Using comparative transcript profiling and association mapping to detect QTLs and diagnostic SNP markers for maturity corrected resistance to *Phytophthora infestans* in potato (*Solanum tuberosum* L.)

Inaugural-Dissertation

Zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Meki Shehabu Muktar

aus Dilla,

Äthiopien

Köln, 2014

Die vorliegende Arbeit wurde am Max-Planck-Institut für Pflanzenzüchtungsforschung in der Abteilung für Pflanzenzüchtung und Genetik (Direktor Prof. Dr. Maarten Koornneef) angefertigt.

Max Planck Institute for
Plant Breeding Research

Berichterstatter	**PD Dr. Christiane Gebhardt**
(Gutachter)	**Prof. Dr. Ulf-Ingo Flügge**
Tag der mündlichen Prüfung	**27.06.2014**

Abstract

Late blight of potato (*Solanum tuberosum* L.) caused by the oomycete *Phytophthora infestans* (Mont.) de Bary, is one of the most important bottlenecks of potato production worldwide. The disease could be better managed by using cultivars with high levels of quantitative or field resistance, which is mostly race non-specific and more durable than single *R*-gene mediated resistance. Breeding for quantitative resistance is however, a complex and challenging process and requires multiple year and location trials. In addition, it is hampered by a correlation between resistance and late maturity. This problem can be tackled by using diagnostic DNA markers, which enhance efficiency and precision of resistance breeding. The aim of this project was the identification of genes for quantitative resistance to *P. infestans* not compromised by late maturity and the discovery of diagnostic single nucleotide polymorphism (SNP) markers. To meet the objectives, the analysis of candidate genes and genome wide association (GWA) mapping of SNP markers were used.

In the first approach, novel candidate genes that were differentially expressed in quantitative resistant versus susceptible potato genotype pools, were selected from transcriptome data generated using SuperSAGE (serial analysis of gene expression) analysis of nine samples comprising three genotype groups with different resistance levels and three infection time points. Twenty-two selected novel candidate genes were subjected to validation of differential expression by quantitative real time PCR (qRT-PCR) and allele specific pyrosequencing. Candidates showing reproducible transcriptional regulation in three independent infection experiments were tested for association with maturity corrected resistance (MCR) to late blight in a population of 184 tetraploid genotypes (CONQUEST2 population). Using mixed linear model (MLM) analysis including kinship and population structure, six novel candidate genes associated with late blight resistance not compromised by late maturity were identified.

In the second approach, GWA mapping was conducted by scanning 8303 SNP markers distributed across the 12 potato chromosomes. The genotyping of the CONQUEST2 population was performed with the Illumina SolCAP SNP potato genotyping array and produced a total of 6286 informative bi-allelic SNPs. Approximately 5600 SNP markers could be used for association analysis. No or very weak relatedness and sub populations were detected in CONQUEST2 population. The proportion of pair wise r^2 values > 0.1 (loci are in

linkage disequilibrium: LD) and > 0.8 (loci are in strong LD) was only 1.57 % and 0.01%, respectively. The trendline of the nonlinear regression curve reached the threshold level for LD, r^2 = 0.1, between 270 and 280 bps, showing a rapid LD decay in the potato genome. This fast LD decay implies that a large number of genome wide markers are required for detecting all quantitative trait loci (QTLs) in potato by GWA mapping. Nevertheless, genome wide marker-trait association analysis, with correction for kinship and population structure in a mixed linear model (MLM), identified few novel SNP markers associated with late blight resistance not compromised by late maturity.

In conclusion, comparative transcript profiling combined with association mapping can be used for the detection of novel late blight QTLs and diagnostic SNP markers that can be used in marker-assisted resistance breeding. The SolCAP potato genotyping array is a useful but limited tool in identification of diagnostic SNP markers for agronomic traits in potato.

Zusammenfassung

Die Kraut- und Knollenfäule der Kartoffel (*Solanum tuberosum* L.), verursacht durch den Oomyzeten *Phytophthora infestans* (Mont.) de Bary, ist weltweit eines der wichtigsten Probleme im Kartoffelanbau. Die Krankheit könnte besser beherrscht werden durch den Anbau von Sorten mit hoher quantitativer oder Feldresistenz, die überwiegend Rassen-unspezifisch ist und eine längere Lebensdauer hat als die durch einzelne R-Gene vermittelte Resistenz. Die Züchtung auf quantitative Resistenz ist jedoch komplex und schwierig und erfordert mehrjährige Feldversuche, die zusätzlich erschwert wird durch die Korrelation zwischen Resistenz und später Reifezeit. Eine mögliche Lösung dieses Problems sind diagnostische DNA Marker, welche Effizienz und Präzision der Resistenzzüchtung erhöhen. Ziel dieses Projekts war die Identifizierung von Genen für quantitative Resistenz gegen *P. infestans* ohne Beeinträchtigung durch späte Reifezeit, sowie die Entdeckung von diagnostischen ‚single nucleotide polymorphism' (SNP) Markern. Um diese Ziele zu erreichen, wurde die Analyse von Kandidatengenen und die genomweite Assoziationskartierung (GWA) von SNP Markern verwendet.

Im ersten Ansatz wurden neue Kandidatengene ausgewählt, die in Gruppen von quantitativ resistenten bzw. anfälligen Kartoffel-Genotypen differenziell exprimiert waren. Die Auswahl erfolgte auf der Basis von Transkriptom-Daten, die mittels SuperSAGE (‚serial analysis of gene expression') Analyse von neun Proben, bestehend aus drei Genotyp Gruppen mit unterschiedlicher Resistenz und drei Infektionszeitpunkten, erhalten worden waren. Die differentielle Expression von zweiundzwanzig der ausgewählten neuen Kandidatengene wurde mittels quantitativer ‚real time' PCR (qRT-PCR) und allelspezifischer Pyrosequenzierung überprüft. Kandidaten, deren transkriptionelle Regulation in drei unabhängigen Infektionsexperimenten reproduzierbar war, wurden auf Assoziation mit reifekorrigierter Krautfäuleresistenz (‚maturity corrected resistance': MCR) in einer Population von 184 tetraploiden Genotypen (CONQUEST2 Population) geprüft. Mit Hilfe eines gemischt linearen Modells (MLM) unter Berücksichtigung von Verwandtschaft und Populationsstruktur wurden sechs neue Kandidatengene identifiziert, die mit Krautfäuleresistenz ohne Beeinträchtigung durch späte Reifezeit assoziiert sind.

