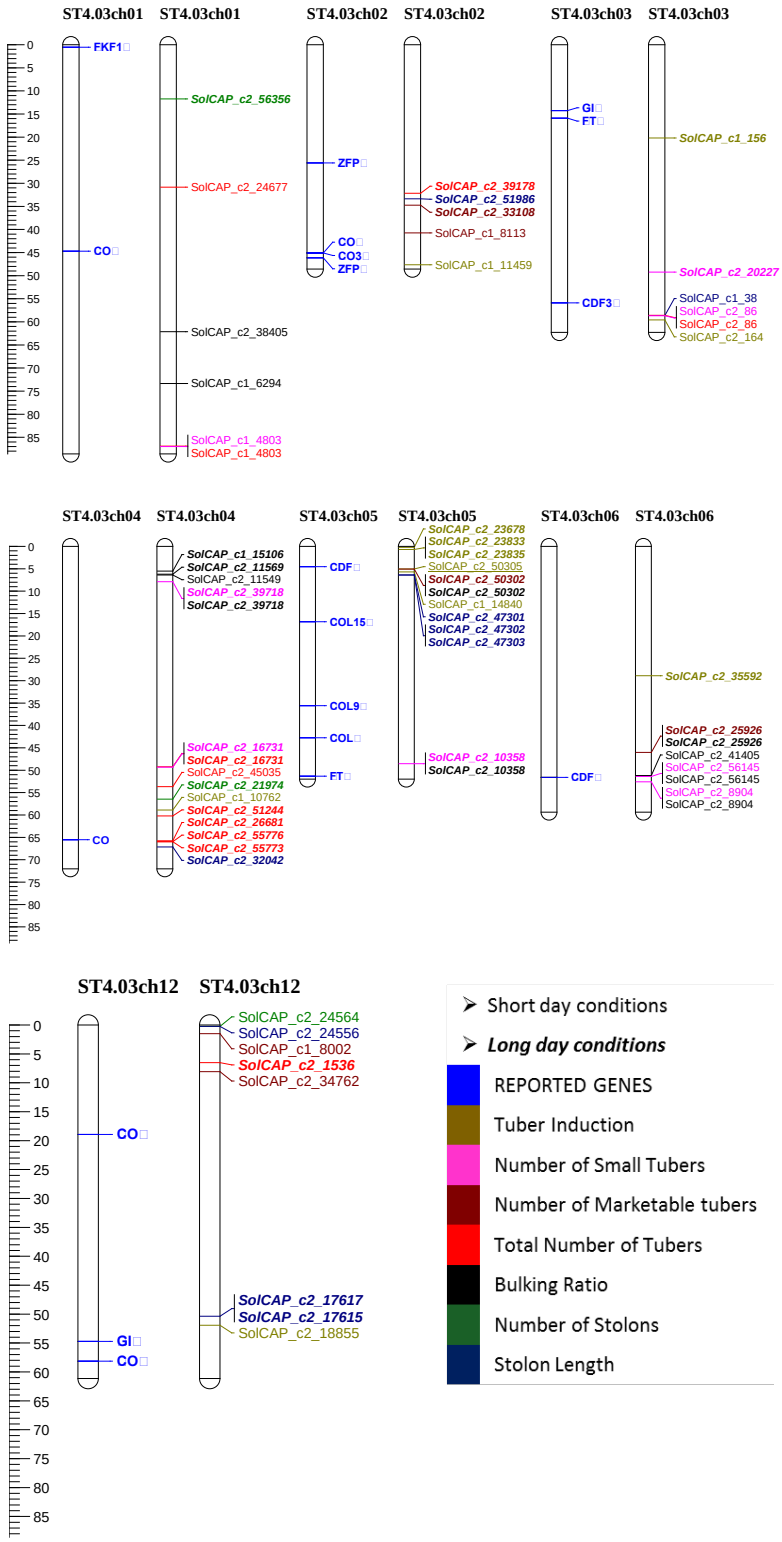

Figure 1. Physical map of the genes underlying photoperiod responses (blue) described in the plant model, *Arabidopsis thaliana*, and the SNP markers associated with tuberization traits responses identified in this study.

Table 2. Genomic position and association statistics of markers significantly associated with LB susceptibility level by Mixed Linear Model

SolCAP ID	Chromosome	Site bp	p Value	Minor Allele Frequency	Allele Effect Estimate	Annotation
solcap_snp_c2_51234	4	60254953	7.40E-05	14.2	-2.15	Cellulose synthase
solcap_snp_c1_15006	4	60255822	7.40E-05	14.2	-2.15	Cellulose synthase
solcap_snp_c1_8560	9	61075452	1.93E-05	11.4	-2.56	Phd finger protein

Table 3. Genotypic combinations at markers associated with levels of LB susceptibility on Chrs. 4 and 9

Chr 4 solcap_snp_c2_51234	Chr 9 solcap_snp_c1_8560	Genotype No.	Mean	Significant Differences
AAAA	AAAA	55	6.1	a
AAAB	AAAA	35	5.4	ab
AABB	AAAA	12	4.7	abc
AAAA	**AAAB**	30	4.3	bcd
AAAB	**AAAB**	20	3.4	cde
AAAA	**AABB**	5	2.0	de
AAAB	**AABB**	3	1.3	e
AABB	**AAAB**	5	1.7	e
AAAB	**ABBB**	1	1.4	-
AABB	**AABB**	1	0.5	-

Significant differences between groups were assayed by Fisher's test.

6. REFERENCES

Bonferroni, C.E. 1936. Teoria statistica delle classi e calcolo delle probabilità. R Istituto Superiore di Scienze Economiche e Commerciali di Firenze 8.

