Identification of Chromosomal Locations in Potatoes Associated with Tuberization Related Traits Affected by Photoperiod Using SNP Markers

by

Maria V. Caraza Salas

Submitted in partial fulfilment of the requirements
for the degree of Master of Science

at

Dalhousie University
Halifax, Nova Scotia
March 2016

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Developing potato varieties adapted to diverse environmental conditions requires an understanding of tuberization related traits. Many of these traits are affected by photoperiod. This study identified genetic markers associated with tuberization related traits. One hundred seventy one tetraploid breeding clones, developed by the International Potato Center, were used. Their tuberization related traits were evaluated under short (12 hours) and long (16 hours) photoperiod exposures, either at 75 or 90 days after planting. Clones were genotyped with the Potato SolCAP SNP array and 4,738 informative SNPs were analyzed for population structure, linkage disequilibrium, and for identifying associations between SNPs and tuberization related traits. In total, 84 significant markers were identified. Genome wide association analysis identified candidate markers for variety development through marker assisted selection. The 10 best clones for each tuberization related trait, based on their genome estimated breeding values, were identified for use in future breeding programs.

LIST OF ABBREVIATIONS USED

BLUP	Best Linear Unbiased Prediction
BR	Bulking Ratio
BW	Bacteria Wilt
CDF	Cycling DOF Factor
Chr	Chromosome
CIP	International Potato Center
cM	Centimorgan
CO	CONSTANS
COL	CONSTANS-like
CRY	Cryptochrome
DAP	Days After Planting
DNA	Deoxyribonucleic Acid
FAO	Food and Agriculture Organization of the United Nations
FT	Flowering Locus T
GEBV	Genome Estimated Breeding Values
GI	GIGANTEA
GS	Genomic Selection
GWAS	Genome Wide Association Analysis
LD	Linkage Disequilibrium
LTVR	Lowland Tropics Virus Resistance
MAF	Minor Allele Frequency
Mb	Mega base pairs
MT	Number of Marketable Tubers
PC	Principal Component
PCA	Principal Component Analysis
PHY	Phytochrome
QTL	Quantitative Trait Loci
RFLP	Restriction Fragment Length Polymorphism
RR	Ridge Regression
SL	Stolon Length
ST	Number of Small Tubers
SN	Stolon Number
SNC	Single Node Cuttings
SNP	Single Nucleotide Polymorphism
SolCAP	Solanaceae Coordinated Agricultural Project
TI	Tuber Induction
TT	Number of Total Tubers
ZFP	Zinc finger protein

ACKNOWLEDGMENTS

CHAPTER 1: INTRODUCTION

An understanding of tuberization traits is critical for developing potato varieties that can adapt broadly to the diverse and changing environmental conditions. Potatoes can be grown in a wide range of climatic conditions, from below sea level to an altitude of over 4000 meters and in over 150 countries around the world (International Potato Center, 2009). The potato crop is the most important vegetable crop according to its production and consumption worldwide (FAOSTAT, 2013). The edible part of potatoes is the underground tuber that represents a rich nutrition for the increasing world population. The meteorological elements affecting the growth, development, production and quality of potatoes in a specific area are mainly air and soil temperatures, photoperiod and crop water use or evapotranspiration (Pereira and Shock, 2006). Photoperiod, or day length, is the interval in 24 hours period during which an organism is exposed to light. It controls several developmental responses in plants by affecting plant growth, development, flowering, plant maturity and tuberization (Jackson, 2009).

Potato tuberization is short-day dependent, meaning that potatoes make fewer or no tubers when the day length is longer than 12 hours (Kloosterman et al., 2013). The International Potato Center (CIP) has developed several potato populations that possess traits such as disease resistance, relative earliness and adaptation to warm, arid conditions in short day environments (Bonierbale et al., 2003). These populations are currently valuable for studying traits such as adaptation to long day environments, typical in countries like Canada. Nevertheless, little is known about the genetic factors controlling potato photoperiodic responses and the genes associated with it (Kloosterman et al., 2013).

1

In order to satisfy the need to improve the adaptation of potatoes to diverse environments around the world, the aim of this study was to identify chromosomal locations associated with tuberization related traits affected by photoperiod. In order to do so, the first objective was to examine the tuberization related traits under contrasting day lengths; secondly, to assess the genetic diversity and genetic relationships using the Potato SolCAP SNP array to finally identify the chromosomal locations associated. Finally, this study aimed at selecting the best parental clones based on genome estimated breeding values. These goals were achieved by using a collection of tetraploid clones generated at CIP to evaluate several tuber traits of the potato plants under long and short photoperiod exposures, in field conditions. Then these clones were genotyped using the Potato SolCAP SNP array that comprises 8,303 single nucleotide polymorphism (SNP) markers. These SNP markers provided the genome with higher coverage (Felcher et al., 2012). Based on the outcome, the population structure, linkage disequilibrium (LD), relatedness and associations with tuberization related traits affected by photoperiod were studied. The association analyses identified a total of 84 SNP markers that are related to tuber induction, number of tubers, bulking ratio and stolon number and length. These association analyses provided important information about the chromosomal locations and potential genes involved in tuberization related traits. By combining genomic predictions and phenotypic data, this study also identified clones with the best predicted breeding values for photoperiod responses. These clones may be used for further breeding programs. Tuberization related traits evaluated in this study were demonstrated to have high accuracies on their predictions; this was important for the effective selection of clones that have the best performance under short and long day conditions.

2

CHAPTER 2: LITERATURE REVIEW

2.1   Potato breeding

The cultivated potato (*Solanum tuberosum* L.) is the world's fourth most important crop after maize, rice and wheat (FAOSTAT, 2013). More than a billion people worldwide eat potatoes as a staple food, and there is a total global crop production of more than 300 million metric tons per year. Potatoes are grown on 19.5 million hectares of land, in 158 countries from latitudes 65°N to 50°S, and at altitudes from under sea level to 4000 meters above sea level (FAOSTAT, 2013). They are a summer crop in the tropical highlands of Bolivia, Peru and Mexico, as well as in the temperate regions of the world such as North America, Europe, northern China and Australia. Potatoes are also cultivated all year round in parts of southern China and Brazil (Storey, 2007).

Potatoes are an important crop for food security to deal with population growth and increased global hunger because they are energy-rich and nutritious, easy to grow on small plots, cheap to purchase, and ready to cook without expensive processing. Over half of the global potatoes are produced in Asia, Africa and Latin America, where it is a major carbohydrate (starch) supplier in the diets of hundreds of millions of people (Storey, 2007). In addition, one hectare of potatoes can yield two to four times the food quantity of grain crops and they produce more food per unit of water than any other major crops (International Potato Center, 2010). It also provides significant amounts of proteins, with a good amino acid balance, vitamins C, B6 and B1, folate, the minerals potassium, phosphorus, calcium, and magnesium and micronutrients iron and zinc (Storey, 2007)

Since 1960, the growth in acreage for potato production has rapidly increased in developing countries. Therefore, the generation of potato varieties, adapted to a diverse range of climates would enable farmers to profitably add potatoes into mixed cropping systems. In addition, since potatoes are being promoted by the Food and Agricultural Organization (FAO) as a famine reducing crop, there is an increasing demand for land for potato cultivation. This trend will also require good varieties with high-yield and disease resistance (CGIAR, 2000).

2.1.1   Origin of potatoes

Potatoes were developed by pre-Colombian cultivators from Andean and Chilean landraces between 8000 and 5000 BC (National Research Council (U.S.) and Advisory Committee on Technology Innovation, 1989). They were first domesticated in the highlands of southern Peru (Spooner et al., 2005). These Andean landraces show tremendous morphological and genetic diversity. These potato landraces are very diverse in tuber color and shapes, and leaf, floral, and growth habit variations. Genetically, cultivated potatoes have a variety of ploidy levels, ranging from diploid ($2n=2x=24$), to pentaploid ($2n=5x=60$) (Spooner and Bryan, 2005). Therefore, they are valuable genetic resources in breeding programs for disease resistance, environmental tolerance and other agronomic traits of interest (Ochoa and Center, 1990).

The first record of cultivated potatoes outside South America was from Gran Canaria in the Canary Islands to Belgium in 1562 (Hawkes and Francisco-Ortega, 1993). Since then, the growing of potatoes spread north-eastwards across Europe, and became adapted to the long summer days of northern Europe. The early European potatoes are

assumed to have been first selected in 17th century from Chilean collections because they were better adapted to the European conditions (Bradshaw and Ramsay, 2009). However, there are other potatoes selected from South America that are registered to be sent from Cuzco in Peru to Spain in mid-16th century (Glendinning, 1983). Hence, it may be assumed that the earliest introductions of cultivated potatoes to Europe came from both the Andes and coastal Chile (Spooner and Bryan, 2005) . These varieties later spread out in European countries and later to North America in the 19th century (Bradshaw and Ramsay, 2009).

2.1.2   Potato crop

The potato (*Solanum tuberosum* L) is an annual, herbaceous, dicotyledonous and vegetatively propagated plant. It can also grow as a perennial in selected environments and be propagated through botanical seeds, called true seeds (Sarkar, 2008). The life cycle of vegetatively propagated potato plants can be divided into five stages: sprout development, plant establishment, tuber initiation, tuber bulking, and tuber maturation. Stage I, sprout development, occurs once the seed tubers have broken dormancy and start to sprout. Stage II, plant establishment, refers to the growth of stems, leaves and roots from sprouts until the initiation of tubers occurs. Stage III, tuber initiation, is when the tips of stolons (underground modified stems) hook and begin to swell, resulting in initiation of tubers. Stage IV, tuber bulking, is the critical growth period for tuber size enlargement; this stage refers to the linear tuber growth phase. Finally, Stage V, tuber maturation, is where the tuber dry matter content reaches its maximum level and tuber skin is set. The last three stages are a part of the process called tuberization (Johnson and Powelson, 2008). Tuberization is a coordinated, morphological process occurring on stolons. Timing of these

growth stages vary, depending on genetic and environmental factors, such as temperature, soil type, moisture, cultivar, and geographic location (Johnson and Powelson, 2008). The potato growing season is typically 90-120 days, but can be as short as 75 days in the lowland subtropics (where temperature of the season is high) and as long as 180 days in the high Andes (where temperature of the season is low) (Bradshaw and Ramsay, 2009).

2.1.2.1 Effect of photoperiod on tuber development

The length of a day, time from the sunrise to the sunset, is called photoperiod, or day length. This period varies with seasons on the planet's surface, depending on the latitude. The seasons occur in predictable ways every year and living organisms always respond to these seasonal changes accordingly (Moore, 1920). Using day length as a cue, species may change their growth, physiological stage and development, in accordance with sensed photoperiod and changes in climate. In potatoes, timing of reproduction, including flowering and tuberization, is also controlled by photoperiod (Thomas and Vince-Prue, 1997).

Photoperiod determines the amount of light and darkness a plant is exposed to. The term "short day" indicates that the day length is 12 hours or less of light; the term "long day" indicates that there are more than 12 hours of light in the day. Plants can be divided into three major groups on the basis of their flowering response to these day-lengths. Group 1 includes plants that flower when the day length is longer than 12 hours in summer and they are called long day plants. An example of this group of plants is spring wheat. Group 2 comprises plants that flower in the fall when the day length is shorter than 12 hours and they are called short day plants. This group includes crops, such as potato and rice, which

6

originated from tropical regions with naturally short day lengths. Group 3 includes day-neutral plants in which day length does not affect floral transition. This group comprises tomatoes, corn and some day-neutral strawberries (Abelenda et al., 2014). Day length is sensed by the leaf tissues, which produce a mobile signal that is transported to other parts of the plant through signal transduction pathways. In the case of potatoes, the signal is transported to the shoot apex or underground stems, to induce a flowering transition or a tuberization transition, respectively. These important aspects of plant development are regulated by seasonal fluctuations in day length and have a direct impact on the formation of tubers in potatoes (Haverkort, 2007).

While photoperiod affects all the traits in plant growth and development, it has a significant impact on potato tuberization. It has been reported that, compared with short photoperiods, long photoperiods increase stem elongation and stem weight, as well as the number of leaves (Haverkort, 2007). Tuberization starts with the stolon-to-tuber transition. This process is photoperiod-dependent, since it is induced by short days and low temperatures. Tuberization in potatoes is also accompanied by the flowering process; therefore, it is highly regulated and related to the flowering regulatory pathways (Abelenda et al., 2011, 2014; Driver et al., 1943). Previous research found that tuber yield is promoted by short days (such as 8 hr of light) with low night temperatures (10-12 °C), while long days (16 hr or longer of light) with high night temperatures (15-19 °C) result in fewer tubers formed (Abelenda et al., 2014; Gregory, 1956). Tuberization in the wild potato species is sensitive to day length since they require short days; however, the effect of photoperiod on tuberization in the cultivated potatoes is less evident, but not absent (Haverkort, 2007).

The photoperiod response in plants involves three main parts: photoreceptors, a circadian clock and resulting signaling transduction pathways from the clock, specific to either flowering or tuberization (Simpson, 2003). Photoreceptors are protein molecules that absorb light signals and are located in the leaves. The circadian clock is the biological clock in plants that controls the rhythmicity in the behavior of the plant, controlling mainly leaf movements, cell elongation rates and stomatal aperture reaction to light. The process of photoperiod response in plants starts first when the light is sensed by expanded leaves, absorbed by phytochromes (PHY) A to E and cryptochromes (CRY) 1 and 2 (located in the chloroplasts) (Martínez-García et al., 2002). The circadian clock recognizes the duration of day and night; it allows the plant to generate two main signals, florigen (during the day time) and tuberigen (during the night time), in relation to flowering and tuberization. The first mobile signal, florigen, is transported from the leaves to the vegetative shoot apex, to induce flowering, while the second signal, tuberigen, is transported to the underground stolon tips, to induce tuber formation. These two signals are transported via the phloem system (Jackson, 1999). Furthermore, they are regulated in separate pathways (flowering and tuberization), with different insights, suggesting that related photoperiodic pathways may control their synthesis (Navarro et al., 2011).

Short photoperiods promote tuber formation in all potato cultivars, especially in wild *Andean* landraces, such as *S. tuberosum* spp. *andigena* and *S. demissum*, that are strictly dependent on short days for tuberization (Abelenda et al., 2014). In contrast with cultivated potatoes, many wild species are classified as late tuberizing, especially under the long summer days typical of North American or European climates (Bradeen and Kole, 2011). Nevertheless, these modern potato genotypes are widely cultivated around the world

under different climatic conditions, resulting in a better adaptation to long days for tuber formation transition.

2.1.2.2 Genetic factors associated with photoperiod responses

Much research has focused on the transition from vegetative growth to the flowering stage (reproductive growth) in plants. This transition is controlled by endogenous signals and environmental factors, such as light. The molecular mechanism of the photoperiodic control of flowering has been explained in the plant model *Arabidopsis thaliana*, which is a long day plant (Cheng and Wang, 2005; Lagercrantz, 2009; Martínez-García et al., 2002; Song et al., 2012; Turck et al., 2008). However, not much is known about the molecular mechanisms of short day plants (Kojima et al., 2002).

Tuberization and flowering processes share a number of common regulatory genes (Rodríguez-Falcón et al., 2006). These genes are mainly members of the FLOWERING LOCUS T (FT), CONSTANS (CO), and CYCLING DOF FACTOR (CDF) gene families. A study using *Arabidopsis* showed that the FT genes play a central role in flowering (Cheng and Wang, 2005; Turck et al., 2008), as the florigenic signal, FT protein, has been found in the phloem for the control of flowering time. Orthologues of FT have been shown to regulate flowering in diverse plant species, including short-day plant rice (Kojima et al., 2002), day-length neutral tomatoes (Navarro et al., 2011), and biannual sugar beets (Pin et al., 2010). In potatoes, the homolog of FT (StSP6A) has recently been demonstrated to be the mobile tuberigen involved in tuber formation (Navarro et al., 2011).

The CO gene family functions along with CONSTANS-like (COL) genes to play a role in the photoperiodic flowering pathways. Research has indicated that members of this

family are clock-regulated genes in *Arabidopsis thaliana* and *Chrysanthemum lavandulifolium* (Fu et al., 2015; Salazar et al., 2009). They produce a protein in the leaves that functions as an external coincidence sensor for light. CONSTANS protein is a zinc finger transcriptional regulator that accelerates flowering, through the induction of FT gene expression in the leaf vasculature, under light (Song et al., 2012). Potato homologs of CO are called StCO and are also involved in tuberization. These genes work along with members of the CDF gene family. Members of the CDF gene family, as well as their homologs, Dof transcription factors, are involved in the repression of CO transcription, binding to the CO promoter during the morning. Dof proteins are members of a major family of plant transcription factors; when associated with light reception, they play an important role in CDF gene expression (Yanagisawa, 2002). CDF genes are negatively regulated (degraded) by the circadian clock-associated FLAVIN-BINDING KELCH and GIGANTEA (Fu et al., 2015). The CDF gene family was recently reported in potatoes (StCDF) to be a group of important regulators for plant life cycle length and tuber initiation development (Kloosterman et al., 2013). Sequence analysis of the StCDF1 gene showed the presence of three allelic variants in potatoes: StCDF1.1, StCDF1.2 and StCDF1.3 (Kloosterman et al., 2013). With the availability of the published potato genome sequence (Xu et al., 2011), these and other genes reported mainly in *Arabidopsis*, can be studied for their functions related to flowering and tuberization.

2.1.3 Potato breeding interest

The main breeding targets in potato variety development are yield, tuber quality and disease resistance. Additionally, traits such as compact plants, large tubers, and smooth

tuber shape are also of interest (Douches et al., 1996). Other external quality traits required for fresh market and processed potatoes include tuber size and shape, eye depth, skin color, and lack of blemishes due to bruising and diseases. Internal quality traits include dry matter content, nutritional quality, flavor, starch quantity and quality, and lack of defects such as hollow heart and internal necrosis (Jansky, 2009).

An important germplasm resource for breeding programs are the wild potatoes grown in a large range of altitudes and adapted to a greater variety of habitats around the world (Hawkes, 1990). These wild species (*S. tuberosum* spp. *andigena*) contain genes which encode numerous traits not found in cultivated *Solanum tuberosum* and they represent an especially rich source of disease-resistance and tuber-quality (Spooner and Bamberg, 1994). The wild and cultivated selections of potatoes are extensively collected in gene banks throughout the world. One of the most important potato gene banks is located at the International Potato Center (CIP), in Lima, Peru. Potato clones coming from the collection generated at CIP are part of a valuable resource of germplasm available for potato breeding purposes.

2.2   Marker technology

DNA markers can be used to identify individual organisms with specific traits. These markers then can be used for breeding to improve cultivars. They can also be used to track loci and genome regions in the offspring. In addition, DNA markers allow for the detection of specific sequence differences between two or more individuals. The significant expansion of DNA sequence databases has opened the opportunity for the identification of sequence variations through single nucleotide polymorphism (SNP) markers. SNP markers

are individual nucleotide base differences between DNA sequences of the individuals and can represent differences between individuals or within populations. SNP markers occur at varying frequencies, depending on the species and genome region (Langridge and Chalmers, 2005). The sequencing of genomes of various crops has allowed for a dramatic increase in the number of SNP markers available; as a consequence, it is possible now to conduct genome wide studies using a large number of SNP markers (over several thousands) that covers the chromosomes at higher resolution.

2.2.1 Genetic markers used in potatoes

The slow progress of genetic linkage analysis and breeding for cultivated potatoes is the result of their high heterozygosity, their tetraploid nature and their lack of useful genetic markers (Bonierbale et al., 1988; De Koeyer et al., 2011). The potato genome is 844 Mb (Xu et al., 2011) and has 12 chromosomes. In order to detect natural DNA variations, several molecular tools have been developed over the last 35 years, starting with hybridization-based analysis of restriction fragment length polymorphisms (RFLPs) (Botstein et al., 1980). Thereafter, the research advanced to PCR-based marker systems, such as amplified fragment length polymorphism (AFLP), simple sequence repeats (SSRs) and lately, the analysis of SNP markers (Gebhardt, 2005). Another type of DNA variations characterized by DNA markers are Indels, they are insertion or deletion events in chromosomes. These DNA variations provide information in both length and sequence polymorphisms because they are caused by the addition or removal of DNA sequences.

The most direct and informative method for detecting DNA variation is sequence analysis. New sequencing technologies allow for the parallel scoring of thousands of SNPs

in a clone, which facilitates genomic studies for all organisms, including potatoes (Bryan, 2011). SNPs and indels are valuable molecular markers, due to their abundance and relative stability in the genome, and can be applied to identify genes underlying important traits (Edwards and McCouch, 2007). However, the molecular pinpointing of genes to tag potato traits requires an increase in genetic resolution. This can be achieved by analyzing a large number of SNP markers that cover most of the genome. In the potato genome, SNPs are found with high frequency every 15-21 bp (Bryan, 2007) and can be detected in amplicons generated from genomic DNA of heterozygous individuals. Moreover, SNP markers also allow for the estimation of the allele dosage in tetraploid individuals (Rickert et al., 2003).

One major concern about using markers in potatoes is their heterozygosity, since the DNA variations are observed within individual genotypes and between genotypes (Uitdewilligen et al., 2013). Many diploid potato clones, as well as cultivated tetraploid potatoes have been used for genetic mapping. Diploid potato genotypes, at a single locus with 2 alleles, have three possible allele dosages (AA, AB and BB). However, in tetraploid genotypes, DNA variants occur in five possible allele dosages: nulliplex (AAAA), simplex (ABBB), duplex (AABB), triplex (AAAB) and quadruplex (BBBB) (Uitdewilligen et al., 2013). Although difficult to analyze, high-density genotyping tools and software are now available for whole genome profiling, such as Infinium SNP genotyping (http://solcap.msu.edu/potato.shtml) or genotyping by sequencing (Elshire et al., 2011).

2.2.2   Potato SolCAP SNP array

A consortium led by SolCAP (http://solcap.msu.edu/) identified a large number of SNPs from the commercial potato cultivars Atlantic, Premier Russet and Snowden

(Hamilton et al., 2011). By comparing their sequences, 69,011 SNPs in potatoes were identified. In order to validate these SNPs, Felcher et al. (2012) developed linkage maps for two diploid mapping populations (DRH and D84) and compared those maps with the assembled potato genome sequence. Over 4,400 markers were mapped (1,960 in DRH, 2,454 in D84, 787 in both), resulting in map sizes of 965 cM in DRH and 792 cM in D84. From these studies, 8,303 SNP markers were selected to develop the Potato SolCAP SNP array (Felcher et al., 2012). When looking at these SNPs, 3,018 SNPs are found within candidate genes; 536 SNPs are from previously mapped genetic markers; and the remaining 4,749 SNPs, are new SNPs dispersed across the chromosomes.

The Illumina Infinium system is based upon a single nucleotide extension in DNA synthesis; the platform has a pool of 250,000 beads, which are oligo-nucleotides specific to each SNP that the array identifies. This pool of beads is then assembled randomly onto the chip. The Potato SolCAP SNP array contains 8,303 SNPs on the Illumina Infinium chip, and each chip can analyze 24 samples, and it is available for scientists worldwide. After the samples are hybridized to the chip, a few major steps should be followed. First, the bead chips are loaded into the Illumina iScan Reader (Illumina, 2005). Then the equipment, Illumina iScan, reads the intensities of the fluorescent dyes (red and green) of each sample associated with the two alleles of the SNP by using a laser to excite the red or green fluorophore of the single base extension product on the beads. Then, the iScan Reader (Illumina, 2005) records the high-resolution images of the light emitted from the fluorophores. These intensities are expressed as Cartesian coordinates (X, Y) by the software. After normalization, Genome Studio Software transforms the intensities to a combined SNP intensity $R=(X+Y)$ and an intensity ratio theta$=(2/\pi)*\text{arctan}(Y/X)$ (Hackett

et al., 2013). The theta score gives information about the dosage of each allele for the samples analyzed, these theta values range from 0 to 1. With the theta values, the Illumina system builds a redundancy database since every SNP is read in the sample an average of 15 to 30 times. At the end, the average of the theta values is calculated for each SNP to generate the array reads and the call is generated as the average of those calls generated for each SNP (Felcher et al., 2012). Data generated in Genome Studio is then exported for further analysis.

A wide range of research has been carried out using the Potato SolCAP SNP array, such as the study of the population structure and linkage disequilibrium in diploid potatoes and genome wide association mapping using tetraploid potatoes (Massa et al., 2015; Stich et al., 2013). In addition, several linkage and QTL mapping experiments have been performed (Hackett et al., 2013), including the assessment of the genetic underpinnings of late blight resistance and tuberization (Lindqvist-Kreuze et al., 2014). Furthermore, a retrospective analysis of potato breeding (Hirsch et al., 2013) at the genome level has been performed by using the Potato SolCAP SNP array to genotype a panel with release dates ranging from years 1857 to 2011, to understand the genetic basis of diversification and trait improvement.

2.2.3   Tetraploid SNP calling

SNP calling in tetraploid potato clones can be achieved by using some specifically designed software. Genome Studio software generates SNP theta scores, which are used to determine the allele dosage. The Potato SolCAP SNP array has been designed for diploid species; however, most cultivated potatoes are tetraploid in nature. Assessing the genotype

calling in tetraploid individuals needs to recognize for each SNP locus, one of five possible genotypes (AAAA, AAAB, AABB, ABBB, and BBBB). Genome Studio Software (Illumina, 2005) is an option for SNP calling (Stich et al., 2013) but it assumes a diploid model and three markers of AA, AB, BB classes for each SNP (Hirsch et al., 2013). Therefore, this software is not suitable for use with tetraploids and other approaches have to be applied.

Converting the continuous signal scores (theta values) to discrete genotype classes can be achieved by a number of different approaches, which include: 1) pre-determined dosage cluster calling boundaries (http://solcap.msu.edu/potato_infinium.shtml); 2) mixture models, FitTetra (Voorrips et al., 2011) or Hackett (Hackett et al., 2013) and, 3) cluster analysis by NbClust (Charrad et al., 2014). The first approach, pre-determined boundaries, has genotype specific boundaries available, that can be used for SNP calling of tetraploid samples, for 5,031 SNPs out of the 8,303 SNP markers, analyzed with the chip (Illumina, 2005). In addition, raw theta values data that range from 0 to 1 could be used for further purposes, as they are generated before being assigned to a specific genotype.

The second approach, mixture models, uses the allele signal ratio which fits a mixture of five normal distributions to the allele signal ratios, with each distribution representing one of the five possible genotype classes. This approach constrains the means of the five distributions by the corresponding allele ratio; at the end, the assignment of components to genotype classes is automatic. In addition, to help identify each distribution, the relationship between allele ratios and means of the distributions are considered, even when the distributions overlap significantly. An algorithm implemented in the R software,

FitTetra (Voorrips et al., 2011), uses a mixture model to assign the genotypes to tetraploid samples. This method uses data from bi-allelic markers generated by genotyping assays that produce intensifying signals for both alleles (theta values) as the Potato SolCAP SNP array does. This R-package rejects markers that do not allow reliable genotyping for the majority of samples; it assigns a missing value to samples that cannot be scored into one of the five possible genotypes with a 95% of confidence. Moreover, another method developed by Hackett et al. (2013) uses a normal mixture model as well to infer SNP dosage from the intensity ratios data. This method takes into account that an ideal SNP, where all five possible dosages can be observed, is expected to consist of theta scores centered on 0.0, 0.25, 0.5, 0.75 and 1.0. In addition, it retained SNPs where the trimmed range between the five groups was greater than or equal to 0.1. This threshold was chosen by the evaluation of the theta values to establish how severe the spatial trends needed to be to affect a visual classification. They also excluded SNPs with missing theta scores (Hackett et al., 2013).

The third approach, the clustering method, is the process of partitioning a set of objects into groups called clusters to finally have the objects within a group more similar to each other than to objects in different groups. This clustering algorithm depends on some assumptions in order to define the subgroups present in a data set. The R-package NbClust (Charrad et al., 2014) was developed to determine the best cluster scheme. This hierarchical clustering is a divisive type where all observations start in one cluster, and splits are performed recursively as one moves down the hierarchy. It is necessary to define the distance and the agglomeration criteria. For the former, the Euclidean distance, is the usual square distance between the two vectors. The agglomeration method, available in NbClust,

is the average method where the distance between the two clusters is the mean of the distances between the pair of points analyzed; this method tends to form clusters with the same variance and in particular, small variance.

2.3   Genome wide association study

Genome-Wide Association Study (GWAS) is a technique that evaluates associations between phenotype and genotype to identify functional DNA variants that closely resemble the phenotype (Oraguzie, 2007). This technique is a tool for high resolution genetic analysis that uses a population that needs to be both phenotypically and genotypically characterized with a large collection of SNP markers well distributed along the genome (Bradeen and Kole, 2011). Mapping for GWAS is called association mapping which is similar to generating quantitative trait loci (QTL) mapping in the identification of novel functional variations that explain the phenotype. The main difference between these two techniques is that QTL mapping uses structured populations in a single segregating population; while GWAS uses samples/populations that may not be genetically related (called unstructured population), and the number of markers required for a GWAS should be much larger than those for QTL analysis (Oraguzie, 2007). An unstructured population in a GWAS represents many more recombination events that often come from offspring of many generations from a common ancestor. Therefore, it gives the potential of analyzing a wide genetically diverse population in order to have a greater resolution (Rafalski, 2010).

In the ideal genome wide association study, the DNA variants are directly associated with the phenotypic variation. In this case, physically linked DNA polymorphisms (such as SNPs) within the gene(s), or in the chromosomal locations

flanking the gene(s), show association with the phenotype. Physically linked DNA polymorphisms are commonly transmitted through successive meiotic generations, and these non-random associations of alleles at different loci are called linkage disequilibrium (LD) (De Koeyer et al., 2011). A limited number of association mapping experiments have been performed in potatoes (Achenbach et al., 2009; Gebhardt et al., 2004; Urbany et al., 2011). Most of these were based on genotyping or sequencing of candidate genes.

Traits determined by the action and interaction of two or more genes are called quantitative traits, and they are considered complex traits. In potatoes, traits related to tuber life-cycle, such as dormancy, tuberization and flowering are considered quantitative traits. To study these complex traits, dense genome coverage is required. Efforts to map quantitative traits in potatoes have been advanced by improvements in both marker development technologies and in analytical statistical methods (Bryan, 2011). GWAS is not the only approach available for identification of markers for a specific trait. Other approaches, such as the candidate gene approach, have been employed to map disease resistance, tuberization, tuber dormancy, and cold sweetening traits (De Koeyer et al., 2011). The candidate gene approach has also been successful in identifying marker–trait associations to traits with clear and well-known biochemical basis, such as the starch synthesis in maize (Wilson et al., 2004). However, candidate genes are biased because they are selected based on only information available from genetic, biochemical or physiological studies (Hall et al., 2010).

GWAS are more comprehensive in nature as they permit the interrogation of the entire genome, rather than focusing on small candidate regions of the genome. Moreover, there are no prior assumptions about the genetic associations of the causal variants (Pearson

and Manolio, 2008). Since GWAS in potato requires high-density genetic markers, the development of the Potato SolCAP SNP array represents the first tool available for GWAS in potatoes (Stich et al., 2013).

2.3.1   Molecular marker analysis

Associations between DNA variants and phenotypic data are only detectable when the frequency of the trait alleles in the population is sufficient for statistical analysis. Therefore, the power of association tests depends on the number of individuals phenotyped and genotyped (the more the better), the allele frequencies and the quality of the phenotypic data (Bryan, 2011). The Potato SolCAP SNP array was utilized for the examination of population structure and genetic diversity of a collection of potato varieties that were important progenitors in Europe (Stich et al., 2013). This study evaluated how informative the Potato SolCAP SNP array was for 36 European potato varieties and determined that three-quarters of the SNPs were polymorphic in the European cultivars. Another study, using the same array, detected low levels of heterozygosity in a collection of 250 wild potato species (Hirsch et al., 2013). This study found a similar portion of informative SNPs from the Potato SolCAP SNP array. Based on their findings there is a need to use a diverse panel, rather than individual accessions, to increase allelic diversity in potatoes (Hirsch et al., 2013).

The Potato SolCAP SNP array has emerged as an increasingly valuable technology, allowing the assessment of potato genetic diversity in a high-throughput fashion (Foster et al., 2010). The genetic diversity of a population, or the level of biodiversity, refers to the total amount of genetic variation in the genetic makeup of a species (De Koeyer et al.,

2011). This diversity serves as a way for populations to adapt to changing environments (Nevo, 2001). Millions of years of adaptations to various ecological and geographic areas have created significant genetic diversity among the wild potato species, including genomic divergence and creation of polyploid complexes (De Koeyer et al., 2011). It is an essential component of plant breeding strategies for determining the extent and distribution of genetic diversity.

Understanding genetic diversity of the populations to conduct a GWAS study is essential in order to determine the resolution of the study and to better understand the results to reduce false positives. Genetic diversity can be measured at various levels, and there are several statistical methodologies available for its assessment, including genetic structure, relatedness and the linkage disequilibrium (Flint-Garcia et al., 2003). The following section describes how each is analyzed.

2.3.1.1 Population structure

The identification of associations can show mixed results in plant systems, due to population structure. Therefore, it is important to describe and identify the presence of population stratification, in order to avoid nonfunctional or false positive associations (Flint-Garcia et al., 2003). The most common way to detect the population structure is by applying the approach available through the software STRUCTURE (Pritchard et al., 2000). This software assigns to each clone in a population a membership probability within a Bayesian framework using the marker information. The Bayesian framework captures the genetic population structure by describing the molecular variation in each subpopulation. These group memberships are thereafter used to assign the genotypes to different

population sub-groups and used as a factor in the marker-trait associations. The model used in STRUCTURE assumes that there are K populations (may be unknown) characterized by a set of allele frequencies at each locus. Using probability, each individual is assigned to a sub-population, or more than one sub-population, if their genotypes indicate that they are admixed. The model assumes that markers are not in linkage disequilibrium (described below) within sub-populations, so it cannot handle markers that are physically close together.

Another way to classify the population structure is to apply Principal Component Analysis (PCA). The availability of a large amount of genotype information in a population makes PCA widely used to detect and quantify genetic structure of populations (Ma and Amos, 2012). In this analysis, DNA markers are treated as features, and based on these features the PCs are generated. The top PCs explain the highest variations due to the population structure in the sample. Thereafter, the clones are projected in groups covered by the top PCs. Finally, clones from the same sub-population are found to be part of a cluster inside of the population. To graphically visualize these clusters, the PC-plot (the pattern of the scatter plot of the top PCs) is used to infer population relationships or within-population structures. The genetic similarities of populations are then inferred from the Euclidean distances between their clusters in the PC-plot as described by Jolliffe (2002).

2.3.1.2 Relatedness

The hidden relationships between individuals are accounted for by their levels of relatedness expressed through a kinship matrix (K) (Yu et al., 2006). K is an n x n matrix (n=number of individuals) of relative kinship coefficients that defines the degree of genetic

covariance between every pair of individuals. Marker-based relative kinship estimates are necessary for quantitative inheritance studies since it provides information on the levels of relatedness between individuals in a population (Ritland, 1996; Yu et al., 2006). This is very important when analyzing complex traits, since the model used needs to take into account the information from the *K* matrix. The resulting values for kinship rank from 0, which indicates no relationship, up to 1 for a totally related pair to systematically account for multiple levels of relatedness among individuals in a population.

2.3.1.3 Linkage disequilibrium

GWAS for a trait could be achieved by using DNA markers, usually SNPs, or haplotypes that are composed of several linked SNP markers. The number of SNPs required for a GWAS depends on the patterns of linkage disequilibrium (LD) in the population or collection of samples. LD is a function of the recombination frequency between polymorphic loci and the number of meiotic generations. In the presence of LD, haplotype association is likely to be more powerful since the allele frequencies within that haplotype will determine if that site will be useful for detecting an association with the trait (Garner and Slatkin, 2003). LD decays with distance since the process of recombination shuffles genetic material between chromosomes and the further apart two markers are, the more probable it is that they will not segregate together (Gebhardt, 2011). It has been proven that phenotype-genotype associations are better detected by individual SNP analysis, since wide variations in LD decay across the genome make any generalizations very difficult (Rafalski, 2010). The power to assess marker-trait associations depends on the rate of decay of LD between loci. Fine mapping requires a fast LD decay, within 1 Kb,

to reduce the size of the windows of evaluation in the genome. On the other hand, when the LD decay is slow, whole genome scans are preferred to first identify the location of QTL in the genome (Mackay and Powell, 2007). LD decay estimated in potatoes 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 and this density could be achieved with the availability of the Potato SolCAP SNP array.

There are several methods to measure linkage disequilibrium. The most commonly used for bi-allelic loci is the r^2 measure. This standard measure is equivalent to the covariance and the correlation between alleles at two different loci (Hill and Robertson, 1968). The estimation of LD can be affected by the population structure and relatedness since differences in allele frequencies may also produce LD between unlinked loci, due to the presence of individuals from different genetic origins within the population studied. It is well known that population structure and relatedness can lead to false associations (Myles et al., 2009). Fortunately, recent studies corrected the biased calculation of r^2 estimate due to the sample structure and relatedness between genotyped individuals by adding these parameters to the model to calculate r^2 (Mangin et al., 2012).