Im zweiten Ansatz wurde eine GWA Kartierung durchgeführt mit Hilfe von 8303 SNP Markern, die über die zwölf Kartoffelchromosomen verteilt waren. Die Genotypisierung der

CONQUEST2 Population erfolgte mit dem ‚Illumina SolCAP potato genotyping array' und ergab 6286 informative, bi-allelische SNP Marker, von denen etwa 5600 für die Assoziationsanalyse verwendbar waren. In der CONQUEST2 Population wurde keine oder nur sehr schwache Verwandschaft und Populationsstruktur gefunden. Der Anteil von paarweisen r^2 Werten > 0.1 (zwei Loci befinden sich im Kopplungsungleichgewicht, ‚linkage disequilibrium: LD) und > 0.8 (Loci befinden sich in sehr starkem LD) betrug nur 1.57% bzw. 0.01%. Die Trendlinie der nichtlinearen Regressionskurve erreichte den LD Grenzwert r^2 = 0.1 zwischen 270 und 280 Basenpaaren. Dies zeigte, dass LD im Kartoffelgenom mit dem physikalischen Abstand rasch abnimmt. Diese rasche LD Abnahme bedeutet, dass eine große Zahl genomweiter Marker erforderlich ist, um alle ‚quantitative trait loci' (QTL) mittels GWA zu erfassen. Nichtsdestotrotz identifizierte die genomweite Marker-Merkmal Assoziationsanalyse, unter Verwendung von MLM mit einer Korrektur für Verwandtschaft und Populationsstruktur, einige neue SNP Marker, die mit Krautfäuleresistenz ohne Beeinträchtigung durch späte Reifezeit assoziiert sind.

Schlussfolgerung: Die vergleichende Analyse von Transkriptmengen in Kombination mit Assoziationskartierung kann für die Identifizierung von neuen QTLs und von diagnostischen SNP Markern verwendet werden, die für eine Marker-gestützte Resistenzzüchtung genutzt werden können. Der ‚SolCAP potato genotyping array' ist ein nützliches, aber auch begrenztes Instrument für die Identifizierung von SNP Markern, die für agromische Merkmale der Kartoffel diagnostisch sind.

Contents

1. General Introduction

1.1. The potato

Potato (*Solanum tuberosum* L.) is a member of the Solanaceae, an economically important family that includes tomato, pepper, eggplant, petunia and tobacco (Xu et al., 2011). Potato produces a subterranean swollen stem, called tuber. The tuber is consumed directly as a staple food or vegetable, processed into chips, french fries, and many other food products. Potato is a major source of carbohydrate energy in the diets of hundreds of millions of people. It is also contain about 2% protein and provide an excellent source of lysine, vitamins C, B6 and B1a, the minerals potassium, phosphorus, calcium, and magnesium, and the micronutrients iron and zinc (Bradshaw and Bonierbale, 2010). Potato is a source for the industrial production of starch, and alcohol. Potato starch is the starting material for the preparation of more than 500 different commercial products (Bradshaw and Bonierbale, 2010), for example for the production of biodegradable plastics (Milbourne et al., 2007). It also serves in unconventional uses as a source of biofuel or as bioreactor for producing pharmaceutical compounds in transgenic plants (Gebhardt et al., 2014).

Potato was originated in the area of present day southern Peru and extreme northwestern Bolivia (Fig. 1.1) (Bradshaw and Bonierbale, 2010), and was domesticated 7, 000 to 13, 000 years ago (Milbourne et al., 2007). The Spanish introduced the potato to Europe in the second half of the 16th century and subsequently distributed throughout the world by European mariners (Bradshaw and Bonierbale, 2010; Milbourne et al., 2007). In the 19th century, potato was taken to East Africa by European missionaries and colonialists. In 1858 a German man took potato to Ethiopia, a country with the greatest potential for potato production among African countries (Haverkort et al., 2012). Currently, potato is the world's most important food crop after wheat and rice. China is the number one potato producer in the world followed by Russia and India (Gebhardt, 2013).

The potato species have different polyploidy levels; from diploid to hexaploid, with a basic haploid chromosome number of x=12 (Bradshaw and Bonierbale, 2010; Milbourne et al., 2007). The cultivated potato, *Solanum tuberosum* ssp. *tuberosum*, is tetraploid (2n=4x=48) with autotetrasomic inheritance, in which the four sets of homologous chromosomes pair

randomly during meiosis (Milbourne et al., 2007). Potato reproduces by sexual means, develop flowers and set true seeds in berries following natural pollination by insects. This sexual reproduction creates high diversity by recombining the variants of genes that arose by mutation. In addition, potato displays inbreeding depression on selfing. Thus, potato is highly heterozygous plant. However, it also reproduces asexually by setting tubers. The genetically unique seedlings that grow from true seeds produce tubers that can be replanted as seed tubers and hence distinct clones can be established and maintained by asexual (vegetative) reproduction. Most potato cultivars are propagated through seed tubers and are genetically uniform (Bradshaw and Bonierbale, 2010).

The potato genome was sequenced using a doubled monoploid *S. tuberosum* group Phureja DM1-3 516 R44. The genome size was estimated to 844 Mbp comprising more than 39, 000 protein coding genes, of which more than 800 are disease resistance genes (Xu et al., 2011).

1.2. *Phytophthora infestans*

Phytophthora infestans is a plant pathogen belongs to the class oomycetes, microorganisms with a similar morphology to fungi. Oomycetes differ from true fungi as their cell walls contain cellulose and lack chitin. They have diploid mycelium in part of their life cycle and biflagellate zoospores (Pereira et al., 2012). Taxonomically, oomycetes are more related to organisms such as brown algae and diatoms (Haas et al., 2009).

P. infestans originated in the Toluca valley of Mexico, where many different strains evolve together with wild potato relatives (Fig. 1.1) (Yoshida et al., 2013; Pereira et al., 2012). In 1845, *P. infestans* reached Europe, spreading rapidly from Belgium to other European countries and then to Great Britain and Ireland. The impact of the epidemic reached catastrophic levels in Ireland, where the population was more dependent on potato for their subsistence than in other parts of Europe. The subsequent great famine killed around 1 million people, and an additional million were migrated to other countries. Ever since triggering the Irish famine, *P. infestans* has continued expanding its territory throughout the world where potato and tomato are growing and remains the most destructive pathogen on the world's most important food crops (Yoshida et al., 2013).

P. infestans first identified by de Bary in 1847 and because of the devastating nature of the pathogen, he named it *Phytophthora infestans* (= infectious plant destroyer) and the resulting disease is called late blight (Fry, 2008). *P. infestans* is an extraordinarily virulent and adaptable pathogen (Haas et al., 2009; Fry, 2008). It reproduces by asexual and sexual means. In asexual cycle, sporangia are produced on sporangiophores that grow from infected tissues. The sporangia are readily dehiscent and can be aerially dispersed to other plant tissues. Sporangia in free water germinate either via a germ tube at higher temperatures (optimum around 20–25 C), or by releasing wall-less zoospores at lower temperatures (optimum between 10 and 15° C) (Kamoun and Smart, 2005). The biflagellated zoospores are motile for a short time before encysting. Encysted zoospores germinate directly via a germ tube to penetrate leaf or stem tissue. It is capable of completing an asexual cycle, from infection to the production of sporangia, in less than 5 days, and the sporangia can be washed off the leaves and fall onto the soil where their spores subsequently infect tubers (Fry and Goodwin, 1997). Under favourable weather conditions, *P. infestans* is capable of destroying a potato crop in a matter of days, resulting in total loss of the crop unless control measures are implemented correctly (Fry, 2008). *P. infestans* also reproduces sexually in the presence of two different thalli, called A1 and A2, and produce oospores, resistant sexual spores that survive in the soil and in crop residues (Judelson and Blanco, 2005; Kamoun and Smart, 2005).