Bonierbale, M., Amorós, W., and Landeo, J. 2003. Improved resistance and quality in potatoes for the tropics. In: Proceedings of the 26th international horticultural congress, August 11–17. Toronto, Canada. *Acta Horticulturae* 619:15–22.

Charrad, M., Ghazzali, N., Boiteau, V., and Niknafs, A. 2014. NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *Journal of Statistical Software* 61 1–36.

Crossa, J. et al. 2010 Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics* 186: 713–724.

Cubillos, A.G., and Plaisted, R.L. 1976. Heterosis for yield in hybrids between *S. tuberosum* ssp. *Tuberosum* and ssp. *Andigena*. *American Potato Journal* 53: 143–150.

D'hoop, B., Paulo, M., Mank, R., Eck, H., and Eeuwijk, F. 2008. Association mapping of quality traits in potato (*Solanum tuberosum L.*). *International Journal of Plant Breeding* 161: 47–60.

De Mendiburu, F., and Simon R. 2015. Agricolae. In Statistical Procedures for Agricultural Research (R package).

Desrousseaux, D., Sandron, F., Siberchicot, A., Cierco-Ayrolles, C., and Mangin, B. 2013. Linkage disequilibrium corrected by the structure and the relatedness (LDcorSV), B. Mangin, ed. (R CRAN).

Elston, R.C. 1963. A weight free index for the purpose of ranking of selection with respect to several traits at a time. *Biometrics* 19(1): 85–97.

Endelman, J.B. 2011. Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. *Plant Genome* 4: 250–255.

Evanno, G., Regnaut, S., and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* 14: 2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x

Ewing, E. E. 1992. Critical photoperiods for tuberization: a screening technique with potato cuttings. *Am Potato J* 55: 43–53.

Felcher, K., Coombs, J.J., Massa, A.N., Hansey, C.N., Hamilton, J.P., Veilleux, R.E., Buell, C.R., and Douches, D.S. 2012. Integration of two diploid potato linkage maps with the potato genome sequence. *PLoS ONE* 7.

Flint-Garcia, S.A., Thornsberry, J.M., and Buckler, E.S.I.V. 2003. Structure of linkage disequilibrium in plants. *Annual Review of Plant Biology* 54: 357–374.

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.

Hackett, C., McLean, K., and Bryan, G.J. 2013. Linkage analysis and QTL mapping using SNP dosage data in a tetraploid potato mapping population. *PLoS ONE* 8.

Hamblin, M.T., Buckler, E.S., and Jannink, J.L. 2011. Population genetics of genomics based crop improvement methods. *Trends Genet* 27: 98–106.

Hirsch, C.D., Hamilton, J., Childs, K., Cepela, J., Crisovan, E., Vaillancourt, B., Hirsch, C.N., Habermann, M., Neal, B., and Buell C.R. 2014. Spud DB: A resource for mining sequences, genotypes, and phenotypes to accelerate potato breeding. *The Plant Genome* 7(1): 1–12.

Kennedy, B.W., Quinton, M., and van Arendonk, J.A. 1992. Estimation of effects of single genes on quantitative traits. *J. Anim. Sci*. 70: 2000–2012.

Kloosterman, B., Gomez, M.D., Oortwijn, M., De Boer, J.M., Kowitwanich, K., Van Eck, H.J., Visser, R.G.F., Bachem, C.W.B., Abelenda, J.A., Prat, S., Horvath, B., and Smaczniak C. 2013. Naturally occurring allele diversity allows potato cultivation in northern latitudes. *Nature* 495: 246–250.

Lipka, A.E., Tian, F., Wang, Q., Peiffer, J., Li M., Bradbury, P.J., Gore, M.A., Buckler, E.S., Zhang, Z. 2012. GAPIT: Genome association and prediction integrated tool. *Bioinformatics* 28: 2397–2399.

Mangin, B., Siberchicot, A., Nicolas, S., Doligez, A., This, P., and Cierco-Ayrolles, C. 2012. Novel measures of linkage disequilibrium that correct the bias due to population structure and relatedness. *Heredity* 108: 285–291.

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

Mihovilovich, E., Sanetomo, R., Hosaka, K., Ordoñez, B., Aponte, M., and Bonierbale, M. 2015. Cytoplasmic diversity in potato breeding: case study from the International Potato Center. *Mol Breeding* 35: 137. DOI 10.1007/s11032-015-0326-1.

Plaisted, R.L .1987. Advances and limitations in the utilization of Neotuberosum in potato breeding. In: GJ Jellis and DE Richardson (eds.), *The production of new potato varieties*. Cambridge, UK: Cambridge University Press, pp. 186–196.

Pritchard, J.K., Stephens, M., and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155: 945.

Simko, I., Costanzo, S., Haynes, K., Christ, B., and Jones, R. 2004. Linkage disequilibrium mapping of a Verticillium dahliae resistance quantitative trait locus in tetraploid potato (*Solanum tuberosum*) through a candidate gene approach. *International Journal of Plant Breeding Research* 108: 217–224.

Snapp, S.S. 1999. Mother and baby trials: A novel trial design being tried out in Malawi. *Target Newsletter of the Southern Africa Soil Fertility Network* 17: 8.

Staaf, J., Vallon-Christersson, J., Lindgren, D., Juliusson, G., Rosenquist, R., Höglund, M., Borg, Å., and Ringnér, M. 2008. Normalization of Illumina Infinium whole- genome SNP data improves copy number estimates and allelic intensity ratios. *BMC Bioinformatics* 9: 409–409.