2.3.2   Phenotypic data collection

Phenotypic evaluation of association mapping populations in replicated field trials normally results in more reliable phenotypic data (Li et al., 2008). In a GWAS, existing phenotypic data from national lists, gene banks and breeding companies, in addition to

present day trials, can be utilized (D'hoop et al., 2008). However, the use of a group of selected germplasm within designed trials can provide more reliable results.

In association analyses of quantitative traits, there is an implicit assumption that the phenotype data follow a normal distribution when evaluating QTLs. The violation of this assumption can result in the identification of false-positive associations (Goh and Yap, 2009). In the case that the phenotypic trait is not normally distributed, the approach to follow is to perform transformation of the data to approximate the distribution to normality (Labbe and Wormald, 2005). If GWAS use a combination of two or more populations, the traits are preferably transformed in the same manner to enable the comparison of genetic effects. However, traits evaluated in different populations usually are not distributed in the same manner; therefore, they cannot be transformed in the same way (Peng et al., 2007). A literature search of this topic showed that transforming a trait to ensure normality using a Box-Cox transformation is highly recommended in order to avoid false-positive linkages (Labbe and Wormald, 2005). The Box-Cox transformation method identifies an appropriate exponent "λ", to use to transform data into a normal distribution. The "λ" value indicates the power to which all data should be raised (Box and Cox, 1964). It is generally acknowledged that deviation from normality can reduce the power of the study; therefore, the effect of the normalization directly affects the results of the GWAS (Goh and Yap, 2009).

2.4   Genomic selection

The prediction of complex plant traits, such as growth, yield, and adaptation to stress, is one of the main challenges of the crop breeding community world-wide. Such

complex traits are controlled by many genes, with minor effects limiting their breeding improvements. Genomic selection (GS) is a breeding selection method that does not dissect the main trait looking at the genetic causes (phenotype-genotype associations); instead, it considers them as a black box, using the genome profile without understanding the underlying biology (Cabrera-Bosquet et al., 2012). In order to remove the need to search for significant QTL marker loci associations individually, GS accounts for several predictors simultaneously. The advantages that GS offers are the acceleration of breeding cycles with the enhancement of the rate of annual genetic gain per unit of time and cost. In addition, it helps to replace the phenotypic selection or marker-assisted breeding protocols, based on whole genome predictions in which phenotyping updates the model to build up the prediction accuracy (Desta and Ortiz, 2014).

The advantage of GS in plant breeding is that preferred clones for breeding can be selected even if their phenotypes have yet to be observed. This was proven in a wheat study, in which researchers used GS in two extensive datasets (Crossa et al., 2010). This study not only identified clones with best predicted breeding values; but, they found that *genotype by environment* interaction is an important component of genetic variability, since the estimates of marker effects can be different across environmental conditions (Crossa et al., 2010). Ridge regression is a statistical analysis that was one of the first methods proposed for genomic selection (Hoerl and Kennard, 2000). This method has no limit on the number of markers and also improves the numerical stability when markers are highly correlated (Hoerl and Kennard, 2000). The performance of breeding lines is predicted by ridge regression, based on their kinship, population stratification and marker information.

The accuracy of the prediction methods are compared by cross-validation, using another population or a subset of the same population.

Genomic selection will generate the genome estimated breeding values (GEBV) for an inference population based on the reference population which is phenotyped and genotyped. Genomic selection requires marker information for the inference population, in order to predict the phenotypes of these clones in the field. Afterwards, the correlation between the predicted phenotypes and the real phenotypic outcomes indicate whether or not the model employed is useful for predicting that specific trait in the population. To determine whether genomic selection can be effective for predicting the performance of lines with yet-to-be observed phenotypes, a strong correlation between the predicted and real values is expected (Crossa et al., 2010).

CHAPTER 3: MATERIALS AND METHODS

3.1 Plant materials

3.1.1 Potato clones used for marker-trait association studies

This study used a collection of advanced tetraploid potato lines generated by the International Potato Center (CIP) breeding program, in two different experiments. Experiment 1 consists of 130 tetraploid breeding lines (Table 1). These clones represent four of CIP's advanced breeding populations, primarily clones from the Lowland Tropics Virus Resistance (LTVR) population (Mihovilovich et al., 2007), adapted to the subtropical lowlands with virus resistance, medium-maturity, and heat tolerance, clones from a population derived from a wide range of late blight resistance sources known as B3 (Li et al., 2012), pre-breeding clones, and clones resistant to bacteria wilt (BW). Since few of these clones are adapted to long-day environments, three cultivars [Desiree (Europe), Atlantic (North America) and CTO-33 (USA)] were also included as controls.

3.1.2 Potato clones used for marker validation

For validating marker-trait associations detected in Experiment 1, 66 tetraploid clones from CIP's stress-tolerant germplasm were utilized in Experiment 2 (Table 1). Experiment 2 included clones from the LTVR and BW populations, as described earlier, clones with heat tolerance, clones with different levels of late blight resistance, and clones with Leaf Miner Flying resistance. Experiment 2 also included 25 of the tetraploid clones from Experiment 1 (Table 1).

3.1.3 Potato clones used for genotyping

For genotype calling purposes, 233 potato clones from Table 1 were analyzed using the Potato SolCAP SNP array (Felcher et al., 2012). In addition to the 171 clones phenotyped from Experiments 1 and 2, 11 diploid samples and 51 tetraploid samples were included as part of the genotyping panel to assess the SNP quality, but they were not phenotyped (Table 1).

Table 1. Source of the tetraploid clones included in Experiment 1, Experiment 2 and for genotyping.

Population	Number of clones genotyped	Number of clones phenotyped			
		Experiment 1	Experiment 2	Experiment 1 and 2	*Total*
A	5	0	0	0	0
B3	15	15	0	0	15
BW	11	8	3	0	11
Controls	10	3	9	3	9
Intermediate LT-LB	12	0	11	0	11
LBHT-1	28	0	13	0	13
LMF	2	0	2	0	2
LTVR	136	101	28	22	107
Diploids	11	0	0	0	0
Pre-breeding	3	3	0	0	3
TOTAL	**233**	130	66	25	**171**

BW: bacteria wilt resistant, LT-LB: low tropics and late blight resistant, LBHT: late blight resistant and heat tolerant, LMF: leaf miner fly resistant, LTVR: lowland tropics virus resistant.

3.2 Phenotypic analysis

3.2.1 Field trials

Plants from Experiments 1 and 2 were grown at the lowland subtropics La Molina Station of CIP (Lima, Peru; 12° 3' 0" South) under field conditions. The station is located

29

at an altitude of 280 meters and is considered a coastal desert. The planting times were September 2010 for Experiment 1, and October 2012 for Experiment 2. Seed tubers were provided by CIP. There were two hilling procedures, the first one was 25 days after planting and the second one was 35 days after planting. Plants were watered on a regular basis. There were eight applications of a mix of pesticides (Avit, Sunfire, Gladiador, Enziprom, Quimifol Boro) for the control of pests and diseases. Fertilizers were used to improve the soil quality with Nitrogen (N), Phosphorous (P) and Potassium (K). The fertilizer was applied three times: when preparing the soil (180-120-160), at planting time (90-120-160) and the last one with the hilling (90-00-00). The same conditions were maintained for both experiments.

3.2.2 Experimental design

The experimental design used for each experiment was a randomized complete block design, split-split plot with three replications. The two splits refer to the harvest date and the day-length utilized. There were three factors employed in the two experiments: potato clone, harvest date and day-length. There were two harvest dates: 75 and 90 days after planting (DAP), and two day-lengths, 12 hours (short-day) and 16 hours (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 $\mu mol\ m^{-2}\ s^{-1}$ to extend the photoperiod from 12 hours (Figure 1) to 16 hours (Figure 2). The seed tubers of each clone were planted in a plot of six plants.

Figure 1. Short photoperiod field trial in Experiment 2.

Figure 2. Long photoperiod field trial in Experiment 2, showing high pressure sodium vapor lamps used in the field to extend the photoperiod.

3.2.3 Trait data collection

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. These data were collected at each of the two harvest dates of 75 and 90. 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 three field grown plants at 41, 59 and 74 DAP. The cuttings were maintained in a mist chamber for 10 days; later, this trait was assessed by rating the extent of tuber induction in SNC, from 1 (no induction) to 9 (strong induction). The number of tubers was measured and categorized into number of small tubers, number of marketable tubers, and the total number of tubers counted. In the small tuber size category, tiny and small tubers (<50 g) were included; in the marketable category, the medium and large tubers (≥50 g) were included. Tubers were collected from 6 plants per plot. Bulking ratio refers to the tuber enlargement at a specific time after planting. It was calculated as the ratio of the number of tubers with medium and large sizes over the total number of tubers of all sizes; i.e., tiny + small + medium + large, multiplied by the ratio of tuberized plants, and expressed as a percentage. Stolon number was ranked on a scale from 1 to 9, with 1 assigned to the absence of stolons and 2 to 9 for very few to numerous stolons. Stolon length was measured and ranked on a scale from 1 to 9 for short to long stolons (Table 2). These variables were evaluated in both Experiments 1 and 2.

Table 2. Overview of phenotypic traits, data types, scales, days after planting (DAP) where traits were evaluated under both photoperiods, and the abbreviations (Abbr.) used for each trait.

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
Number of small tubers	Continuous	Number of harvested small tubers (TINY + SMALL) per plot	75 90	ST75 ST90
Number of marketable tubers	Continuous	Number of harvested marketable tubers (MEDIUM + LARGE) per plot	75 90	MT75 MT90
Total number 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 Number*	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

*Indicates traits that were not evaluated under short day conditions in Experiment 1.

3.2.4 Statistical analysis

Statistical analysis for phenotypic data was performed using Proc Mixed of SAS software (SAS Institute, 1999) for a split-split plot design with three replications in the experimental field.

Mixed model:

$$y_{ijklm} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_l + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\alpha\delta)_{il} + (\beta\gamma)_{jk} + (\beta\delta)_{jl} + (\gamma\delta)_{kl}$$

$$+ (\alpha\beta\gamma)_{ijk} + (\alpha\beta\delta)_{ijl} + (\alpha\gamma\delta)_{ikl} + (\beta\gamma\delta)_{jkl} + (\alpha\beta\gamma\delta)_{ijkl} + \varepsilon_{ijklm}$$

Where y represents the phenotypic response, α is the photoperiod ($i = 12$, 16 hr), β is the harvest day ($j = 75$, 90 DAP), γ are the replications ($k = 1, 2, 3$), and δ is the clones (n=130, n=66).

Phenotypic analysis was conducted to generate adjusted mean values for each clone in each photoperiod response trait using the R-package Agricolae (De Mendiburu and Simon, 2015). These adjusted means were the result of the statistical averages that were corrected to compensate for data imbalances and these means removed outliers present in data sets. In addition to the calculation of the mean values for each trait, this study determined the relationship among the 7 traits. Multivariate analysis of traits was conducted, taking into account the effects of all variables on the responses, in order to group similar traits and similar clones. The means, variances and correlations among the 7 traits were calculated using Minitab 17 (Minitab, 2010), in order to determine the effect of photoperiod and DAP on the phenotypic responses and the relationship amongst the response variables.

3.2.5 Heritability analysis

The measurement of the repeatability of the phenotypic data was calculated from the ratio between genotypic and phenotypic variance. These variance components were estimated by a mixed linear model with genotype as random terms for each photoperiod and DAP evaluated. For this purpose, the previously used mixed linear model was performed using the R-package "lme4" (Bates et al., 2014) with the function *lmer*. Based on the variance components, the heritability of the 7 tuberization related traits was calculated with the following equation: $H^2 = \dfrac{\sigma_g^2}{\sigma_g^2 + \dfrac{\sigma_e^2}{n}}$

Where H^2 is the heritability, σ_g^2 represents the variance component for the genotypic main effect, σ_e^2 represents the variance component for the residuals and n is the number of replications (n=3) (Holland et al., 2003).

3.3   Genotypic analysis

3.3.1   SNP array

DNA from all 233 clones was extracted from young leaf tissues using the CTAB method (Doyle and Dickson, 1987). The extracted DNA was quantified using a fluorometer and adjusted to a concentration of 50 ng/µL. SNP genotyping of all the 233 DNA samples, with the Potato SolCAP SNP Array (Felcher et al., 2012), was performed by the Saskatoon Research Centre, Agriculture and Agri-Food Canada, Saskatoon, Canada, in two groups from Experiment 1 and 2. The genotyping was carried out using an Illumina iScan Reader equipped with the Infinium HD Assay Ultra by Gen-Probe. This equipment read the intensities of the fluorescent dyes associated with the two variations (two alleles) of the SNP (Illumina, 2005). After normalization, the output from the Genome Studio software transformed the intensities of the dyes to intensity ratio theta values (Staaf et al., 2008). These values provided the information about the dosage of each allele in each sample analyzed and it was exported from the Genome Studio software (Illumina) to a .txt file for further analysis.

3.3.2   SNP genotype calling

Genotype calling was carried out using the theta values obtained from the Genome Studio analysis described earlier in section 2.2.2. Since the potato samples are tetraploids,

and the Genome Studio software only classifies genotypes for diploids at each SNP, alternative methods to classify tetraploid genotypes (containing two alleles in each SNP) into the five possible genotypes (AAAA, AAAB, AABB, ABBB, BBBB) were used. Four methods were used to assign tetraploid genotypes: 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). These four methods of tetraploid genotype calling and the raw theta values were used to run the association analysis described below, in order to identify the most appropriate method for this study.

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 around 0.0, 0.25, 0.5, 0.75 and 1, respectively. A SNP was retained in the set of informative SNP markers for further analysis if the difference between 2 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.

3.3.3   Genetic diversity analyses

The genetic diversity of all the samples used in this study was analyzed using qualified polymorphic SNP markers to calculate the genetic relationships and the

population structure of all the clones. 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. To input data, it used a multivariate normal (MVN)-expectation maximum (EM) algorithm. This algorithm represents a general approach to calculating maximum likelihood estimates of unknown parameters when data are missing (Poland et al., 2012) and the imputed value is the population mean for that marker. In order to investigate the population structure, STRUCTURE software (Pritchard et al., 2000) was used on the basis of a subset of 120 SNP markers. These 120 SNP markers were selected so that 10 of them were distributed along each of the 12 chromosomes of the potato genome and data was handled as a tetraploid. STRUCTURE software was run with a burn-in period of 20,000 iterations three repetitive times, using a burning time of 1,000,000. The structure was determined by assuming 1 to 10 subpopulations. The most probable number of populations was determined by plotting the natural logarithm (ln) likelihood against the number of subpopulations.

Principal component analysis (PCA) was performed to determine the grouping of the clones into a smaller number of sub-populations. It was computed using the entire subset of informative SNPs by eigenvalue decomposition of the marker-based relationship matrix. For this purpose, R-package rrBLUP (Endelman, 2011) was employed and the number of Principal Components (PC) was determined by the number of PCs that accounted for more than 5% of the total spectrum. From these analyses, it was determined whether there were any patterns of relationship among the clones.

To determine the extent of the marker-trait associations and identify the chromosomal locations associated with tuberization related traits, the linkage disequilibrium (LD) between loci across the potato genome was quantified by the squared correlation coefficient r^2 using R-package LDcorSV (Desrousseaux et al., 2013). 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, the squared correlation between paired markers was plotted against the distance between pairs of markers in base pairs. These results were useful in determining which markers were in LD with the associated SNP for a specific trait.

3.4   Association analysis

To calculate the relationship between a marker and a phenotype, a genome-wide association analysis (GWAS) was performed using the R-package rrBLUP (Endelman, 2011) with all the qualified SNP markers and all the 7 traits. 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.

This model included PCs and population parameters previously determined (P3D) (Zhang et al., 2010). The P3D parameter was equivalent to efficient mixed-model association (EMMA), which can correct a wide range of sample structures by explicitly

accounting for pairwise relatedness between clones using high-density markers to model the phenotype distribution. By visualizing the P-value distribution in a quantile–quantile plot (Wilk and Gnanadesikan, 1968), the confirmation of the expected results with the obtained results was evaluated. To control Type I error (false positives) across the entire experiment (experiment-wise error rate, EWER) 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.

3.5   Genomic selection

For the genomic selection (GS) analysis, 130 breeding lines belonging to Experiment 1 were used as the training population and the 42 clones which were only in Experiment 2 were used as a validation population, where the phenotype data were removed to evaluate the genome estimated predictions. A genome estimated breeding value (GEBV) was calculated using the R-package rrBLUP (Endelman, 2011). This analysis was based on a mixed model of the form:

$$Y = \mu + Xg + e$$

Where Y is the vector of phenotypic means, μ is the overall mean of the training set, X is the marker matrix, g is the marker effects matrix, and e is the vector of residual effects. This model has a design matrix for the genetic values that were partitioned to allow for the designation of the validation population; therefore, the phenotypes from the validation population were removed. For generating GEBV, the model was supplied with the relationship matrix calculated from the markers with the *A.mat* function. The analysis was

run with 100 iterations to obtain the correlation accuracy, which is determined by the correlation between the predicted trait values and the observed trait values.

CHAPTER 4: RESULTS

4.1   Phenotypic analysis

The collection of phenotypic responses to short and long day length evaluated at 75 and 90 DAP during the 2010 and 2012 field trials for a total of 171 potato genotypes (Table 1) with three replicates for 7 tuberization related traits (Table 2) is described in the following section. These 171 potato clones were evaluated in two experiments of 130 and 66 clones and the analysis of variance components for the three factors (day length, DAP and clone) indicated that there were significant differences in the phenotypic responses at the 5% level of significance (Tables A1, A2).

4.1.1   Experiment 1

In this experiment, 130 clones (Table 1) were used. To ensure normality of the adjusted means, the normal probability plot (NPP) of each trait was obtained; if the p-value was greater than 0.05, normality was satisfied. When normality was not achieved, the power of transformation was calculated using Box-Cox transformation, available in Minitab; thereafter, the data was transformed, as needed (Table A3). Under short day length, the number of marketable tubers in both levels (MT75, MT90), as well as the bulking ratio (BR75, BR90), showed normal distributions; tuber induction (TI41), number of small tubers (ST75, ST90) and the total number of tubers (TT75, TT90) required transformation to achieve normality and all of the variables, achieved normality after transformation. It was difficult to achieve normality for TI59 and TI74 since Box-Cox transformation did not work; therefore, non-transformed data was used for further analysis.

Under long day length, only one level of stolon length (SL90) showed normal distribution and the rest of the variables required transformation. After performing Box-Cox transformation, most of the variables achieved normality, except for one level of tuber induction (TI41), one level of the number of marketable tubers (MT90), and the bulking ratio (BR75, BR90). In those cases, non-transformed data was used for further analysis (Table A3). After ensuring the variables' normality, the analysis of tuberization traits' distributions was performed (Table 3).

Table 3. Phenotypic analysis of Experiment 1 under short and long day length evaluated for 130 clones.

Response variable	Day length	DAP	Abbr.	Mean (±s.e.)	StDev	Variance	%CV
Tuber Induction	12	41	TI41	3.11(±0.05)	0.58	0.33	18.52
		59	TI59	5.44(±0.16)	1.78	3.18	32.78
		74	TI74	6.57(±0.14)	1.62	2.61	24.60
	16	43	TI41	3.30(±0.04)	0.40	0.16	12.26
		59	TI59	4.85(±0.11)	1.27	1.61	26.18
		78	TI74	5.40(±0.12)	1.41	1.99	26.11
Number of Small Tubers	12	75	ST75	19.95(±0.81)	9.22	85.05	46.22
		90	ST90	20.46(±0.94)	10.71	114.59	52.33
	16	75	ST75	23.90(±1.40)	15.97	255.17	66.84
		90	ST90	41.45(±1.97)	22.44	503.36	54.13
Number of Marketable Tubers	12	75	MT75	20.22(±0.63)	7.20	51.86	35.62
		90	MT90	20.42(±0.60)	6.82	46.55	33.41
	16	75	MT75	10.49(±0.80)	9.12	83.25	86.96
		90	MT90	15.43(±0.85)	9.67	93.48	62.68
Total Number of Tubers	12	75	TT75	40.17(±1.03)	11.78	138.81	29.33
		90	TT90	40.88(±1.10)	12.60	158.73	30.82
	16	75	TT75	34.38(±1.97)	22.41	502.15	65.18
		90	TT90	56.87(±1.97)	22.42	502.82	39.43
Bulking Ratio	12	75	BR75	50.28(±1.32)	15.08	227.50	30.00
		90	BR90	50.54(±1.32)	15.05	226.48	29.78
	16	75	BR75	23.31(±1.58)	18.01	324.43	77.28
		90	BR90	28.25(±1.61)	18.35	336.55	64.94
Stolon Number*	16	75	SN75	3.82(±0.10)	1.17	1.38	30.75
		90	SN90	3.86(±0.13)	1.49	2.23	38.67
Stolon Length*	16	75	SL75	4.12(±0.13)	1.49	2.23	36.25
		90	SL90	4.62(±0.15)	1.74	3.04	37.60

*Variables were not evaluated under short day conditions. Mean (±s.e.) represents the mean ± the standard error, StDev represents the standard deviation, and %CV represents the coefficient of variation.

4.1.1.1 Short photoperiod

The variance of tuberization related traits was analyzed in order to understand the trait distribution in the population and for an accurate GWAS analysis (Figures 3, A1-A4). The phenotype distributions for the 130 clones analyzed in Experiment 1 and grown under short day length were normal for most of the tuberization related traits (ST, MT, TT and BR), which indicates that they are suitable for GWAS analysis. However, for the tuber induction, Figure A1 indicated that the variability of the trait was not well distributed when evaluated 41 DAP (TI41). The second and third time points evaluated showed a better distribution.
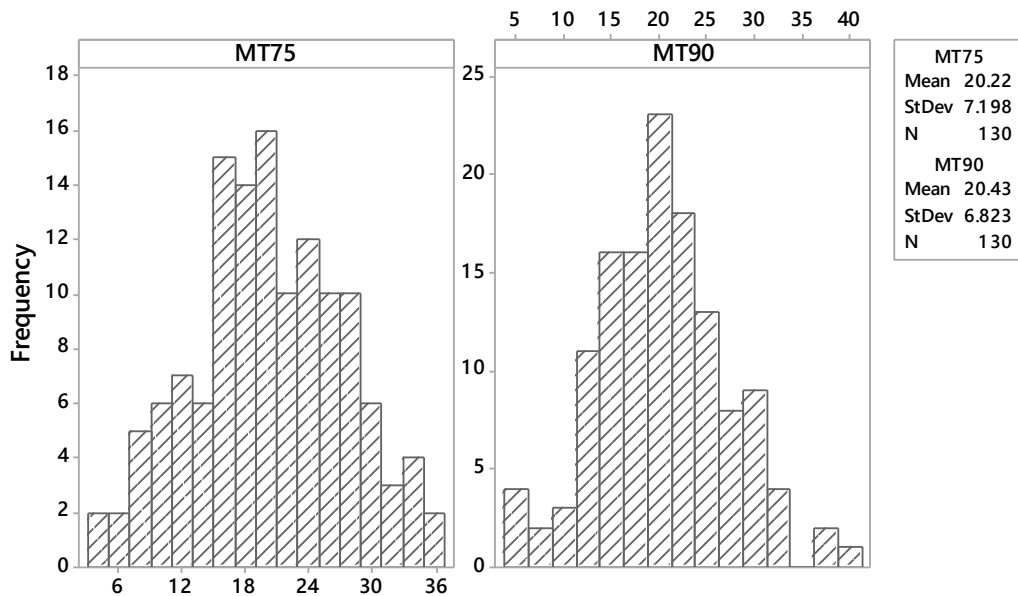
Figure 3. Distribution of the number of marketable tubers evaluated 75 and 90 DAP in the 130 clones from Experiment 1 grown under short day conditions. Histograms show the normality of the phenotype distribution.

Under short day conditions, figures 3 and A1-A4 showed that there is good variability in the phenotypic responses. Under short day conditions, the values for the number of marketable tubers ranged from 0 to 40 marketable tubers when evaluated 90

DAP. Short photoperiod promoted tuberization in clones evaluated in Experiment 1 and the maximum bulking ratio was of 75% (Figure A4).

4.1.1.2 Long photoperiod responses

For an accurate GWAS analysis, the variance of tuberization related traits under long day conditions was analyzed, in order to understand the trait distribution in the population (Figures 4, A5-A10). Under long day conditions, the tuberization related traits showed a normal distribution for tuber induction in TI59 and TI74; the number of small and total tubers (ST and TT) and; the number and stolon length (SN and SL). The number of marketable tubers, as well as the bulking ratio (MT and BR), had a good variability in their responses, but the distribution was not completely normal. On the other hand, the first level of tuber induction (TI41) did not demonstrate a good variability compared to the other two levels of TI59 and TI74. Thus, long photoperiod delayed TI in the growing season.

Tuberization related traits in clones from Experiment 1 showed a different distribution when evaluated under long day conditions. The tuber number and bulking ratio showed a short tailed distribution on the left side (Figures 4, A6-A8). This truncated appearance on the left indicates that the phenotypic variables approached zero very quickly and most of the values regarding the tuberization related traits are closer to zero, when evaluated under long day conditions. The presence of a bimodal distribution, in MT90 as well as BR90, indicates that for some clones, these traits are largely affected by day length (Figure A8).
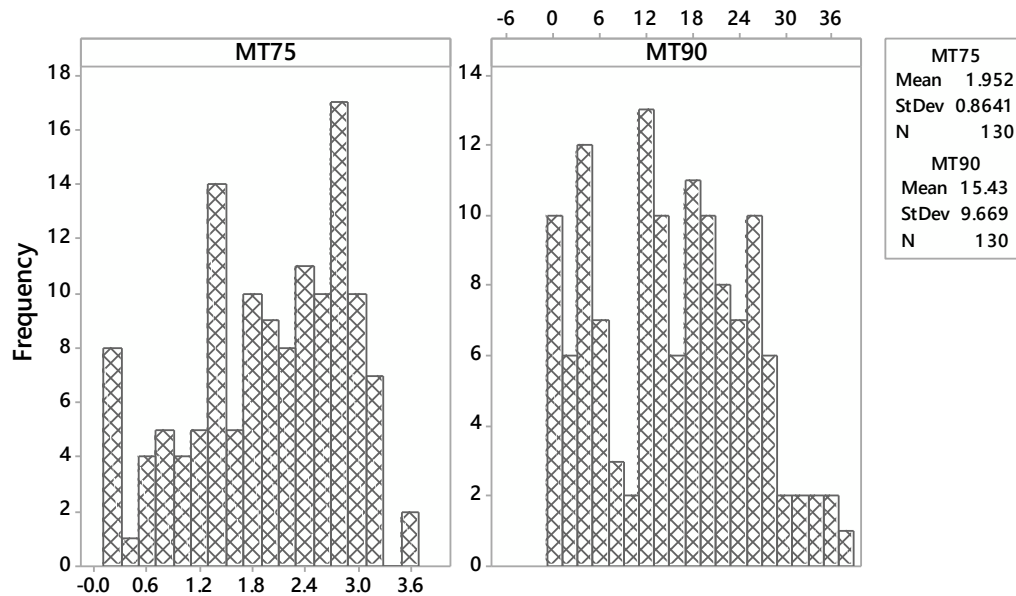
Figure 4. Distribution of the number of marketable tubers evaluated 75 and 90 DAP in the 130 clones from Experiment 1 grown under long day conditions. Histograms show the phenotype distribution.

4.1.1.3 Differences and correlations

Interval plots from the adjusted means were generated in order to document how the tuberization related traits were affected by photoperiod at every DAP evaluation (Figure 5). There are significant differences in all five phenotypic responses between the short and long photoperiods. Stolon number and length were not compared since they were not evaluated under short day conditions. Tuber induction (TI) increased as DAP were extended; this TI was also higher when clones were exposed to a short photoperiod. Therefore, long photoperiod exposed clones had a significantly reduced tuber induction. The number of small tubers (ST) did not increase as the DAP were extended under short day conditions; however, under long day conditions, this number increased significantly. Clones exposed to short day conditions showed a lower ST compared to the clones exposed to a long photoperiod. The number of marketable tubers (MT) was significantly higher in

45

clones exposed to a short photoperiod; under long day conditions, clones showed a lower MT. The increase in DAP did not affect MT in clones exposed to a short photoperiod; however, this increase in time had an effect on the production of marketable tubers under long day conditions. This response might be due to the fact that under long day conditions, potato clones require more time to produce tubers which are acceptable to the market. However, this increase in the number of marketable tubers was still less than the number of marketable tubers harvested in clones exposed to a short photoperiod. The total number of tubers (TT) did not show significant differences when the DAP were extended under short day conditions, but it did when tubers were exposed to a long photoperiod. When evaluated 90 DAP, TT increased under long day conditions and this increase is mainly due to the high production of small tubers. The bulking ratio (BR) trait better described how the tuberization was affected by photoperiod. Under short day conditions, BR did not show significant differences when DAP increased; however, this ratio rose slightly in the clones grown under long photoperiod, when DAP was extended. In addition, the BR interval plot (Figure 5) also showed a substantial decrease in tuberization when the clones were exposed to a long photoperiod.

Correlations were calculated between pairs of tuberization related traits evaluated in the clones from Experiment 1, grown under short and long day conditions (Table 4). Under short day conditions, ST had a positive correlation with TT and a negative correlation with BR, while MT showed a positive correlation with TT and BR. At the same time, BR had a negative correlation with the stolon number (SN). Under long day conditions, TI correlated positively with MT and BR. ST showed a positive correlation with MT and TT; however, it had a negative correlation with BR. MT had a positive

correlation with TT, BR and a strong negative correlation with SN and stolon length (SL). TT showed a positive correlation with BR and a negative correlation with SL. Finally, SN showed a positive correlation with SL. When analyzing the correlations between the short and long photoperiod tuberization responses, TI and BR were the only traits that had significant positive correlations for different day lengths. In addition, TI under short day conditions was correlated with MT under long day conditions.
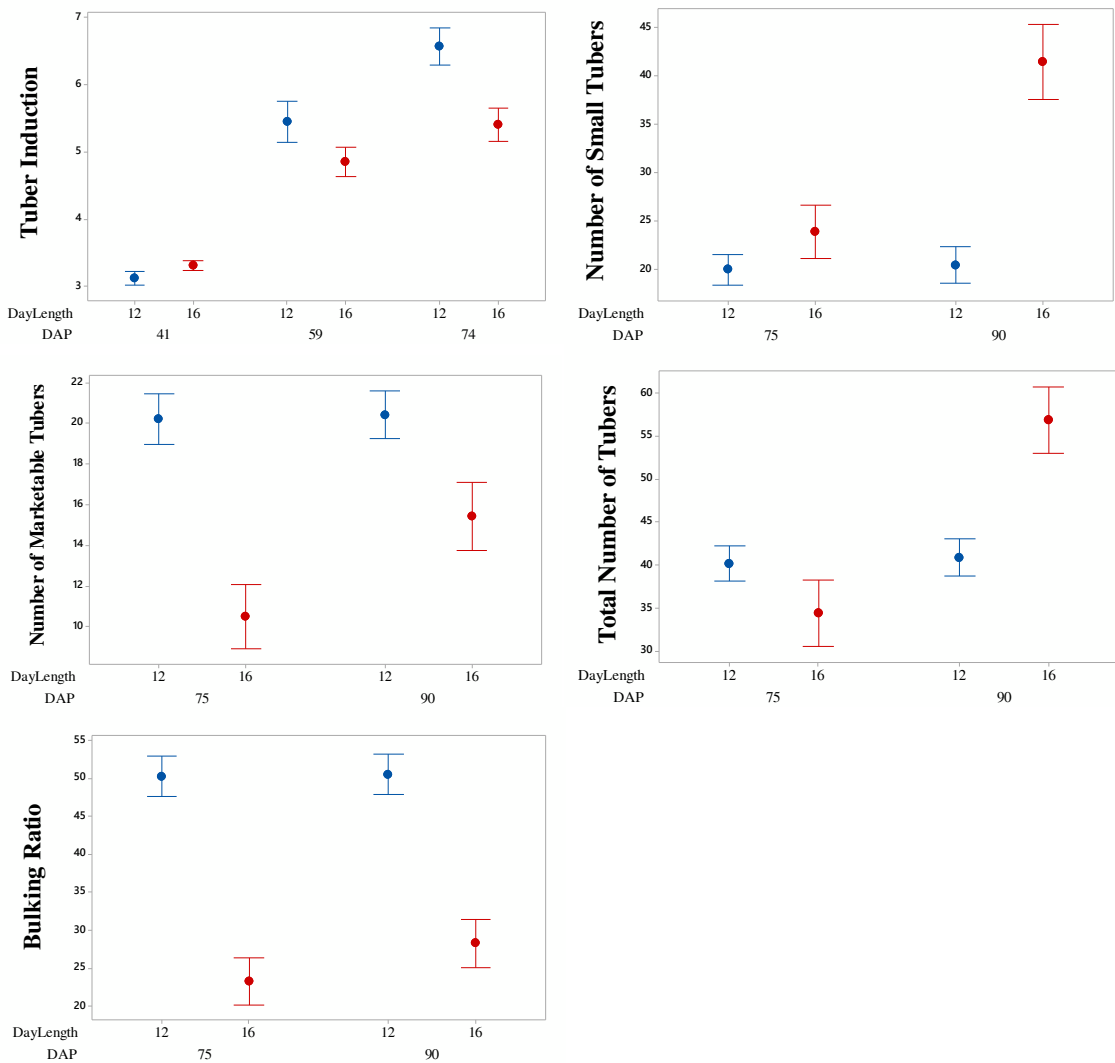
Figure 5. Interval plots of the tuberization related traits evaluated in the clones from Experiment 1 grown under short (blue) and long (red) day conditions. Individual standard deviations (Table 3) were used to calculate the intervals.

Table 4. Correlations among the tuberization related traits in Experiment 1.

| | | Short Daylength | | | | | | | | | | | Long Daylength | | | | | | | | | | | | | | | |
| | | Tuber Induction | | | Number of Small Tubers | | Number of Marketable Tubers | | Total Number of Tubers | | Bulking Ratio | | Tuber Induction | | | Number of Small Tubers | | Number of Marketable Tubers | | Total Number of Tubers | | Bulking Ratio | | Stolon Number | | Stolon Length | |
Variable	DAP	41	59	74	75	90	75	90	75	90	75	90	41	59	74	75	90	75	90	75	90	75	90	75	90	75	90
Short Daylength																											
Tuber Induction	41	1.0																									
	59	0.2	1.0																								
	74	0.2	0.6	1.0																							
Number of Small Tubers	75	-0.1	-0.1	0.0	1.0																						
	90	0.0	-0.1	0.0	0.9	1.0																					
Number of Marketable Tubers	75	-0.1	0.0	0.1	0.0	0.0	1.0																				
	90	0.0	0.0	0.0	0.0	0.0	0.7	1.0																			
Total Number of Tubers	75	-0.1	0.0	0.0	0.8	0.7	0.6	0.5	1.0																		
	90	-0.1	-0.1	0.0	0.7	0.8	0.4	0.5	0.9	1.0																	
Bulking Ratio	75	0.1	0.2	0.1	-0.7	-0.6	0.6	0.4	-0.2	-0.2	1.0																
	90	0.0	0.0	0.0	-0.7	-0.8	0.3	0.5	-0.3	-0.4	0.7	1.0															
Long Daylength																											
Tuber Induction	41	0.3	0.2	0.1	0.0	0.1	0.1	0.0	0.0	0.1	0.1	-0.1	1.0														
	59	0.4	0.6	0.4	-0.2	-0.1	-0.1	-0.1	-0.2	-0.1	0.2	0.0	0.3	1.0													
	74	0.4	0.6	0.5	-0.2	-0.1	-0.1	0.0	-0.2	-0.1	0.2	0.1	0.2	0.7	1.0												
Number of Small Tubers	75	0.1	0.4	0.4	0.2	0.1	0.3	0.2	0.3	0.2	0.1	-0.1	0.1	0.3	0.3	1.0											
	90	0.0	-0.1	0.0	0.4	0.4	0.1	0.1	0.4	0.4	-0.2	-0.2	0.0	-0.1	-0.2	0.1	1.0										
Number of Marketable Tubers	75	0.3	0.5	0.4	-0.2	-0.2	0.3	0.2	0.0	0.0	0.4	0.3	0.2	0.5	0.5	0.6	-0.2	1.0									
	90	0.2	0.5	0.4	-0.2	-0.2	0.4	0.3	0.0	0.0	0.5	0.3	0.2	0.4	0.5	0.7	-0.2	0.8	1.0								
Total Number of Tubers	75	0.2	0.5	0.4	0.0	0.0	0.4	0.2	0.2	0.1	0.2	0.1	0.1	0.4	0.5	0.9	0.0	0.8	0.8	1.0							
	90	0.1	0.1	0.2	0.3	0.3	0.3	0.2	0.4	0.3	0.0	-0.1	0.1	0.1	0.1	0.4	0.9	0.1	0.2	0.3	1.0						
Bulking Ratio	75	0.2	0.5	0.3	-0.4	-0.3	0.3	0.2	-0.1	-0.2	0.6	0.4	0.1	0.4	0.5	0.2	-0.4	0.9	0.7	0.5	-0.1	1.0					
	90	0.2	0.4	0.3	-0.4	-0.4	0.3	0.2	-0.2	-0.2	0.6	0.4	0.1	0.4	0.5	0.4	-0.5	0.8	0.9	0.6	-0.1	0.8	1.0				
Stolon Number	75	0.0	-0.2	-0.1	0.4	0.4	0.1	0.1	0.4	0.4	-0.3	-0.3	0.1	-0.2	-0.3	-0.2	0.3	-0.5	-0.4	-0.3	0.1	-0.5	-0.5	1.0			
	90	0.0	-0.3	-0.2	0.4	0.4	0.0	0.0	0.3	0.4	-0.3	-0.3	0.1	-0.2	-0.3	-0.3	0.3	-0.5	-0.5	-0.4	0.1	-0.5	-0.6	0.9	1.0		
Stolon Length	75	-0.1	-0.2	-0.2	0.3	0.3	0.0	0.0	0.3	0.3	-0.2	-0.2	-0.1	-0.2	-0.3	-0.3	0.3	-0.5	-0.5	-0.5	0.1	-0.5	-0.6	0.8	0.8	1.0	
	90	-0.1	-0.3	-0.2	0.3	0.3	0.0	0.0	0.2	0.2	-0.3	-0.2	0.0	-0.3	-0.4	-0.4	0.3	-0.6	-0.5	-0.5	0.0	-0.5	-0.6	0.8	0.9	0.8	1.0

Correlations between tuberization related traits evaluated in the 130 clones from Experiment 1 grown under short and long day conditions are shaded. Significant correlations have values ≤-0.5 (negative correlation) or ≥0.5 (positive correlation). P-values are significant at 5 % level of significance.