Once inside the plant cells, *Phytophthora* uptake nutrients from the living cell and secrete different types of enzymes and effector proteins that further facilitate the nutrient access. The effector proteins inhibit enzymatic activity and enable protection against plant-derived hydrolytic enzymes (Fry, 2008). The effectors classified into two main categories, depending on their site of action. Apoplastic effectors remain in the apoplastic space where they interact with extracellular host molecules. Cytoplasm effectors translocated into the plant cytoplasm. There are two main families of cytoplasm effectors; the RXLR and Crinkler (CRN) protein families (Nowicki et al., 2012; Haas et al., 2009; Fry, 2008).

P. infestans has a genome size of 240 Mb encoding more than 18,000 genes, with three quarters of the genome consisting of repetitive DNA. A large number of genes codes for effector proteins, many of which are delivered inside plant cells to promote host colonization,

for instance by suppressing plant immunity. RXLR proteins, the main class of host-translocated effectors, are encoded by about 550 genes (Haas et al., 2009).

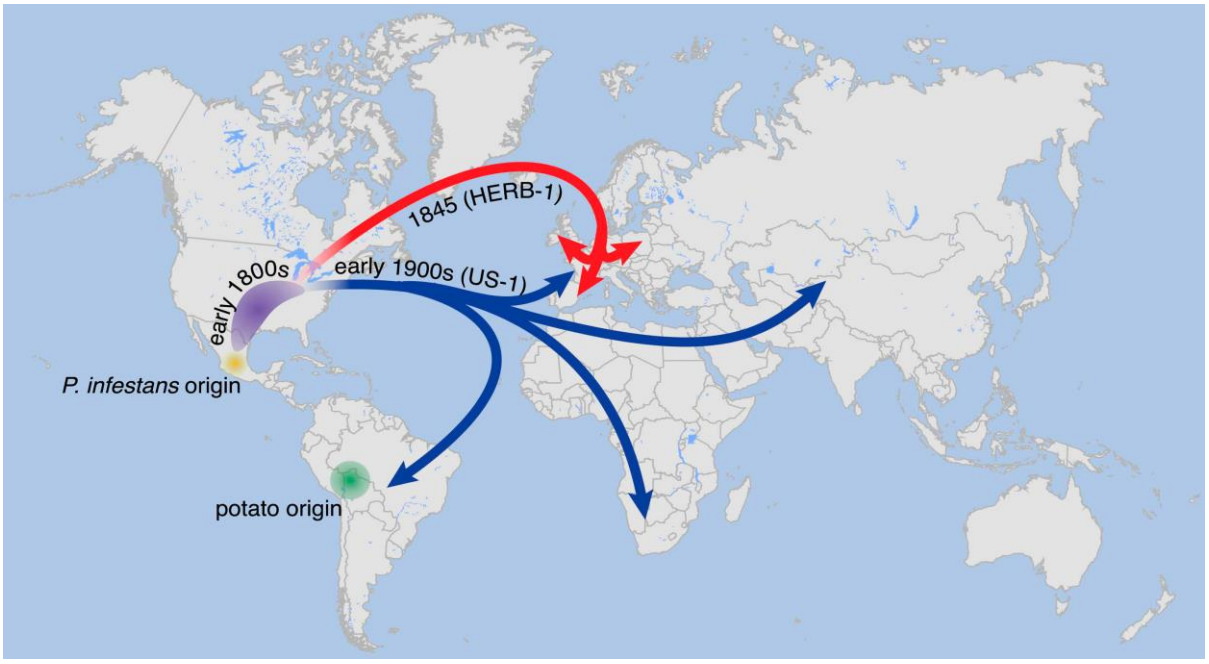


Fig. 1.1. Places of origin for potato and *Phytophthora infestans*. Suggested paths of *P. infestans* migration are indicated by arrow; the path of the strain (HERB-1) that triggered the Irish famine in 1845 shown by red, while the path for the modern strain (US-1) is shown by blue (Yoshida et al., 2013).

1.3. Interaction between potato and *Phytophthora infestans*

Up on the arrival of pathogens on plant surfaces, plants detect the presence of pathogen on their surface by perceiving both chemical and physical signals of pathogen origin, and react rapidly to the attempted infection (Nowicki et al., 2012). The plant-pathogen interactions can be incompatible or compatible. In incompatible interaction, the host plant can successfully defend itself against the pathogen. Usually the plant has a gene for resistance (*R*-gene) that recognizes the corresponding specific gene for avirulence (*Avr*) of the pathogen. In compatible interaction, the pathogen infects the plant, but there is often a big difference in how effectively the plant can defend itself against the pathogen. Even when conditions for infection and disease development are favourable, a plant may develop no disease, only mild disease, or severe disease, depending on the specific genetic makeup of the plant and of the pathogen that attacks it.

1.3.1. Incompatible interaction

Incompatible interaction takes place in the so called qualitative resistance, which also called monogenic resistance and is controlled by a sing *R*-gene. This type of resistance is known by complete arrest of pathogen growth. Most plant *R*-genes are members of a super family of genes that have a nucleotide binding site (NBS) (Rigden et al., 2000) and a leucine-rich repeat (LRR) domain (Dangl & Jones, 2001). *R*-genes operate by detecting pathogen effectors, proteins secreted by the pathogen to facilitate the infection process. If these effector proteins are detected by an *R*-gene they are called avirulence proteins (Avr-proteins). Direct recognition of an Avr-protein by the product of an *R*-gene is rare; more commonly the *R*-gene "guards" a host protein which is the virulence target of the effector and the response is triggered when the activity of the effector on the host is recognized. Recognition of the Avr-protein or its activity by the plant *R*-gene leads to elicitation of a strong defense response accompanied by a hypersensitive response (HR), a form of programmed cell death that localises the pathogen and prevents further spread (Gleason et al., 2008; Jones & Dangl, 2006).

However, pathogens continuously adapt their effectors to evade recognition by the *R*-genes. The plant-pathogen co-evolution processes are well represented by the 'zigzag' model (Jones & Dangl, 2006). In phase 1, PAMPs or MAMPs (Pathogen/Microbe Associated Molecular Patterns) are recognized by PRRs (Pattern Recognition Receptors), resulting in PTI (PAMP-Triggered Immunity). In phase 2, successful pathogens deploy effectors that contribute to pathogen virulence. These effectors suppress PTI resulting in ETS (Effector- Triggered Susceptibility). In phase 3, an NB-LRR protein specifically recognizes a given effector, resulting in ETI (Effector-Triggered Immunity). ETI is an accelerated and amplified PTI, resulting in disease resistance and, usually, a HR at the infection site. In phase 4, natural selection drives pathogens to avoid ETI either by shedding or diversifying the recognized effector, or by acquiring additional effectors that suppress ETI. Again, natural selection results in new resistance specificities so that ETI can be triggered (Jones & Dangl, 2006) and the boom and bust cycle continues.

So far, many race specific *R*-genes resistance to *P. infestans* have been identified from wild potato species and introgressed into varieties (Gebhardt and Valkonen, 2001). 11 *R* genes

(*R1* to *R11*) were identified in *S. demissum*, a wild species originating from Mexico (Gebhardt and Valkonen, 2001). Another three *R* genes, *RB*/*Rpi-blb1*, *Rpi-blb2*, and *Rpi-blb3* were identified from *S. bulbocastanum*, and have been mapped to chromosomes VIII, VI, and IV, respectively (Park et al. 2009; Naess et al. 2001). Another one, *Rpi-vnt1.1*, was identified from *S. venturii* (Foster et al., 2009). All of the *R* genes identified so far are defeated by virulent races of *P. infestans*.