Visker, M.H.P.W., Keizer, L.C.P., Van Eck, H.J., Jacobsen, E., Colon, L.T., and Struik, P.C. 2003. Can the QTL for late blight resistance on potato chromosome 5 be attributed to foliage maturity type? *Theor Appl Genet* 106: 317–325. DOI 10.1007/s00122-002-1021-2

Voorrips, R., Gort, G., and Vosman, B. 2011. Genotype calling in tetraploid species from bi-allelic marker data using mixture models. *BMC Bioinformatics* 12.

Yu, J., Pressoir, G., Briggs, W.H., Vroh, B.I., and Yamasaki, M., et al. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* 38: 203–208.

Yuen, J.E., and Forbes, G.A. 2009. Estimating the level of susceptibility to Phytophthora infestans in potato genotypes. *Phytopathology* 99: 782–786.

APPENDIX

Output 3, Activity 3.1: Evaluate ("phenotype") panel of breeding lines from CIP for tuberization and bulking-based maturity under natural short-day and extended-day length conditions under field conditions in Peru

- Phenotypic data

The genetic material used in the project consisted of a panel of 162 advanced breeding clones generated by CIP's breeding program. Data were generated in two experimental trials under field conditions at CIP's headquarters station in La Molina (Lima, 12° 3' 0" South) during the warm season of spring. The station is located at 280 masl and is considered a coastal desert environment. The first trial planted in September 2010 comprised 127 breeding clones and three long-day length adapted varieties used as checks. The second trial planted in October 2012 included 35 breeding clones not present in the first trial and 9 long- and short-day length adapted varieties used as checks. Maximum temperatures ranged 22°–26°C during the growing season in 2010 and 25°–29°C during 2011. The design employed in the two experiments consisted of three factors: potato clone, harvest date, and day length. There were two harvest dates, 75 and 90 DAP, and two day lengths, 12 hr (short day) and 16 hr (long day). The short-day length represents natural growing conditions and the long-day length was implemented by using high pressure sodium vapor lamps with a light intensity of 0.9314 µmol m^{-2} s^{-1} to extend the day length from 12 to 16 hr. The seed tubers of each clone were planted in a plot of six plants. Phenotypic data were collected for 7 tuberization-related traits: tuber initiation, number of small tubers, number of marketable tubers, total number of tubers, bulking ratio, stolon number, and stolon length (Table S1). These data were recorded at each of the two harvest dates of 75 and 90 DAP. Tuber initiation, which is an indirect assessment of intensity of tuberization stimulus, was measured using single node cuttings (SNC) (Ewing 1978). SNC were taken from 3 field-grown plants of each genotype under each day-length treatment at 41, 59, and 74 DAP. Bulking ratio was calculated as the ratio of the number of tubers with medium and large sizes over the total number of tubers of all sizes, multiplied by the ratio of tuberized plants, and expressed as a percentage.

Table S1. Measurements taken to assess tuber induction and bulking of genotypes in CIP's potato panel in two experiments Peru

Trait	Type	Scale	DAP	Abbr.
Tuber induction	Ordinal	Minimum 1 (no growth of the buried bud) to 9 (shortened or round or slightly elongated sessile tuber) with increments of 1	41 59 74	TI41 TI59 TI74
No. of small tubers	Continuous	Number of harvested small tubers (TINY + SMALL) per plot	75 90	ST75 ST90
No. of marketable tubers	Continuous	Number of harvested marketable tubers (MEDIUM + LARGE) per plot	75 90	MT75 MT90
Total no. of tubers	Continuous	Number of harvested total tubers (TINY + SMALL+ MEDIUM + LARGE) per plot	75 90	TT75 TT90
Bulking ratio	Continuous	Expressed as a percentage	75 90	BR75 BR90
Stolon no.	Ordinal	Minimum 1 (no stolons) to 9 (countless number of stolons) with increments of 1	75 90	SN75 SN90
Stolon length	Ordinal	Minimum 1 (no stolons) to 9 (extremely long) with increments of 1	75 90	SL75 SL90

The check cultivars were several long-day length-adapted varieties ('Atlantic', 'Desiree', 'Granola', 'Kufri Jyoti', 'Spunta'), a photoperiod-insensitive lowland tropical breeding line (DTO-33) (Ewing 1992), and three short-day length Peruvian varieties: 'Tomasa Tito Condemayta' ('Tomasa'), 'Perricholi', and 'Yungay'.

Phenotypic analysis was conducted to generate adjusted mean values for each clone in each variable of tuberization-related traits using the R-package Agricolae (De Mendiburu and Simon 2015). Adjusted mean values were used for GWAS and GS.

Output 3, Activity 3.2: "Genotype" panel of breeding lines from CIP and determine LD extent and structure of the component populations

- "Genotype" panel of breeding lines

SNP genotyping of the panel of 162 advanced breeding clones and nine check varieties was done with the potato SolCAP SNP array (Felcher et al. 2012), by the Saskatoon Research Centre, AAFC, Saskatoon, Canada. The genotyping was carried out using an Illumina iScan Reader equipped with the Infinium HD Assay Ultra by Gen-Probe. The dosage of each allele in each sample was called using the Genome Studio software after normalization and transformation of the intensities of the fluorescent dyes (red and green) associated with the two SNP alleles retrieved from the equipment reads, to intensity ratio theta values (range from 0 to 1) (Staaf et al. 2008). The information was exported from the Genome Studio software (Illumina) to a .txt file for further analysis.