4.1.2   Experiment 2

In this experiment, 66 clones (Table 1) were used. To ensure normality of the adjusted means, the normal probability plot (NPP) of each trait was obtained and evaluated; if the p-value was greater than 0.05, normality was satisfied. When normality was not achieved, the power of transformation was calculated using Box-Cox transformation available in Minitab; thereafter, data was transformed as needed (Table A4). Tuber induction under short and long day length required transformation to achieve normality at its first level (TI41), while TI59 and TI74 remained with the original data for further analysis. The number of small tubers, as well as the total number of tubers and the stolon number required transformation under both photoperiods. After transformation, they all achieved normality and the transformed data was used for the analysis. For the number of marketable tubers, as well as the bulking ratio, data was normal under short and long day length and no transformation was required. Finally, the stolon length data under short day conditions was not normal and transformation was not successful; therefore, the original data was considered for further analysis, as well as SL90 under long day conditions. Under long day conditions, SL75 achieved normality through transformation. After ensuring the variables' normality, the analysis of tuberization trait distributions was performed (Table 5).

Table 5. Phenotypic analysis of Experiment 2 under short and long day length evaluated for 66 clones.

Response variable	Day length	DAP	Abbr.	Mean (±s.e.)	StDev	Variance	%CV
Tuber Induction	12	41	TI41	4.32(±0.19)	1.55	2.41	35.91
		59	TI59	6.20(±0.22)	1.75	3.06	28.20
		74	TI74	6.92(±0.19)	1.50	2.25	21.69
	16	41	TI41	4.25(±0.12)	0.96	0.92	22.53
		59	TI59	5.96(±0.18)	1.47	2.17	24.73
		74	TI74	6.25(±0.17)	1.35	1.81	21.52
Number of Small Tubers	12	75	ST75	43.09(±3.56)	28.94	837.28	67.15
		90	ST90	39.82(±2.95)	23.99	575.41	60.24
	16	75	ST75	50.77(±3.05)	24.74	612.15	48.73
		90	ST90	46.30(±3.33)	27.03	730.58	58.37
Number of Marketable Tubers	12	75	MT75	29.06(±1.09)	8.85	78.27	30.44
		90	MT90	28.14(±1.12)	9.07	82.24	32.23
	16	75	MT75	18.41(±1.21)	9.86	97.23	53.56
		90	MT90	29.85(±1.25)	10.19	103.79	34.13
Total Number of Tubers	12	75	TT75	72.17(±3.29)	26.70	712.85	37.00
		90	TT90	67.98(±2.90)	23.55	554.45	34.64
	16	75	TT75	69.15(±2.64)	21.45	460.19	31.02
		90	TT90	76.17(±3.22)	26.19	686.02	34.39
Bulking Ratio	12	75	BR75	44.62(±1.93)	15.66	245.17	35.09
		90	BR90	44.90(±1.99)	16.16	261.18	35.99
	16	75	BR75	29.35(±2.10)	17.05	290.66	58.09
		90	BR90	42.70(±2.11)	17.15	294.03	40.16
Stolon Number	12	75	SN75	2.29(±0.14)	1.15	1.31	49.97
		90	SN90	2.19(±0.14)	1.14	1.30	52.10
	16	75	SN75	2.92(±0.17)	1.40	1.96	47.91
		90	SN90	3.61(±0.20)	1.59	2.54	44.11
Stolon Length	12	75	SL75	2.26(±0.19)	1.51	2.29	67.02
		90	SL90	2.17(±0.19)	1.51	2.28	69.49
	16	75	SL75	3.01(±0.21)	1.67	2.79	55.50
		90	SL90	3.85(±0.23)	1.84	3.38	47.74

Mean (±s.e.) represents the mean ± the standard error, StDev represents the standard deviation, and %CV represents the coefficient of variation.

4.1.2.1 Short photoperiod responses

The variance of tuberization related traits was analyzed in order to understand the trait distribution in the population, for accurate GWAS analysis (Figures 6, A11-A16). The phenotype distributions for the 66 clones analyzed in Experiment 2 and grown under short day length showed a good distribution for five of the seven tuberization related traits (TI, ST, MT, TT and BR), which indicates good trait variability for GWAS analysis. The stolon number (SN), as well as the stolon length (SL), showed a normal distribution with some interruptions. These data were still considered for further analysis based on their good distribution. However, for the second level of tuber induction (TI59), the histogram showed a bimodal distribution (Figure A11), indicating that the variability of the trait was highly affected by photoperiod.
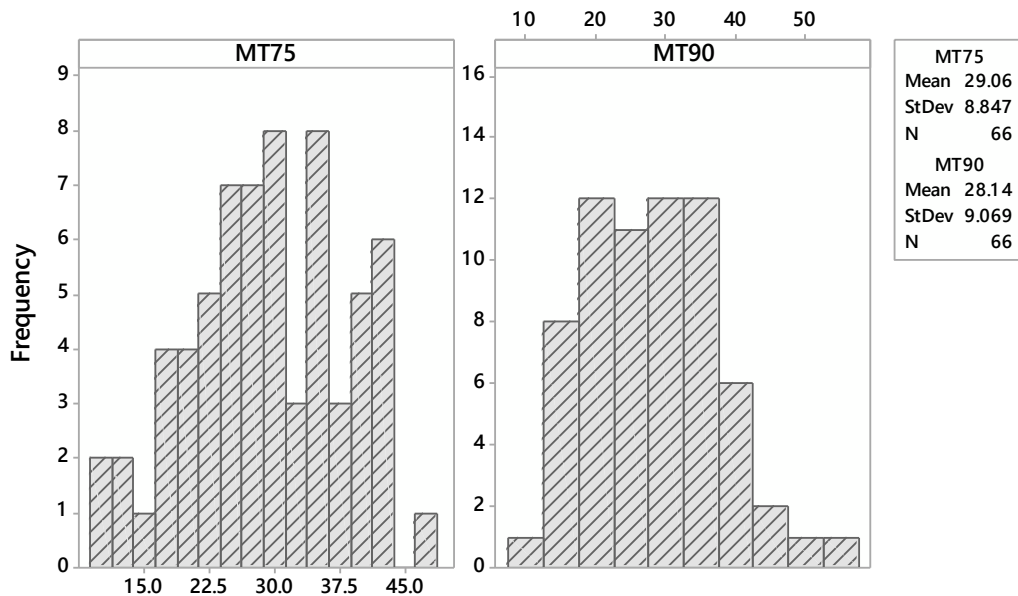
Figure 6. Distribution of the number of marketable tubers evaluated 75 and 90 DAP in the 66 clones from Experiment 2, grown under short day conditions. Histograms show the normality of the phenotype distribution.

4.1.2.2 Long photoperiod responses

In order to have an accurate GWAS analysis, the variance of tuberization related traits was analyzed to ensure the understanding of the trait distribution in the population. The phenotype distributions for the 66 clones analyzed in Experiment 2 and grown under long day length showed a good distribution for six of the tuberization related traits (TI, ST, MT, TT, BR and SN), which indicates a good trait variability for the GWAS analysis (Figures 7, A17-A22). The 7th trait, stolon length (SL), showed a distribution with an interruption when evaluated 75 DAP, but this data was still used for further analysis based on the good distribution. However, for the second level of tuber induction (TI59), the histogram showed a bimodal distribution that indicated that the variability of the trait was not well represented (Figure A17). For the rest of the phenotypic variables, data responses were considered for the GWAS analysis.
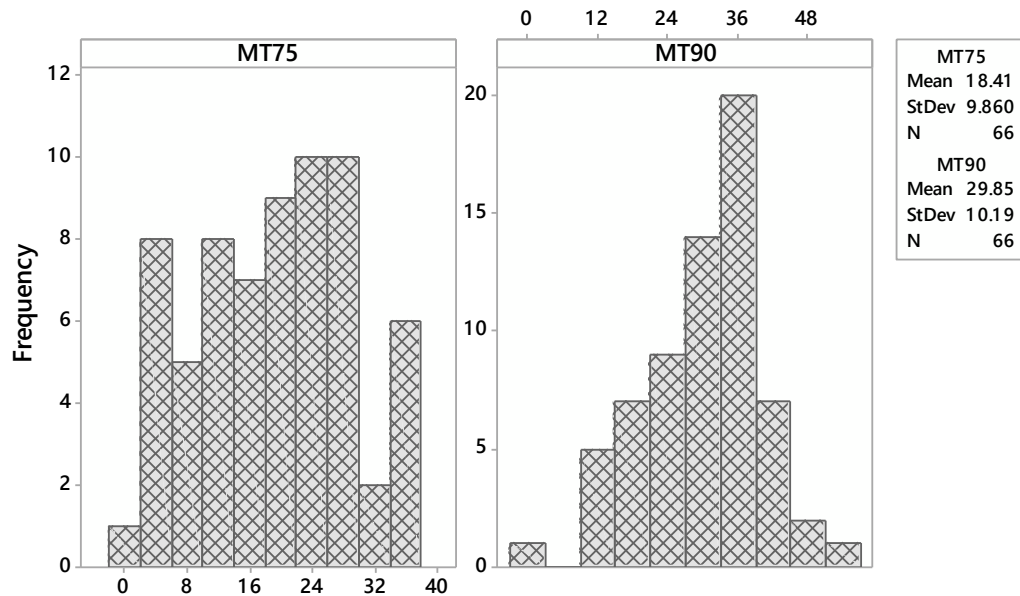
Figure 7. Distribution of the number of marketable tubers evaluated 75 and 90 DAP in the 66 clones from Experiment 2, grown under long day conditions. Histograms show the normality of the phenotype distribution.

4.1.2.3 Differences and correlations

Interval plots from the adjusted means were generated in order to show how the tuberization related traits in clones from Experiment 2 were affected by photoperiod at every DAP evaluation (Figure 8). Most of the phenotypic responses had the same behavior as in Experiment 1 regarding their responses to both photoperiods. However, these differences in the responses of tuberization related trait under short and long day lengths were diminished. Tuber induction increased as DAP were extended and under long day conditions; TI was slightly lower compared to short photoperiod, but these differences were not significant. In the same way, ST had slightly differences when the two photoperiods were compared; in this case, long day length promoted the number of small tubers slightly more. The number of marketable tubers evaluated 75 DAP had significant differences when both photoperiods were compared; however, at 90 DAP, these differences were not significant and the clones under long photoperiod produced slightly more tubers when compared to short day MT. The total number of tubers, as well as BR, reflected the results from ST and MT; there were no significant differences between the two day lengths' responses. SN, as well as SL, increased with long photoperiod exposure. Under long day conditions, they also increased when DAP were extended (Table 5, Figure 8).

The correlations of tuberization related traits were calculated in all 66 clones from Experiment 2 grown under short and long day conditions (Table 6). Under short day conditions, TI had a negative correlation with SN and SL. ST was positively correlated with TT. BR had a negative correlation with ST and TT, but a positive correlation with MT. Finally, SN was positively correlated with SL.
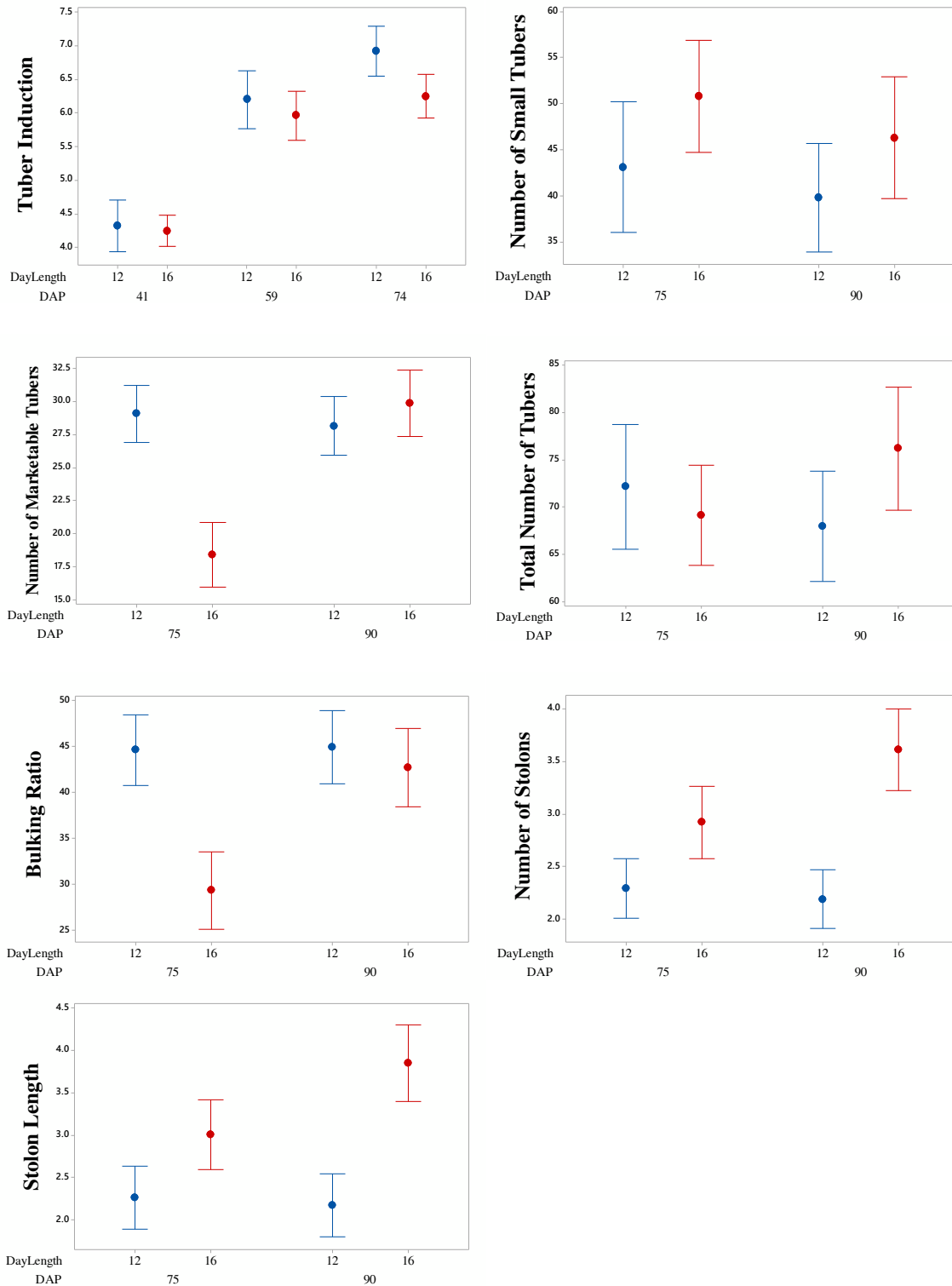
Figure 8. Interval plots of the tuberization related traits evaluated in the clones from Experiment 2 grown under short (blue) and long (red) day conditions. Individual standard deviations (Table 5) were used to calculate the intervals.

Table 6. Correlations among the tuberization related traits in Experiment 2.

Short Daylength (rows) / Long Daylength (rows) as indicated in the left margin.

Variable	DAP	Short Daylength															Long Daylength														
		Tuber Induction			Number of Small Tubers		Number of Marketable Tubers		Total Number of Tubers		Bulking Ratio		Stolon Number		Stolon Length		Tuber Induction			Number of Small Tubers		Number of Marketable Tubers		Total Number of Tubers		Bulking Ratio		Stolon Number		Stolon Length	
		41	59	74	75	90	75	90	75	90	75	90	75	90	75	90	41	59	74	75	90	75	90	75	90	75	90	75	90	75	90
Short Daylength																															
Tuber Induction	41	1.0																													
	59	0.5	1.0																												
	74	0.4	0.8	1.0																											
Number of Small Tubers	75	-0.1	0.0	0.0	1.0																										
	90	-0.1	0.0	0.0	0.9	1.0																									
Number of Marketable	75	-0.2	0.1	0.1	-0.4	-0.4	1.0																								
	90	-0.3	-0.1	-0.2	-0.1	-0.2	0.6	1.0																							
Total Number of Tubers	75	-0.1	0.0	0.0	1.0	0.9	-0.1	0.0	1.0																						
	90	-0.2	-0.1	-0.1	0.9	0.9	-0.1	0.1	0.9	1.0																					
Bulking Ratio	75	0.0	0.1	0.1	-0.9	-0.8	0.6	0.3	-0.7	-0.7	1.0																				
	90	-0.1	-0.1	-0.1	-0.7	-0.8	0.5	0.6	-0.6	-0.6	0.8	1.0																			
Stolon Number	75	-0.5	-0.6	-0.7	0.3	0.3	0.0	0.4	0.3	0.4	-0.2	-0.1	1.0																		
	90	-0.4	-0.5	-0.6	0.3	0.4	-0.1	0.3	0.4	0.5	-0.3	-0.2	0.8	1.0																	
Stolon Length	75	-0.3	-0.5	-0.6	0.3	0.2	-0.1	0.3	0.2	0.3	-0.3	-0.1	0.9	0.8	1.0																
	90	-0.3	-0.5	-0.6	0.2	0.2	0.0	0.3	0.2	0.4	-0.2	0.0	0.8	0.9	0.9	1.0															
Long Daylength																															
Tuber Induction	41	0.5	0.3	0.1	-0.1	-0.1	-0.2	-0.3	-0.2	-0.2	0.1	0.0	-0.3	-0.3	-0.2	-0.3	1.0														
	59	0.7	0.7	0.7	-0.2	-0.2	-0.1	-0.2	-0.2	-0.2	0.1	0.0	-0.6	-0.6	-0.5	-0.4	0.4	1.0													
	74	0.4	0.7	0.7	0.0	0.0	-0.1	-0.4	-0.1	-0.1	0.0	-0.2	-0.6	-0.5	-0.5	-0.5	0.3	0.7	1.0												
Number of Small Tubers	75	-0.1	-0.1	0.0	0.7	0.6	-0.2	0.1	0.7	0.6	-0.6	-0.4	0.3	0.3	0.3	0.3	-0.2	-0.1	-0.2	1.0											
	90	0.0	0.0	0.0	0.8	0.8	-0.3	-0.1	0.8	0.8	-0.8	-0.7	0.3	0.3	0.2	0.2	-0.2	-0.1	-0.1	0.8	1.0										
Number of Marketable	75	0.2	0.4	0.3	-0.5	-0.5	0.5	0.2	-0.4	-0.4	0.6	0.4	-0.5	-0.5	-0.5	-0.4	0.1	0.4	0.3	-0.5	-0.5	1.0									
	90	0.0	0.3	0.3	-0.2	-0.2	0.5	0.4	0.0	0.0	0.3	0.3	-0.1	-0.2	-0.1	-0.1	0.0	0.2	0.2	-0.1	-0.3	0.5	1.0								
Total Number of Tubers	75	0.0	0.1	0.1	0.5	0.4	0.0	0.2	0.6	0.5	-0.4	-0.3	0.1	0.1	0.1	0.1	-0.2	0.0	-0.1	0.9	0.7	-0.1	0.2	1.0							
	90	-0.1	0.1	0.2	0.8	0.8	-0.1	0.0	0.8	0.8	-0.7	-0.6	0.2	0.3	0.2	0.2	-0.2	0.0	0.0	0.8	0.9	-0.3	0.1	0.8	1.0						
Bulking Ratio	75	0.3	0.4	0.3	-0.6	-0.6	0.4	0.1	-0.6	-0.6	0.7	0.5	-0.5	-0.6	-0.5	-0.4	0.3	0.3	0.3	-0.8	-0.7	0.9	0.4	-0.5	-0.6	1.0					
	90	0.2	0.3	0.2	-0.6	-0.7	0.4	0.2	-0.6	-0.6	0.7	0.6	-0.3	-0.4	-0.3	-0.3	0.2	0.3	0.2	-0.6	-0.8	0.7	0.7	-0.4	-0.6	0.8	1.0				
Stolon Number	75	-0.5	-0.5	-0.5	0.4	0.4	0.0	0.4	0.4	0.5	-0.4	-0.1	0.9	0.8	0.8	0.7	-0.3	-0.6	-0.6	0.4	0.4	-0.6	-0.2	0.2	0.3	-0.6	-0.4	1.0			
	90	-0.4	-0.5	-0.5	0.4	0.4	0.0	0.3	0.4	0.5	-0.4	-0.1	0.8	0.8	0.8	0.7	-0.3	-0.6	-0.5	0.4	0.4	-0.6	-0.1	0.2	0.4	-0.6	-0.4	0.9	1.0		
Stolon Length	75	-0.4	-0.5	-0.6	0.3	0.2	0.0	0.4	0.3	0.4	-0.2	0.0	0.8	0.7	0.8	0.7	-0.2	-0.6	-0.6	0.4	0.3	-0.4	-0.1	0.2	0.3	-0.5	-0.3	0.9	0.8	1.0	
	90	-0.5	-0.4	-0.5	0.3	0.3	0.0	0.4	0.4	0.5	-0.3	-0.1	0.8	0.8	0.8	0.7	-0.3	-0.5	-0.5	0.4	0.3	-0.4	0.0	0.2	0.4	-0.5	-0.3	0.8	0.9	0.9	1.0

Correlations between tuberization related traits evaluated in the 66 clones from Experiment 2 grown under short and long day conditions are shaded. Significant correlations have values ≤-0.5 (negative correlation) or ≥0.5 (positive correlation). P-values are significant at 5 % level of significance.

Under long day conditions, TI presented negative correlations with SN and SL, as shown under short day length. The number of small tubers was negatively correlated with MT and BR, but positively correlated with TT. The number of marketable tubers was positively correlated with BR, as shown in all the cases before, but they were both negatively correlated with SN and SL. The total number of tubers was negatively correlated with BR, as shown under short day conditions, and SN and SL were positively correlated as well. When analyzing the correlations between short and long photoperiod tuberization responses, the number of correlations increased in comparison to correlations in Experiment 1, as all the tuberization related traits were correlated positively with each other in both photoperiods. In addition, TI under short day conditions was negatively correlated with SN and SL under long day conditions. ST under short day length was negative correlated with MT and BR under a long day length, and positively correlated with TT under long day length. Finally, SN and SL were positively correlated with each other under both photoperiods (Table 6).

4.1.3   Heritability of the tuberization related traits

The repeatability of the phenotypic tuberization related traits, shown as heritability (H^2), was calculated from the ratio between genotypic and phenotypic variance in Experiment 1 (Table 7a) and Experiment 2 (Table 7b). In Experiment 1, the heritability was high for all the tuberization related traits and bulking ratio had the highest heritability in clones grown under long day conditions (0.83), followed by the number of marketable tubers in clones grown under long day conditions as well (0.82). Tuber induction, as well as the total number of tubers, showed a high H^2 as well. The number of marketable tubers

in clones grown under short day conditions was the variable with the lowest heritability; however, 0.45 is still considered a good heritability.

Table 7a. Estimated variance and heritability for phenotypic traits from Experiment 1.

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
Number 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
Number 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 number 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 number	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 in the 130 clones.

In Experiment 2, the heritability was higher for all the tuberization related traits in comparison with heritabilities in Experiment 1. Tuber induction in clones grown under short day conditions had the highest heritability (0.93). The number of small tubers, total

number of tubers and stolon length had high heritability (0.94, 0.90 and 0.88), followed by

bulking ratio and stolon number (0.87, 0.84). The number of marketable tubers in clones

grown under short day conditions had the lowest heritability (0.54); overall, all the

tuberization related traits had high heritability and therefore, were used for further GWAS.

Table 7b. Estimated variance and heritability of phenotypic traits from Experiment 2.

Trait	Day length	DAP	V_g	V_e	H^2
Tuber induction	12	41	2.22	0.63	0.91
		59	2.86	0.64	0.93
		74	2.09	0.48	0.93
	16	41	0.61	0.85	0.68
		59	1.93	0.98	0.86
		74	2.09	0.48	0.93
Number of small tubers	12	75	785.10	157.90	0.94
		90	499.40	231.80	0.87
	16	75	514.40	334.30	0.82
		90	647.20	261.60	0.88
Number of marketable tubers	12	75	49.76	80.81	0.65
		90	45.38	117.82	0.54
	16	75	81.64	43.53	0.85
		90	73.68	93.73	0.70
Total number of tubers	12	75	643.60	209.20	0.90
		90	468.40	238.80	0.85
	16	75	361.70	336.60	0.76
		90	594.00	275.40	0.87
Bulking ratio	12	75	203.10	126.80	0.83
		90	202.90	188.00	0.76
	16	75	249.30	110.00	0.87
		90	246.00	131.60	0.85
Stolon number	12	75	0.95	1.11	0.72
		90	1.01	0.91	0.77
	16	75	1.60	1.01	0.83
		90	2.10	1.21	0.84
Stolon length	12	75	2.04	0.87	0.88
		90	1.69	1.54	0.77
	16	75	2.35	1.29	0.84
		90	2.91	1.56	0.85

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 in the 66 clones.

4.2   Genotypic analysis

The 171 clones included in the panels evaluated in Experiments 1 and 2, along with the 11 diploid controls and 51 extra tetraploid clones (Table 1), were genotyped with the Potato SolCAP SNP array (Felcher et al., 2012). After the genome-wide SNP genotyping was performed, the theta scores were extracted for further analysis using the Illumina Genome Studio software (Illumina, 2005).

4.2.1 SNP genotype calling

Four different SNP genotype calling methods were compared in order to identify the one with higher quality. The first method used was FitTetra (Voorrips et al., 2011) and this R-package automatically 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 SNPs were informative in assigning genotypes. These SNPs were well distributed across the potato genome. The second method, described by Hackett et al. (Hackett et al., 2013), assigned the genotypes and identified a total of 5,282 informative SNPs. The third method used the R-package NbClust (Charrad et al., 2014); it assigned genotypes and identified a total of 5,487 informative SNPs. The fourth method used the SolCAP boundaries that are available from the consortium (http://solcap.msu.edu/potato_infinium.shtml); this method assigned SNP calls for a total of 2,033 informative SNPs. In addition to these four SNP calling genotype methods, this study also included all 8,303 SNPs in the form of raw theta values for the comparison (Table 8).

Table 8. Summary of total number of informative SNPs from 4 individual methods across the potato chromosomes.

Chromosome	SolCAP	Genotyping Method			
		FitTetra	Hackett	NbClust	SolCAP boundaries
ST4.03ch00*	132	55	59	60	25
ST4.03ch01	760	500	511	546	186
ST4.03ch02	687	459	515	539	189
ST4.03ch03	621	398	456	463	189
ST4.03ch04	743	459	494	510	186
ST4.03ch05	541	303	353	373	123
ST4.03ch06	605	408	424	467	197
ST4.03ch07	647	414	458	465	178
ST4.03ch08	512	338	372	370	161
ST4.03ch09	576	366	420	425	160
ST4.03ch10	440	235	264	273	97
ST4.03ch11	502	281	338	343	141
ST4.03ch12	455	250	282	302	87
NA©	1082	272	336	351	114
Total	8,303	4,738	5,282	5,487	2,033

*ST4.03ch00 lists the markers that are located on unanchored scaffolds of the reference genome. ©NA refers to SNPs which positions have not been assigned.

In order to select the best genotype calling method suitable for this study, a GWAS was performed for BR evaluated 90 DAP on the 130 clones from Experiment 1 under short day conditions. From these results (Table 9), four SNPs in common were found to be associated with the BR trait and their SNP genotype calling histograms were compared. These comparisons were performed in order to check the quality of SNPs associated, as well as the accuracy of their genotype calls.

Table 9. SNPs associated with Bulking Ratio 90 DAP for each genotype calling method.

Marker	Chr. no.	Position	Genotyping Method				
			FitTetra	Hackett	NbClust	SolCAP boundaries	Raw 8,303
SolCAP_c2_38405	1	62145049	3.77	3.81	3.72	4.46	3.49
SolCAP_c2_34547	1	84727244	ns	ns	ns	*	3.09
SolCAP_c2_39282	4	1633476	3.01	ns	ns	*	ns
SolCAP_c2_35998	4	70939686	ns	*	3.26	*	ns
SolCAP_c1_13135	6	51183281	*	*	3.02	*	ns
SolCAP_c2_41407	6	51183805	ns	ns	ns	3.17	ns
SolCAP_c2_41405	6	51484815	ns	ns	ns	3.17	ns
SolCAP_c1_3001	6	51925054	*	3.58	*	*	ns
SolCAP_c2_8904	6	52599348	4.18	4.23	4.41	*	4.30
SolCAP_c2_8966	6	52859976	ns	ns	3.05	ns	3.01
SolCAP_c2_9001	6	52947389	ns	*	*	*	3.40
SolCAP_c2_9002	6	52947838	ns	ns	ns	*	3.53
SolCAP_c2_9005	6	52947949	ns	*	3.04	*	3.34
SolCAP_c2_9009	6	52951567	ns	*	3.05	*	ns
SolCAP_c2_22750	9	31697694	*	*	*	*	3.05
SolCAP_c2_46921	10	5681818	*	*	*	*	3.24
SolCAP_c1_14083	11	4317208	ns	*	ns	3.03	ns
SolCAP_c1_5716	11	18376160	ns	ns	3.05	*	ns
SolCAP_c2_4978	11	20780633	3.39	3.29	3.36	*	ns
SolCAP_c2_44634	11	24839121	3.19	3.29	3.36	3.36	3.29
SolCAP_c2_31290	12	4217966	3.39	3.91	3.92	*	3.08
SolCAP_c2_12917	NA	NA	*	3.15	ns	*	3.28
SolCAP_c2_55484	NA	NA	ns	3.09	*	*	ns

ns: non-significant association (p>0.05); * SNP marker did not pass the filters and were not included in the association analysis.

The five different genotyping methods were compared using the SNP marker SolCAP_c2_44634 (Figure 9) identified as being associated with BR90. The SNP calling results did not show significant differences in the genotype calls when the five methods were compared. However, it was evident that every method handled the genotype calling in different ways when the proximity of the genotype groups was close, indicated by the assignment of NA (not assigned) when it was difficult to call the genotype.
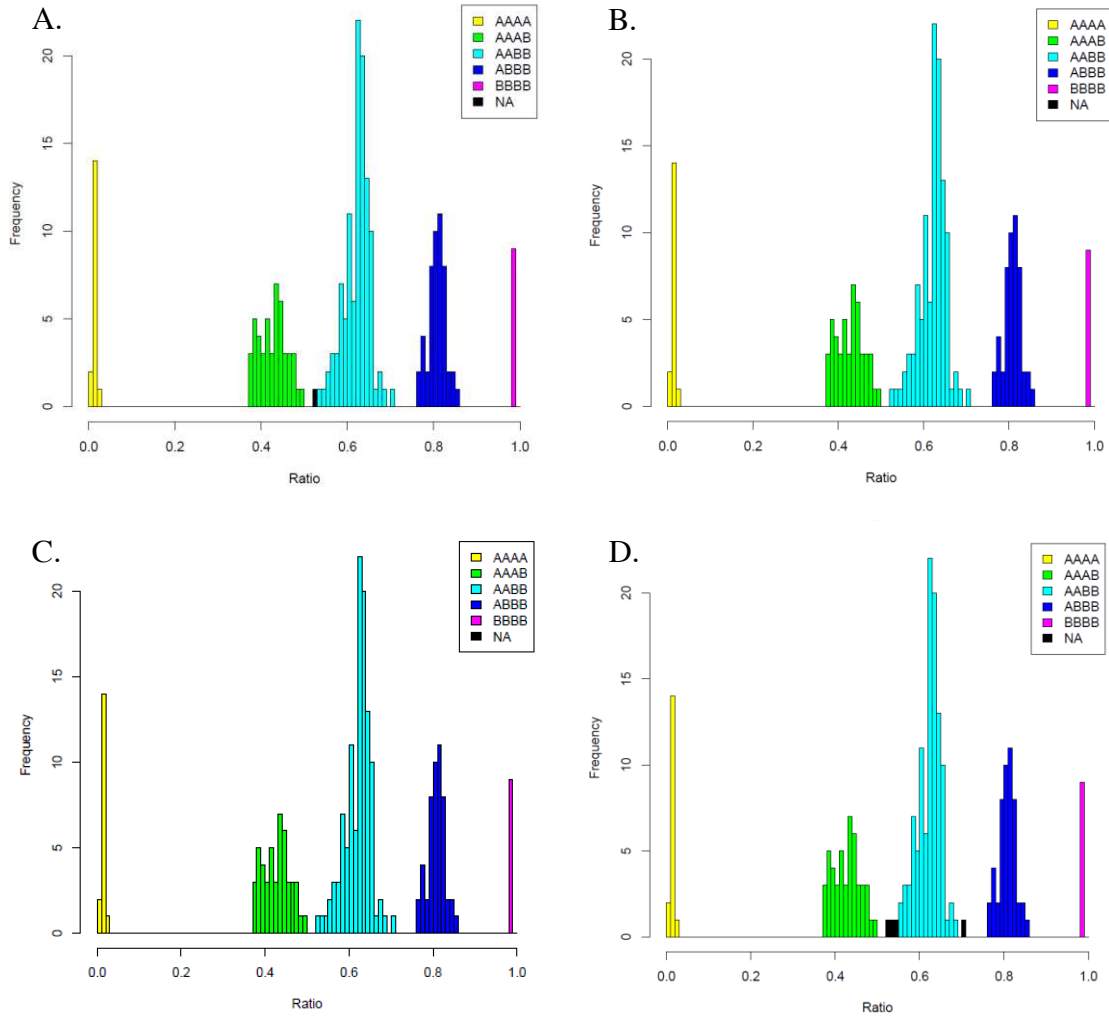
Figure 9. Histograms of the genotype calls generated using FitTetra (A), Hackett (B), NbClust (C), and SolCAP boundaries (D), for the SolCAP_c2_44634 SNP.

From Figure 9, all the methods were shown to have a good calling when genotypes were assigned for the SolCAP_c2_44634 SNP. Nevertheless, when looking at the rest of the SNPs that showed association with BR90 under short day conditions, some wrong assignments in the genotype calling were found in all the methods except FitTetra. These wrong SNP genotype callings could have misled the association analysis. For example, when analyzing SolCAP_c1_3001 SNP that was shown to be associated by using the Hackett genotyping method, this histogram showed how the genotype calls were clearly

not the most accurate (Figure 10A). This histogram showed that groups AABB and ABBB do not have a clear cluster separation and some of these genotype calls are not correct. In the case of NbClust method, one of the SNPs associated was SolCAP_c2_35998; this SNP calling histogram (Figure 10B) clearly shows a miscalling for the genotypes because it can only distinguish two genotypes.
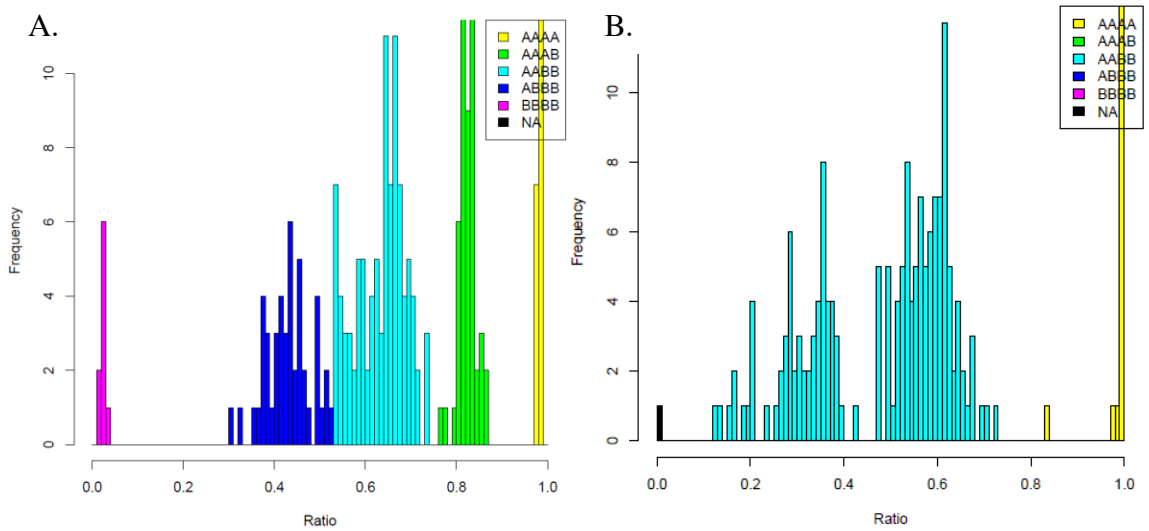
Figure 10. Genotypes miscalls for SolCAP_c1_3001 by Hackett (A), and SolCAP_c2_35998 by NbClust (B).