1.3.2. Compatible interaction

In compatible interaction, the pathogen can infect the host plant, but the infection and disease severity vary from plant to plant depending on the level of resistance the plant has. This type of resistance is called quantitative resistance and mediated by multiple genes or quantitative trait loci (QTLs) with each providing a partial increase in resistance. Quantitative resistance is known by slowdown of pathogen growth and disease progression. Compared to *R*-gene mediated resistance, quantitative resistance is characterized by a partial and durable effect of resistance that is generally race-nonspecific but pathogen species-specific.

Little is known about the biology or the genetic architecture of quantitative traits. Pathogenesis-related (PR) genes, phenylalanine ammonia lyase (PAL), proteinase inhibitors, defense signaling genes and genes with sequence similarity with *R* genes and protein kinases have been reported playing role in quantitative resistance (Kou and Wang, 2010; Gebhardt and Valkonen, 2001). According to Kou and Wang (2010), host pattern recognition receptor (HPRR)-type genes and a number of defense-responsive genes have suggested as contributing to quantitative resistance and some of them mediate broad spectrum resistance. Quantitative resistance may also be contributed by *R* genes that have residual effects against virulent pathogens or defeated *R* genes (Gebhardt, 2013).

Quantitative resistance is very complex type of resistance. Quantitative traits do not follow patterns of Mendelian inheritance. Instead, their phenotypes typically vary along a continuous gradient depicted by a bell curve. Quantitative traits are controlled by multiple genes or QTLs. Plants with the same phenotype can carry different alleles at each of many genes or QTLs or plants with identical QTL genotypes can show different phenotypes when grown under different environments. Furthermore, the effect of one QTL can depend on the allelic constitution of the plant at other QTL. For these reasons, one cannot infer the genotype

from the phenotype, and one must construct specialized genetic stocks and grow them in precisely controlled environments (Clair, 2010).

However, the genetic variation of most quantitative traits likely involves a small number of major genes or QTLs, a larger number of loci with moderate effects, and a very large number of loci with minor effects. The effects of the major QTLs can be studied via segregation analysis as well as evolutionary and selection history. The numerous genes with small effects, however, cannot be investigated individually (Clair, 2010).

A number of resistance QTLs, which confer quantitative/field resistance against *P. infestans* have been identified in potato (Nowicki et al., 2012; Gebhardt and Valkonen, 2001). Out of the various identified QTLs, one major resistance QTL, located on potato chromosome V co-localizing with *R1*, is an important one (Gebhardt and Valkonen, 2001). Unfortunately, this QTL is associated with late plant maturity, reducing its potential value (Gebhardt and Valkonen, 2001). Another major QTL for *P. infestans* resistance not compromised by late maturity is a gene in chromosome XI, encoding allene oxide synthase 2 (*StAOS2*) (Pajerowska-Mukhtar et al., 2009; 2008).

1.4. Association mapping

The conventional linkage mapping relies on linkage disequilibrium (LD) generated by crossing two parental lines. Recombination in meiosis that leads to double haploids (DHs), F2 or recombinant inbred lines (RILs) reduces the association between a given QTL and markers distant from it. Unfortunately, derivation of these populations requires relatively few meiosis, such that even markers that are far from the QTL (e.g. 10 cM) remain strongly associated with it. Such long-distance association hampers precise localization of the QTL (Xu, 2006). These problems are worsening in polyploid crops, like tetraploid potato, with high heterozygous and tetrasomic segregation. Furthermore, self incompatibility in diploid potatoes complicates more the use of conventional linkage mapping (Gebhardt, 2007). Association mapping via LD (non-random association of alleles at different loci) offers promise in this area (Oraguzie and Phillip, 2007).

Association mapping examines the joint inheritance of functional polymorphisms and physically linked molecular markers in a set of genotypes with unknown ancestry (Gebhardt

et al., 2014). It involves searching for genotype-phenotype correlations among unrelated individuals following two main approaches. The first approach is the candidate gene approach that uses pre-specified genes selected based on a priori knowledge of their function, such as their involvement in pathogenesis, defense and defense related pathways (Gebhardt et al., 2007). The second approach is genome wide association mapping (GWA) that scans the whole genome for polymorphism without prior information about the genomic regions governing the variation of the target trait (Soto-Cerda and Cloutier, 2012; Stich et al., 2013). The scanned DNA markers analyzed statistically for co-segregation with phenotypic traits to find markers linked with a quantitative trait locus (QTL) contributing to the trait (Stich et al., 2008; Yu et al., 2006).

In contrast to the conventional linkage mapping, association mapping exploits the existing LD in a population that undergone historical recombination events in the past (Soto-Cerda and Cloutier, 2012). This makes association mapping more applicable to a much wider germplasm base as compared to the conventional linkage mapping.

However, the use of a collection of genotypes has its own consequences for the analysis. In a collection of genotypes, significant marker trait association might not be due to physical linkage between markers and QTLs. In such type of population, population structure and familial relatedness potentially create LD between unlinked loci (Pritchard 2001; Stich et al. 2005; Yu et al. 2006), leading to false positives or spurious associations (Stich et al., 2008; Yu et al., 2006).

Different methods have been suggested to avoid such non-functional, spurious associations between a phenotype and an unlinked gene (Stich et al., 2008; Malosetti et al., 2007; Yu et al., 2006). Pritchard et al. (2000) proposed a Bayesian method of testing association that depends on the inferred ancestries of individuals. The computer program STRUCTURE uses computationally intensive methods to partition individuals into populations given molecular marker data. Another method is based on principal component analysis (PCA) across a large number of biallelic control markers with a genome wide distribution (Price et al., 2006). The PCA summarizes the variation observed across all markers into a smaller number of underlying component variables. These can be interpreted as relating to separate, unobserved, subpopulations from which the individuals in the dataset (or their ancestors) originated (Xu,

2006). Yu et al. (2006) proposed a mixed-model approach that combines population structure and family relatedness to correct false positives. In this approach, both a marker-based relationship matrix (K) and population structure (Q) are included in a mixed model for association analysis of a single trait in a single environment. Stich et al. (2008) proposed a two-step-approach, in which adjusted entry means of the phenotypic data analysis is followed by association analysis. Malosetti et al. (2007) proposed association mapping approach based on the mixed models with attention to the environmental variation.

1.5. Marker assisted selection

Marker assisted selection (MAS) refers to the use of DNA markers that are tightly-linked to target gene or quantitative trait loci (QTLs) governing the trait as a substitute for or to assist phenotypic screening. By determining the allele of a DNA marker, plants that possess particular genes or (QTLs) may be identified based on their genotype rather than their phenotype. DNA-based genetic markers essentially detect point mutations, insertions, deletions or inversions in allelic DNA fragments, which differentiate the individuals of the same species (Gebhardt, 2007). The first molecular marker was restriction fragment length polymorphisms (RFLP) followed by PCR based markers such as amplified fragment length polymorphisms (AFLP), simple sequence repeats (SSR) and single nucleotide polymorphisms (SNP) markers (Gebhardt, 2007).