Four methods able to convert the continuous signal scores (theta values) to five discrete genotype classes found in 4x genotypes (AAAA, AAAB, AABB, ABBB, and BBBB) were used for SNP genotype calling: pre-determined boundaries (SolCAP); mixture models, of FitTetra (Voorrips et al. 2011) and Hackett (Hackett et al. 2013); and NbClust cluster analysis (Charrad et al. 2014). Several genotyping control procedures were performed for all methods to select informative SNPs based on the criteria of removing missing data, identifying polymorphisms of SNPs and defining the range of theta values for each SNP. These control procedures eliminated the SNPs with more than 5% of missing data in the population, and also removed those which were non-informative, monomorphic, or problematic. An ideal SNP for which all possible dosages can be observed is expected to consist of theta scores in five clusters (AAAA, AAAB, AABB, ABBB, BBBB), centered on 0.0, 0.25, 0.5, 0.75, and 1, respectively. An SNP was retained in the set of informative SNP markers if the difference between two clusters was equal to or greater than 0.1. This parameter was established in order to have well defined clusters to finally translate theta values of each SNP into actual genotype scores. To determine which SNP calling method was the best to develop our genotypic data for further studies (i.e., patterns of LD, population structure, genomic relationships, GWAS, GS), the SNP quality and accuracy of genotype calls were compared between methods based on the sharpness of cluster separations and correct assignments in the genotype calling by analyzing several SNP markers associated with a trait.

- LD extent

The extent of LD between loci across the potato genome was quantified by the squared correlation coefficient (r^2) using R-package LDcorSV (Desrousseaux et al. 2013) to enable assessment of the resolution of the marker-trait associations on the chromosomic regions associated with LB resistance and tuberization-related traits identified by GWAS.

This LD measurement was corrected by the kinship relationships of genotyped individuals and the structure of the samples (Mangin et al. 2012). To determine how fast LD decays across the genome, r^2 between paired markers was plotted against the distance between pairs of markers in base pairs.

- Relationship matrix

The calculation of the relatedness (marker-based relationship) among clones was performed by using the function A.mat from the R-package rrBLUP (Endelman 2011) that calculated the additive relationship matrix.

- Population structure

To investigate possible structure in the panel of 171 genotypes, the Bayesian model-based clustering method of Pritchard et al. (2000), implemented in the software Structure v2.1 (http://pritch.bsd.uchicago.edu), was used. The model with admixture and with uncorrelated allele frequencies was applied with the assumed number of populations (K) varying from 1 to 10, three replicate runs per K value, and a burning period length of 20,000 iterations, followed by data collection on 10^6 runs. This method allocates individual genotypes to groups in such a way that Hardy-Weinberg equilibrium and linkage equilibrium are valid within groups, whereas these forms of equilibrium are absent between groups. The ΔK based on the rate of change in the log probability of data between successive K values (Evanno et al. 2005) was used to define the number of populations.

Output 3, Activity 3.3: Analyze phenotypic, SNP, and candidate gene data to locate QTL for response to photoperiod in the panel of breeding lines

- GWAS

GWAS was performed using the R-package rrBLUP (Endelman 2011) with all the qualified SNP markers and all the tuberization and bulking-related traits recorded under long- and short-day length conditions. This was done to unravel the genetic architecture of these traits and identify significantly associated SNP markers in the panel of advanced breeding clones. This analysis was based on a mixed model (Yu et al. 2006) for controlling population structure and relatedness:

$$y = Xb + Qw + S\alpha + Z\mu + \varepsilon;$$

where y represents the phenotype, Xb is the vector of non-genetic fixed effects mean, Qw represents the fixed effect of the population structure, Sα is the fixed effect of the marker, Zμ is the random effect of the covariance between clones, and ε is the vector of residual effects.

To control the Type I error (false positives) across the entire experiment (experiment-wise error rate) when determining the associations, Bonferroni correction (Bonferroni 1936) was considered chromosome-wise for determining the *p*-value threshold. The *p*-value equal to 0.05 was corrected by dividing it by the number of markers per chromosome.

- Candidate gene analysis

This was performed by Canadian partner Dr. De Koeyer of AAFC. Targeted capture and re-sequencing of 15,295 exons, including 3225 exons of genes related to the photoperiod response pathway, was performed. DNA samples from 16 breeding lines were sequenced in this study. SNPs were identified and this information was used to develop 43 high-resolution melting (HRM) assays targeting photoperiod genes.

Three allele-specific and 3 HRM probe assays were developed based on the Kloosterman et al. (2013) publication on the CDF1 gene. One of the CDF1 allele-specific assays designed to distinguish alleles 1 and 2 versus allele 3 was tested on the panel of genotypes grown in the long- and short-day trials in Lima, Peru.

Results

- Phenotypic data

The repeatability of the phenotypic tuberization and bulking-related trait data, shown as heritability (H^2), was calculated from the ratio between genotypic and phenotypic variance (Table S2). Moderate to high heritability (H^2 range = 0.45–0.85) indicated significant genotypic variation in the population for almost all traits except for tuber induction (H^2 range = 0.11–0.33) assessed at 41 DAP. This low heritability indicated the lack of enough genotypic variation for tuber induction in the panel at that evaluation time, and that environmental variation accounted for the greatest portion of the phenotypic variation observed. On the other hand, the heritability values obtained for the other traits, including tuber induction assessed later in the season, implied a good correlation between phenotype and genotype, so that SNP markers with an effect on these traits detected by GWAS will be more reliable.

A table (Table A1. Adjusted mean values) with the adjusted mean values for each trait is available on the project's website https://research.cip.cgiar.org/confluence/display/cclf/Home.