When applying the SolCAP established boundaries for genotype calling, due to the lack of boundary information for all the SNPs from the array, the number of informative SNP markers decreased dramatically. In addition, the boundaries determined are strict and do not allow for adjustments; therefore, some clear genotypes are assigned as no call (NA) (Figure 11). SolCAP_c2_38405 is one of the associated SNPs using the SolCAP boundaries genotyping. From Figure 11, it is evident that there are 5 different well defined clusters; however, due to the rigid establishment of the boundaries, the method does not generate calls for obvious genotypes in the test.
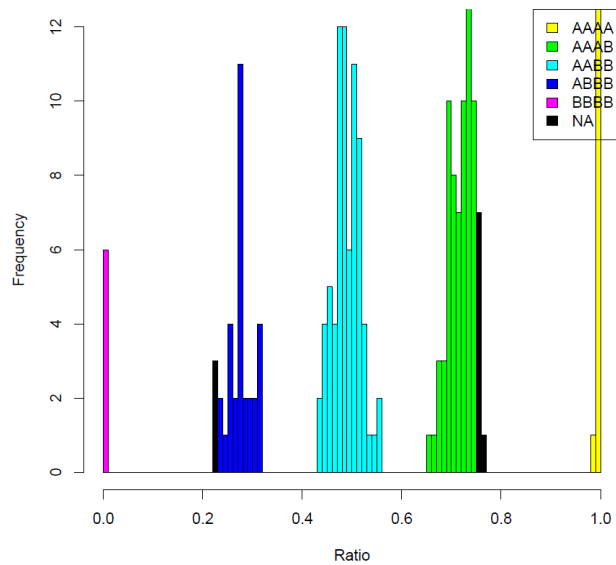
Figure 11. Histogram of genotypes calls for SolCAP_c2_38405 SNP using SolCAP boundaries available.

Based on the above comparisons, it is concluded that the most suitable method, for SNP genotype calling, among the four analyzed for this study was FitTetra. This method was shown to have reliable genotype calls and identified a good number (4,738) of informative SNP markers that give reasonable genome coverage for the study (Figure 12). Therefore, assigned genotypes generated by FitTetra for the 171 clones included in Experiments 1 and 2 were used for further analysis.
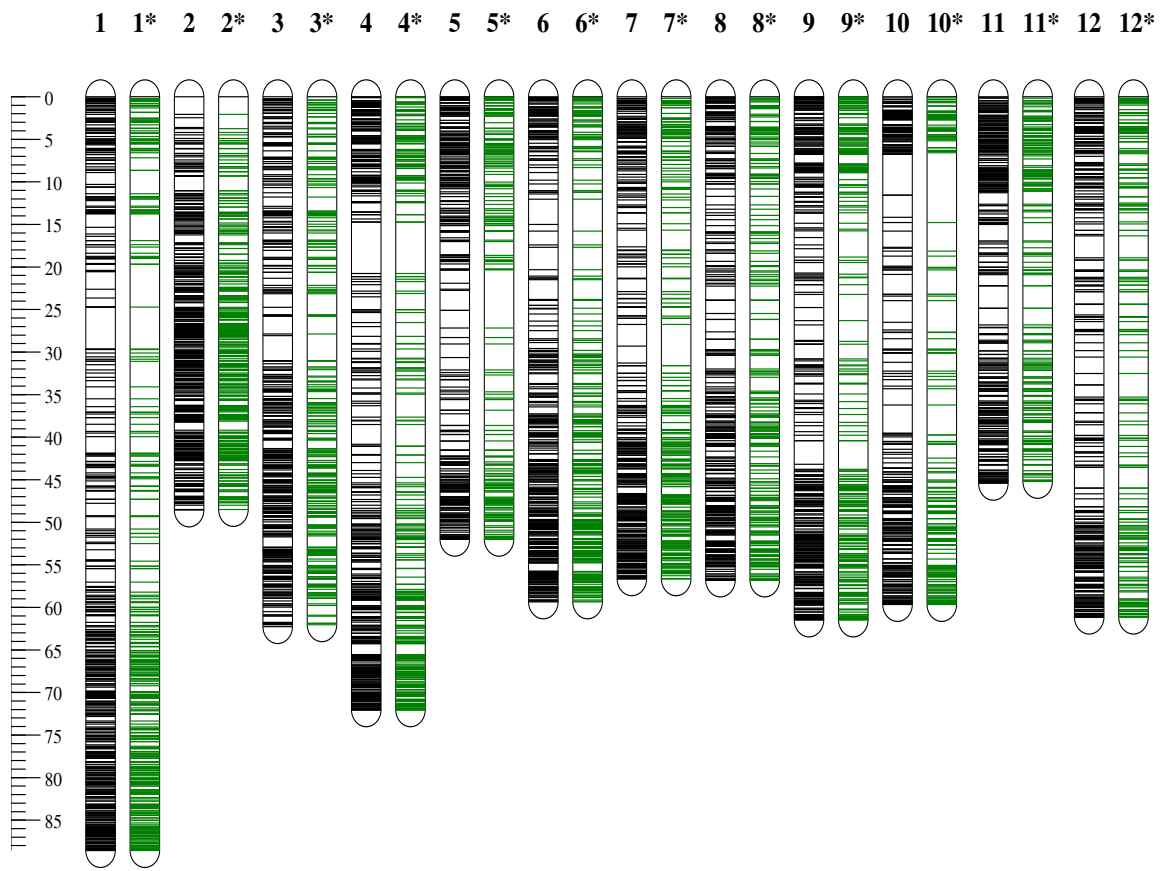
Figure 12. Distribution of 8,303 (black, 1-12) and 4,738 (green, 1*-12*) SNP markers on the 12 potato chromosomes. The scale shows the physical distance in Mb. Map positions are according to Felcher et al. (2012).

4.2.2   Genetic diversity analysis

Detailed information regarding relatedness and hidden population structure, as well as the extension of LD, are important prerequisites for association analysis. A total of 4,738 informative SNPs classified by FitTetra were used in the following sections.

4.2.2.1 Kinship

The covariance among individuals was described by the kinship based on the genetic similarities among individuals. From the genomic relationship matrix generated, a heat map was plotted in order to graphically see the level of relatedness among the individuals. The level of relatedness was found to be very low (close to zero) in this panel of 171 clones (Figure 13).

A low level of relatedness plays an important role in genome-wide association studies. When mapping a phenotype, it is important to know whether its response has variation correlated with the genetic relatedness among individuals. In this set of clones, there are no complex patterns of genetic relatedness among the individuals, which implies that there is no strong phenotypic-genotypic covariance. This low level of relatedness is an important element because the identification of associations across the genome with the phenotype will not describe the genetic relatedness among individuals. Otherwise, a high level of relatedness would be a problem when mapping traits where the variation of the phenotype is highly correlated with allele frequency differences (Flint-Garcia et al., 2005).
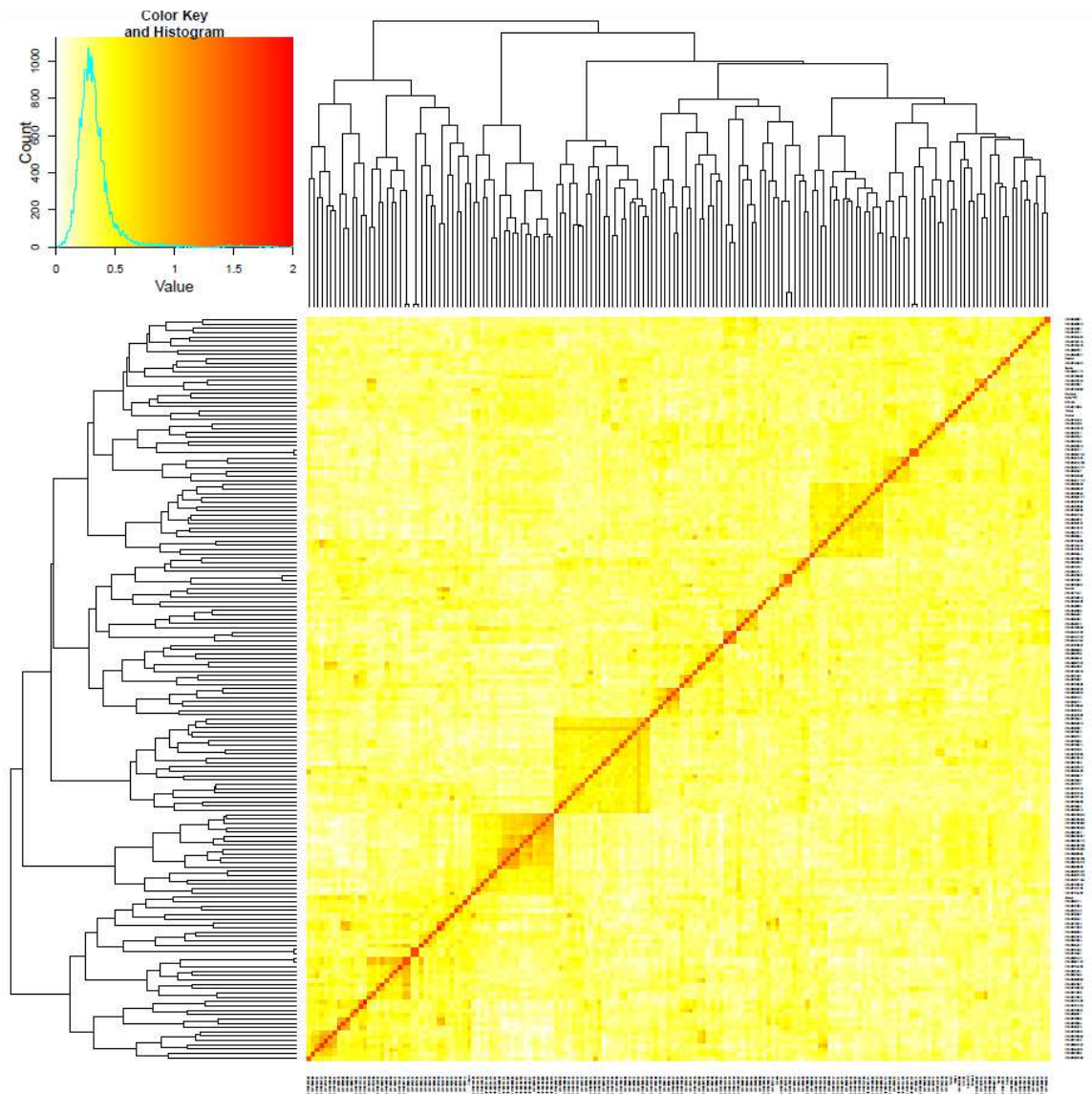
Figure 13. Heat map of the values from kinship matrix, using 4,738 SNP markers in 171 genotyped clones.

4.2.2.2 Structure and PCA analyses

The population structure was evaluated first with STRUCTURE software by using a subset of 120 SNP markers. These markers were chosen to represent physically distal independent markers, 10 SNP markers per chromosome, distributed along the 12 chromosomes. The results showed a continuous increase of the goodness of fit statistic [LnP(D)], which stands for the logarithm of the likelihood averaged over 20,000 iterations

versus the number of groups (K), increasing from the assumption of 1 to 10 groups (Figure 14). This results indicates that no optimal K was found, suggesting an unstructured or loosely structured population, when using this subset of markers. The most probable number of subpopulations was determined by plotting the natural logarithm (Ln) likelihood against the number of subpopulations (Figure 15). From Figure 15, the ΔK shows a clear peak at the true value of K, which for this study was two. Therefore, the most likely number of subpopulations determined from the STRUCTURE output was two subpopulations (K=2). From this finding, the barplot of K=2 is shown in Figure 16. This graph is the results of sorting the genotypes according to the probabilities (Q-values) of each genotype belonging to one of the two inferred subpopulations. The clones show a more prominent representation of subpopulation 2 (Q2).
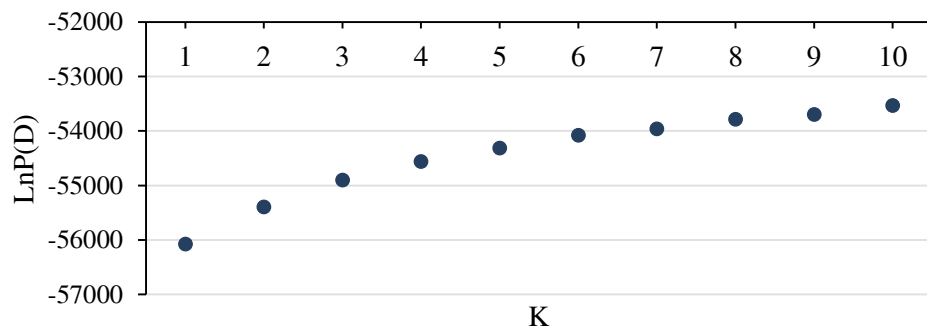
Figure 14. Goodness of fit, LnP(D), versus number of groups, K, plot for 171 clones from Experiment 1 and 2.
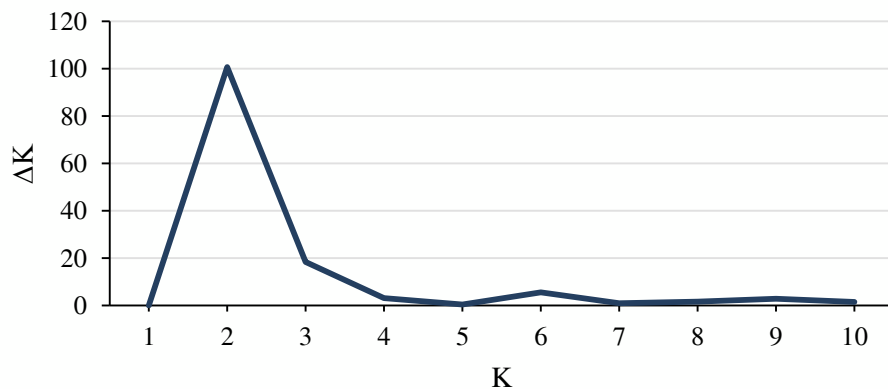
Figure 15. Calculation of delta K by $\Delta K = M([L''K])/s[L(K)]$, K=2.
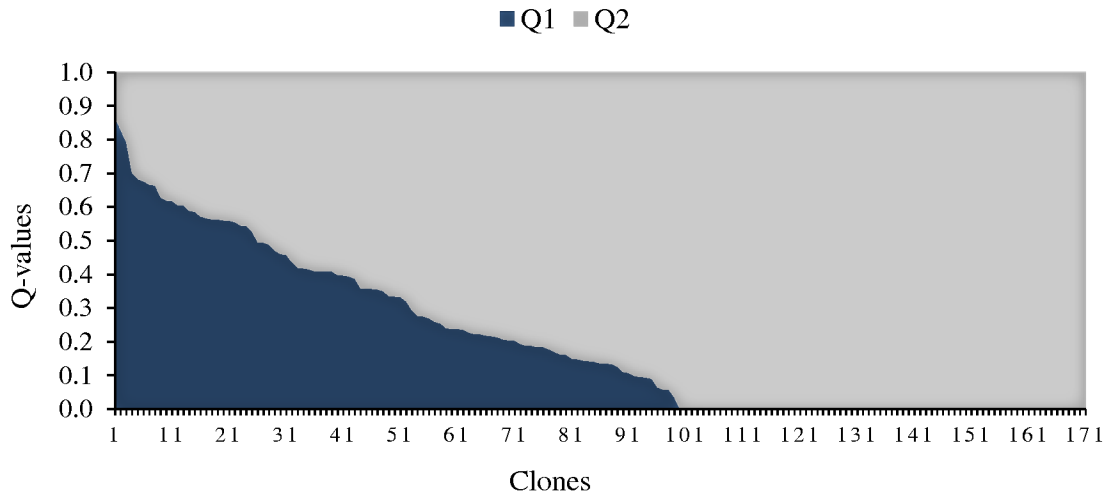
Figure 16. Barplot of population structure for 171 genotypes with two inferred subpopulations (K=2), with genotypes ordered according to Q-value to belong to subpopulation 1 (blue) or subpopulation 2 (light blue).

However, using the entire set of 4,738 SNP markers, a clear unstructured population was evidenced (Figures 17 and 18). By performing a PCA, this study revealed that the proportion of the variance explained by the first PC was slightly more that 5% of the total variance (Figure 17). This finding accounts for an unstructured population. Figure 18 illustrates the principal components of the analysis of the 171 genotypes, visualized using a scatterplot. This scatterplot has 2 axis based on the number of PC determined (PC1 and PC2). The color is used in an informative way based on the first two PCs of the PCA, recoded on the red and green color scales. The differences between the two PCs are represented in two complimentary ways: by the distances and colors. Most of the clones can be observed as scattered without any clustering trend. There is one clone that appears to be slightly separated and colored yellow and another group of potato clones colored green (Figure 18). The analysis of STRUCTURE and PCs showed the lack of population stratification in this study which makes it suitable for a GWAS.

69
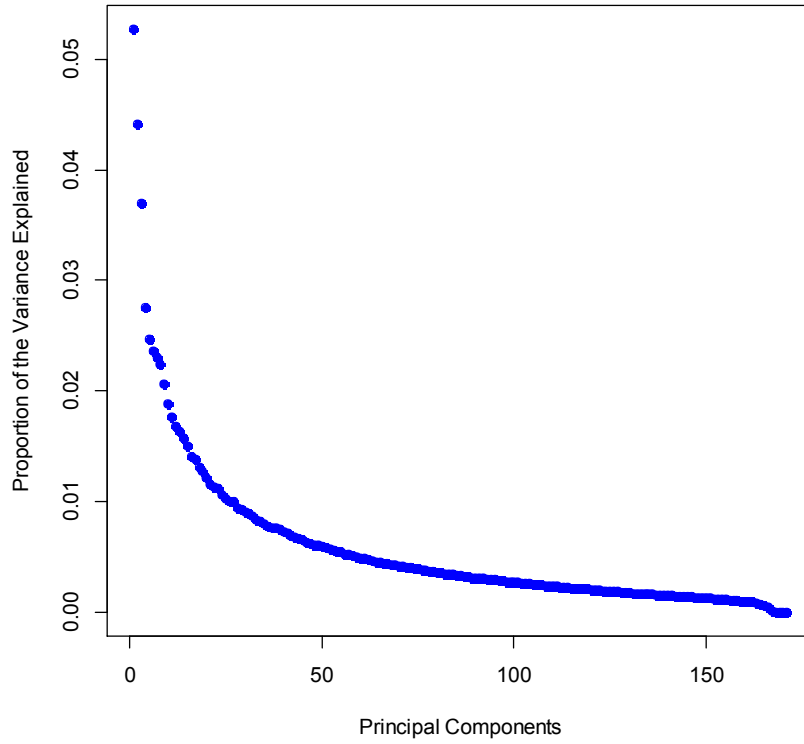
Figure 17. Proportion of the variance explained by each of the 171 PCs analyzed.
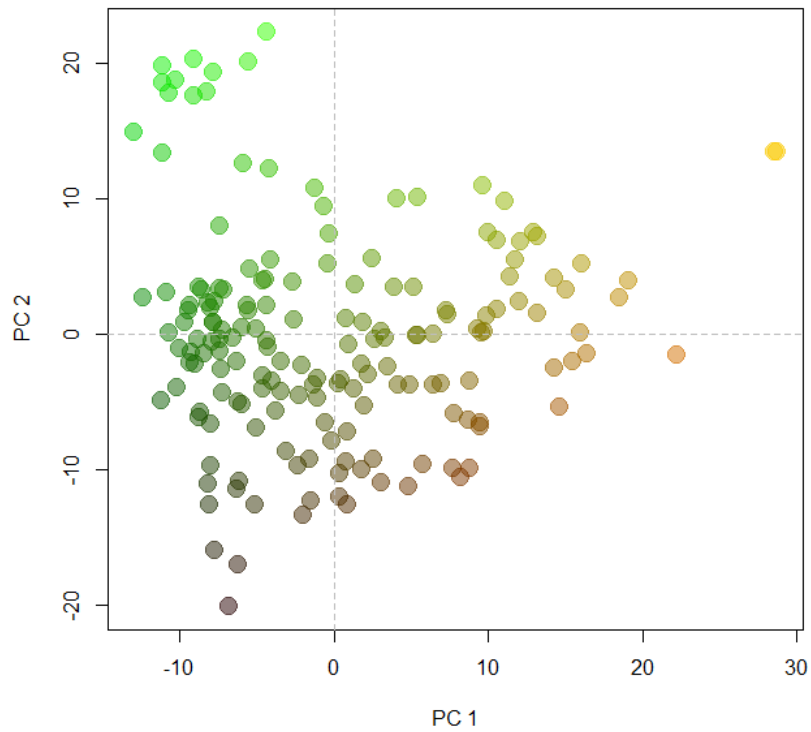
Figure 18. Scatterplot of the principal component analysis of 171 clones evaluated in Experiments 1 and 2.

4.2.2.3 Linkage disequilibrium

The genome wide LD was studied by using 4,738 SNP markers screened in a set of 171 potatoes breeding lines (Table 1) using the LD statistic r^2. The analysis suggested that LD decays below the 0.2 threshold when the genetic distance exceeds 0.3 Mb (Figure 19); this is the physical window size in linkage equilibrium flanking the causal polymorphism of the association in a GWAS.
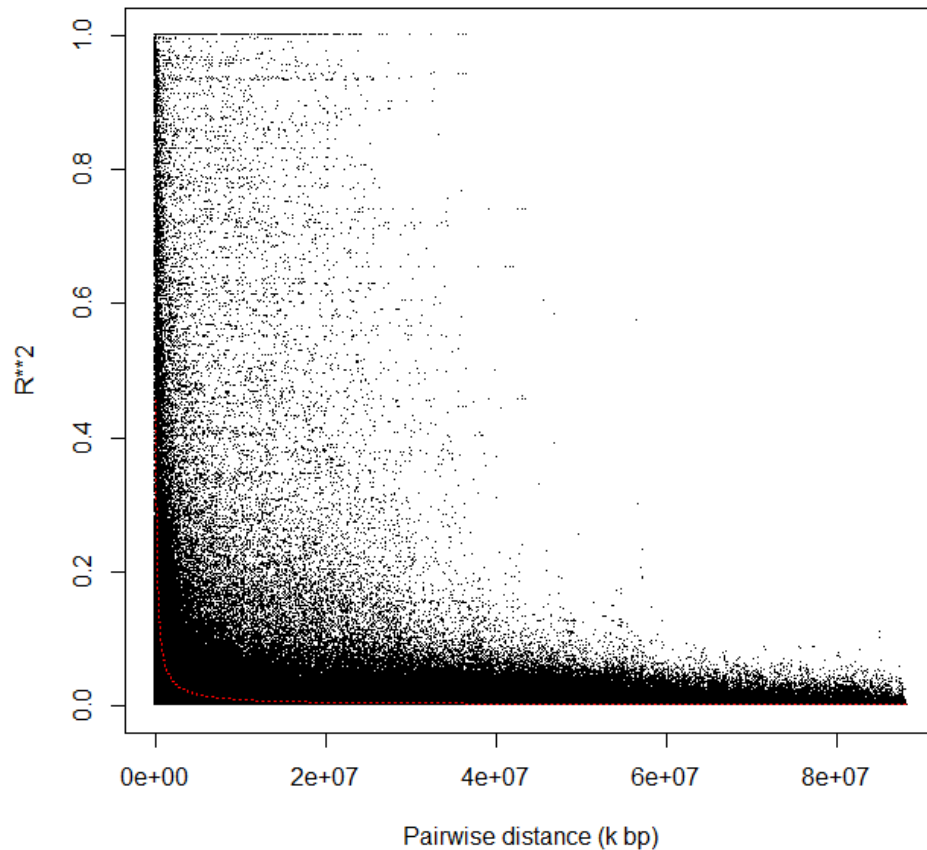
Figure 19. Linkage disequilibrium measure r^2 plotted vs. the physical map distance between all pairs of SNP markers calculated for all 171 tetraploid potato clones. The red line indicates the non-linear regression of r2 vs. the physical map distance between the SNP markers.

4.3 Genome wide association study

The identification and development of DNA markers that can be applied early in the breeding programs requires the genetic dissection of factors that control the tuber

development affected by photoperiod based on a genome wide linkage with molecular markers. This study identified regions underlying tuberization related traits affected by photoperiod based on marker effects estimated by GWAS analysis in two panels of breeding lines belonging to Experiment 1 and 2. SNP markers identified in Experiment 1, in most cases, did not fully match with the associations identified in the Experiment 2. The determination of the association significance was based on a p-value of 0.05; applying the Bonferroni correction, this p-value was divided by 300, which is in average, the number of informative SNP markers per chromosome. A total of 84 SNP markers, distributed in the 12 chromosomes, were identified to be associated with tuber induction, number of tubers, bulking ratio and stolon number and length.

The most significant SNPs detected for tuber induction were 22 SNP markers identified on chromosomes (Chr) 5, 7, 9 and 12 (Table A5). Under short day conditions, associated SNPs identified in Chr 5 were SolCAP_c2_50305 and SolCAP_c1_14840 within a region of 0.6 Mb; additionally, on Chr 12, SolCAP_c2_18855 was associated with 59 and 74 DAP in Experiment 2. One of the most important SNPs identified as associated with tuber induction under short day length was SolCAP_c2_40879 on Chr 9. This marker has been annotated for a CO gene (PGSC0003DMG400011378), well-known to be involved in photoperiodic responses (Martínez-García et al., 2002). When the tuber induction was evaluated under long day conditions, the main associations were also detected on Chr 5: one region involving two SNP markers 41 bp apart (SolCAP_c2_23833 and SolCAP_c2_23835), and SolCAP_c2_50305, which was also associated with this trait in short day conditions. In Chr 7, there were two SNPs highly associated, SolCAP_c2_33489 and SolCAP_c2_19826, in a region of 0.4 Mb.

In the case of the number of tubers, there were a total of 41 significant SNP markers identified according to the size of tuber evaluated: small (ST), marketable (MT), and total (TT) number of tubers (Table A6). For ST, the most relevant SNP markers were identified on chromosomes 6 and 11. Under short day conditions, associated SNPs identified in Chr 6 were SolCAP_c2_56145 and SolCAP_c2_8904 within a region of 0.1 Mb. Under long day conditions, the most significant SNP markers associated with ST were SolCAP_c2_13355 and SolCAP_c1_4328, 0.05 Mb apart on Chr 11. For MT, the most relevant SNP markers associated were identified on chromosomes 5, 6, 11 and 12. Under short day conditions, SolCAP_c2_6000 and SolCAP_c2_20947 were identified on Chr 11, and SolCAP_c1_8002 and SolCAP_c2_34762 on Chr 12. Under long day conditions, SolCAP_c2_50302 on Chr 5 was highly associated, as well as SolCAP_c2_25926 on Chr 6. When determining the associations for TT, the most relevant SNP markers associated were identified on chromosomes 1, 4 and 9. Under short day conditions, SolCAP_c2_24677 and SolCAP_c1_4803 were significantly associated on Chr 1, and SolCAP_c2_45035 on Chr 4. Under long day conditions, SolCAP_c2_51244, SolCAP_c2_26681, SolCAP_c2_55776 and SolCAP_c2_55773 were identified significant on Chr 4 in a region of 5 Mb. In addition, SolCAP_c2_3997 and SolCAP_c1_4228 on Chr 9 were associated within a region of 0.3 Mb.

For bulking ratio, there were 16 SNP markers identified on chromosomes 4, 5, 6 and 11 (Table A7). Under short day conditions, SolCAP_c2_11549 was detected on Chr 4; SolCAP_c2_41405, SolCAP_c2_56145 and SolCAP_c2_8904 were detected on Chr 6 in a region of 0.1Mb. Additionally, SolCAP_c2_4978 and SolCAP_c2_44634 were associated on Chr 11. Under long day conditions, SolCAP_c1_15106 and

SolCAP_c2_11569 were found associated on Chr 4; and SolCAP_c2_50302 and SolCAP_c2_10358 were detected on Chr 5.

Finally, for the number and length of stolons, there were 10 and 16 SNP markers associated, respectively. The most important associations were found on chromosomes 5, 7, 10 and 12. For stolon number under short day conditions, SolCAP_c2_33489 in Chr 7 was important, as well as SolCAP_c2_24564 in Chr 12. Under long day conditions, SolCAP_c2_27806 and SolCAP_c2_27808, 200 bp apart, were associated on Chr 10 (Table A8). For the stolon length, under short day conditions, SolCAP_c2_24556 was found associated on Chr 4. Under long day conditions, SolCAP_c2_47301, SolCAP_c2_47302 and SolCAP_c2_47303 were found associated on Chr 5 within a region of 130 bp; SolCAP_c2_27806 and SolCAP_c2_27808 on Chr 10 within a region of 100 bp and, SolCAP_c2_17617 and SolCAP_c2_17615 were associated on Chr 12 in a region of 200 bp (Table A9).

The comparison of the results from both experiments revealed differences in the associations detected, with a trend for Experiment 1 to deliver more associations that Experiment 2, since the number of individuals evaluated in Experiment 1 was higher. Figure 20 shows the associated loci detected with all the tuberization related traits for both experiments on the 12 potato chromosomes. Many of these SNPs were clustered within small genomic regions with significant effects on several traits, especially on chromosomes 4, 5, 6 and 11. SNP markers identified in this study are the starting point for marker-based selection of cultivars with good tuber production and adaptation to photoperiod.
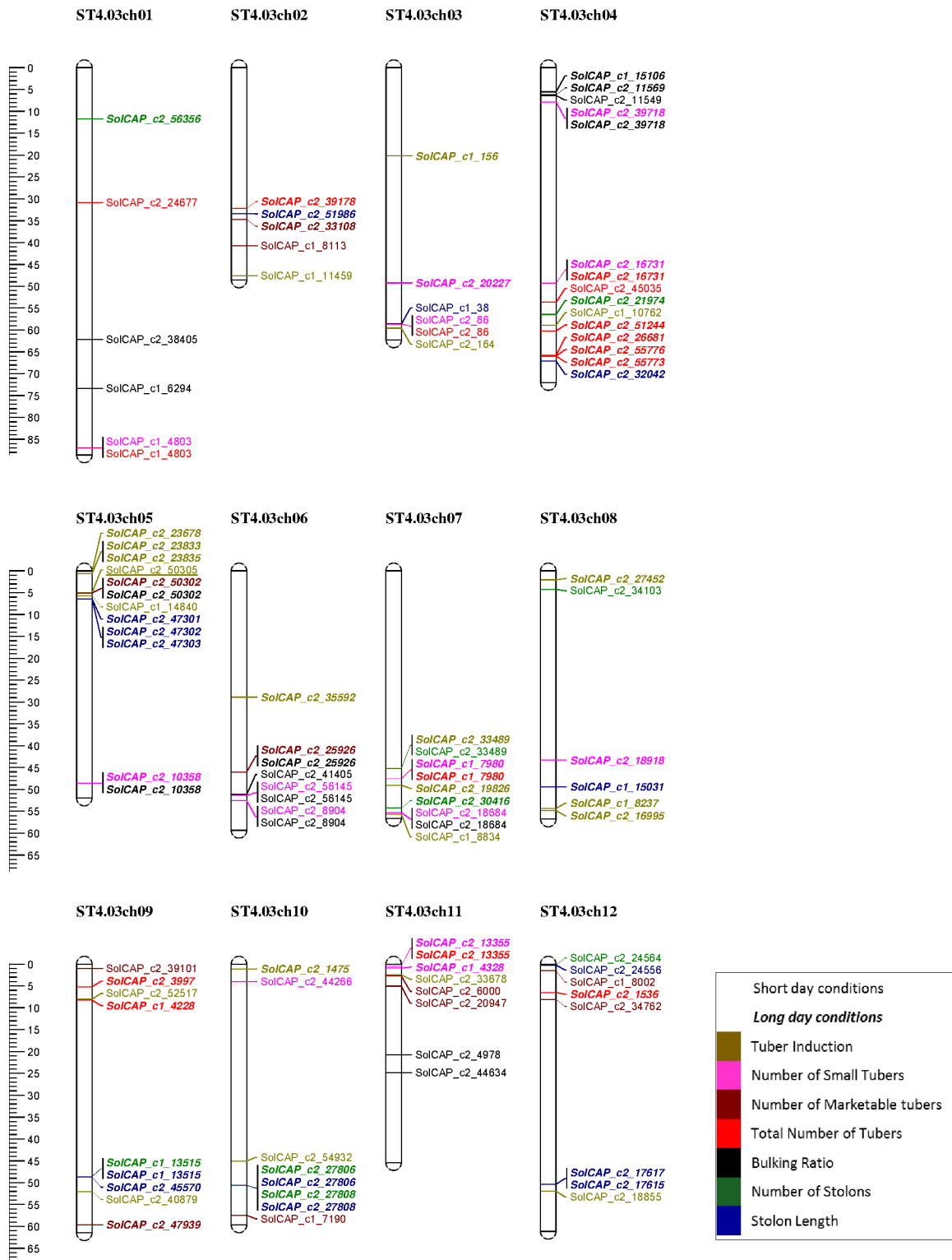
Figure 20. Physical map of 84 SNP markers associated with tuberization related traits, obtained from Experiments 1 and 2, under short and long day conditions.

4.4   Genomic selection

This study evaluated the potential of GS for tuberization related traits under short and long day-length environmental conditions. The prediction model used a training population of 130 breeding lines from Experiment 1, and 41 breeding lines as the validation population, exclusively from Experiment 2. By using the values from the training population, the model employed predicted the genotypic value for the inference population. The correlations were calculated between the predicted genotypic value and observed phenotypes for the prediction population, as well as the cross-validation accuracy. This analysis was performed separately for short (Table 10) and long days (Table 11). The R-package rrBLUP, generated accuracies of the predictions for two methods, GAUSS and RR.

Table 10. Cross-validation accuracies for tuberization related traits under short day conditions.

Trait	DAP	Training Population n=130		Inference Population n=41	
		RR	GAUSS	RR	GAUSS
Tuber Induction	41	0.92	0.98	0.18	0.17
	59	0.89	0.97	0.34	0.39
	74	0.91	0.98	0.15	0.22
Number of Small Tubers	75	0.84	1.00	0.21	0.18
	90	0.87	1.00	0.43	0.40
Number of Marketable Tubers	75	0.89	0.93	0.15	0.19
	90	0.86	0.93	0.38	0.40
Number of Total Tubers	75	0.84	0.99	0.05	0.01
	90	0.81	1.00	0.25	0.24
Bulking Ratio	75	0.89	0.94	0.50	0.49
	90	0.87	0.92	0.57	0.55

In some cases for the tuberization related traits under short day-length (Table 10), the accuracy with GAUSS was higher than RR; however, some other cases showed the contrary. Across the 11 phenotypic traits evaluated, the one with the highest values for prediction was bulking ratio, with an accuracy of 0.57, followed by the number of small and marketable tubers, with an accuracy of 0.43 and 0.40, respectively. Therefore, these traits were predicted quite well. The total number of tubers was the trait with the lowest correlations between the predictions and the real phenotypic data; this leads to the conclusion that TT might not be as good as trait for genome prediction.

Under long day conditions, there were 13 tuberization related traits evaluated for prediction (Table 11). The trait with the highest values for prediction was tuber induction, with an accuracy of 0.42, followed by bulking ratio, with 0.40. Other traits with good accuracy for prediction were the number of marketable tubers, as well as the number and length of stolons. In general terms, the accuracy of the phenotypic traits evaluated in this study were lower when the breeding lines were exposed to long photoperiod.

Since the accuracy of the predictions was high under both short and long day lengths, the generated GEBV were employed in the selection of the best clones in the inference population. The first 10 clones with the highest RR GEBV for all the tuberization related traits are shown in Tables 12 and A10-14 for short and long photoperiod. From these tables, the selection of clones that are expected to have a good performance under both photoperiods can be determined. Based on BR90 GEBV (Table 12), clones CIP-394223.19 and CIP-301023.15 are expected to have a good performance under both photoperiods, and they are recommended for use in future breeding programs.