MAS may greatly increase the efficiency and effectiveness for breeding compared to conventional breeding. The main advantages of MAS compared to conventional phenotypic selection are; simpler compared to phenotypic screening, selection may be carried out at seedling stage and single plants may be selected with high reliability. In addition, it saves time and labour by substituting difficult or time-consuming field trials that need to be conducted at particular times of year or at specific locations, or are technically complicated (Collard and Mackill, 2008). Furthermore, selection based on DNA markers may be more reliable due to the influence of environmental factors on field trials. In some cases, using DNA markers may be more cost effective than the screening for the target trait. Another benefit from using MAS is that the total number of lines that need to be tested may be reduced. Since many lines can be discarded after MAS at an early generation, this permits a more effective breeding design (Collard and Mackill, 2008).

A diagnostic marker should predict the phenotype of interest in a breeding population. Ideally, a diagnostic marker originates directly from the allelic variant of the gene that is causal for the phenotypic effect. The smaller the physical distance between a marker locus and the factor causal for the trait of interest, the larger the LD is between marker and trait alleles, which increase the diagnostic value of the marker (Gebhardt, 2013).

MAS has been successfully implemented for *P. infestans* resistance in potato, mainly for a single *R*-gene based resistance (Tiwari et al., 2013). For quantitative resistance, a prominent and reproducible QTL on potato chromosome V was reported (Gebhardt and Valkonen, 2001). However, the same resistance QTL overlaps with a major QTL for plant maturity. StAOS2 is a marker for a major QTL in chromosome XI (Pajerowska-Mukhtar et al., 2009) conferring maturity-corrected resistance (MCR) to *P. infestans* and has been successfully implemented in MAS (Draffehn et al., 2013).

1.6. The objectives of the study

The objective of this study was the identification of genes (QTLs) and diagnostic SNP markers associated with maturity corrected resistance (MCR), relative area under the disease progress curve (rAUDPC) and plant maturity (PM). Under the long day conditions in the growing seasons in central and northern Europe, quantitative resistance to *P. infestans* is correlated with late plant maturity. Late plant maturity is an undesirable characteristic due to its incompatibility with agricultural practices. Therefore, the main objective of this study was the identification of genes for quantitative resistance to *P. infestans* not compromised by late plant maturity (MCR) and the discovery of diagnostic SNP markers. We also aimed to analyze population structure and genome wide linkage disequilibrium (LD) and LD decay in a panel of 184 tetraploid potato clones (CONQUEST2 population).

2. Selection and validation of novel candidate genes for association with quantitative resistance to *Phytophthora infestans* in potato

2.1. Introduction

Late blight of potato (*Solanum tuberosum* L.) caused by the oomycete pathogen *Phytophthora infestans* (Mont.) de Bary, is one of the most important bottlenecks of potato and tomato (*Solanum lycopersicum*) production worldwide (Gebhardt, 2013; Guenthner et al., 2001). Since causing the Irish famine in the 1840s, *P. infestans* has continued to destroy potato throughout the world and remains the most destructive disease of the third largest food crop (Yoshida et al., 2013). *P. infestans* attacks foliage, stems and tubers of the plant causing loss of crop quantity and quality. If not controlled, late blight epidemics can completely destroy the crop yield.

Although the use of chemical fungicides is an option to control *P. infestans*, it is environmentally unsafe and costly (Guenthner et al., 2001). In addition, continuous use of fungicides causes the emergence of new resistant strains. Hence, improvement of the genetic resistance has been a priority option to manage late blight in potato and tomato. A number of *R*-genes conferring resistance to specific races of *P. infestans* have been identified and introgressed into advanced cultivars (Gebhardt, 2013; Foster et al., 2009; Ballvora et al., 2002). However, this type of resistance was quickly overcame by new races of the pathogen having virulence alleles compatible with the resistance alleles in the plant. Instead, most previous results showed that *P. infestans* could be better managed by using cultivars with high levels of field resistance or quantitative resistance, which are more durable and mostly race non-specific (Rietman et al., 2012; Fry, 2008). Quantitative resistance is controlled by many genes and quantitatively more resistant cultivars have certain level of compatible type interaction allowing some infection and growth of the pathogen, which reduces the selection pressure on the pathogen (Gebhardt, 2013). In addition, the pathogen has to undergo many mutations to overcome the several different genes involved in quantitative resistance (Draffehn et al., 2013; Gebhardt, 2013). Breeding for quantitative resistance is however, complex and challenging and requires multiple year and location trials. Quantitative

resistance is controlled by many minor genes and also by the environment. In addition, it is more complicated by epistasis, which is interaction between quantitative trait loci (QTLs) (Jannink and Jansen, 2001). Furthermore, it is hampered in potato by a correlation between resistance and late plant maturity. In potato, genotypes with high levels of field resistance to late blight often are late maturing, which is an undesirable agronomic character (Visker et al., 2003). The reason for the correlation is not known, but it is either due to pleiotropic effects (i.e. both traits are controlled by the same gene) or due to linkage between the genes controlling the traits (Bormann et al., 2004). Quantitative resistance breeding can be facilitated by using diagnostic DNA markers, which are directly at or genetically linked to the genes or QTLs playing a role in resistance. A marker at the QTL for resistance is preferable as diagnostic marker compared to a linked marker. A marker physically tightly linked to the causal gene can also be considered diagnostic (Gebhardt, 2013). Hence, the identification of QTLs and diagnostic DNA markers for *P. infestans* resistance that are not compromised by late maturity was the aim of this project.

In this project, two approaches were followed to identify QTLs and diagnostic DNA markers associated or linked with late blight resistance. The first was the candidate gene approach that uses pre-specified genes selected based on a priori knowledge of their function, such as their involvement in pathogenesis, defense and defense related pathways. The second approach was genome wide association mapping that scans the whole genome for polymorphism without prior information about the genomic regions governing the variation of the target trait. This approach is discussed in chapter III.

The candidate gene approach takes advantages of the knowledge of disease resistance in plants, genes and proteins that are likely to be involved in resistance. These make the approach powerful and precise for QTL detection. However, it is biased as it hypothesizes about the identity and function of the genes underlying the trait. Furthermore, the hypothesis is limited to the knowledge of the hypothesizer, which increases the chance of missing some genes that might have role in resistance (Gebhardt et al., 2007). The limitation of the candidate gene approach can be complemented through comparative transcript profiling that leads to the identification of candidates differentially regulated between quantitatively more

resistance and susceptible genotype pools (Draffehn et al., 2013; Gyetvai et al., 2012; Kreuze et al., 2010; Ros et al., 2005).

So far, some QTLs for key traits were identified in potato through the candidate gene approach. These QTLs include resistance to late blight (Pajerowska-Mukhtar et al., 2009; Gebhardt and Valkonen, 2001), resistance to *Globodera pallida*, chip color and tuber starch content (Gebhardt et al., 2007), tuber flesh color and cooking type (Kloosterman et al., 2010), and cold sweetening (Menendez et al., 2002). Among QTLs identified in potato through candidate gene approach, the best example is a major QTL in chromosome XI for late blight resistance (Pajerowska-Mukhtar et al., 2008). The QTL encodes allene oxide synthase 2 (StAOS2) and explained about 30 % of the genetic variation (Pajerowska-Mukhtar et al., 2009).