- "Genotype" panel of breeding lines

The FitTetra method proved to be most reliable for genotype calling because miscalling of genotypes or otherwise lack of genotype calling assignments for obvious genotypes were encountered with all methods except this one. FitTetra classified all markers in dosage scores (0, 1, 2, 3, or 4), which reflected the Potato SolCAP SNP array design. From the total of 8,303 SNPs analyzed, 4,738 were identified as informative SNP markers, providing a reasonable genome coverage for our study (Fig. S1). Therefore, genotypes for the 162 advanced breeding clones and nine check varieties were assigned by FitTetra.

A table (Table A2. Matrix of genotypic data) comprising genotypic data of 4,738 SolCAP SNPs for 162 advanced breeding clones and nine check varieties is available on the project's website https://research.cip.cgiar.org/confluence/display/cclf/Home.

- LD extent

By means of 4,738 informative SNP markers obtained from the Potato SolCAP SNP array, we estimated LD in the panel of 171 advanced clones to extend to a distance of 0.3 Mb (~0.3 cM). The scatter plot of r^2 values versus physical distances (bp) between all pairs of SNP markers was used to display the rate at which LD declined across the genome (Fig. S2). The estimation of r^2 in this study was corrected for the bias in LD extent due to population structure and relatedness between individual genotypes by applying the corrected r^2 measure (Margin et al. 2012) implemented in LDcorSV-package.

Table S2. Estimated variance and heritability for phenotypic traits

Trait	Day Length	DAP	V_g	V_e	H^2
Tuber induction	12	41	0.04	0.59	0.11
		59	2.39	1.58	0.75
		74	2.21	0.78	0.85
	16	41	0.05	0.22	0.33
		59	1.17	0.74	0.76
		74	1.59	0.76	0.81
No. of small tubers	12	75	66.11	57.76	0.77
		90	89.36	77.68	0.78
	16	75	199.90	167.50	0.78
		90	335.70	486.20	0.67
No. of marketable tubers	12	75	32.67	88.29	0.53
		90	25.65	95.42	0.45
	16	75	68.05	43.74	0.82
		90	71.30	66.15	0.76
Total no. of tubers	12	75	104.00	144.60	0.68
		90	121.70	131.90	0.73
	16	75	420.90	230.40	0.85
		90	315.20	566.30	0.63
Bulking ratio	12	75	145.80	247.00	0.64
		90	131.30	330.90	0.54
	16	75	268.00	170.40	0.83
		90	272.60	185.00	0.82
Stolon no.	16	75	1.05	1.02	0.76
		90	1.82	1.20	0.82
Stolon length	16	75	1.77	1.33	0.80
		90	2.59	1.47	0.84

V_g represents the variance component for the factor genotype, V_e is the variance component of the residuals and H^2 is the estimated heritability.

This correction, and the fact that our study used a greater number of genome-wide distributed SNPs compared with other studies (Gebhardt et al. 2004; Simko et al. 2004), gave better support to the extent estimated for potato here.

This rate of LD suggests that genome scanning (i.e., use of a great number of markers distributed across the genome, as applied in this study) is appropriate for GWAS to achieve good precision in the localization of QTL. In turn, this will facilitate the identification of closely associated SNP markers for developing reliable MAS.

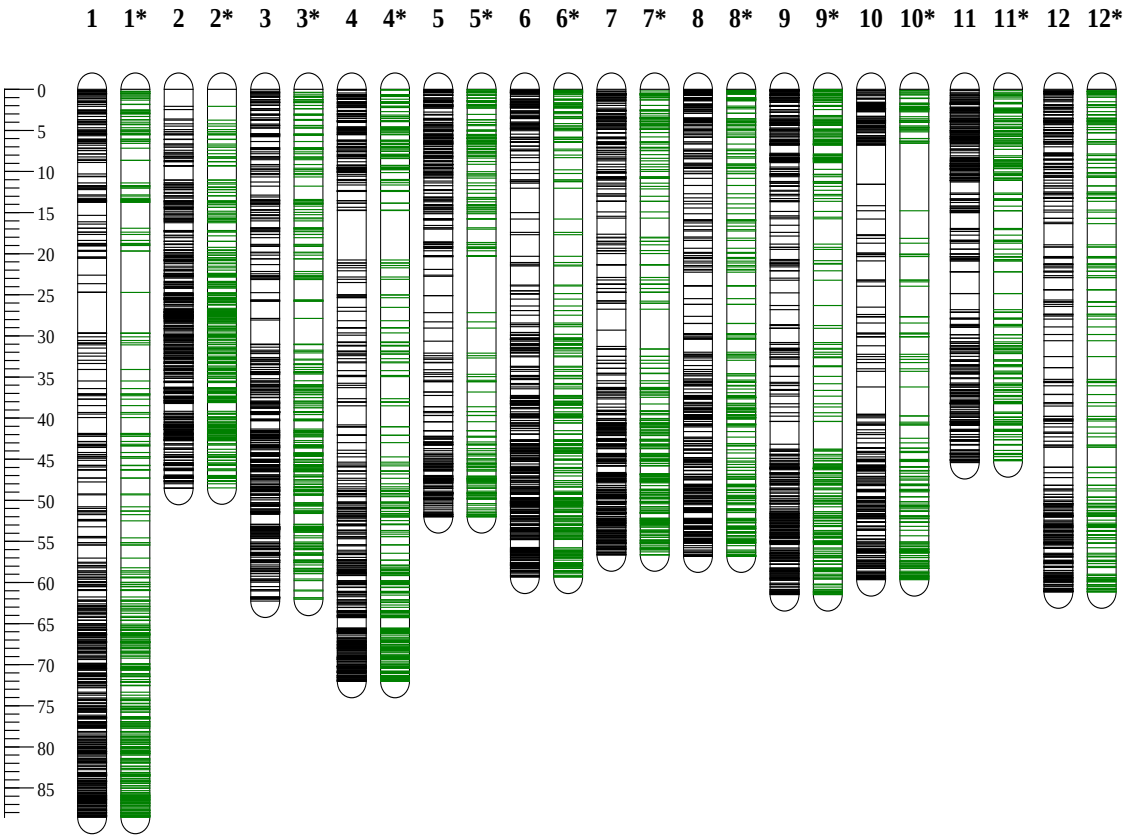

Figure S1. Distribution of 8,303 (black, 1-12) and 4,738 (green, 1*-12*) SNP markers on the 12 potato chromosomes. The scale shows the physical distance in Mb. Map positions are according to Felcher et al. (2012).