Table 11. Cross-validation accuracies for tuberization related traits under long day conditions.

Trait	DAP	Training Population n=130		Inference Population n=41	
		RR	GAUSS	RR	GAUSS
Tuber Induction	41	0.90	0.95	0.14	0.15
	59	0.87	0.99	0.38	0.42
	74	0.84	0.94	0.24	0.26
Number of Small Tubers	75	0.95	1.00	-0.18	-0.18
	90	0.94	1.00	0.01	0.02
Number of Marketable Tubers	75	0.93	0.98	0.31	0.33
	90	0.96	0.99	0.25	0.25
Number of Total Tubers	75	1.00	1.00	-0.11	-0.05
	90	0.98	1.00	-0.07	-0.05
Bulking Ratio	75	0.93	0.99	0.38	0.37
	90	0.96	0.99	0.42	0.40
Stolon Number	75	0.91	1.00	0.06	0.11
	90	1.00	1.00	0.29	0.31
Stolon Length	75	0.95	1.00	0.23	0.27
	90	0.95	1.00	0.30	0.32

Table 12. Top 10 clones for bulking ratio (BR90) based on RR GEBV under short or long day conditions.

Clones with best GEBV under short day conditions		Clones with best GEBV under long day conditions	
Clone	RR GEBV	Clone	RR GEBV
CIP-394223.19	8.16	CIP-394223.19	8.50
CIP-301023.15	7.69	CIP-397077.16	7.20
CIP-301024.14	6.75	CIP-301026.23	5.13
CIP-301024.95	6.27	CIP-394223.9	4.28
CIP-300065.4	2.94	Granola	2.59
CIP-301029.18	2.39	Spunta	2.35
CIP-300055.32	1.62	CIP-300065.4	1.67
CIP-394223.9	1.54	CIP-301023.15	1.66
CIP-300137.31	1.22	CIP-300135.14	1.15
CIP-398208.29	0.58	CIP-300137.31	1.09

CHAPTER 5: DISCUSSION

This study phenotypically evaluated a set of 171 potato clones generated by CIP in two different field experiments to identify the chromosomal locations underlying tuberization related traits affected by photoperiod using SNP markers. A genome wide association analysis was performed using the Potato SolCAP SNP array. In addition, this study evaluated the new genomic selection approach applied on tuberization related traits in potatoes to accurately select superior progenitors based on their genome estimated breeding values.

This study mainly focused on the evaluation of the influence of day length on the tuberization related traits in potatoes, a plant originating in short day regions. Photoperiod has been previously reported in potatoes to affect stem and leaf sizes, as well as stolon and tuber initiation (Haverkort, 2007). The effects of extending the photoperiod on potatoes have been shown to delay the stolon and tuber initiation, as well as to reduce the tuber size and number, resulting in inconsistency of tuber growth and lower numbers of tubers (Ewing, 1978). The populations in this study were phenotyped for TI, ST, MT, TT, BR, SN and SL. In all seven phenotypes, this study demonstrated that the day length (12 hours vs. 16 hours) affected the tuberization related traits. TI was enhanced in Experiment 1 under short days in concordance with a previous study which stated that as photoperiod increases, the tuberization becomes irregular, then delayed and finally, inhibited (Haverkort, 2007; Lagercrantz, 2009; Moore, 1920; Simpson, 2003). The same result was observed when comparing the number of tubers in both photoperiods. Under long day length, the number of small tubers increased, as well as the overall number; however, short days showed an

increase in the number of marketable tubers, which suggests that the potato clones studied were more adapted to short days compared to long days.

Another factor evaluated in the phenotypic trials was the earliness of tuberization because the two experiments were evaluated 75 and 90 DAP. The study clearly showed that when DAP was extended from 75 to 90, the production of tubers, as well as the tuber initiation, was increased. This finding supports the study from Kooman et al. (1996), in which they demonstrated that early tuber initiation led to earlier maturing crops. The differences between short and long photoperiod responses were shortened when the sampling time was extended. At early harvesting (75 DAP), the differences between short and long day lengths were greater, with regard to MT. It is well known that long nights (short days) promote tuber formation in potato cultivars (Abelenda et al., 2014), and this study obtained similar results. However, Experiment 2 showed some discrepancies in these findings (Figure 8). The phenotypic differences in all the tuberization related traits, between short and long day lengths, were diminished in Experiment 2. These discrepancies may be related to the fact that field Experiment 2, performed in 2012, was planted a month later in comparison to Experiment 1 in 2010. This change meant an increase in the seasonal temperature. This external environmental factor is well known to affect the growth and development of potatoes (Haverkort, 2007; Pereira and Shock, 2006) and it also explains the differences in the performance of the seven tuberization related traits measured in this study. Therefore, this increase in temperature during the time of potato growth in Experiment 2 may have affected the tuberization in such a way that the differences between short and long photoperiods were reduced. This study demonstrated that the higher

temperature stimulated more rapid development, making the differences between short and long days non-significant for most of the tuberization related traits, except for SN and SL.

This study estimated and compared the heritability of the tuberization related traits and the correlations between the traits. The phenotypic data analysis demonstrated several correlations between the seven traits. A comparable connection between tuberization related traits existed, as shown by the high correlation between tuber induction and the number of tubers. The consistently high correlations were between tuber induction and the number of marketable tubers, as well as bulking ratio. This was expected since it has been previously demonstrated that early tuber induction results in an increase in tuber production (Haverkort, 2007). Most of the tuberization related traits were positively correlated; for instance, the number of small tubers with the total number of tubers and the number of marketable tubers with the bulking ratio. However, tuberization is negatively correlated with the stolon number and length (Table 4 and 6). An explanation for this finding is that the higher number and length of stolons describes the lack of tuber formation, since the potato clones were not able to transit from the stolon stage to the tuber stage (Ewing, 1978).

The heritability estimates obtained for both experiments were high on average (Tables 7a and 7b). This may be due to the fact that this study evaluated two environments (short and long day length). Therefore, the genetic variation expressed in heritability was better estimated, due to the balancing effect between the two different photoperiods. The heritability estimates in Experiment 2 were higher. These values could be inflated since this population was smaller and had fewer differences between photoperiodic responses, but they had higher variability, which could be due to the increase in temperature in the experimental trial. Phenotypic values of traits with high heritability estimates are expected

to have stronger associations with genetic markers (Holland et al., 2003; Massa et al., 2015), since most of their variations are dependent on the genetic effects, which are important for GWAS and GS.

Genome wide association analysis and genomic selection are now possible in potatoes, with the availability of the Potato SolCAP SNP array (Felcher et al., 2012). This SNP array provides a marker density sufficient to generate genetic maps to identify associations for agronomic traits in potatoes (Douches et al., 2014). Several methods for assigning potato clones to a genotype class have been studied: FitTetra (Voorrips et al., 2011); NbClust (Charrad et al., 2014); the method described by Hackett (Hackett et al., 2013) and the available SolCAP boundaries (http://solcap.msu.edu/potato_infinium.shtml). This assignment is necessary to take full advantage of the potato array technology. This study compared the four approaches available for genotype calling in tetraploid potatoes. These results from the total number of informative SNP markers with assigned genotypes showed that the SolCAP boundaries method did not provide sufficient information (only 25% were genotyped) for these SNP markers (Table 8). Even though these SNP calling boundaries were previously established and reliable, they dramatically decreased the amount of information generated from the Potato SolCAP SNP array. On the other hand, the FitTetra method was demonstrated to be the most suitable SNP calling method for the population studied. This approach is based on a mixture model (Voorrips et al., 2011). Using the allele signal ratio (theta values), this method fitted a mixture of five normal distributions, representing the five possible genotype classes. This model automatically assigned the genotype classes because it modeled the component means as a function of the allele ratios for each SNP. This study identified 4,738 informative SNP markers that

were well distributed in the chromosomes of potatoes (Figure 12). The genotype calls from this method are reliable since it uses the relation between the allele ratios and means of the distributions of each genotype, even when the distributions of these classes overlap considerably. This feature ensures a minimum level of genotype miscalling.

Another important aspect in GWAS and GS is the study of the genetic diversity of the populations used. This study evaluated the genetic diversity across the 171 clones in three different aspects: relatedness between individuals, population structure and LD. The level of relatedness in the population of study was found to be considerably low. This is convenient for a GWAS, since genetic relationships can create false signals when determining the associations (Yu et al., 2006). The second aspect, population structure, was evaluated using two different methods. The first method used a subset of SNP markers (120 SNPs), distributed along the 12 chromosomes, in order to determine the structure of the population. This method determined that the clones studied can be grouped into two subpopulations. However, these results were not definitive, since 120 SNPs are a very low fraction of the total of 4,738 informative SNPs available for the study. The second method used all the SNP markers to perform a PCA (Ma and Amos, 2012). The PCA analysis indicated that the PCs accounted for a very low percentage of variability in the population. Both population structure results illustrated that the population studied was suitable for GWAS, since it was considered to be relatively unstructured, meaning that the potato clones are unrelated or the genetic relationship among them are very low. The third aspect evaluated for the genetic diversity was the LD. In potatoes, LD decay has been shown to vary from less than 1 cM (Gebhardt et al., 2004), to up to 10 cM (Simko et al., 2004). This wide range of LD decay in the previous studies has been based on a limited amount of

marker information. With the availability of the Potato SolCAP SNP array, this study determined the LD decays on 0.3 Mb (Figure 19). However, a previous study indicated that the r^2 reached a value of 0.2 within about 0.1 Mb distance (Simko et al., 2006), which suggested a faster LD decay, compared to this study. The difference in results relies on the fact that SNPs on the Potato SolCAP SNP array are distributed genome-wide, with larger physical distances, compared to the SNP markers used in Simko's study. In addition, for the correction of the r^2 estimation, this study included the sample structure, as well as the relatedness between genotyped individuals which is a novel measurement (Mangin et al., 2012). This was a positive addition, since it has been demonstrated that individuals from different genetic origins within the population studied could mislead the calculation of the LD, based on their differing allele frequencies. Therefore, the results obtained in this study are considered more accurate.

The GWAS model applied in this research used a mixed model, accounting for the population structure, based on the 4,738 SNP markers and the relatedness between the individuals in the population studied. For this purpose, the study employed two panels in order to validate the associations identified. The first panel identified SNP markers that, in most cases, did not fully match with the associations identified in the validation panel. However, all associations can be considered valid, since most of them are in nearby genomic regions. Understanding the genetic base to explain how the potato crop responds to photoperiod involves the study of different tuberization related traits, since photoperiod affects tuberization directly. This study first created a listing of 22 genes responsible for photoperiod responses that were well described in the plant model, *Arabidopsis thaliana* (Table A15). After collecting these data, the sequences of these selected genes were

downloaded from https://www.arabidopsis.org/servlets/ to finally BLAST them on the potato genomics resource website (http://solanaceae.plantbiology.msu.edu/blast.shtml). From this procedure, the positions of the genes involved in photoperiodic responses from *Arabidopsis thaliana* were located in the potato genome. They were then compared with the identified chromosomal locations associated with tuberization related traits affected by photoperiod. This comparison was focused on analyzing the proximity of these genes to associations identified in this study. For this purpose, a physical genetic map was generated, to overlap the SNP markers identified in this research with the reported genes in *Arabidopsis thaliana* (Figure 21).
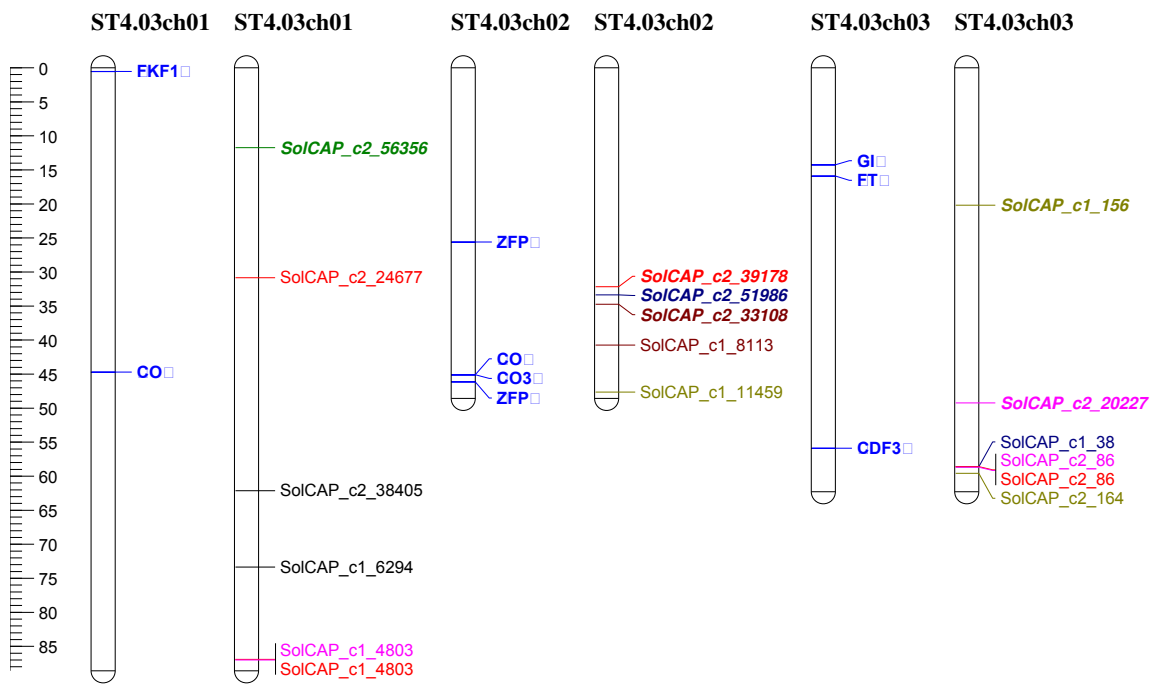
Figure 21. Physical map of the genes underlying photoperiod responses (blue) described in the plant model, *Arabidopsis thaliana* and the SNP markers identified in this study to be associated with tuberization traits responses under short and long day lengths.

Figure 21. Continuation: Physical map of the genes underlying photoperiod responses (blue) described in the plant model, *Arabidopsis thaliana* and the SNP markers identified in this study to be associated with tuberization traits responses under short and long day lengths.

Figure 21. Continuation: Physical map of the genes underlying photoperiod responses (blue) described in the plant model, *Arabidopsis thaliana* and the SNP markers identified in this study to be associated with tuberization traits responses under short and long day lengths.

Based on this information, the proximity of identified SNPs associated with previously reported genes underlying photoperiod responses suggests that the SNP markers may be detecting associations with some of these genes. In addition, there were other chromosomal locations annotated for genes which may control these tuberization related traits, such as: SolCAP_c1_10762, associated with TI under short day conditions, is annotated for a photoperiod responsive protein in Chr 4; SolCAP_c2_27452, associated with TI under long day conditions, is annotated for a FLAVIN-BINDING KELCH protein in Chr 8, known to down-regulate CDF genes, and most importantly; SolCAP_c2_40879, associated with TI under short day conditions, is annotated for a CO gene in Chr 9. Further

studies will be required to validate the genes associated with these tuberization related traits affected by photoperiod.

One major aspect in the identification of markers associated with single large effects is that few markers often will not explain the phenotypic response of interest. Fortunately, GS offers the opportunity of estimating all markers' effects by using novel statistical methods (Jannink et al., 2010). Endelman (2011) developed an approach to calculate the GEBV based on SNP marker datasets, calculated kinships and phenotypic data, which in the case of this study were the tuberization related traits. GS had proven to be useful in other crops, such as wheat and maize (Crossa et al., 2010), using a model that included marker information and pedigrees. Conclusions from these studies indicated that GS selection in plant breeding is an effective strategy. They based this conclusion on the high correlations they obtained between observed and predicted values; therefore, this approach can be an effective strategy for selecting among lines whose phenotypes have yet to be observed. Our study demonstrated that GS is also useful for predicting most of the phenotypes for the validation population under short and long day length, since the correlations were relatively high under both short and long day length (Tables 10 and 11). Under short day conditions, bulking ratio was demonstrated to be a good trait for prediction; on the other hand, the total number of tubers had the lowest prediction accuracy. This result shows that the total number of tubers cannot be well-predicted since the markers used do not fully describe the performance of a clone. The total number of tubers includes all tubers, regardless of their size, which makes it an unstable trait. Under long day conditions, tuber induction was the trait with the highest prediction accuracy, followed by bulking ratio. In general terms, tuber induction and bulking ratio have been demonstrated to be good traits

to be predicted, reflecting on their high accuracies. The predictions for tuberization related traits when exposed to both photoperiods, and their accuracies were mostly appropriate for GS. Therefore, these traits can be predicted by GS and used in breeding for earliness and adaptation to short or long day-length environments. Prediction accuracies can be improved by increasing the size of the training population (Crossa et al., 2010), but it cannot be expected to have a significant improvement, due to the heritability of some traits associated with yield components, such as those evaluated in this study.

A better understanding of the influence of photoperiod over tuberization has been described in this study, as well as the genetic factors involved in these responses to day length. For further studies, the chromosomal locations associated with tuberization related traits in potatoes grown under short and long day length provide a starting point for the identification of genes involved in these responses. In addition, this study also evaluated how the recently described GS could lead to new advances in potato breeding in diverse locations since it showed good accuracy in the prediction of tuberization related traits in potatoes.

CHAPTER 6: CONCLUSIONS

This study determined that long photoperiods reduce significantly the tuber induction as well as the production of marketable tubers, and it increases the number of small tubers as well as the number and length of stolons. The Potato SolCAP SNP array was successfully used to examine the genetic diversity as well as genetic relationships in tetraploid potato clones. By using this information, this study also performed a genome wide association analysis to detect chromosomal locations to explain tuberization related traits in potatoes affected by photoperiod. Using the Potato SolCAP SNP array, this study selected a total of 4,738 SNP markers for various analyses. A total of 84 SNP marker-trait associations in the potato genome from a collection of 171 potato clones were identified. Several associated SNP markers were clustered within small genomic regions and some are at the same locations as previously reported. These SNPs may provide insight on photoperiod responses in potatoes. This study also can be viewed as a case study for genome wide association analysis in potatoes. The outcome of this work demonstrated that genome wide association analysis is a good approach for complex traits even if it is unable to capture minor gene effects. This study also demonstrated how genomic selection can better predict breeding values for the tuberization related traits and be useful in the selection of clones for potato breeding purposes. As well, the identification of the best clones as parents could be used immediately for breeding programs.

APPENDIX

Table A1. Analysis of variance of day length, DAP and clones for tuberization traits related in Experiment 1.

Parameter	Tuber Induction	Number of Small tubers	Number of Marketable Tubers	Number of Total Tubers	Bulking Ratio
Day length	<.0001*	<.0001*	<.0001*	<.0001*	<.0001*
DAP	<.0001*	<.0001*	<.0001*	<.0001*	0.5050
Clone	<.0001*	0.0010*	<.0001*	<.0001*	<.0001*
Day length*DAP	<.0001*	<.0001*	<.0001*	<.0001*	0.7790
Day length*Clone	<.0001*	0.9180	<.0001*	0.3610	0.1210
DAP*Clone	<.0001*	0.9940	1.0000	1.0000	1.0000
Day length*DAP*Clone	<.0001*	1.0000	1.0000	1.0000	1.0000

*P-values are significant at 5 % level of significance.

Table A2. Analysis of variance of day length, DAP and clones for tuberization traits related in Experiment 2.

Parameter	Tuber Induction	Number of Small tubers	Number of Marketable Tubers	Number of Total Tubers	Bulking Ratio	Stolon Number	Stolon Length
Day length	<.0001*	<.0001*	<.0001*	0.0240	<.0001*	<.0001*	<.0001*
DAP	<.0001*	0.0010*	<.0001*	0.2200	<.0001*	<.0001*	<.0001*
Clone	<.0001*	<.0001*	<.0001*	<.0001*	<.0001*	<.0001*	<.0001*
Day length*DAP	<.0001*	0.5740	<.0001*	<.0001*	<.0001*	<.0001*	<.0001*
Day length*Clone	<.0001*	<.0001*	<.0001*	<.0001*	<.0001*	0.0170	<.0001*
DAP*Clone	<.0001*	0.1480	0.0070*	0.4630	0.0520	0.8910	0.7890
Day length*DAP*Clone	<.0001*	0.3080	0.1870	0.2570	0.8230	0.9730	0.9990

*P-values are significant at 5 % level of significance.

Table A3. Data normalization in Experiment 1.

Response variable	Day length	DAP	Normality test p-value	Box-Cox transformation λ	Normality test after transformation	Stage of Variable for further analysis
Tuber Induction	12	41	<0.005	0.00	<0.005**	Transformed
		59	<0.005*	-0.50	<0.005	Original
		74	0.013	1.00		Original
	16	41	<0.005*	1.00		Original
		59	<0.005	0.00	0.119	Transformed
		74	<0.005	0.00	<0.005**	Transformed
Number of Small Tubers	12	75	<0.005	0.50	0.876	Transformed
		90	<0.005	0.00	0.717	Transformed
	16	75	<0.005	0.50	0.771	Transformed
		90	<0.005	0.50	0.616	Transformed
Number of Marketable Tubers	12	75	0.795			Original
		90	0.645			Original
	16	75	<0.005	0.34	<0.005**	Transformed
		90	0.007*	0.50	<0.005	Original
Total Number of Tubers	12	75	0.011	0.50	0.066	Transformed
		90	0.007	0.50	0.333	Transformed
	16	75	0.017	0.50	0.052	Transformed
		90	0.013	1.00		Original
Bulking Ratio	12	75	0.017	1.00		Original
		90	0.028	1.00		Original
	16	75	<0.005*	0.36	<0.005	Original
		90	0.007*	0.50	<0.005	Original
Stolon Number	16	75	0.044	0.50	0.068	Transformed
		90	<0.005	0.50	0.487	Transformed
Stolon Length	16	75	0.007	0.50	0.119	Transformed
		90	0.115			Original

* Refers to normality test that were not achieved by the p-value, but they passed the pen test for normality.
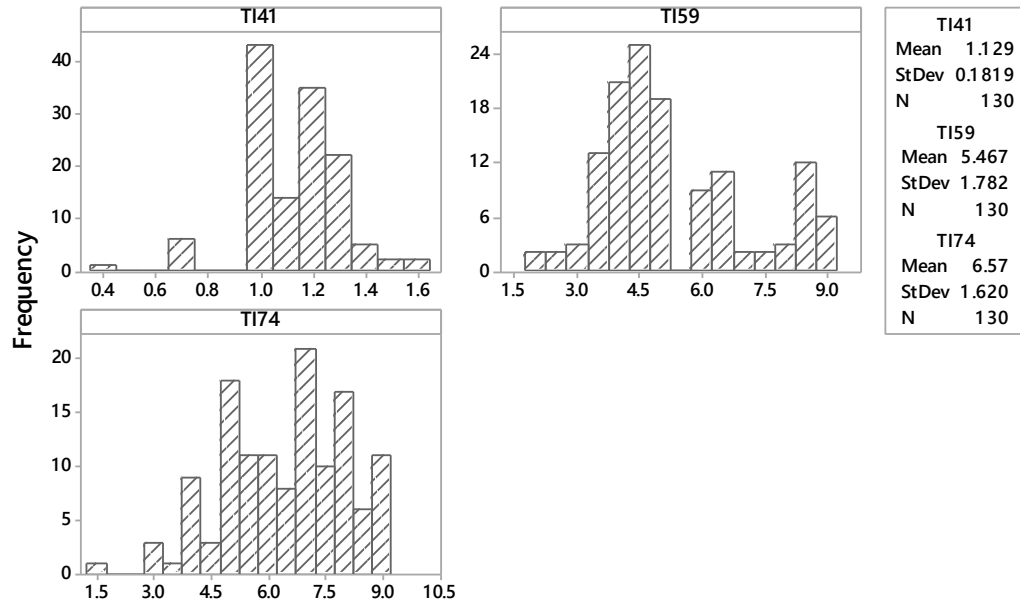
Figure A1. Distribution of tuber induction evaluated 41, 59, and 74 DAP in the 130 clones from Experiment 1 grown under short day conditions. Histograms show the phenotype distribution.
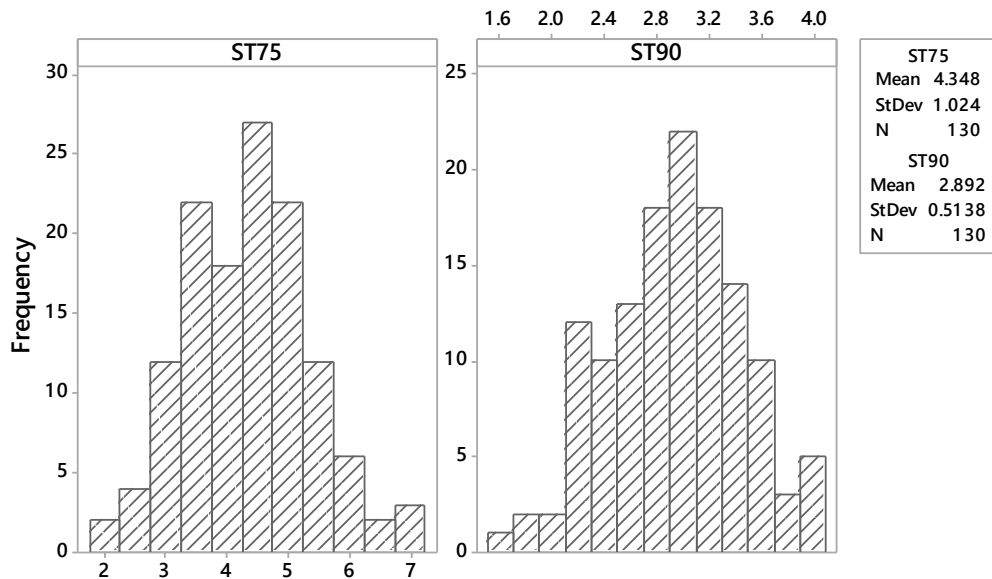
Figure A2. Distribution of the number of small tubers evaluated 75 and 90 DAP in the 130 clones from Experiment 1 grown under short day conditions. Histograms show the normality of the phenotype distribution.
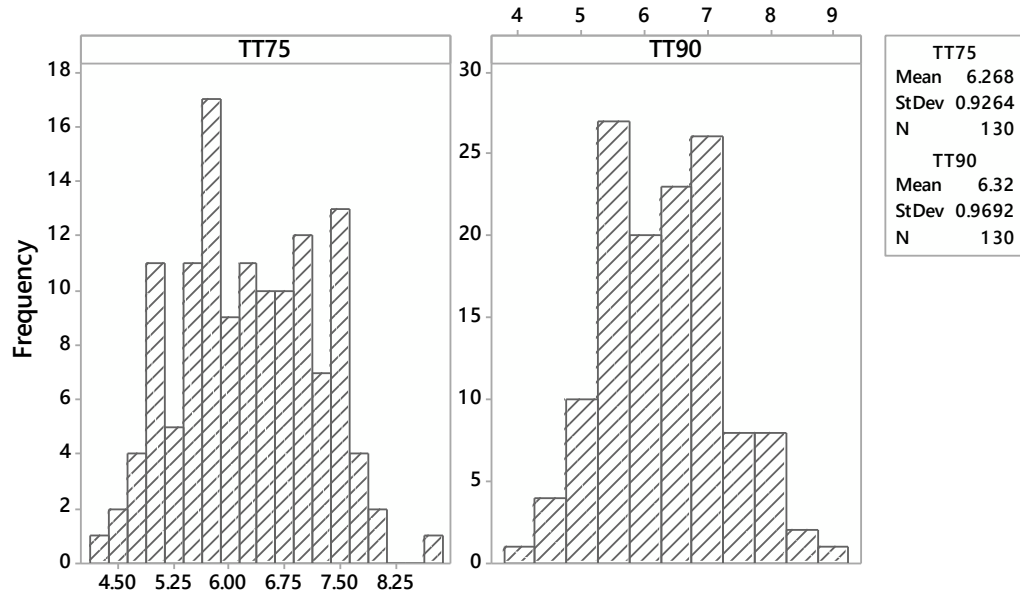
Figure A3. Distribution of the total number of tubers evaluated 75 and 90 DAP in the 130 clones from Experiment 1 grown under short day conditions. Histograms show the normality of the phenotype distribution.

Figure A4. Distribution of the bulking ratio evaluated 75 and 90 DAP in the 130 clones from Experiment 1 grown under short day conditions. Histograms show the normality of the phenotype distribution.

Figure A5. Distribution of the tuber induction evaluated 41, 59, and 74 DAP in the 130 clones from Experiment 1 grown under long day conditions. Histograms show the normality of the phenotype distribution.

Figure A6. Distribution of the number of small tubers evaluated 75 and 90 DAP in the 130 clones from Experiment 1 grown under long day conditions. Histograms show the normality of the phenotype distribution.

95

Figure A7. Distribution of the total number of tubers evaluated 75 and 90 DAP in the 130 clones from Experiment 1 grown under long day conditions. Histograms show the normality of the phenotype distribution.
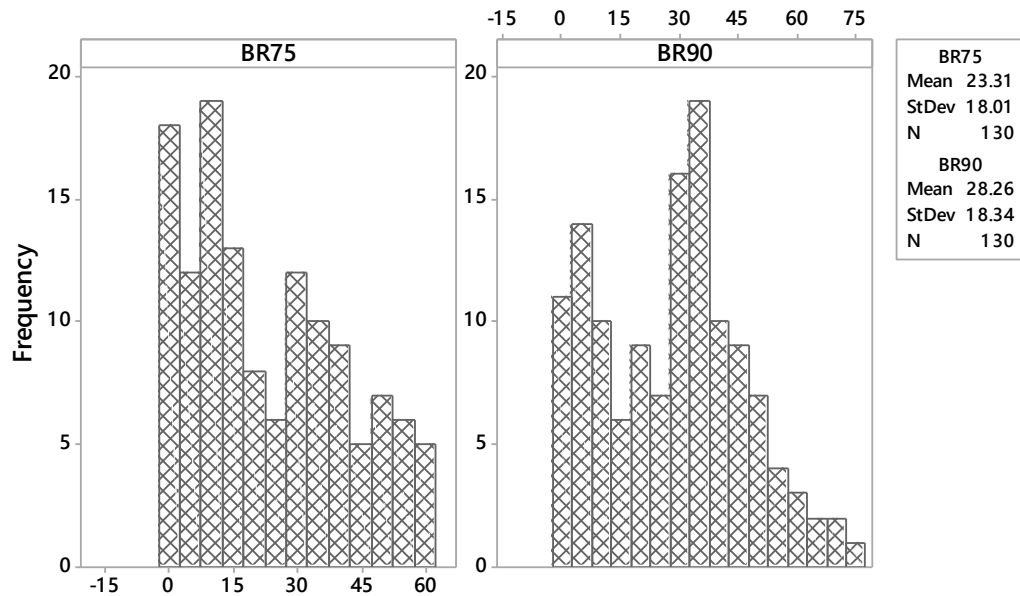
Figure A8. Distribution of the bulking ratio evaluated 75 and 90 DAP in the 130 clones from Experiment 1 grown under long day conditions. Histograms show the normality of the phenotype distribution.
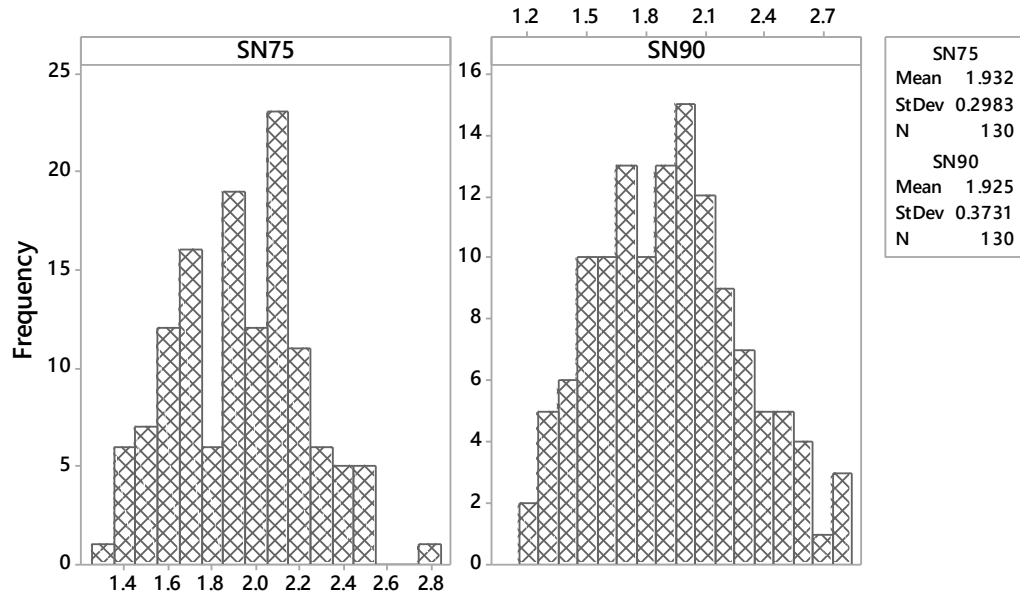
Figure A9. Distribution of the stolon number evaluated 75 and 90 DAP in the 130 clones from Experiment 1 grown under long day conditions. Histograms show the normality of the phenotype distribution.
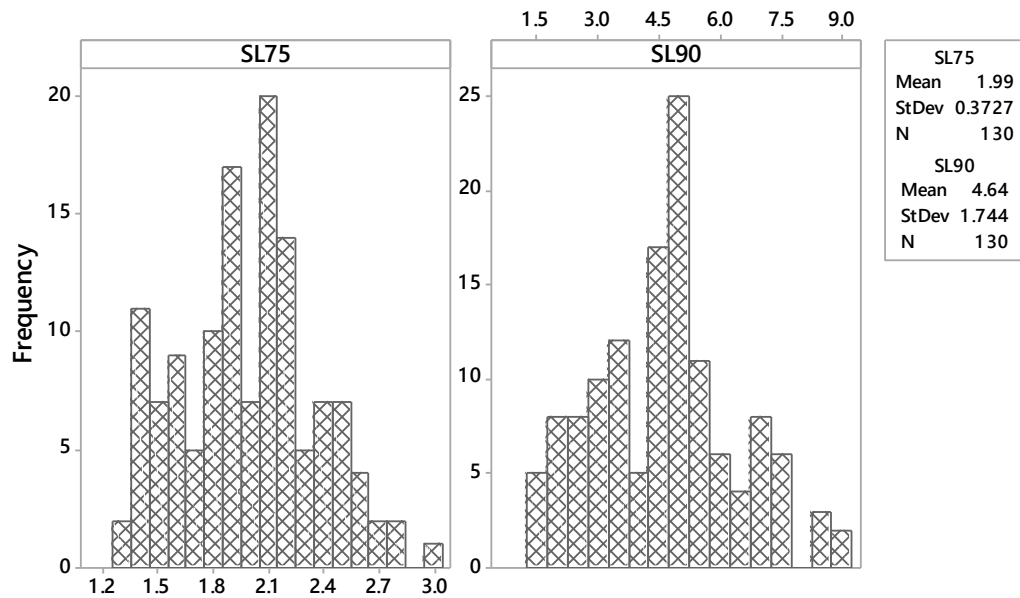
Figure A10. Distribution of the stolon length evaluated 75 and 90 DAP in the 130 clones from Experiment 1 grown under long day conditions. Histograms show the normality of the phenotype distribution.

Table A4. Data normalization in Experiment 2.