Many genes alter their level of expression in response to environmental stress conditions like pathogen attack. By profiling the transcripts in RNA pools from pathogen infected genotypes contrasting for disease phenotype or marker data, differentially expressed genes associated with resistance can be identified (Draffehn et al., 2013; Gyetvai et al., 2012; Kloosterman et al., 2010). The differential gene expression can either derive from a polymorphism located physically near the QTL or indirectly from a distant location on the genome (Kloosterman et al., 2010). Differentially expressed genes in potato due to *P. infestans* infection have been reported (Draffehn et al., 2013; Gyetvai et al., 2012; Thümmler and Wenzel, 2005).

In this study, novel candidate genes, which were differentially expressed in quantitative resistant versus susceptible potato genotype pools, were selected from transcriptome data generated by SuperSAGE (serial analysis of gene expression) analysis. The SuperSAGE analysis was performed on nine samples comprising three genotype groups with different resistance levels and three infection time points (Draffehn et al., 2013). SuperSAGE is a digital gene expression profiling method and permits the simultaneous detection and quantification of the transcriptome both from the host and pathogen in pathogen infected samples (Matsumura et al., 2003). SuperSAGE generates 26 base pairs (bp) tag from each transcript in the sample, and the tag frequency is proportional to the amount of the corresponding transcript. In addition, the 26 bp tags can be used as gene specific primers (Draffehn et al., 2013; Matsumura et al., 2003). The selected candidates were subjected to

validation for differential expression by qRT-PCR and allele specific pyrosequencing. Candidates showing reproducible transcriptional regulation in three independent infection experiments as well as with reference to SuperSAGE expression data were tested for association with maturity corrected resistance (MCR) to late blight in a population of tetraploid breeding clones.

2.2. Materials and Methods

2.2.1. Materials

2.2.1.1. *Potato genotypes used for linkage mapping*

A total of 111 half sib F1 segregating plants derived from the crosses Phy14 x Phy20 and Phy16 x Phy20 were used for linkage analysis. The genotypes were evaluated for late blight resistance in 2010 under field conditions and phenotypic data for AUDPC (area under disease progress curve), MCR (maturity corrected resistance) and PM (plant maturity) were provided by the breeding company Pflanzenzucht GmbH & Co. Ka (SAKA). DNA was extracted from leaf samples of each genotype as described below (2.2.2.1.1).

2.2.1.2. *Potato genotypes used for association mapping*

For association mapping, the CONQUEST2 population was used, which consisted of 184 tetraploid potato clones (Pajerowska-Mukhtar et al., 2009). The CONQUEST2 population was originally from two breeding companies, Böhm-Nordkartoffel Agrarproduktion (BNA) and SAKA. The population was phenotyped in the field for rAUDPC, MCR and PM and adjusted entry means were calculated as described previously by Pajerowska-Mukhtar et al. (2009). DNA of the CONQUEST2 population maintained in Max Planck potato research group (Cologne, Germany) was used for candidate genes amplicon sequencing analysis.

2.2.1.3. *Potato genotypes used for candidate genes expression analysis*

Twenty-four tetraploid potato genotypes were selected based on their MCR to complex isolates of *P. infestans* and plant maturity date. Twelve genotypes were quantitatively more resistant while the other twelve were quantitatively more susceptible. The genotypes were half-sib offspring from two crosses with two different resistant parents (Phy14 and Phy16) and one common susceptible parent (Phy20). The genotypes, phenotype and their pedigree are shown in Table 2.1.

Table 2.1. Tetraploid potato genotypes used for differential candidate genes expression analysis. Their pedigree, phenotype for plant maturity (PM), relative area under the disease progress curve (rAUDPC) and maturity corrected resistance (MCR) and genotype at the *StAOS2* locus.

Genotype	PM	rAUDPC	MCR	Cross	StAOS2-snp691	StAOS2-snp692
SL001	7	-4.28	-84.1	PapPhy20 x PapPhy14	AAGG	CGGG
SL029	1	125.52	-117.7	PapPhy20 x PapPhy14	AAAA	CCCC
SL049	5	83.16	-51.1	PapPhy20 x PapPhy14	AAAA	CCCC
SL262	5	79	-55.3	PapPhy20 x PapPhy14	AAAA	CCCC
SL101	5	16.39	-117.9	PapPhy20 x PapPhy14	AAAA	CCCC
SL311	5	-1.77	-136	PapPhy20 x PapPhy16	AAAA	CCCC
SL312	5	55.23	-79	PapPhy20 x PapPhy16	AAAA	CCCC
SL318	5.7	22.91	-92.3	PapPhy20 x PapPhy16	AGGG	CCGG
SL331	7	24.41	-55.4	PapPhy20 x PapPhy16	AGGG	CGGG
SL411	6	16.7	-90.3	PapPhy20 x PapPhy16	AAGG	CGGG
SL412	4	22.59	-138.9	PapPhy20 x PapPhy16	AAAA	CCCC
SL427	5.3	25.82	-100.3	PapPhy20 x PapPhy16	AGGG	CGGG
SL028	5.3	198.02	71.9	PapPhy20 x PapPhy14	AAAA	CCCC
SL078	3	252.34	63.6	PapPhy20 x PapPhy14	AAGG	CGGG
SL116	4	223.02	61.5	PapPhy20 x PapPhy14	AAAA	CCCC
SL125	1	287.16	43.9	PapPhy20 x PapPhy14	AAAA	CCCC
SL173	5	220.39	86.1	PapPhy20 x PapPhy14	AAAA	CCCC
SL191	5.3	207.05	81	PapPhy20 x PapPhy14	AAAA	CCCC
SL192	5.7	212.14	96.9	PapPhy20 x PapPhy14	AAGG	CGGG
SL207	5.7	189.32	74.1	PapPhy20 x PapPhy14	AAAA	CCCC
SL322	4	226.25	64.7	PapPhy20 x PapPhy16	AAGG	CGGG
SL330	5	190.52	56.3	PapPhy20 x PapPhy16	AAGG	CGGG
SL368	5	218.93	84.7	PapPhy20 x PapPhy16	AAGG	CGGG
SL410	5	208.98	74.7	PapPhy20 x PapPhy16	AAAA	CCCC

2.2.2. Methods

2.2.2.1. *Molecular techniques*

2.2.2.1.1. DNA extraction

Potato leaves were harvested and immediately frozen in liquid nitrogen. Frozen samples were freeze-dried (Christ Gefriertrocknung GmbH, Osterode) overnight and about 100 mg freeze-dried leaf samples were transferred to 1.5 ml Eppendorf tube containing one stainless steel

bead. Samples were grinded to powder in a tissue mile (Retsch GmbH, Haan) for 1 min at 30hz (2x). Genomic DNA was extracted using the DNeasy® Plant mini Kit (Qiagen) according to the manufacturer's instructions. The concentration of the DNA was measured using the Nanodrop ND-1000 spectrophotometer (Peclab, Erlangen) and the integrity of the DNA was checked by running 1µl DNA mixed with 2µl DNA loading dye on a 0.8 % agarose gel. The quantity of the genomic DNA was determined by comparing the band intensity on the gel with a reference λ50 (50 ng/µl) marker with concentrations of 100 ng, 200 ng, 300 ng and 400 ng load on the gel. The concentration of the DNA was adjusted to 50 ng/µl with double distilled water (ddH_2O) and stored at -20 °C until used.