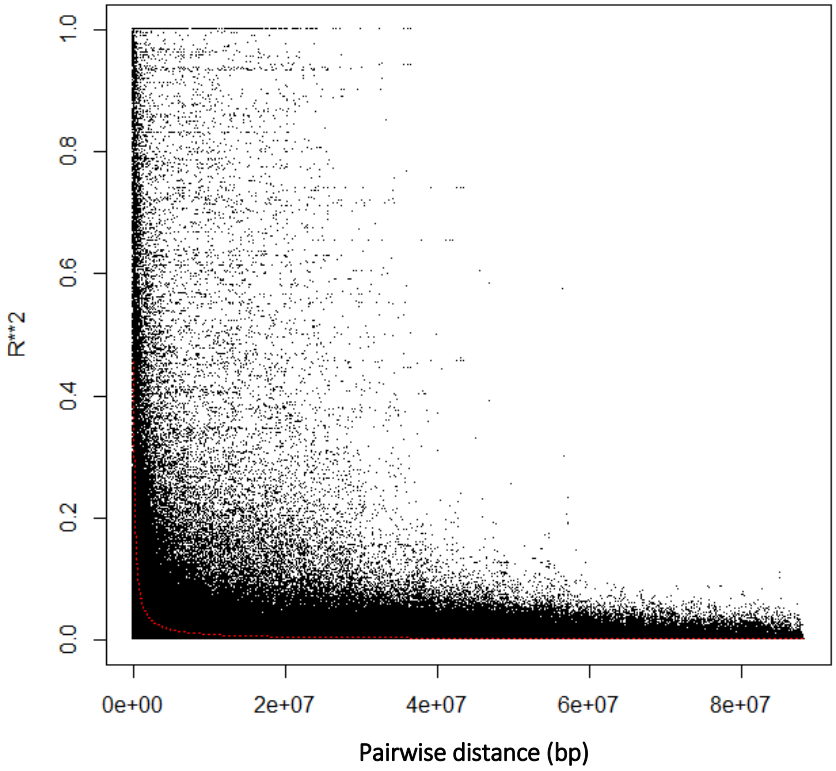

Figure S2. Linkage disequilibrium measure: r^2 plotted vs. the physical map distance (bp) between all pairs of SNP markers assayed in potato panel.

- Population structure and genetic relatedness among individuals

We found K=2 as the most likely number of subpopulations for the panel of 171 advanced clones as determined by Δ*K* (Fig. S3), and many family relationships within sub-populations, such as second and first cousin-level as well as half and full sibs (range 0.15–0.4). The presence of population structure along with diverse levels of familial relatedness indicated that a model that accounts for these complex levels of relatedness should be applied for association analysis

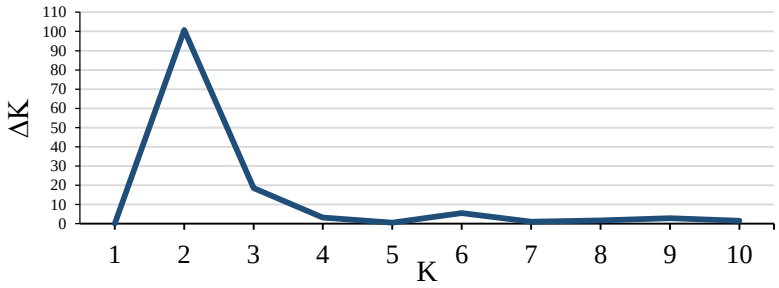

Figure S3. Number of subpopulations (K) in CIP potato panel as determined by ΔK (STRUCTURE v2.3.1 (http://pritch.bsd.uchicago.edu).

- GWAS

Based on the presence of population structure (2 populations) and family relatedness, GWAS was performed using a mixed model (Yu et al. 2006) that includes this information in the variance–covariance model for the genotypic effects. The number of genome-wide significant SNPs detected for tuber induction, number of tubers—small (ST), marketable (MT), and total (TT), bulking ratio and stolon number and length were 22, 41 (ST=14, MT=11, TT=16), 16, 10, and 16, respectively. SNP-trait association signals for tuber induction were identified on Chrs. 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, and 12; for tuber number on Chrs. 1 (TT, ST), 2 (TT, MT), 3 (TT, ST), 4 (TT), 5 (MT ST), 6 (ST), 7 (TT, ST), 8 (ST), 9 (TT, MT), 10 (MT, ST), 11 (TT, MT, ST), and 12 (TT, MT); for bulking ratio on Chrs. 1, 4, 5, 6, 7, and 11; for stolon number on Chrs. 1, 4, 7, 8, 9, 10, and 12; and for stolon length Chrs. 2, 3, 4, 5, 8, 9, 10, and 12 (Fig. S4).