Response variable	Day length	DAP	Normality test p-value	Box-Cox transformation λ	Normality test after transformation	Stage of Variable for further analysis
Tuber Induction	12	41	<0.005	-0.50	0.031	Transformed
		59	<0.005*	1.00		Original
		74	<0.005*	3.00	<0.005	Original
	16	41	0.024	0.00	0.099	Transformed
		59	<0.005*	0.00	<0.005	Original
		74	0.013*	0.50	0.013	Original
Number of Small Tubers	12	75	<0.005	0.00	0.531	Transformed
		90	<0.005	0.00	0.970	Transformed
	16	75	<0.005	0.00	0.267	Transformed
		90	<0.005	0.50	0.034	Transformed
Number of Marketable Tubers	12	75	0.756			Original
		90	0.517			Original
	16	75	0.232			Original
		90	0.074			Original
Total Number of Tubers	12	75	<0.005	0.00	0.596	Transformed
		90	<0.005	0.00	0.589	Transformed
	16	75	0.007	0.00	0.256	Transformed
		90	<0.005	0.00	0.259	Transformed
Bulking Ratio	12	75	0.119			Original
		90	0.237			Original
	16	75	0.228			Original
		90	0.167			Original
Stolon Number	12	75	<0.005	0.50	0.021	Transformed
		90	<0.005	0.00	<0.005**	Transformed
	16	75	<0.005	0.50	0.013	Transformed
		90	0.005	0.50	0.065	Transformed
Stolon Length	12	75	<0.005*	-0.50	<0.005	Original
		90	<0.005*	-0.50	<0.005	Original
	16	75	<0.005	0.00	<0.005**	Transformed
		90	0.020*	0.50	0.018	Original

* Refers to normality test that were not achieved by the p-value, but they passed the pen test for normality.
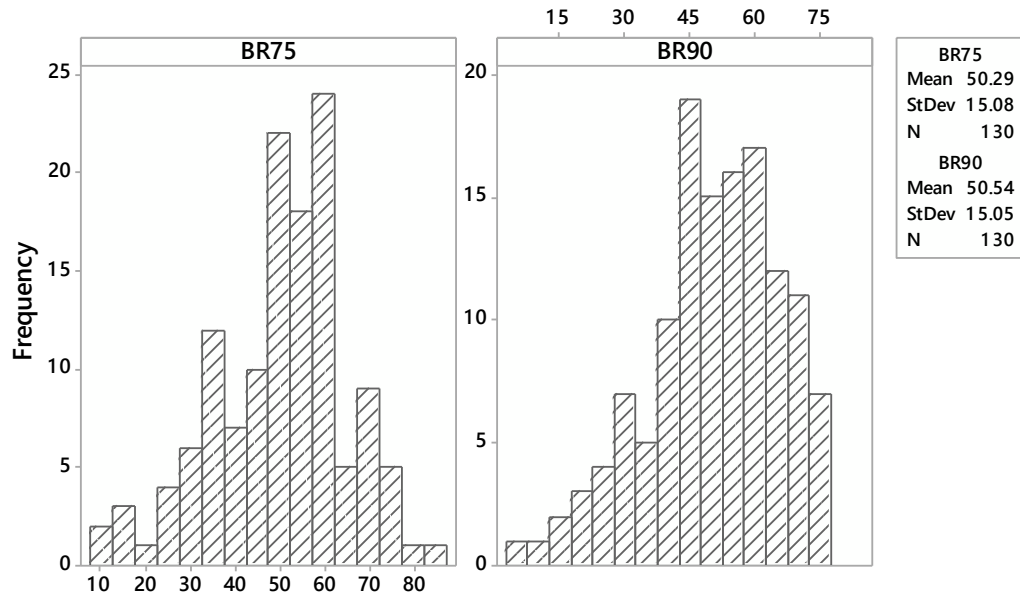
Figure A11. Distribution of tuber induction evaluated 41, 59, and 74 DAP in the 66 clones from Experiment 2 grown under short day conditions. Histograms show the normality of the phenotype distribution.
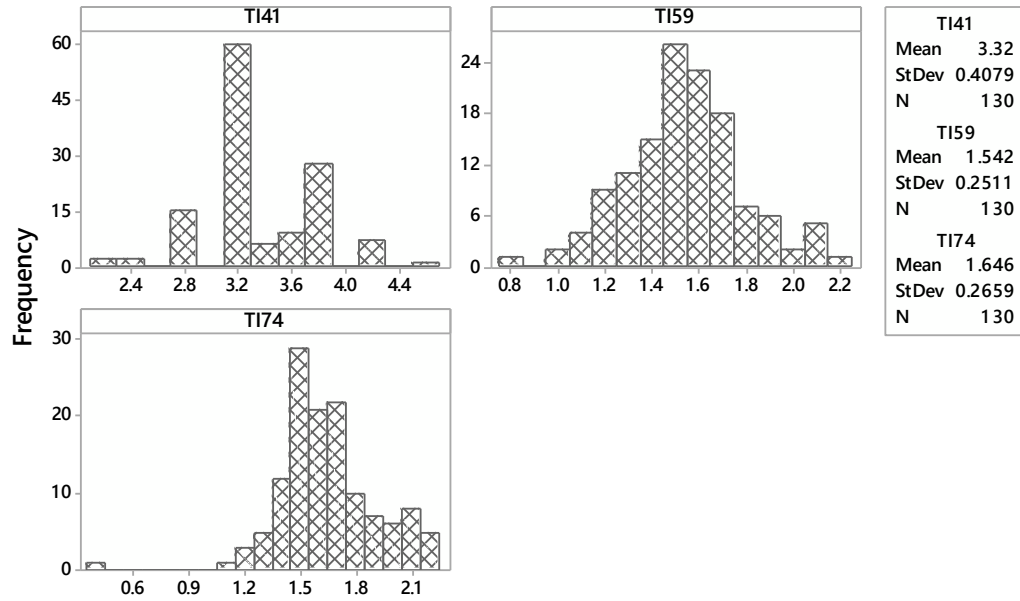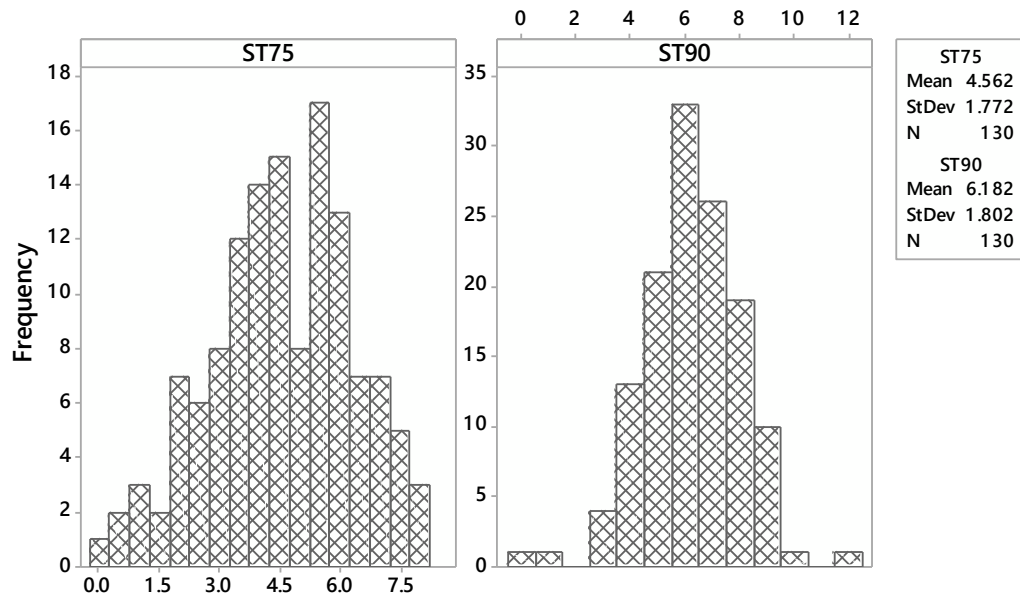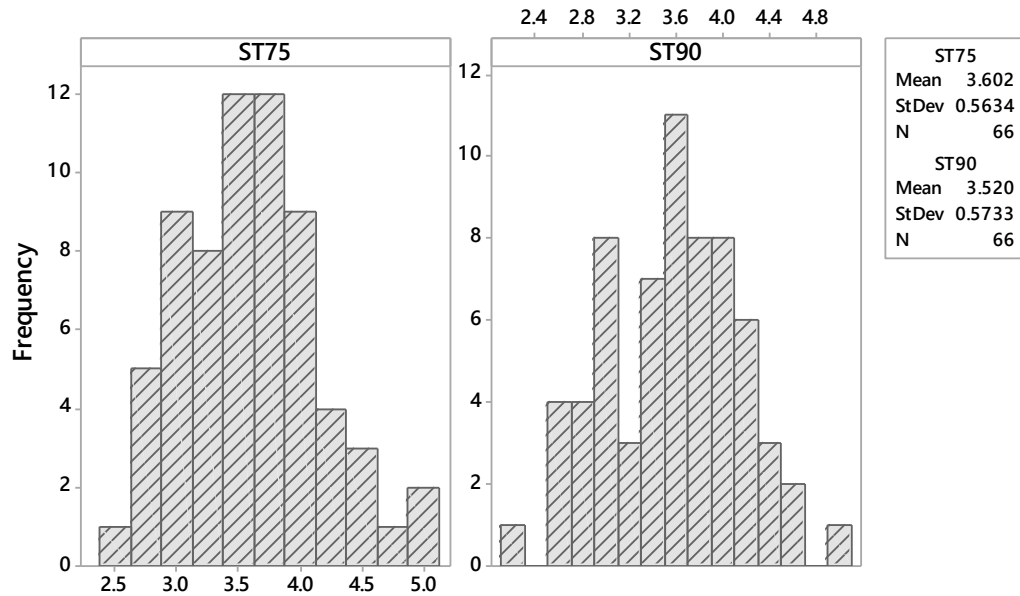

Figure A12. Distribution of the number of small tubers evaluated 75 and 90 DAP in the 66 clones from Experiment 2 grown under short day conditions. Histograms show the normality of the phenotype distribution.
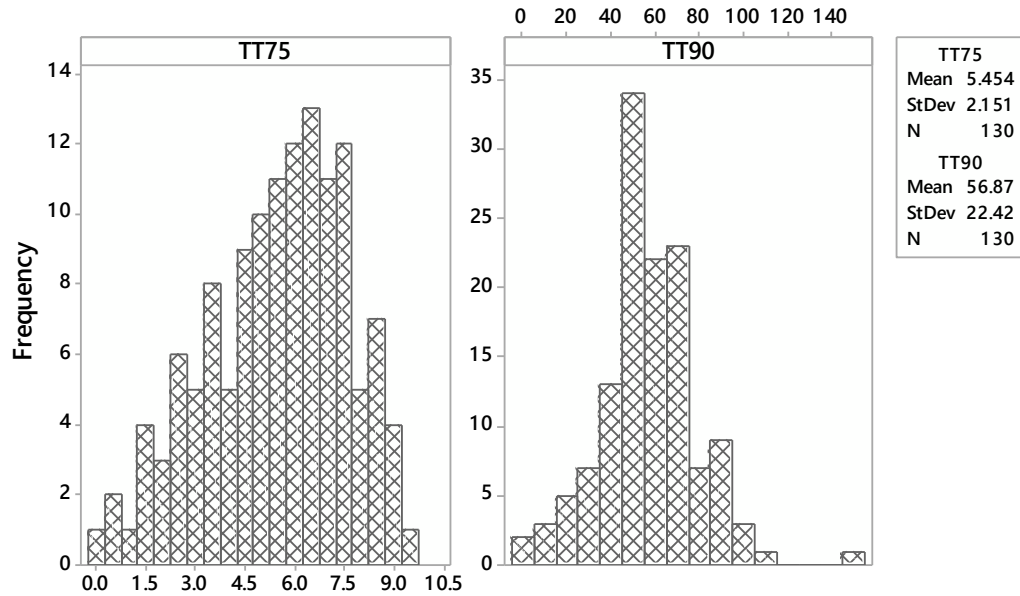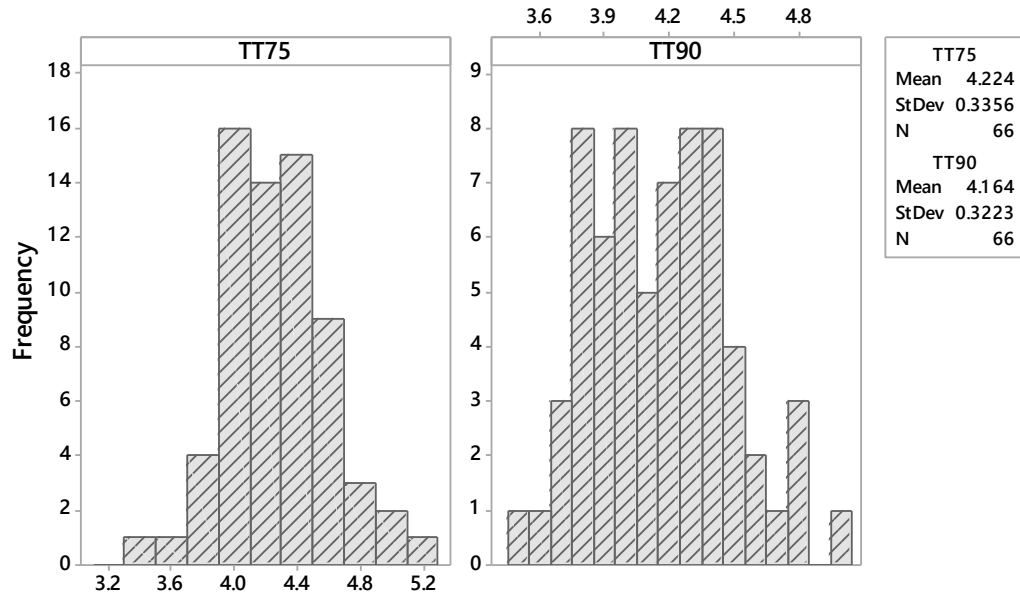
Figure A13. Distribution of the total number of tubers 75 and 90 DAP in the 66 clones from Experiment 2 grown under short day conditions.
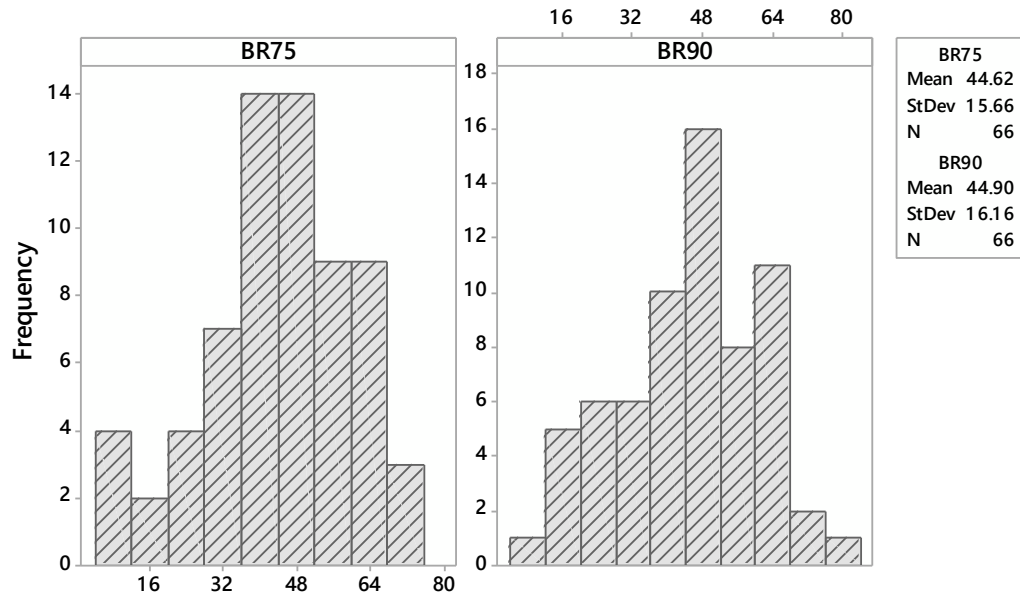
Figure A14. Distribution of the bulking ratio 75 and 90 DAP in the 66 clones from Experiment 2 grown under short day conditions.
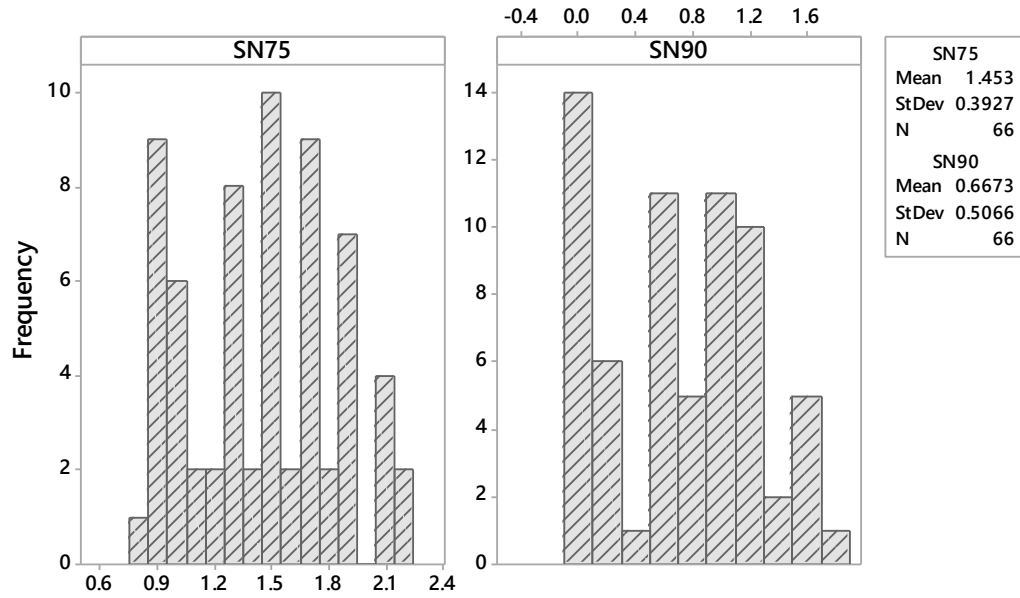
Figure A15. Distribution of the stolon number 75 and 90 DAP in the 66 clones from Experiment 2 grown under short day conditions.
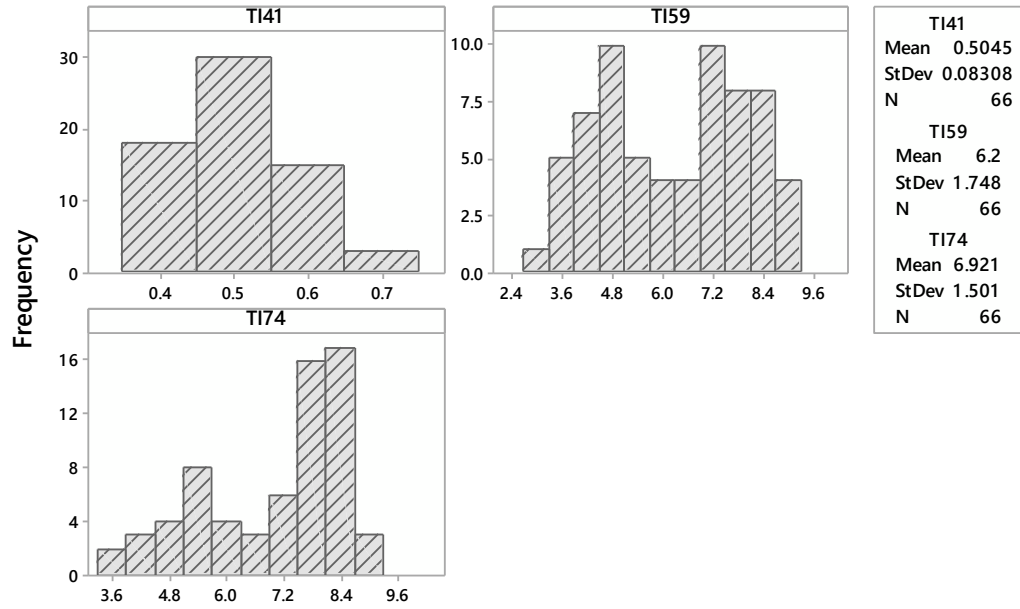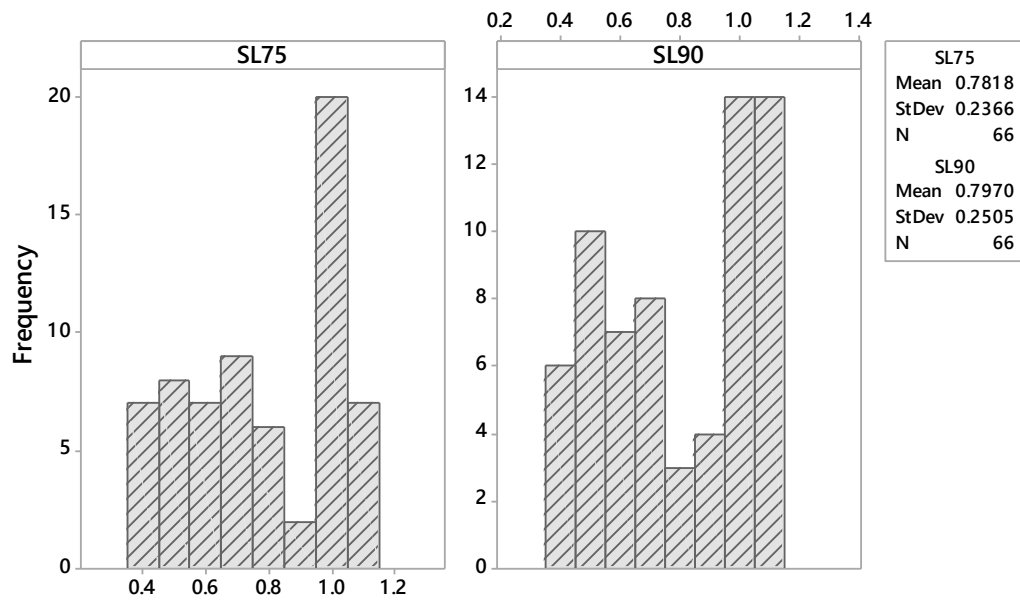