2.2.2.1.2. Total RNA extraction from *P. infestans* infected potato leaves

About 100 mg frozen *P. infestans* infected potato leaves was transferred to collection microtube containing one stainless steel bead inside. The leaf transferring was carried out in a box contain liquid nitrogen. The leaf in collection microtube was grinded to fine powder using the tissue mile (Retsch GmbH, Haan). From the grinded leaf powder, total RNA was extracted simultaneously both from potato and *P. infestans* using TRIzol RNA isolation kit (Ambion). The total RNA was purified from DNA contamination using the DNA-free™ Kit (Ambion). The RNA concentration was measured using the Nanodrop ND-1000 spectrophotometer (Peclab, Erlangen) and the quality of the RNA was determined based on the ratio of absorbance A260nm/A280nm. In addition, the integrity of the RNA was checked by running 2µl total RNA mixed with 2µl DNA loading dye on a 1.5% agarose gel. Total RNA was stored at -80 °C until used.

2.2.2.1.3. Synthesis of first strand cDNA

First strand cDNA was synthesized following the protocol of Maxima H Minus First Strand cDNA Synthesis Kit (Thermo Scientific). About 2 µg total RNA was mixed with 0.25 µl oligo $(dT)^{18}$ primer and 1µl of 10 mM dNTP Mix and reverse transcribed into cDNA by Maxima H Minus Enzyme Mix in a 20 µl total reaction volume. The synthesized cDNA was stored at -20 °C until used in further analysis. First strand cDNA was synthesized from RNA of each genotype and also for the pooled RNA. The cDNA from each genotype was used to quantify *Phytophthora* growth in qRT-PCR while the cDNA from the pooled RNA was used in candidate genes differential expression analysis in qRT-PCR and pyrosequencing.

2.2.2.1.4. Polymerase chain reaction (PCR)

Standard PCR reaction was carried out in 25 µl total volume containing 2.5 µ of 10x standard reaction buffer (10 mM Tris-HCL pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$, 0.1% Triton X-100l), 2 mM dNTP Mix (2.5 µl), 25 mM $MgCl_2$ (2.5 µl), 10 µM each forward and reverse primers (0.5 µ1, each), 0.2 U Taq DNA Polymerase, ddH_2O and 50 ng of template DNA. For standard PCR, either homemade Taq-Polymerase or the commercial Taq DNA Polymerase (Ampliqon) was used. The PCR reaction was run with 2 min initial denaturation at 94 °C, followed by 40 to 50 cycles with 30 sec denaturation at 94 °C, 30 sec annealing at temperature specific to each primer pair (Table 2.2 and 2.3), and extension (1 min/Kb) at 72 °C. At the end of the cycles, the reaction was run for 10 min at 72 °C. Reactions were generally performed in a Labcycler (SensoQuest GmbH, Göttingen).

2.2.2.1.5. Quantitative real time PCR (qRT-PCR)

The qRT-PCR was performed on a Mastercycler^{ep} realplex (Eppendorf, Hamburg) or on CFX384 Touch Real-Time PCR (Bio-Rad) using Power SYBR® Green PCR Master Mix (Applied Biosytems) with 2.5 µl first strand cDNA. PCR cycling conditions were as follows: Initial denaturation at 95 °C for 10 min followed by 40 cycles of 95 °C for 15s, annealing at temperature specific to each primer pair (Table 2.2) for 30s and extension at 72 °C for 45s. At the end of the PCR, reactions were tested for undesired primer dimer formation by melting curve analysis. For standard curve analysis, a serial dilution of the cDNA was prepared as 1:10, 1:100 and 1:1000.

For candidates expression analysis, first strand cDNA was diluted 1:5, and 2.5 µL of the dilution was used as template for quantitative amplification. The primer pair used for each tested candidate is showed in Table 2.2. Expression levels of the candidate genes were normalized against transcripts of the SAND gene (Draffehn et al., 2013; Exposito- Rodriguez et al., 2008; Czechowski et al., 2005).

For growth quantification of *P. infestans* on each genotype, 2.5 µL of the undiluted first strand cDNA was used. Growth was estimated by quantifying the transcript level of 60S ribosomal gene L23a (PITG_02694) (Draffehn et al, 2013), using the primer combination 5_-CGCCTGACCGCTGACTACGA-3_ and 5_-GCGAGAGTGCGATGACGATG-3_.

Table 2.2. Primers used in candidate's expression analysis in qRT-PCR.

Annotation	Forward perimer 5' - 3'	Reverse primer 5' - 3'	Amplicon size	Ta
Arabinogalactan protein (Agp)	CTACCCTATTCGTTCCTACCGT	CCTCAAACTCCTCATCTCAATG	202	61
Eukaryotic translation initiation factor 4e type (EIF)	GATGCTTCCTTGCGCGAC	CATACACATCAGAAGAACTACAGGC	245	60
Clathrin coat assembly protein AP17 (CCAP)	CTGCACTTCTCTTCCGGCTG	GCCTCTGGAGTTGACAAAAGA	212	53
Avr9/Cf-9 rapidly elicited protein 20 (AC9)	AGCCCTTTACACAGCTCGTC	AGTTGAATTAATGGAGAAACAGCTT	157	58
Subtilisin-like protease (SLP)	GGAAGTGCCAAAGAATACCA	TCCCCAGGGAAGTATAATG	217	55
Pectin methyl esterase (PME)	CGTGTTGACGAATTTCAGTGAT	GGCCAAGTATAAGCATAAATTCTC	110	60
Squalene monooxygenase (SMO)	GGCAAAACTAAGTGAACCG	AGTTTCAAGAGTTTAGTCCAAC	242	53
Photosystem II core complex proteins (PSP)	CCTGCTTTGCTTTGGGTAG	AATCATATGGCCAAACGC	250	57
Conserved gene of unknown function (Cgu)	GTTATCCACCTCCTGGC	AACTTGGCTGCAGGATG	256	54
Polyubiquitin (Pubq)	CTTCTGAATGTCCTGTGTCTG	GCAGCAAAAACAAAAGAAC	140	54
Kiwellin (Kiw)	CTTGGGACTATTGCGATGGT	AGTTGAGCGGGCATTGAT	217	58
Salicylic acid-binding protein (SABP)	GGTACTGATCATACGGCA	TCTCATCACACATTCTTCAA	189	52
Chloroplast protease (Chp)	CAAAGCAAGGGAAACAACTCA	GACTGGAGATGAATTCCGTGCTAT	289	60
Heat shock cognate 70 kDa protein (HSP70)	TCCAATTATTGCCAAGATGTACCA	TCAAGCCAAAATACGCAAAAGT	243	60
SAND	CTGCTTGGAGGAACAGACG	GCAAACAGGACCCCTGAATC	163	58
60S ribosomal protein L23a (RL23a)	CGCCTGACCGCTGACTACGA	GCGAGAGTGCGATGACGATG	164	64

Ta = annealing temperature; acronym of each candidate is shown in parenthesis

2.2.2.1.6. Pyrosequencing

Primers for pyrosequencing were designed using Pyrosequencing Assay Design Software (Biotage). Pyrosequencing requires three primers, a pair of forward and reverse primers and a third sequencing primer (Ronaghi et al., 1996). The forward and reverse primers were used to amplify 100 to 300 bp DNA fragment containing the SNP of interest by standard PCR and a third sequencing primer was used to sequence a short region around the SNP of interest by pyrosequencing. One of the PCR primers (forward or reverse) was biotin labeled depending on the direction of the sequencing primer. When the sequencing primer was forward, the reverse primer was biotin labeled at the 5' end. The forward primer was biotin labeled at the 5' end when a reverse sequencing primer was used. To amplify the genomic region containing the SNP of interest, the standard PCR was used (2.2.2.1.4), except that the PCR was run for 50 cycles to avoid any primers leftover that might interfere in pyrosequencing reaction. Only a fragment with a single band corresponded to the size of the amplicon and with good intensity, free from primer dimers was used for further pyrosequencing analysis.