There were SNPs associated with more than one tuberization component trait or showing consistency over time and/or across day-length treatments. These were SolCAP_c2_50302 at 5 Mbp position on Chr. 5 found associated with tuber bulking ratio and marketable tuber number at 90 DAP under long-day length; SolCAP_c2_10358 at 48.5 Mbp position on Chr. 5 associated with tuber bulking ratio and small tuber number at 90 DAP under long-day length; SolCAP_c2_25926 at 46 Mbp position on Chr. 6 associated with bulking ratio and marketable tuber number at 75 DAP under long-day length; SolCAP_c2_56145 and SolCAP_c2_8904 at 51.4 and 52.6 Mbp, respectively, also on Chr. 6 associated with bulking ratio and small tuber number at 75 and 90 DAP under short-day length; SolCAP_c2_33489 at 45.1 Mbp position on Chr. 7 associated with tuber induction at 59 and 74 DAP under long-day length and also with stolon number at 90 DAP under short-day length; SolCAP_c1_7980 at 47.6 Mbp position and SolCAP_c2_18684 at 55.3 Mbp position, both on Chr. 7 also, associated with small and total tuber number at 75 DAP under long-day length and bulking ratio and small tuber number at 75 DAP under short-day length, respectively; SolCAP_c2_13515 at 48.6 Mbp position on Chr. 9 associated with stolon number and length at 75 and 90 DAP, respectively, under long-day length; SolCAP_c2_27806 and SolCAP_c2_27808 at 50.6 Mbp position on Chr. 10, both associated with stolon number and length at 90 DAP under long-day length; and SolCAP_c2_13355 at 0.44 Mbp position on Chr. 11 associated with small and total tuber number at 90 DAP under long-

day length. SolCAP_c2_50305 at 5 Mbp position on Chr. 5, associated with tuber induction at 59 DAP, was the only SNP that show consistency across day length.

- Analyze candidate gene data to locate QTL for response to photoperiod in the panel of breeding lines

A total of 49 bi-allelic SNPs were detected by analyzing 43 HRM assays targeting photoperiod genes developed from SNP markers identified in 3 of the most important genes—Flowering Locus T (FT), Cycling DOF Factor 1 (CDF1), and Mother of FT (MFT)—in the CIP breeding clones. Development of allele-specific assays for these photoperiod genes and their analysis in the panel of breeding lines have not yet been performed.

One of the HRM-based CDF1 allele-specific assays developed on the CDF1 gene (Klosterman et al. 2013) differentiated the allele reported to be associated with early maturity from other alleles in the CIP panel. This assay also identified additional alleles at this locus in our study.

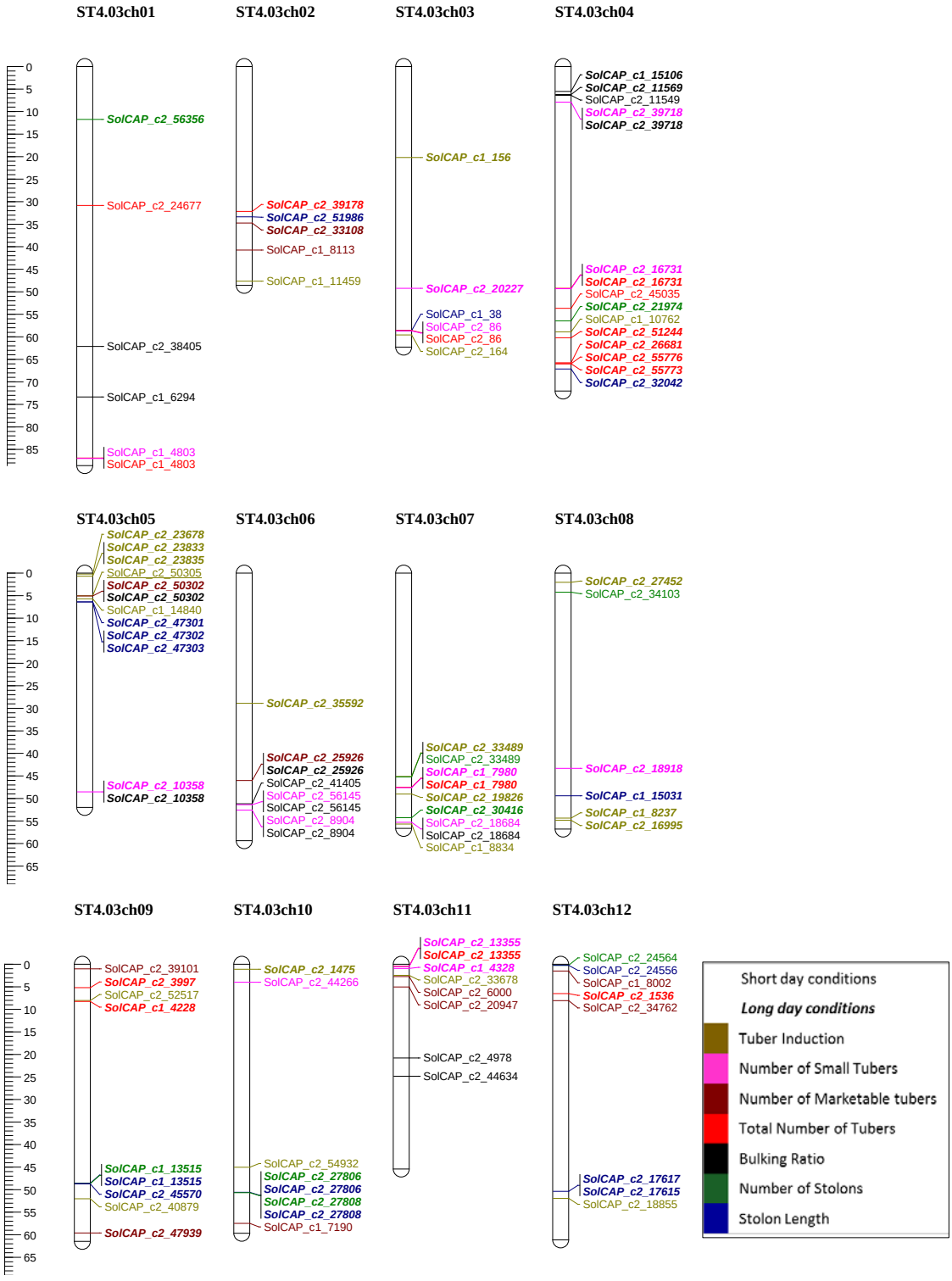

Figure S4. Physical map of 84 SNP markers associated with tuberization-related traits in CIP's potato panel.