Figure A16. Distribution of the stolon length 75 and 90 DAP in the 66 clones from Experiment 2 grown under short day conditions.
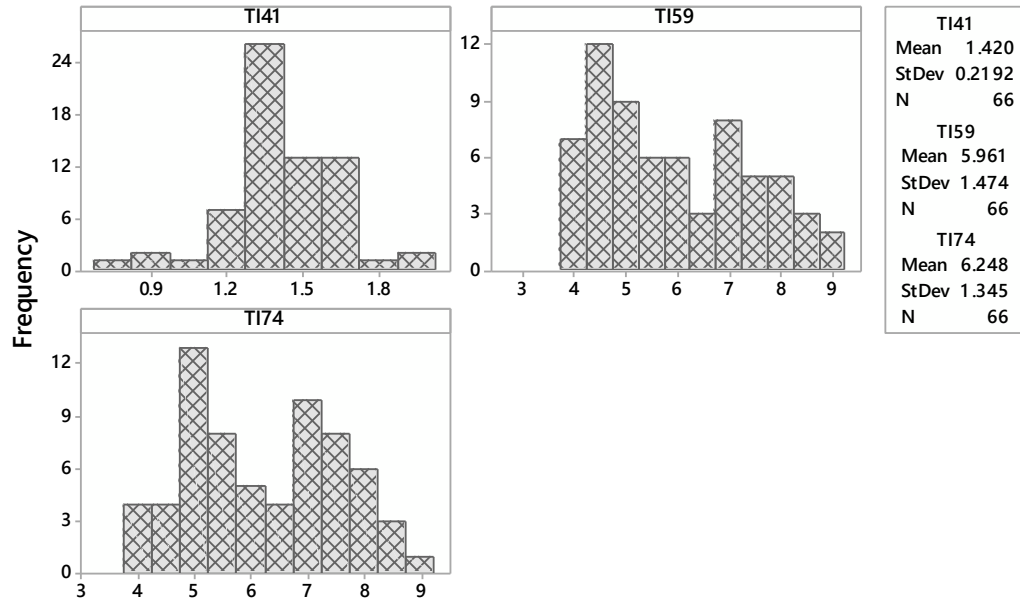
Figure A17. Distribution of the tuber induction 41, 59, and 74 DAP in the 66 clones from Experiment 2 grown under long day conditions.
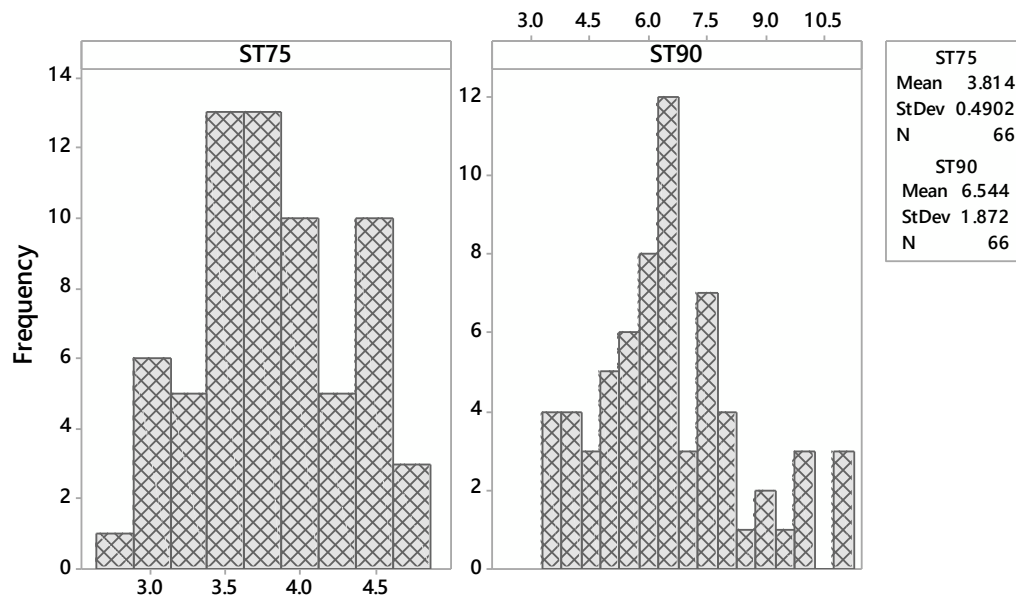
Figure A18. Distribution of the number of small tubers 75 and 90 DAP in the 66 clones from Experiment 2 grown under long day conditions.
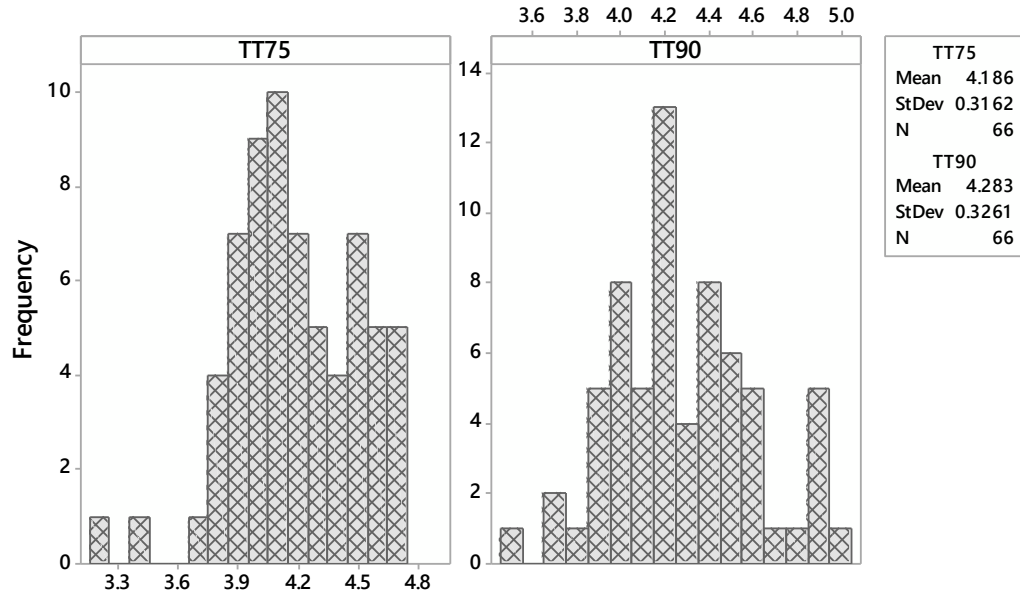
Figure A19. Distribution of the total number of tubers 75 and 90 DAP in the 66 clones from Experiment 2 grown under long day conditions.
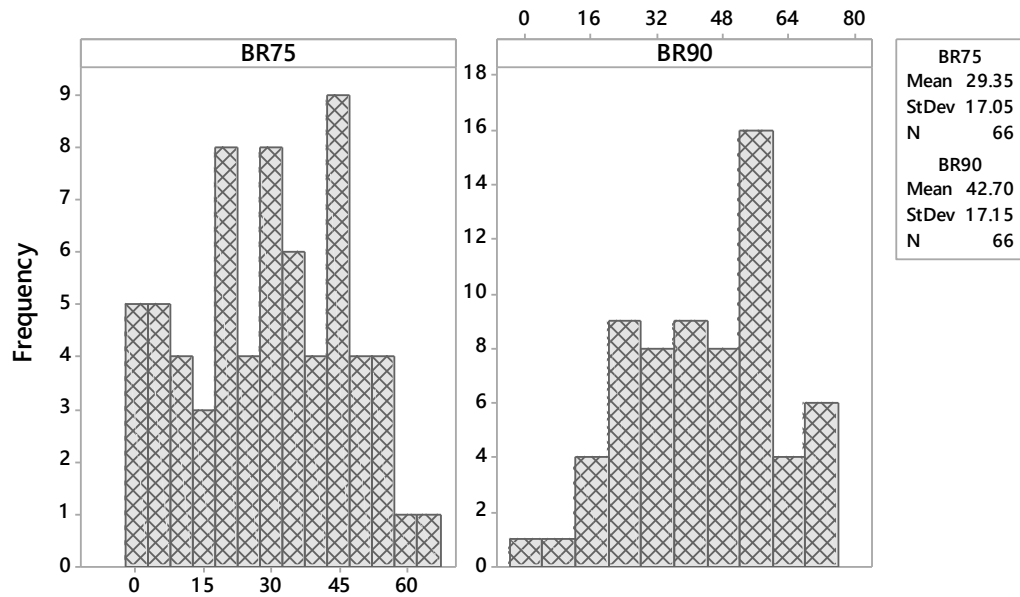
Figure A20. Distribution of the bulking ratio 75 and 90 DAP in the 66 clones from Experiment 2 grown under long day conditions.
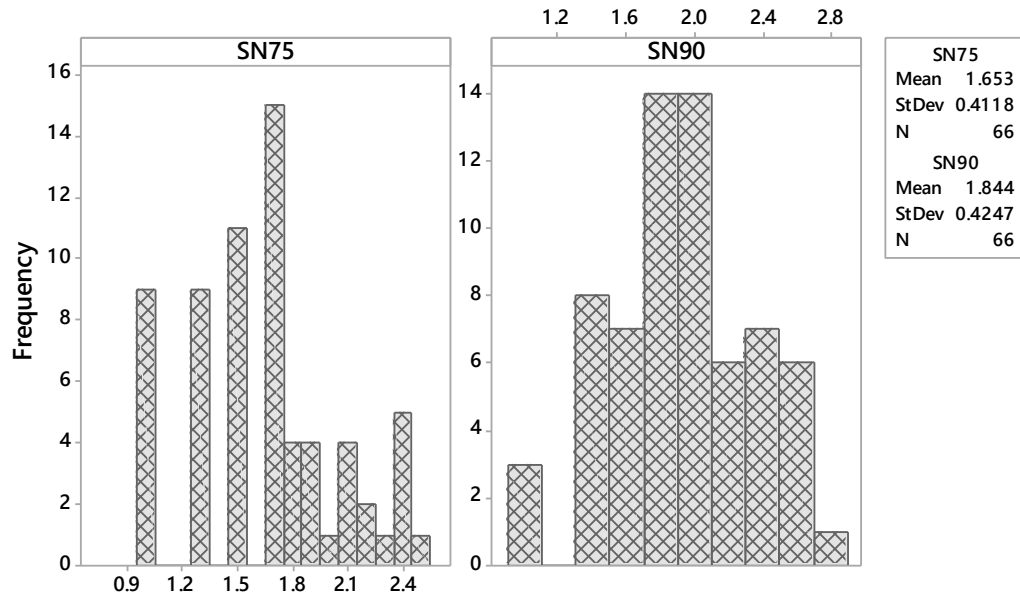
Figure A21. Distribution of the stolon number 75 and 90 DAP in the 66 clones from Experiment 2 grown under long day conditions.
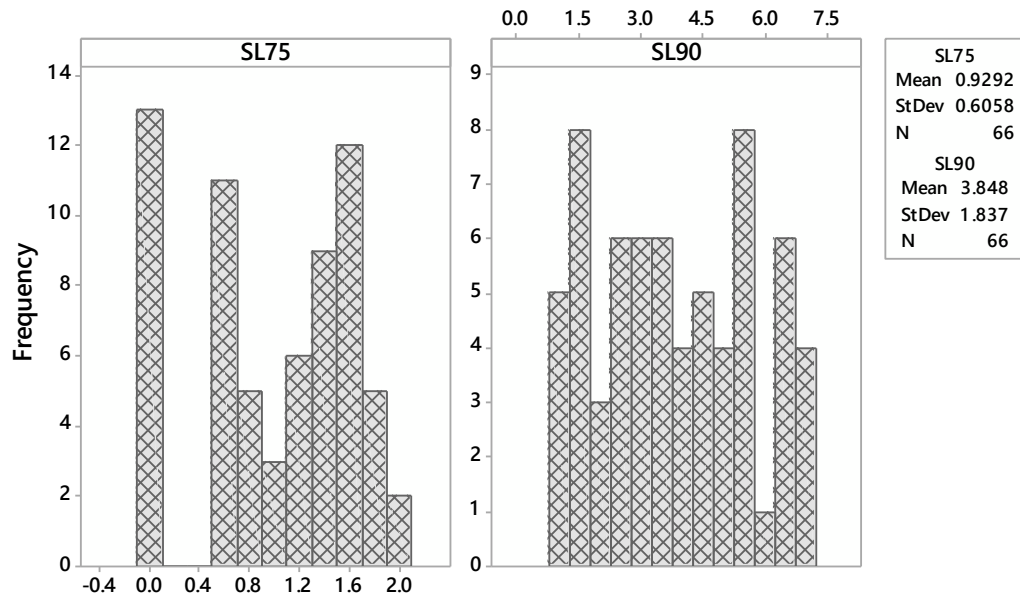
Figure A22. Distribution of the stolon length 75 and 90 DAP in the 66 clones from Experiment 2 grown under long day conditions.
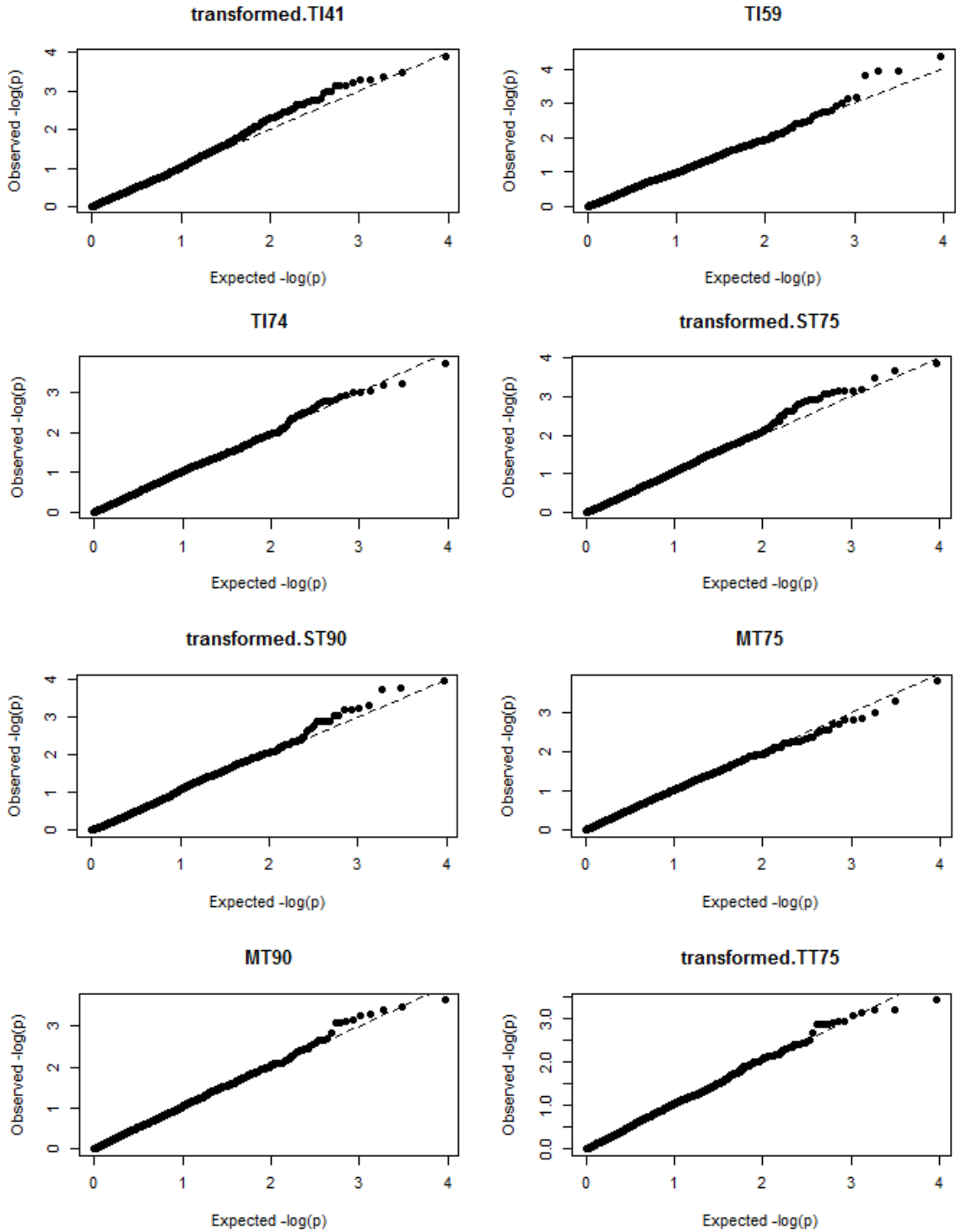
Figure A23. QQ plots for the tuberization related traits in Experiment 1 under short day conditions.
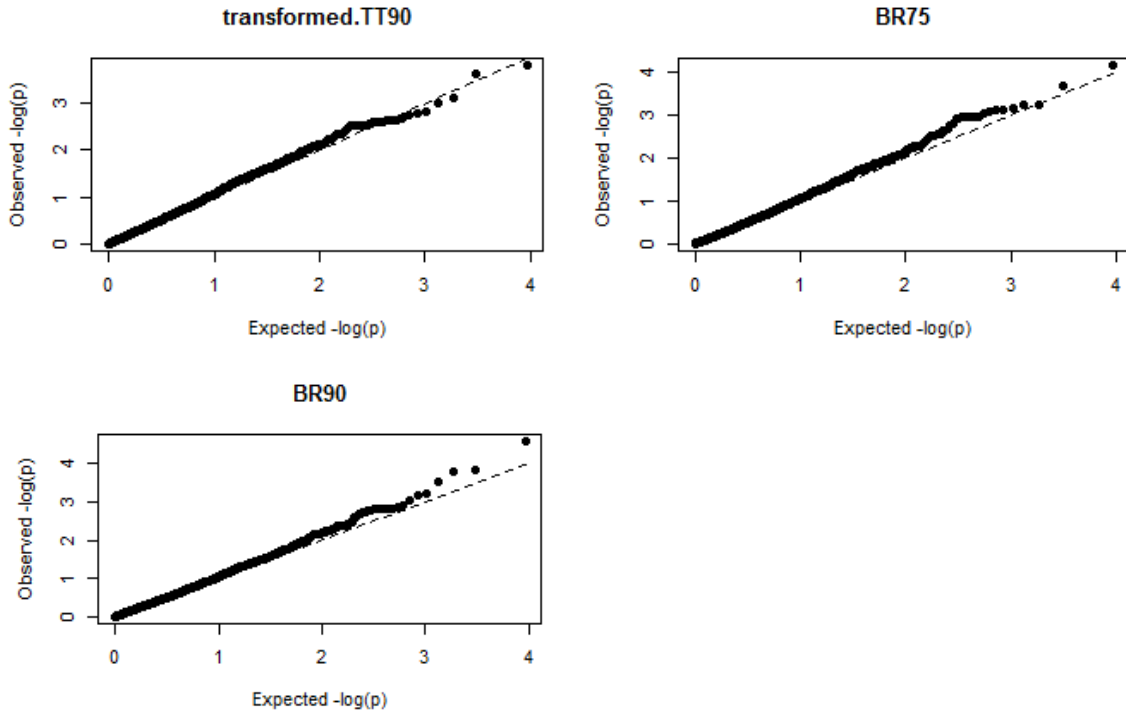
Figure A23. Continuation: QQ plots for the tuberization related traits in Experiment 1 under short day conditions.
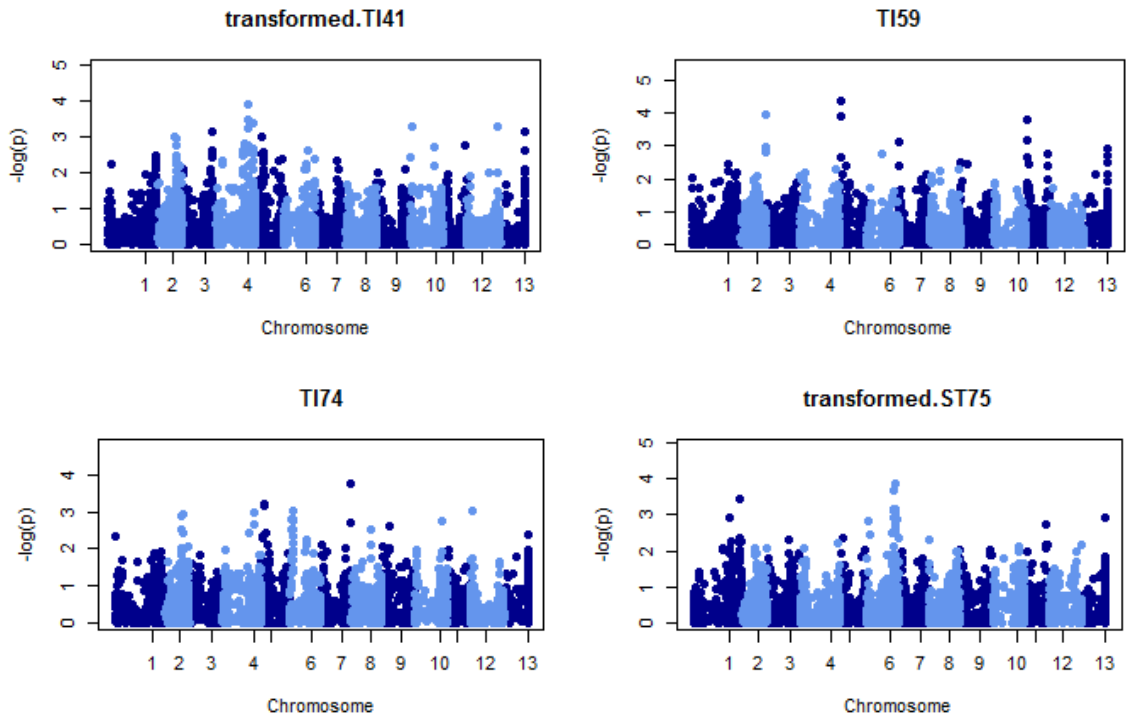
Figure A24. Manhattan plots for the tuberization related traits in Experiment 1 under short day conditions.
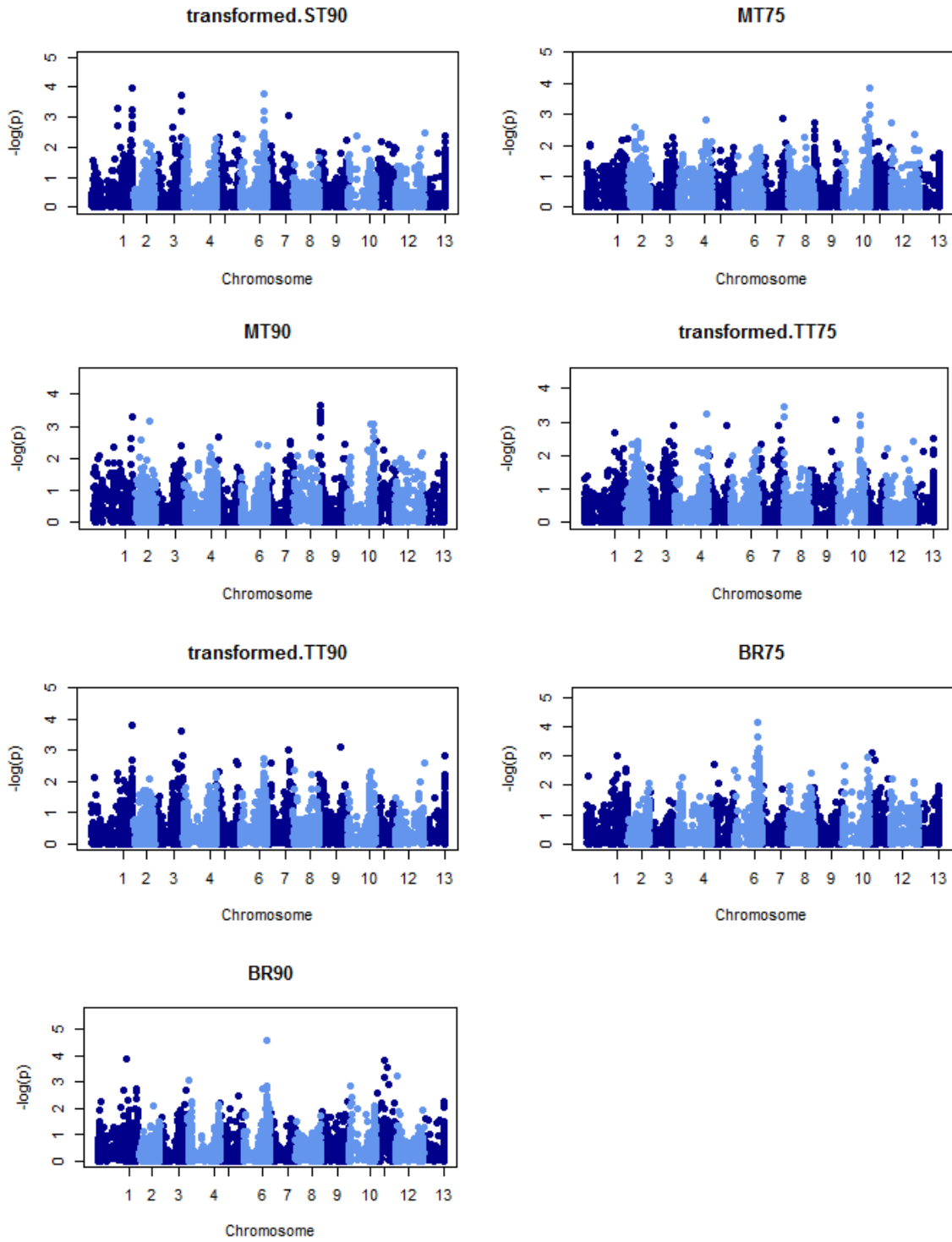
Figure A24. Continuation: Manhattan plots for the tuberization related traits in Experiment 1 under short day conditions.
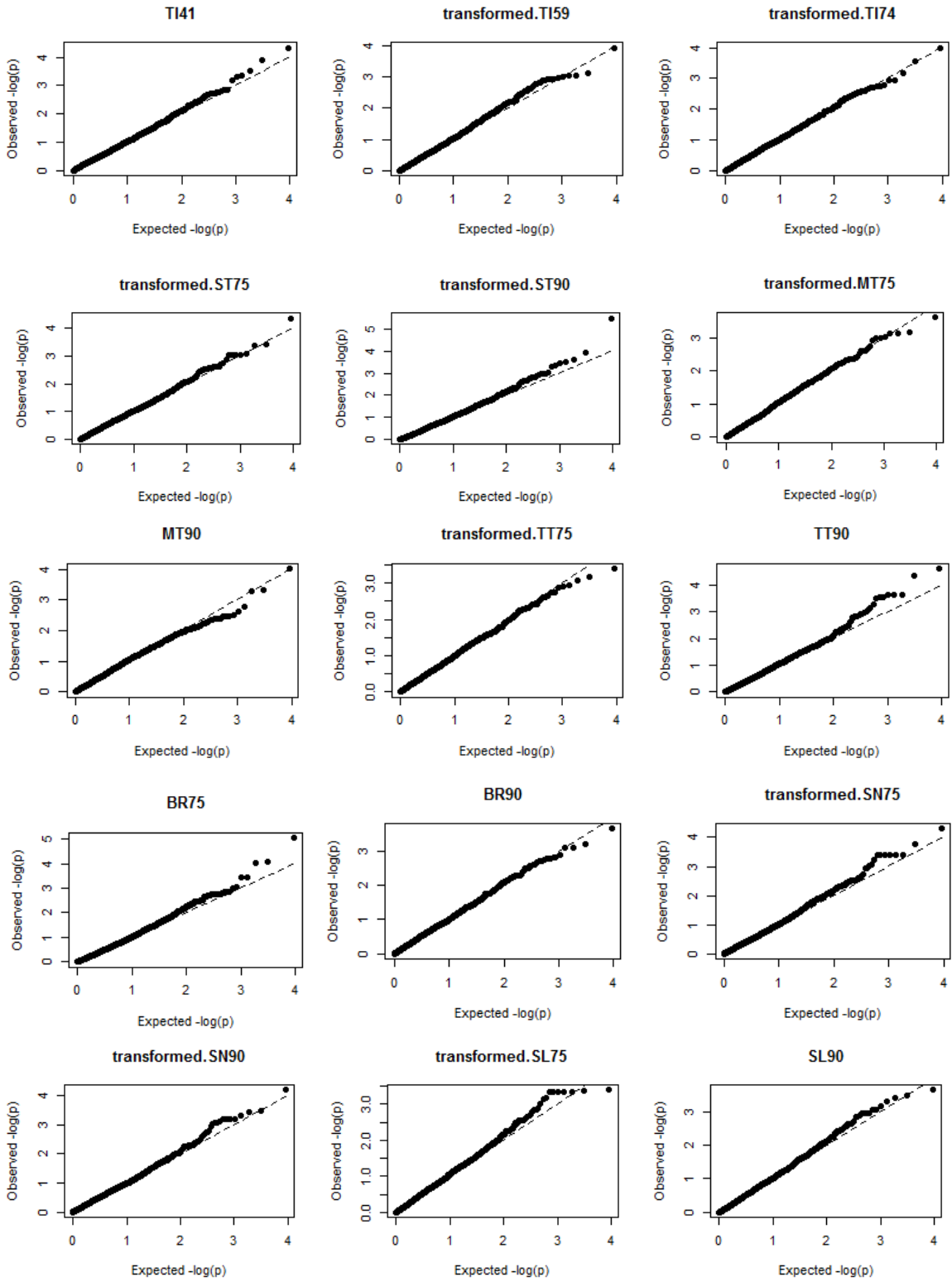
Figure A25. QQ plots for the tuberization related traits in Experiment 1 under long day conditions.
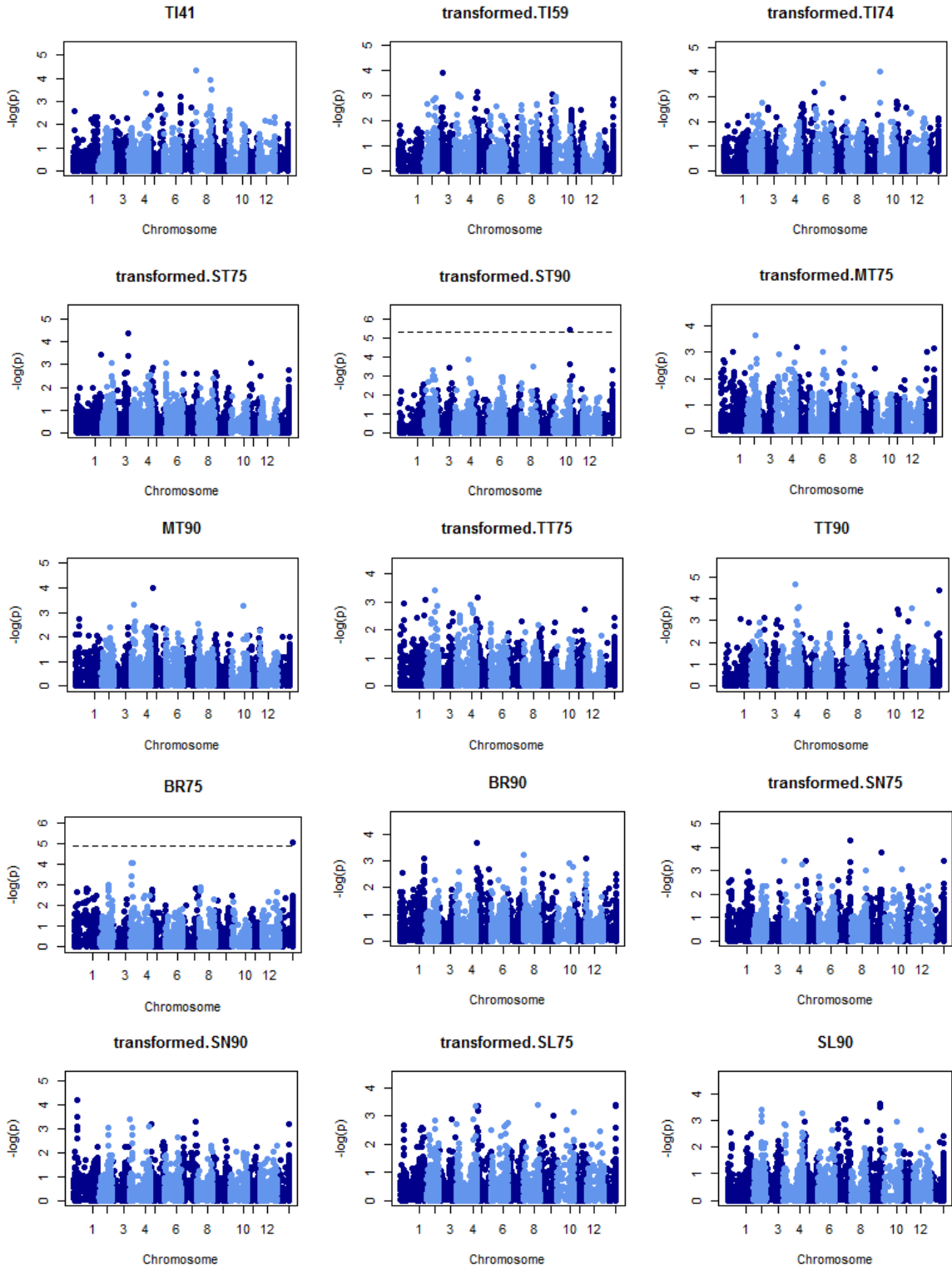
Figure A26. Manhattan plots for the tuberization related traits in Experiment 1 under long day conditions.
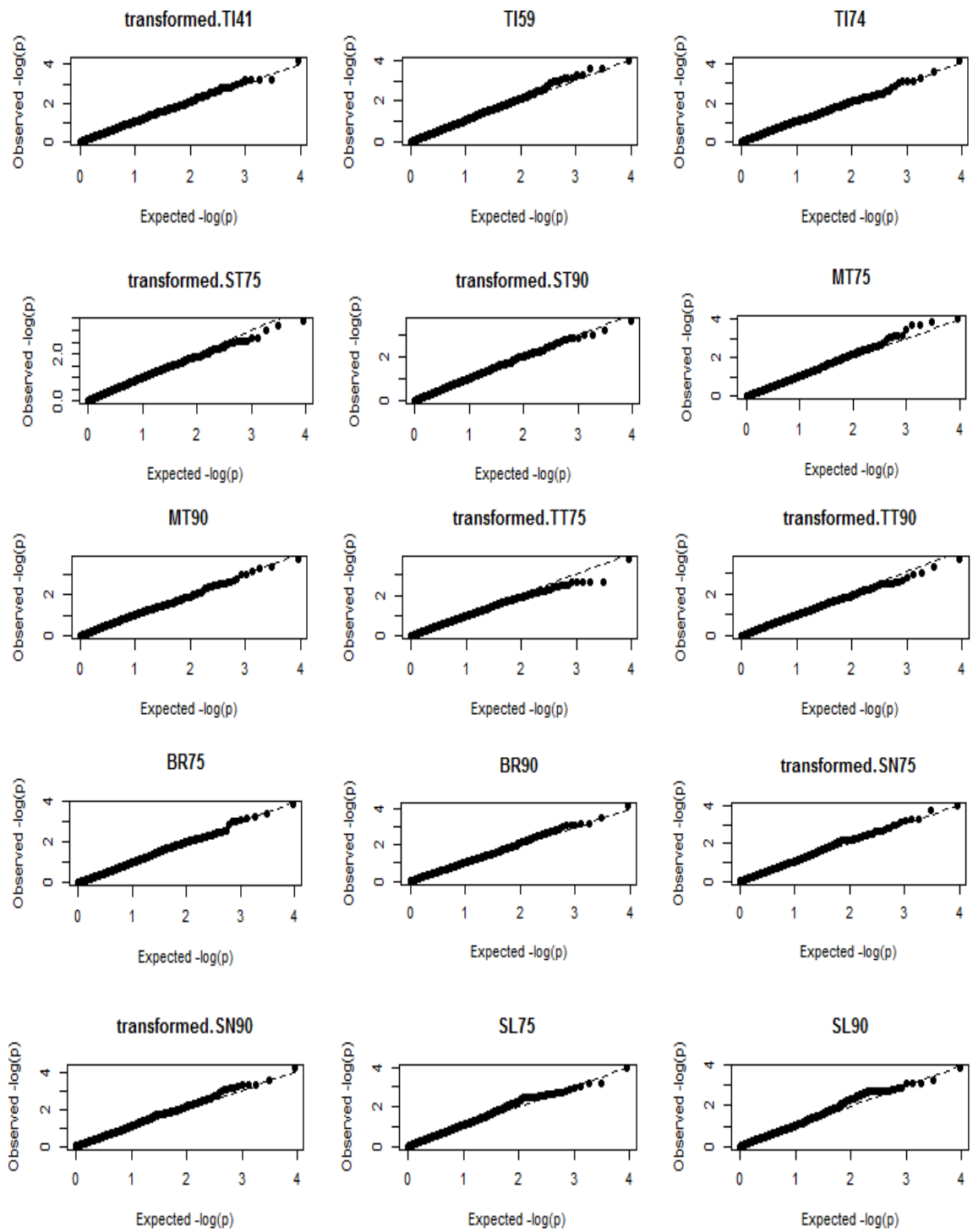
Figure A27. QQ plots for the tuberization related traits in Experiment 2 under short day conditions.
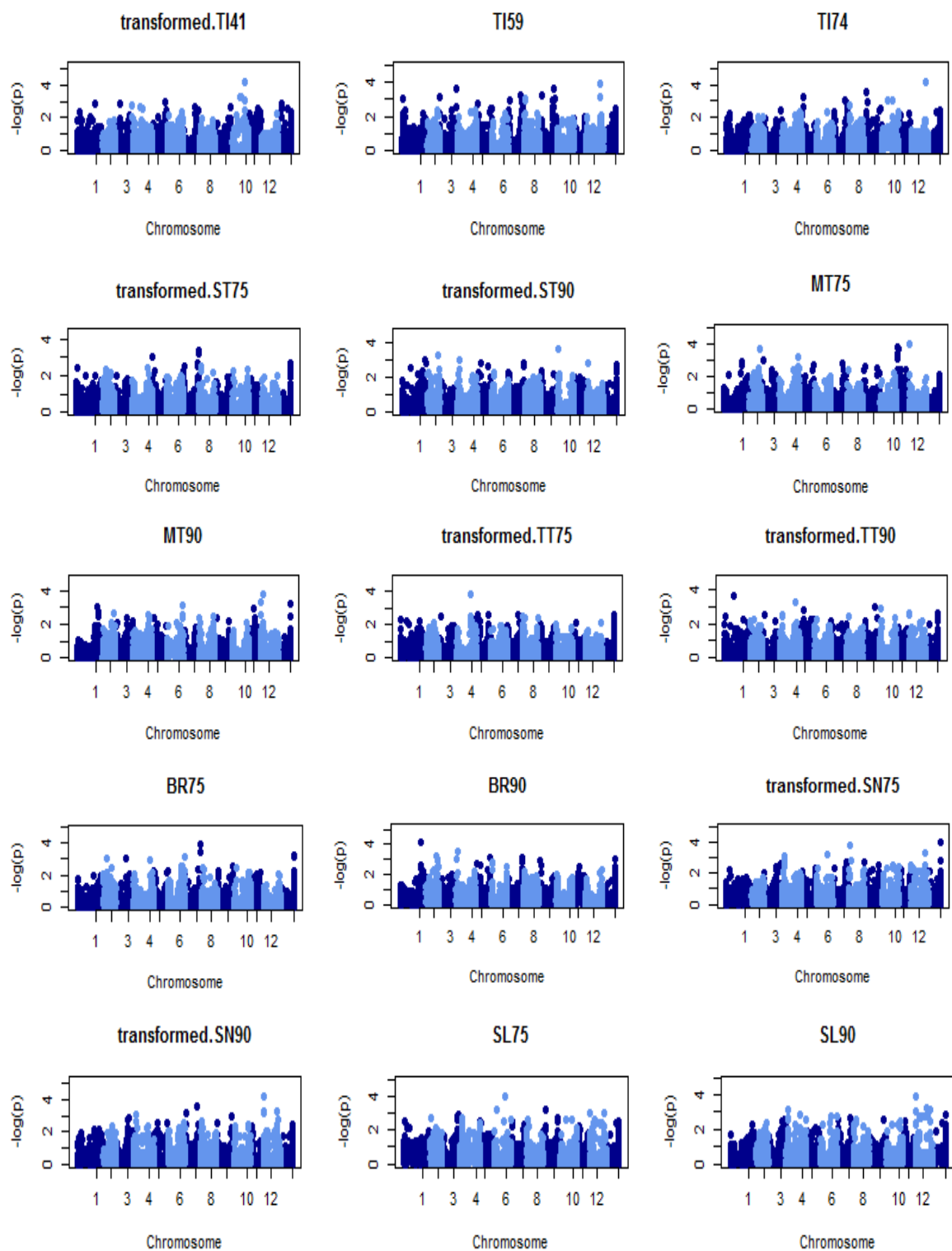
Figure A28. Manhattan plots for the tuberization related traits in Experiment 2 under short day conditions.
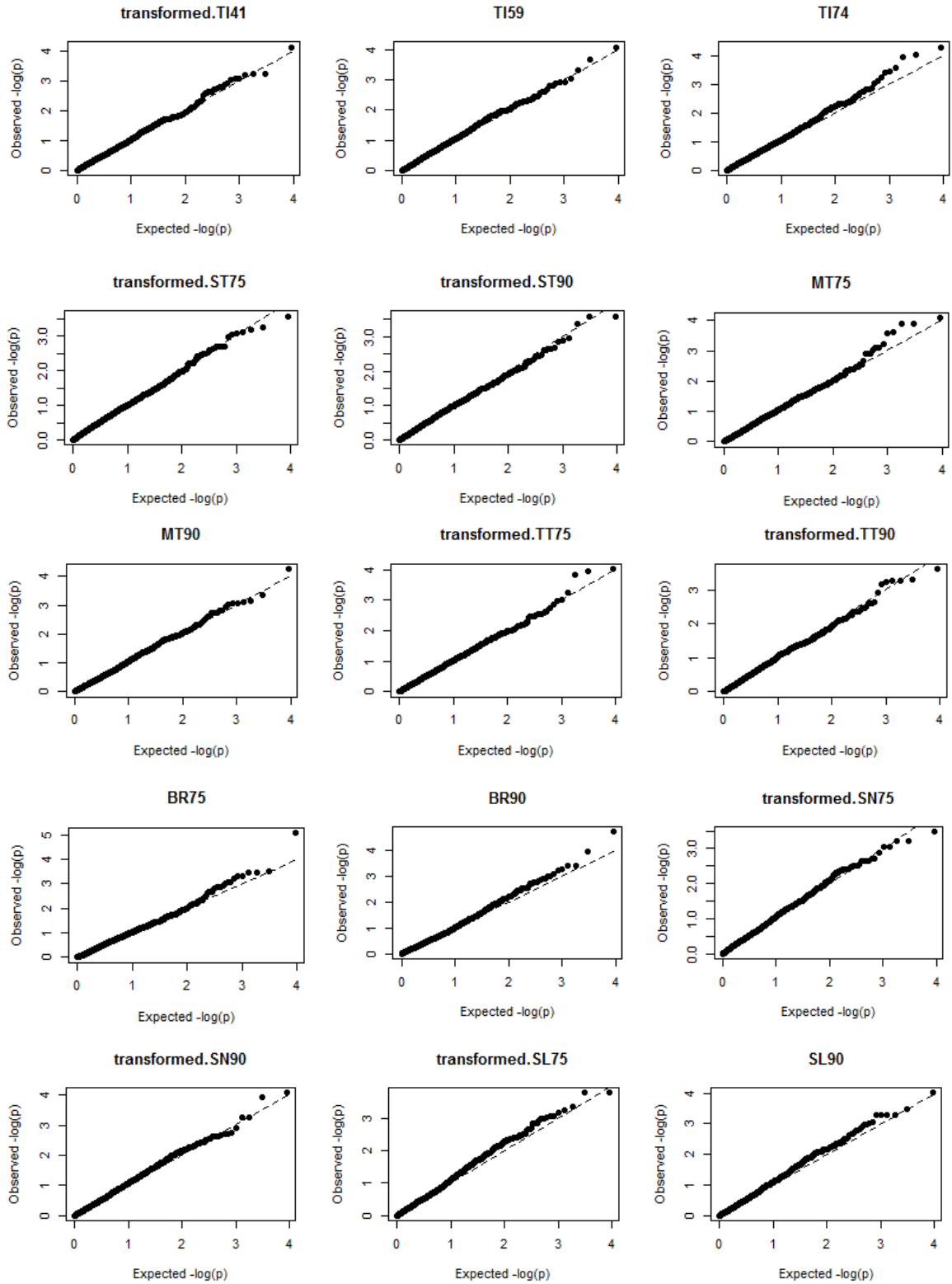
Figure A29. QQ plots for the tuberization related traits in Experiment 2 under long day conditions.
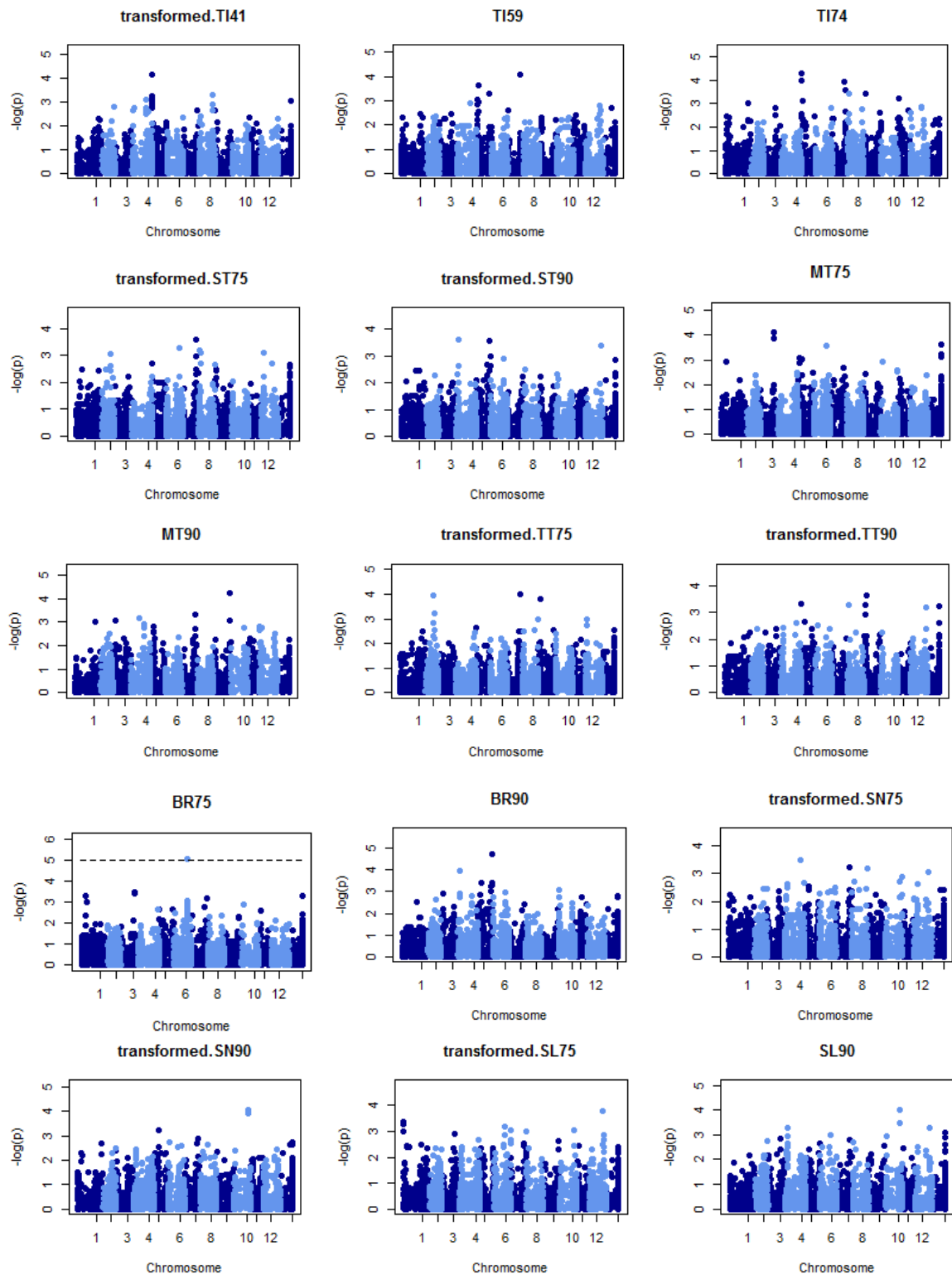
Figure A30. Manhattan plots for the tuberization related traits in Experiment 2 under long day conditions.

113

Table A5. Markers associated with tuber induction.

Marker	DAP	Chr. no.	Position	MAF	SD Exp1	SD Exp2	LD Exp1	LD Exp2	Annotation
SolCAP_c1_11459	59	2	47625747	15	3.94*	0.96	2.01	1.05	Steroid nuclear receptor, ligand-binding
SolCAP_c1_156	59	3	20183296	39	0.86	0.25	3.93*	1.01	Glycoside hydrolase family 28 protein / polygalacturonase (pectinase) family protein
SolCAP_c2_164	59	3	59574996	37	0.12	3.62*	0.64	0.26	Pyridoxin biosynthesis protein PDX1
SolCAP_c1_10762	41	4	58875538	48	3.91*	0.23	1.42	0.72	Photoperiod responsive protein
SolCAP_c2_23678	41	5	224801	49	0.47	0.44	0.45	4.15*	Integral membrane protein
SolCAP_c2_23833	74	5	694521	5	0.00	2.14	0.00	4.02*	Dihydropterin pyrophosphokinase-dihydropteroate synthase
SolCAP_c2_23835	74	5	694562	5	0.00	2.35	0.00	4.30*	Dihydropterin pyrophosphokinase-dihydropteroate synthase
SolCAP_c2_50305	59	5	5052466	9	4.39*	1.32	0.97	3.67*	Methyltransferase
SolCAP_c1_14840	59	5	5692769	8	3.93*	2.03	1.22	2.55	Sucrose transporter
SolCAP_c2_35592	74	6	28886273	21	1.93	0.06	3.55*	0.57	Gene of unknown function
SolCAP_c2_33489	59	7	45146538	11	0.43	3.25	0.17	4.09*	Glucan endo-1,3-beta-glucosidase
SolCAP_c2_33489	74	7	45146538	11	0.86	3.08	0.57	3.97*	Glucan endo-1,3-beta-glucosidase
SolCAP_c2_19826	74	7	49027301	29	0.45	1.06	1.32	3.58*	Pre-mRNA splicing factor
SolCAP_c1_8834	74	7	55678678	12	3.74*	0.10	1.40	0.47	E3 ubiquitin protein ligase upl2
SolCAP_c2_27452	41	8	2045972	47	0.82	0.72	4.34*	1.38	FKBP-type peptidyl-prolyl cis-trans isomerase 3, chloroplastic
SolCAP_c1_8237	41	8	54357787	13	1.57	0.26	3.92*	0.13	Conserved gene of unknown function
SolCAP_c2_16995	41	8	54835391	13	1.00	0.18	3.53*	0.06	BHLH transcription factor JAF13
SolCAP_c2_52517	74	9	8043250	33	0.78	3.60*	0.16	3.42	NG
SolCAP_c2_40879	59	9	52027191	27	0.23	3.59*	1.46	1.18	CONSTANS-like zinc finger protein
SolCAP_c2_1475	74	10	1121595	6	0.78	0.19	4.00*	0.34	Ethylene-responsive transcription factor 3
SolCAP_c2_54932	41	10	45025424	48	1.39	4.15*	0.02	0.55	Serine-threonine protein kinase, plant-type
SolCAP_c2_33678	59	11	2464166	9	3.82*	1.40	1.14	1.81	6-phosphofructokinase 5, chloroplastic
SolCAP_c2_18855	59	12	51931824	27	0.15	3.94*	0.13	1.27	Phospholipid-transporting ATPase
SolCAP_c2_18855	74	12	51931824	27	0.11	4.12*	0.70	1.72	Phospholipid-transporting ATPase

Table A6. Markers associated with number of tubers.