Output 3, Activity 3.4: Analyze phenotypic, SNP, and candidate gene data to locate QTL for LB resistance in the germplasm panel

- Phenotypic data

CIP's corporate database of standard evaluation trials provided historical data from replicated LB-resistance trials between 2005 and 2011. This, and data from an additional trial conducted in 2012 in which CIP's complete panel of 171 clones was exposed in the same location during 2012, were used in this study. Disease severity was recorded visually in all trials, as percentage of foliage affected at 7- to 8-day intervals until the susceptible control cv. 'Tomasa' reached 100% infection. Area under the disease progress curve (AUDPC) values calculated from percentage of foliage affected over time, as appeared in CIP's corporate database, were transformed to the susceptibility scale. This was done using the moderately susceptible control 'Yungay' as a reference genotype with an assigned scale value of 6 based on its performance relative to highly susceptible 'Tomasa' that was present in all trials (as was 'Yungay'). This was also performed with data recorded from the field trial conducted under this project. The formula for transformation was sAUDPCx = (AUDPCx/AUDPCYungay) × 6, where sAUDPCx is the scale value of the genotype in question, AUDPCx is the AUDPC of the genotype in question, and AUDPCYungay is the AUDPC of 'Yungay' (Yuen and Forbes 2009). Phenotypic analysis was conducted to generate adjusted mean values of LB susceptibility level for each clone using the R-package Agricolae (De Mendiburu and Simon 2015). Adjusted mean values were used for GWAS and GS.

- GWAS

The unified mixed linear model (MLM) approach that accounts for multiple degrees of relatedness (population structure and cryptic relationships) as aforementioned was fitted using GAPIT software (Lipka et al. 2012). This software uses computationally efficient and powerful methods, such as EMMAX and CMLM. In MLM model equation, each SNP was tested in turn using a t-test (H0: No additive association between the SNP and trait), and *p*-values were obtained. Significant marker trait associations were defined using the Bonferroni correction (Bonferroni 1936) by dividing the 0.05 *p*-value by the number of markers per chromosome.

Results

- GWAS

See Achievements by Outputs above.

Output 4, Activity 4.1: Estimate model parameters required for applying genomic selection in training population

Historical data and an additional field trial conducted within the project activities were used in this study to evaluate feasibility of applying GS for LB resistance in CIP's breeding population. Likewise, phenotypic data generated from field experiments under long-day length conditions for component traits associated with tuber initiation and bulking maturity were used to identify traits for which GS can be applied to increased accuracy of selection and accelerate selection cycles. All traits as aforementioned for GWAS were evaluated in the panel of 162 advanced breeding clones and nine check varieties. These phenotypic data were used along with genotypic data generated using the Potato SolCAP SNP Array to estimate model parameters, which were used subsequently to calculate GEBVs for selection of candidate breeding clones.

Prediction models for GS were constructed using a training population of 130 breeding lines. For prediction of future outcomes, we used a set of 41 breeding lines (testing set) that was evaluated in a different year but in the same location and season in which the training population was evaluated. Models that take into account SNP markers (M) or genomic relationship matrix (G) information for constructing predictions, and those that differ in the way that markers are incorporated—that is, parametric (Bayesian Ridge Regression and Bayesian Lasso and semi-parametric (RKHS) models—were evaluated. An additional model in which a Kernel matrix is used instead of the G matrix in the RKHS model was also tested. Here the Kernel matrix is a (130 × 130) matrix whose entries are evaluated in the squared-Euclidean distance between genotypes. Predictive ability was estimated as the correlation between phenotypic outcomes and predictions for the 41 clones used as testing population.

Results

The Bayesian RKHS using a Kernel matrix (i.e., a Gaussian Kernel whose entries are evaluated in the squared-Euclidean distance between genotypes (RKHS_Eucl) outperformed other models in its predictive ability for almost all traits, as indicated by the estimated correlations between phenotypic outcomes and predictions. In contrast, the Bayesian Lasso model performed the best for stolon number and Bayesian Ridge Regression for LB resistance (Table S3).

Table S3. Prediction accuracy for genomic selection models assessed for five traits recorded in a training population of 130 breeding lines

Trait	BRR	BL	RKHS	RKHS_EU
Tuber induction	0.41	0.39	0.38	0.43
Stolon Length	0.26	0.25	0.33	0.32
Stolon number	0.23	0.3	0.12	0.25
Bulking 75 DAP	0.38	0.36	0.36	0.39
Bulking 90 DAP	0.43	0.44	0.42	0.44
Marketable tuber number 75 DAP	0.26	0.27	0.27	0.28
Late blight resistance	0.67	0.66	0.66	-

BRR = Bayesian Ridge Regression, BL = Bayesian Lasso

Output 4, Activity 4.3: Select candidate progenitors based on GEBV

A non-molecular based- or otherwise traditional selection index (i.e., Elston index was applied to rank breeding lines' performance using their GEBV).

Results

See Achievements by Outputs above.