Marker	DAP	Size	Chr. no.	Position	MAF	SD Exp1	SD Exp2	LD Exp1	LD Exp2	Annotation
SolCAP_c2_24677	90	TT	1	30837971	21	0.03	3.64*	1.03	1.38	Cytidine triphosphate synthase
SolCAP_c1_4803	90	ST	1	86942383	44	3.98*	0.15	2.15	0.03	SAUR family protein
SolCAP_c1_4803	90	TT	1	86942383	44	3.82*	0.18	1.07	0.13	SAUR family protein
SolCAP_c2_39178	75	TT	2	32148414	8	1.21	0.77	0.81	3.95*	Transcription factor
SolCAP_c2_33108	75	MT	2	34739125	8	0.49	0.48	3.63*	0.98	Sphingosine kinase
SolCAP_c1_8113	75	MT	2	40729948	13	0.25	3.71*	0.33	0.73	Aldehyde dehydrogenase
SolCAP_c2_20227	75	ST	3	49231930	19	1.55	0.03	4.37*	0.28	Kinase interacting protein
SolCAP_c2_86	90	ST	3	58611092	42	3.74*	0.47	0.81	0.44	ArcA2 protein
SolCAP_c2_86	90	TT	3	58611092	42	3.63*	0.82	0.70	0.54	ArcA2 protein
SolCAP_c2_39718	90	ST	4	7898214	8	1.88	2.50	0.14	3.60*	26S proteasome subunit RPN5b
SolCAP_c2_16731	90	ST	4	49259783	5	0.67	1.82	3.92*	1.08	NG
SolCAP_c2_16731	90	TT	4	49259783	5	0.81	1.87	4.64*	1.13	NG
SolCAP_c2_45035	75	TT	4	53671782	21	0.54	3.78*	0.14	0.95	Cytochrome P450 monooxygenase CYP736B
SolCAP_c2_51244	90	TT	4	60198048	17	0.30	0.41	3.56*	0.40	Ci21A protein
SolCAP_c2_26681	90	TT	4	65750963	5	0.50	0.02	3.64*	0.34	Phosphoenolpyruvate carboxykinase
SolCAP_c2_55776	90	TT	4	65971563	5	0.50	0.02	3.64*	0.34	Receptor protein kinase CLAVATA1
SolCAP_c2_55773	90	TT	4	65972517	5	0.50	0.02	3.64*	0.34	Receptor protein kinase CLAVATA1
SolCAP_c2_50302	90	MT	5	5055763	9	0.20	1.63	4.02*	0.87	Methyltransferase
SolCAP_c2_10358	90	ST	5	48541183	49	1.53	1.10	0.51	3.58*	Conserved gene of unknown function
SolCAP_c2_25926	75	MT	6	46018319	48	1.71	1.42	3.03	3.57*	NG
SolCAP_c2_56145	75	ST	6	51385833	34	3.67*	0.10	0.18	0.51	Conserved gene of unknown function
SolCAP_c2_8904	75	ST	6	52599348	25	3.86*	1.12	0.20	1.32	Kinase
SolCAP_c2_8904	90	ST	6	52599348	25	3.79*	0.63	0.65	1.48	Kinase
SolCAP_c1_7980	75	ST	7	47551623	23	0.22	0.29	0.18	3.57*	NG
SolCAP_c1_7980	75	TT	7	47551623	23	0.08	1.12	0.04	4.03*	NG
SolCAP_c2_18684	75	ST	7	55284214	45	0.46	3.37*	0.50	2.36	Lysine decarboxylase
SolCAP_c2_18918	90	ST	8	43308600	21	1.11	1.28	3.52*	1.05	Coiled-coil domain-containing protein
SolCAP_c2_39101	90	MT	9	1020724	7	3.65*	0.08	0.17	0.04	Glutathion S-transferase
SolCAP_c2_3997	75	TT	9	5201143	24	0.44	0.85	1.56	3.82*	P-glycoprotein
SolCAP_c1_4228	90	TT	9	8247387	41	0.06	2.13	0.03	3.62*	BZIP transcription factor bZIP114

Table A6. Continuation: Markers associated with number of tubers.

Marker	DAP	Size	Chr. no.	Position	MAF	SD Exp1	SD Exp2	LD Exp1	LD Exp2	Annotation
SolCAP_c2_47939	90	MT	9	59598694	20	0.23	0.83	0.38	4.27*	Werner syndrome ATP-dependent helicase
SolCAP_c2_44266	90	ST	10	3991078	35	1.42	3.63*	0.05	1.92	Conserved gene of unknown function
SolCAP_c1_7190	75	MT	10	57478874	19	3.84*	0.71	1.02	0.08	Actin-58
SolCAP_c2_13355	90	ST	11	437672	7	0.20	0.78	5.49*	0.26	Ubiquitin-protein ligase
SolCAP_c2_13355	90	TT	11	437672	7	0.30	0.27	3.51*	0.07	Ubiquitin-protein ligase
SolCAP_c1_4328	90	ST	11	940359	5	0.41	0.70	3.63*	0.30	Nonsense-mediated mRNA decay protein
SolCAP_c2_6000	75	MT	11	2725245	28	0.01	3.85*	0.09	1.12	Peroxidase 17
SolCAP_c2_20947	75	MT	11	5051445	41	0.13	3.73*	0.39	0.14	Beta-fructofuranosidase
SolCAP_c1_8002	75	MT	12	1511064	6	0.07	4.05*	0.43	0.41	14-3-3 protein
SolCAP_c2_1536	90	TT	12	6518553	9	0.02	0.60	3.54*	0.89	Galactosyltransferase family protein
SolCAP_c2_34762	90	MT	12	8063612	37	0.05	3.73*	0.73	0.48	Indole-3-acetic acid-amido synthetase GH3.6
SolCAP_c2_24589	90	TT	13	44567126	6	0.42	1.45	4.38*	0.36	50S ribosomal protein L13, chloroplastic

Table A7. Markers associated with bulking ratio.

Marker	DAP	Chr. no.	Position	MAF	SD Exp1	SD Exp2	LD Exp1	LD Exp2	Annotation
SolCAP_c2_38405	90	1	62145049	39	3.85*	0.23	0.43	0.01	Glucose-1-phosphate adenylyltransferase
SolCAP_c1_6294	90	1	73340313	38	1.22	4.17*	0.53	0.80	Long-chain acyl-CoA synthetase 4
SolCAP_c1_15106	75	4	5513361	13	1.59	0.85	4.05*	0.65	Conserved gene of unknown function
SolCAP_c2_11569	75	4	6210370	11	1.74	0.71	4.07*	0.43	L-allo-threonine aldolase
SolCAP_c2_11549	90	4	6380666	24	1.53	3.53*	0.46	0.70	UDP-glucosyltransferase
SolCAP_c2_39718	90	4	7898214	8	2.28	1.13	0.11	3.96*	26S proteasome subunit RPN5b
SolCAP_c2_50302	90	5	5055763	9	0.03	0.23	3.66*	0.71	Methyltransferase
SolCAP_c2_10358	90	5	48541183	49	1.31	1.26	0.39	4.71*	Conserved gene of unknown function
SolCAP_c2_25926	75	6	46018319	48	2.95	1.88	1.74	5.09*	NG
SolCAP_c2_41405	75	6	51183805	34	3.66*	0.13	0.52	0.85	ADP,ATP carrier protein
SolCAP_c2_56145	75	6	51385833	34	4.14*	0.30	0.76	1.15	Conserved gene of unknown function
SolCAP_c2_8904	90	6	52599348	25	4.58*	0.33	1.16	1.25	Kinase
SolCAP_c2_18684	75	7	55284214	45	0.09	3.90*	0.33	3.21	Lysine decarboxylase
SolCAP_c2_4978	90	11	20780633	49	3.81*	1.35	0.08	1.43	Ribosomal protein L41
SolCAP_c2_44634	90	11	24839121	49	3.55*	1.45	0.10	1.56	Electron transporter
SolCAP_c2_11571	75	13	44567126	11	1.99	0.64	5.06*	0.47	L-allo-threonine aldolase

Table A8. Markers associated with stolon number.

Marker	DAP	Chr. no.	Position	MAF	SD Exp2	LD Exp1	LD Exp2	Annotation
SolCAP_c2_56356	90	1	11737575	13	0.03	4.19*	0.26	Conserved gene of unknown function
SolCAP_c2_21974	75	4	56449719	34	1.57	0.41	3.48*	NG
SolCAP_c2_33489	90	7	45146538	11	3.58*	0.05	2.70	Glucan endo-1,3-beta-glucosidase
SolCAP_c2_30416	75	7	54274624	6	1.13	4.30*	1.29	Conserved gene of unknown function
SolCAP_c2_34103	75	8	4235737	37	3.79*	0.23	0.70	Conserved gene of unknown function 2-methyl-6-phytylbenzoquinone
SolCAP_c1_13515	75	9	48611285	36	0.92	3.77*	0.35	Methyltransferase
SolCAP_c2_27806	90	10	50615253	16	2.47	0.09	4.08*	Ethylene-forming-enzyme-like dioxygenase
SolCAP_c2_27808	90	10	50615474	14	2.22	0.01	3.93*	Ethylene-forming-enzyme-like dioxygenase
SolCAP_c2_24564	90	12	243244	40	4.23*	0.31	1.80	Phosphoenolpyruvate carboxylase
SolCAP_c2_34179	75	13	44567126	47	3.99*	0.79	0.56	Ring finger protein

Table A9. Markers associated with stolon length.

Marker	DAP	Chr. no.	Position	MAF	SD Exp2	LD Exp1	LD Exp2	Annotation
SolCAP_c2_51986	90	2	33352248	15	0.08	3.42*	0.09	Fiber protein Fb2
SolCAP_c1_38	75	3	58609665	31	3.93*	1.03	0.89	ArcA2 protein
SolCAP_c2_32042	75	4	67130719	13	0.46	3.36*	0.92	NG
SolCAP_c2_47301	75	5	6395149	10	1.21	3.34*	1.57	Glutathione s-transferase
SolCAP_c2_47302	75	5	6397015	10	1.21	3.34*	1.57	Glutathione s-transferase
SolCAP_c2_47303	75	5	6397142	10	1.21	3.34*	1.57	Glutathione s-transferase
SolCAP_c1_15031	75	8	49396213	12	0.18	3.37*	0.22	Cinnamoyl-CoA reductase
SolCAP_c1_13515	90	9	48611285	36	1.27	3.65*	0.77	2-methyl-6-phytylbenzoquinone methyltransferase
SolCAP_c2_45570	90	9	48721662	28	0.34	3.48*	0.33	NADH-ubiquinone oxidoreductase 24 kDa subunit, mitochondrial
SolCAP_c2_27806	90	10	50615253	16	2.50	0.20	4.01*	Ethylene-forming-enzyme-like dioxygenase
SolCAP_c2_27808	90	10	50615474	14	2.09	0.09	3.50*	Ethylene-forming-enzyme-like dioxygenase
SolCAP_c2_17617	75	12	50382217	20	2.44	1.29	3.80*	Heat shock 70 kDa protein
SolCAP_c2_17615	75	12	50382493	20	2.44	1.29	3.80*	Heat shock 70 kDa protein
SolCAP_c2_24556	90	12	244664	47	3.87*	0.20	1.63	Phosphoenolpyruvate carboxylase
SolCAP_c2_41725	75	13	44567126	17	0.27	3.40*	0.14	Conserved gene of unknown function
SolCAP_c2_47300	75	13	44567126	10	1.21	3.34*	1.57	Glutathione s-transferase

Table A10. Top 10 clones for tuber induction (TI59) based on RR GEBV under short or long day conditions.

Clones with best GEBV under short day conditions		Clones with best GEBV under long day conditions	
Clone	RR GEBV	Clone	RR GEBV
Granola	0.84	CIP-300135.3	0.63
CIP-300135.14	0.73	Granola	0.53
CIP-300135.3	0.70	CIP-301045.74	0.46
CIP-397077.16	0.62	Spunta	0.35
Spunta	0.58	CIP-300055.32	0.25
CIP-301026.23	0.56	CIP-300135.14	0.23
CIP-300137.31	0.22	CIP-301026.23	0.14
CIP-301024.95	0.20	CIP-397077.16	0.09
CIP-398193.158	0.19	CIP-300137.31	0.08
CIP-301045.74	0.15	CIP-300065.4	0.07

Table A11. Top 10 clones for number of small tubers (ST90) based on RR GEBV under short or long day conditions.

Clones with best GEBV under short day conditions		Clones with best GEBV under long day conditions	
Clone	RR GEBV	Clone	RR GEBV
CIP-394223.19	-5.92	CIP-394901.2	-21.33
CIP-301023.15	-5.04	CIP-301056.54	-19.28
CIP-300055.32	-4.41	CIP-394900.1	-17.71
CIP-301056.54	-3.63	CIP-300055.32	-15.39
CIP-301024.14	-3.58	CIP-301055.53	-12.74
CIP-301024.95	-3.04	CIP-394223.19	-12.08
CIP-300137.31	-2.74	CIP-394898.13	-11.09
CIP-394901.2	-2.51	CIP-394223.9	-5.88
CIP-394223.9	-2.18	Tomasa	-5.40
CIP-398208.29	-2.12	CIP-301044.36	-5.37

Table A12. Top 10 clones for number of marketable tubers (MT90) based on RR GEBV under short or long day conditions.

Clones with best GEBV under short day conditions		Clones with best GEBV under long day conditions	
Clone	RR GEBV	Clone	RR GEBV
CIP-301029.18	1.39	CIP-397077.16	6.87
CIP-301023.15	1.34	Granola	4.34
CIP-397077.16	1.19	CIP-300135.14	3.97
CIP-300065.4	0.77	Spunta	2.72
CIP-394223.19	0.35	CIP-301026.23	2.50
CIP-301024.14	0.17	CIP-301045.74	1.30
CIP-398208.29	-0.03	CIP-300065.4	1.27
CIP-394900.1	-0.33	CIP-394223.19	0.81
Spunta	-0.35	CIP-300135.3	0.44
CIP-300135.14	-0.45	CIP-301024.95	-0.33

Table A13. Top 10 clones for total number of tubers (TT90) based on RR GEBV under short or long day conditions.

Clones with best GEBV under short day conditions		Clones with best GEBV under long day conditions	
Clone	RR GEBV	Clone	RR GEBV
CIP-300135.14	4.95	Granola	12.12
CIP-300135.3	2.81	CIP-300135.14	11.83
CIP-301040.63	2.53	CIP-397077.16	9.25
Yungay	1.99	CIP-300135.3	9.04
CIP-301045.74	1.94	CIP-398190.2	7.72
Tomasa	1.77	Yungay	7.11
CIP-397077.16	1.42	Spunta	6.77
CIP-398190.523	0.47	CIP-301024.95	6.55
CIP-301044.36	0.42	CIP-301026.23	5.33
CIP-398190.2	0.28	CIP-301040.63	4.74

Table A14. Top 10 clones for stolon number and length (SN90 and SL90) based on RR GEBV under long day conditions.

Clones with best GEBV for SN90 under long day conditions		Clones with best GEBV for SL90 under long day conditions	
Clone	RR GEBV	Clone	RR GEBV
Spunta	-1.12	CIP-397077.16	-0.98
CIP-301045.74	-0.98	CIP-300135.3	-0.88
CIP-397077.16	-0.87	Tomasa	-0.83
KufriJYOTI	-0.86	CIP-301045.74	-0.70
CIP-394898.13	-0.74	Spunta	-0.64
CIP-301040.63	-0.70	Yungay	-0.58
CIP-301026.23	-0.65	KufriJYOTI	-0.44
Tomasa	-0.63	Perricholi	-0.39
CIP-300137.31	-0.57	CIP-301040.63	-0.36
CIP-398193.158	-0.45	CIP-300137.31	-0.35

Table A15. Genes reported to be involved in photoperiod responses.

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

REFERENCES

Abelenda, J.A., Navarro, C., and Prat, S. (2011). From the model to the crop: genes controlling tuber formation in potato. Current Opinion in Biotechnology *22*, 287-292.

Abelenda, J.A., Navarro, C., and Prat, S. (2014). Flowering and tuberization: a tale of two nightshades. Trends in Plant Science *19*, 115-122.

Achenbach, U., Paulo, J., Ilarionova, E., Lübeck, J., Strahwald, J., Tacke, E., Hofferbert, H.R., and Gebhardt, C. (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.

Bates, D., Maechler, M., Bolker, B.M., and Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. Journal of Statistical Software *1*, arXiv:1406.5823.

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., Amoros, W., and Landeo, J. (2003). Improved resistance and quality in potatoes for the tropics. Acta Horticulturae, 15-22.

Bonierbale, M., Plaisted, R.L., and Tanksley, S.D. (1988). RFLP Maps Based on a Common Set of Clones Reveal Modes of Chromosomal Evolution in Potato and Tomato. Genetics *120*, 1095.

Botstein, D., White, R.L., Skolnick, M., and Davis, R.W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. The American Journal of Human Genetics *32*, 314-331.

Box, G.E.P., and Cox, D.R. (1964). An Analysis of Transformations. Journal of the Royal Statistical Society *26*, 211-252.

Bradeen, J.M., and Kole, C. (2011). Genetics, Genomics and Breeding of Potato (Boca Raton, FL: Science Publishers).

Bradshaw, J., and Ramsay, G. (2009). Potato origin and Production. In Advances in potato chemistry and technology, J. Singh, and L. Kaur, eds. (Boston: Academic Press), pp. 1-26.

Bryan, G.J. (2007). Genomics. In Potato biology and biotechnology advances and perspectives, D. Vreugdenhil, and J. Bradshaw, eds. (Amsterdam ; Boston: Elsevier).

Bryan, G.J. (2011). Mapping Complex Potato traits. In Genetics, Genomics and Breeding of Potato, J.M. Bradeen, and C. Kole, eds. (Boca Raton, FL: Science Publishers), pp. 113-130.

Cabrera-Bosquet, L., Crossa, J., Von Zitzewitz, J., Serret, M.D., and Luis Araus, J. (2012). High-throughput Phenotyping and Genomic Selection: The Frontiers of Crop Breeding Converge. Journal of Integrative Plant Biology *54*, 312-320.

CGIAR (2000). 25 Years of Food and Agriculture Improvement in Developing Countries. In CGIAR-Mandated Food Crops, Consultative Group on International Agricultural Research, ed. (http://www.worldbank.org/html/cgiar/25years/25cover.html: CGIAR).

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.

Cheng, X.F., and Wang, Z.Y. (2005). Overexpression of COL9, a CONSTANS-LIKE gene, delays flowering by reducing expression of CO and FT in *Arabidopsis thaliana*. Plant Journal *43*, 758-768.

Crossa, J., De los Campos, G., Perez, P., Gianola, D., Burgueno, J., Araus, J.L., Makumbi, D., Singh, R.P., Dreisigacker, S., Jianbing, Y.*,* Arief, V., Banziger, M., Braun, H. (2010). Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. Genetics *186*, 713.

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 Koeyer, D., Chen, H., and Gustafson, V. (2011). Molecular Breeding for Potato Improvement. In Genetics, Genomics and Breeding of Potato, J.M. Bradeen, and C. Kole, eds. (Boca Raton, FL: Science Publishers), pp. 41-59.

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).

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

Douches, D., Hirsch, C., Manrique-Carpintero, N., Massa, A., Coombs, J., Hardigan, M., Bisognin, D., Jong, W., and Buell, C. (2014). The Contribution of the Solanaceae Coordinated Agricultural Project to Potato Breeding. Journal of the European Association for Potato Research *57*, 215-224.

Douches, D.S., Maas, D., Jastrzebski, K., and Chase, R.W. (1996). Assessment of Potato Breeding Progress in the USA over the Last Century. Crop Science *36*, 1544.

Doyle, J.J., and Dickson, E.E. (1987). Preservation of Plant Samples for DNA Restriction Endonuclease Analysis. Taxon *36*, 715-722.

Driver, C.M., Hawkes, J.G., and Genetics, C.B.o.P.B.a. (1943). Photoperiodism in the potato (Cambridge, England.: School of Agriculture).

Edwards, J., and McCouch, S. (2007). Molecular markers for use in plant molecular breeding and germoplasm evaluation. In Marker-assisted selection : current status and future perspectives in crops, livestock, forestry and fish, E.P. Guimarães, ed. (Rome: Food and Agriculture Organization of the United Nations), pp. 29-50.

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., and Mitchell, S.E. (2011). A Robust, Simple Genotyping-by- Sequencing (GBS) Approach for High Diversity Species. PLoS ONE *6*, e19379.

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

Ewing, E.E. (1978). Critical photoperiods for tuberization: A screening technique with potato cuttings. American Potato Journal American Potato Journal *55*, 43-53.

FAOSTAT (2013). Crop Production (http://faostat.fao.org/site/567/DesktopDefault.aspx?PageID=567#ancor).

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.

Flint-Garcia, S.A., Thuillet, A.c., Yu, J., Pressoir, G., Romero, S.M., Mitchell, S.E., Doebley, J., Kresovich, S., Goodman, M.M., and Buckler, E.S. (2005). Maize association population: a high-resolution platform for quantitative trait locus dissection. Plant Journal *44*, 1054-1064.

Foster, J., Allan, G., Chan, A., Rabinowicz, P., Ravel, J., Jackson, P., and Keim, P. (2010). Single nucleotide polymorphisms for assessing genetic diversity in castor bean (*Ricinus communis*). BMC Plant Biology *10*, 13.

Fu, J., Yang, L., and Dai, S. (2015). Identification and characterization of the CONSTANS-like gene family in the short-day plant Chrysanthemum lavandulifolium. Molecular Genetics and Genomics *290*, 1039-1054.

Garner, C., and Slatkin, M. (2003). On selecting markers for association studies: Patterns of linkage disequilibrium between two and three diallelic loci. Genetic Epidemiology *24*, 57-67.

Gebhardt, C. (2005). Potato genetics: Molecular Maps and More. In Molecular marker systems in plant breeding and crop improvement, H. Lörz, and G. Wenzel, eds. (Berlin: Springer), pp. 215-224.

Gebhardt, C. (2011). Population Genetics and Association Mapping. In Genetics, Genomics and Breeding of Potato, J.M. Bradeen, and C. Kole, eds. (Boca Raton, FL: Science Publishers), pp. 133-148.

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.

Glendinning, D.R. (1983). Potato Introductions and Breeding up to the Early 20th Century. NPH New Phytologist *94*, 479-505.

Goh, L., and Yap, V.B. (2009). Effects of normalization on quantitative traits in association test. BMC Bioinformatics *10*, 415-415.

Gregory, L.E. (1956). Some Factors for Tuberization in the Potato Plant. American Journal of Botany *43*.

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*.

Hall, D., Tegstrèom, C., and Ingvarsson, P.a.K. (2010). Using association mapping to dissect the genetic basis of complex traits in plants. Briefings in Functional Genomics *9*, 157-165.

Hamilton, J., Hansey, C.N., Whitty, B.R., Stoffel, K., Massa, A.N., Van Deynze, A., De Jong, W.S., Douches, D.S., and Buell, C.R. (2011). Single nucleotide polymorphism discovery in elite North American potato germplasm. BMC Genomics *12*.

Haverkort, A.J. (2007). Potato crop response to radiation and day length. In Potato biology and biotechnology advances and perspectives, D. Vreugdenhil, and J. Bradshaw, eds. (Amsterdam: Elsevier).

Hawkes, J.G. (1990). The potato: evolution, biodiversity and genetic resources (Washington, D.C.: Smithsonian Institution Press).

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

Hill, W., and Robertson, A. (1968). Linkage disequilibrium in finite populations. International Journal of Plant Breeding Research *38*, 226-231.

Hirsch, C.N., Hirsch, C.D., Felcher, K., Coombs, J., Zarka, D., Van Deynze, A., De Jong, W., Veilleux, R.E., Jansky, S., Bethke, P., Douches, D., Buell, C. (2013). Retrospective View of North American Potato (*Solanum tuberosum L.*) Breeding in the 20th and 21st Centuries. Genes|Genomes|Genetics *3*, 1003-1013.

Hoerl, A., and Kennard, R. (2000). Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics *42*, 80-86.

Holland, J.B., Nyquist, W., and Cervantes-Martinez, C. (2003). Estimating and interpreting heritability for plant breeding: An update. Plant Breeding Reviews *22*, 9-113.

Illumina (2005). Genome Studio, an Integrated Platform for Data Visualization and Analysis.

International Potato Center (2009). World Potato Atlas, K. Theisen, ed. (https://research.cip.cgiar.org/confluence/display/wpa/Peru).

International Potato Center (2010). Potato Facts and Figures, CIP, ed. (http://cipotato.org/potato/facts/).

Jackson, S.D. (1999). Multiple Signaling Pathways Control Tuber Induction in Potato. PLANT PHYSIOLOGY -ROCKVILLE PIKE BETHESDA- *119*, 1-8.

Jackson, S.D. (2009). Plant responses to photoperiod (Oxford, UK), pp. 517-531.

Jannink, J.-L., Lorenz, A.J., and Iwata, H. (2010). Genomic selection in plant breeding: from theory to practice. Briefings in Functional Genomics *9*, 166-177.

Jansky, S. (2009). Breeding, Genetics, and Cultivar Development. In Advances in potato chemistry and technology, J. Singh, and L. Kaur, eds. (Boston: Academic Press), pp. 27-62.

Johnson, D.A., and Powelson, M.L. (2008). Potato health management (St. Paul, MN: American Phytopathological Society).

Jolliffe, I.T. (2002). Principal component analysis, 2nd ed.. edn (New York: Springer).

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., Smaczniak, C. (2013). Naturally occurring allele diversity allows potato cultivation in northern latitudes. Nature *495*, 246-250.

Kojima, S., Takahashi, Y., Kobayashi, Y., Monna, L., Sasaki, T., Araki, T., and Yano, M. (2002). Hd3a, a rice ortholog of the Arabidopsis FT gene, promotes transition to flowering downstream of Hd1 under short-day conditions. Plant & Cell Physiology *43*, 1096-1105.

Kooman, P.L., Fahem, M., Tegera, P., and Haverkort, A.J. (1996). Effects of climate on different potato genotypes 2. Dry matter allocation and duration of the growth cycle. European Journal of Agronomy *5*, 207-217.

Labbe, A., and Wormald, H. (2005). On normality, ethnicity, and missing values in quantitative trait locus mapping. BMC Genetics *6*, S52-S52.

Lagercrantz, U. (2009). At the end of the day: a common molecular mechanism for photoperiod responses in plants? Journal of experimental botany *60*, 2501-2515.

Langridge, P., and Chalmers, K. (2005). The Principle: Identification and Application of Molecular Markers. In Molecular marker systems in plant breeding and crop improvement, H. Lörz, and G. Wenzel, eds. (Berlin: Springer).

Li, J., Lindqvist-Kreuze, H., Tian, Z., Liu, J., Song, B., Landeo, J., Portal, L., Gastelo, M., Frisancho, J., Sanchez, L., Meijer, D., Xie, C., Bonierbale, M. (2012). Conditional QTL underlying resistance to late blight in a diploid potato population. International Journal of Plant Breeding Research *124*, 1339-1350.

Li, L., Paulo, M.-J.a., Strahwald, J., Lèubeck, J., Hofferbert, H.-R., Tacke, E., Junghans, H., Wunder, J.o., Draffehn, A., Eeuwijk, F., Gebhardt, C. (2008). Natural DNA variation at candidate loci is associated with potato chip color, tuber starch content, yield and starch yield. Theoretical and Applied Genetics *116*, 1167-1181.

Lindqvist-Kreuze, H., Gastelo, M., Perez, W., Forbes, G., De Koeyer, D., and Bonierbale, M. (2014). Phenotypic stability and genome wide association study of late blight resistance in potato genotypes adapted to the tropical highlands. Phytopathology, 140114150216007.

Ma, J., and Amos, C.I. (2012). Principal Components Analysis of Population Admixture. PLoS ONE *7*, e40115.

Mackay, I., and Powell, W. (2007). Methods for linkage disequilibrium mapping in crops. Trends in Plant Science *12*, 57-63.

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.

Martínez-García, J.F., Virgós-Soler, A., and Prat, S. (2002). Control of Photoperiod-Regulated Tuberization in Potato by the Arabidopsis Flowering-Time Gene CONSTANS. Proceedings of the National Academy of Sciences of the United States of America *99*, 15211-15216.

Massa, A.N., Manrique-Carpintero, N.C., Coombs, J.J., Zarka, D.G., Boone, A.E., Kirk, W.W., Hackett, C.A., Bryan, G.J., and Douches, D.S. (2015). Genetic Linkage Mapping of Economically Important Traits in Cultivated Tetraploid Potato (*Solanum tuberosum L.*). Genes|Genomes|Genetics *115*.

Mihovilovich, E., Alarcon, L., Perez, A.L., Alvarado, J., Arellano, C., and Bonierbale, M. (2007). High levels of heritable resistance to Potato leafroll virus (PLRV) in Solanum tuberosum subsp, andigena. Crop Science *47*, 1091.

Minitab, I. (2010). Minitab 17 (www.minitab.com: Minitab, Inc.).

Moore, B. (1920). The Relative Length of Day and Night. Ecology *1*, 234-237.

Myles, S., Peiffer, J., Brown, P.J., Ersoz, E.S., Zhang, Z., Costich, D.E., and Buckler, E.S. (2009). Association Mapping: Critical Considerations Shift from Genotyping to Experimental Design. The Plant Cell *21*, 2194-2202.

National Research Council (U.S.), and Advisory Committee on Technology Innovation (1989). Lost crops of the Incas little-known plants of the Andes with promise for worldwide cultivation (Washington, D.C.: National Academy Press).

Navarro, C., Abelenda, J.A., Cruz-Oró, E., Cuéllar, C.A., Tamaki, S., Silva, J., Shimamoto, K., and Prat, S. (2011). Control of flowering and storage organ formation in potato by FLOWERING LOCUS T. Nature *478*, 119.

Nevo, E. (2001). Evolution of genome-phenome diversity under environmental stress. Proceedings of the National Academy of Sciences of the United States of America *98*, 6233-6240.

Ochoa, C.M., and Center, I.P. (1990). The potatoes of South America (Cambridge, England: Cambridge University Press).

Oraguzie, N.C. (2007). Association mapping in plants (New York: Springer).

Pearson, T.A., and Manolio, T.A. (2008). How to interpret a genome-wide association study. JAMA *299*, 1335.

Peng, B., Yu, R.K., DeHoff, K.L., and Amos, C.I. (2007). Normalizing a large number of quantitative traits using empirical normal quantile transformation. BMC Proceedings *1*, S156-S156.

Pereira, A., and Shock, C. (2006). Agrometeorology of Some Selected Crops, Potato. In Guide to Agricultural Meteorological Practices, W.M. Organization, ed. (Geneva, Switzerland).

Pin, P.A., Benlloch, R., Bonnet, D., Wremerth-Weich, E., Kraft, T., Gielen, J.J.L., and Nilsson, O. (2010). An Antagonistic Pair of FT Homologs Mediates the Control of Flowering Time in Sugar Beet. Science *330*, 1397-1400.

Poland, J., Endelman, J., Dawson, J., Rutkoski, J., Wu, S., Manes, Y., Dreisigacker, S., Crossa, J., Sánchez-villeda, H., Sorrells, M*.,* Jannink, J. (2012). Genomic Selection in Wheat Breeding using Genotyping-by-Sequencing. Plant Genome *5*, 103.

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

Rafalski, J.A. (2010). Association genetics in crop improvement. Current opinion in Plant Biology *13*, 174-180.

Rickert, A.M., Kim, J.H., Meyer, S., Nagel, A., Ballvora, A., Oefner, P.J., and Gebhardt, C. (2003). First-generation SNP/InDel markers tagging loci for pathogen resistance in the potato genome. Plant Biotechnology *1*, 399-410.

Ritland, K. (1996). Estimators for pairwise relatedness and individual inbreeding coefficients. Genetics *67*, 175-185.

Rodríguez-Falcón, M., Bou, J., and Prat, S. (2006). Seasonal control of tuberization in potato: conserved elements with the flowering response. Annual Review of Plant Biology *57*, 151-180.

Salazar, J., Saithong, T., Brown, P.E., Foreman, J., Locke, J.C., Halliday, K.J., Carré, I.A., Rand, D.A., and Millar, A.J. (2009). Prediction of photoperiodic regulators from quantitative gene circuit models. Cell *139*, 1170-1179.

Sarkar, D. (2008). The signal transduction pathways controlling in planta tuberization in potato: an emerging synthesis. Plant Cell Reports *27*, 1-8.

SAS Institute (1999). SAS 9.3. Manufacturing Systems Europe *2*, 7.

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.

Simko, I., Haynes, K.G., and Jones, R.W. (2006). Assessment of linkage disequilibrium in potato genome with single nucleotide polymorphism markers. Genetics *173*, 2237.

Simpson, G.G. (2003). Evolution of flowering in response to day length: Flipping the CONSTANS switch. BioEssays *25*, 829-832.

Song, Y., Hun, L., Ilha, L., Sang, Y., Imaizumi, T., and Hong, J.C. (2012). CONSTANS and ASYMMETRIC LEAVES 1 complex is involved in the induction of FLOWERING LOCUS T in photoperiodic flowering in Arabidopsis. The Plant Journal *69*, 332-342.

Spooner, D., and Bamberg, J.B. (1994). Potato genetic resources: Sources of resistance and systematics. American Potato Journal *71*, 325-337.

Spooner, D., Nuñez, J., Rodríguez, F., Naik, P.S., and Ghislain, M. (2005). Nuclear and chloroplast DNA reassessment of the origin of Indian potato varieties and its implications for the origin of the early European potato. Theoretical and Applied Genetics *110*, 1020-1026.

Spooner, D.M., and Bryan, G.J. (2005). A Single Domestication for Potato Based on Multilocus Amplified Fragment Length Polymorphism Genotyping. Proceedings of the National Academy of Sciences of the United States of America *102*, 14694-14699.

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.

Stich, B., Urbany, C., Hoffmann, P., and Gebhardt, C. (2013). Population structure and linkage disequilibrium in diploid and tetraploid potato revealed by genome-wide high-density genotyping using the SolCAP SNP array. Plant Breeding *132*, 718-724.

Storey, M. (2007). The Harvested Crop. In Potato biology and biotechnology advances and perspectives, D. Vreugdenhil, and J. Bradshaw, eds. (Amsterdam ; Boston: Elsevier).

Thomas, B., and Vince-Prue, D. (1997). Photoperiodism in plants (London: Academic Press).

Turck, F., Fornara, F., and Coupland, G. (2008). Regulation and identity of florigen: FLOWERING LOCUS T moves center stage. Annual Review of Plant Biology *59*, 573-594.

Uitdewilligen, J.G., Wolters, A.-M.A., Hoop, B.B., Borm, T.J.A., Visser, R.G.F., and van Eck, H.J. (2013). A next-generation sequencing method for genotyping-by-sequencing of highly heterozygous autotetraploid potato. PLoS ONE *8*, e62355.

Urbany, C., Stich, B., Schmidt, L., Simon, L., Berding, H., Junghans, H., Niehoff, K.H., Braun, A., Tacke, E., Hofferbert, H.*,* Lubeck, J., Strahwald, J., Gebhardt, C. (2011). Association genetics in Solanum tuberosum provides new insights into potato tuber bruising and enzymatic tissue discoloration. BMC Genomics *12*.

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

Wilk, M.B., and Gnanadesikan, R. (1968). Probability plotting methods for the analysis of data. Biometrika *55*.

Wilson, L.M., Whitt, S.R., Ibáñez, A.M., Rocheford, T.R., Goodman, M., and Buckler, E.S. (2004). Dissection of Maize Kernel Composition and Starch Production by Candidate Gene Association. The Plant Cell *16*, 2719-2733.

Xu, X., Pan, S., Cheng, S., Zhang, B., Mu, D., Ni, P., Zhang, G., Yang, S., Li, R., Wang, J.*,* Orjeda, G., Guzman, F., Torres, M., Lozano, R., Ponce, O., Martinez, D., De la Cruz, G., Chakrabarti, S.K., Patil, V.U., Skryabin, K.G., Kuznetsov, B.B., Ravin, N.V., Kolganova, T.V., Beletsky, A.V., Mardanov, A.V., Di Genova, A., Bolser, D.M., Martin, D.M., Li, G., Yang, Y., Kuang, H., Hu, Q., Xiong, X., Bishop, G.J., Sagredo, B., Mejía, N., Zagorski, W., Gromadka, R., Gawor, J., Szczesny, P., Huang, S., Zhang, Z., Liang, C., He, J., Li, Y., He, Y., Xu, J., Zhang, Y., Xie, B., Du, Y., Qu, D., Bonierbale, M., Ghislain, M., Herrera, R., Giuliano, G., Pietrella, M., Perrotta, G., Facella, P., O'Brien, K., Feingold, S.E., Barreiro, L.E., Massa, G.A., Diambra, L., Whitty, B.R., Vaillancourt, B., Lin, H., Massa, A.N., Geoffroy, M., Lundback, S., DellaPenna, D., Buell, C.R., Sharma, S.K., Marshall, D.F., Waugh, R., Bryan, G.J., Destefanis, M., Nagy, I.,

Milbourne, D., Thomson, S.J., Fiers, M., Jacobs, J.M., Nielsen, K.L., Sønderkær, M., Iovene, M., Torres, G.A., Jiang, J., Veilleux, R.E., Bachem, C.W., de Boer, J., Borm, T., Kloosterman, B., van Eck, H., Datema, E., Hekkert, Bt., Goverse, A., van Ham, R.C., Visser, R.G. (2011). Genome sequence and analysis of the tuber crop potato. Nature *475*, 189-195.

Yanagisawa, S. (2002). The Dof family of plant transcription factors. Trends in Plant Science *7*, 555-560.

Yu, J., Pressoir, G., H. Briggs, W., Bi, I.V., Yamasaki, M., Doebley, J., McMullen, M.D., S. Gaut, B., M. Nielsen, D., Holland, J.*,* Kresovich, S., Buckler, E. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature Genetics *38*, 203.

Zhang, Z., Ersoz, E., Lai, C.-Q., Todhunter, R.J., Tiwari, H.K., Gore, M.A., Bradbury, P.J., Yu, J., Arnett, D.K., Ordovas, J.M.*,* Buckler, E. (2010). Mixed linear model approach adapted for genome-wide association studies. Nature Genetics *42*, 355-360.