To capture the biotin labeled PCR product by sepharose beads, 15 µl of the PCR product was mixed with 40 µl of binding buffer (Qiagen), 5 µl of sepharose beads solution and 20 µl

water (double distilled and sterilized water). Then the beads containing biotin labeled PCR product was washed with 70% ethanol (for about 3s), denatured into single strand template by 0.2 M NaOH (for about 5s) and rinsed with 10 mM Tris-HCl (pH 7.6) (for about 8s). Finally, the cleaned beads with biotin labeled single stranded PCR product was transferred to pyrosequencing plate containing 1µl 0.25 µM sequencing primer and 40 µl annealing buffer (Qiagen) in each well. In at least one well, water instead of PCR product was used as a negative control. The plate was heated to 80 °C on a heating block for 2 min, then cooled down to room temperature and placed into the pyrosequencing machine (Pyrosequencer PSQ96^{TM} MA). The pyrosequencing was performed using pyromark gold Q96 reagents kit (Qiagen) and a PSQTM 96 MA pyrosequencing instrument (Biotage AB) according to manufacturers' protocols.

Candidates represented by two allelic SuperSAGE tags that differed by one or two bp and showed contrasting differential expression were analyzed for allele specific expression by pyrosequencing using cDNA as template. The primer pair used for each tested candidate is shown in Table 2.3.

Table 2.3. Primers used in candidate's expression analysis by pyrosequencing

Annotation	Primer 5' - 3'	Amplicon size	Ta
Delta (7)-sterol-C5 (6)-desaturase (DSD)	F: CATACATGGTAAGGTGTGGC	215	57
	R:[Btn]CAACGATGGAAACACGAGAC		
	S:ATGTAATGTGCATGATTTG		
Magnesium-protoporphyrin-IX-monomethyl-ester cyclase (Desaturase) (MPP)	F: GGCTTCTGAGTTATTGGCTG	221	51
	R:[Btn]CTGATAATACAAGGGAGTGTCC		
	S:CATGTTCATTGTTGTAAGTT		
Hydroxypyruvate reductase (HPR)	F: [Btn]TGCCCGAGCATTGTGAATTCAA	287	60
	R:CCAGATTGAGAGTGTGGGTAGCT		
	S: ATATGAGTAAAACTTGTGTC		
UPA18	F: CTCCCCCTCCTCCACGAC	203	60
	R:[Btn]CAATCCTTCCCCCAATGTC		
	S:GAAGCGTAGTGGGAAAAT		
Biotin carboxylase carrier protein (BCCP)	F: [Btn]GTTGATGAACGAAATAGAGGCTG	125	61
	R:TTACCCGAACGGTTCTATGGTTT		
	S:TTCAGCAACAACCTCA		
MADS transcription factor (MADS)	F: CCTCGAAGAGCTGAAACTGC	194	59
	R:[Btn]AGGGAAGTTGCCCTTACTGA		
	S:CAAAGATGCATTTGAGG		

Table 2.3. Continued

Table 2.3.

Asparagine synthetase (AspS)	F: [Btn]GGAACATTAGTGGTGCTCAAGAA	237	60
	R:GGAAGGATCAGCACGTCTTTTAGA		
	S:GAATAAAAGATCAAAACCAT		
Receptor-like protein kinase RLPK)	F: GTACCCTGGAGAATCCTAAG	188	52
	R:[Btn]GGACAGTCTCTCATATTGGAG		
	S:AAAAGTTTGTCATGAAACTA		

F = forward primer; R = reverse primer; S = sequencing primer; Ta = annealing temperature; [Btn] = the position of biotin; acronym of each candidate is shown in parenthesis

2.2.2.1.7. Amplicon sequencing and SNP markers scoring

For each candidate gene, specific primers were designed (Table 2.4) using primer-BLAST in NCBI at http://www.ncbi.nlm.nih.gov/tools/primer-blast/ mostly from exons, but introns were also used in few cases. About 400 to 1300 bp DNA fragments were amplified by standard PCR (2.2.2.1.4). An aliquot of the PCR product was loaded on a 1.5 % agarose gel and visualized after electrophoresis using ethidium bromide staining for checking the presence of a single band corresponded to the size of the amplicon. The amplicon size was determined by comparing with 1kb DNA ladder. Before sequencing, PCR products were treated with ExoSAP-IT® (USB Affymetrix, Santa Clara California, USA) following the suplier's instructions and diluted to 15 to 20 ng/µl, according the requirements of the MPIZ DNA sequencing facility (MPIZ Genome Center). Purified PCR products were sequenced by Sanger sequencing at the Max Planck Genome Centre (Cologne, Germany).

SNPs were detected by aligning sequences using Lasergene software Seqman and SNP markers were scored in dosage dependent manner for the five possible genotype groups (nulliplex = AAAA, simplex = AAAB, duplex = AABB and triplex = ABBB and quadraplex = BBBB) using DAx software. The software determines the three heterozygous groups (AAAB, AABB and ABBB) based on the relative peak height of each nucleotide in the trace file.

Table 2.4. Primers used in amplicon sequencing

Annotation (Acronym)	Primer name	Forward perimer 5' - 3'	Reverse primer 5' - 3'	Amplicon size	Ta
Hydroxypyruvate reductase (HPR)	HPR	TGCAGCAAGAAGAATTGTGG	ATAACGGGAGAATGGGATCAA	1091	59
	DSD1	TAGCCCCTCTTCTTTTTCC	CATTTGGTCCATCCGTTTTC	1040	57
Delta (7)-sterol-C5 (6)-desaturase (DSD)	DSD2	AGCTAGCTAATTCATGATGTCCAGT	CCACCAACGATGGAAACACG	538	61
Pectin methyl esterase (PMR)	PMR	GTAACTAATTTCATGCAGCCG	AGGGAAACTCTTCTGCACCG	410	60
Receptor-like protein kinase (RLPK)	RLPK	CGATTCACATTTCGTACCATGCT	CGAAGATGGATTGGAATCACTAG	1123	61
Clathrin coat assembly protein AP17 (CCAP)	CCAP	TTCTCTGGCACCTCAGTTCTG	TTGAATCCCCCACCACTCC	886	61
	SMO1	ACTAAGTGAACCGCAAGCCA	TATCTTTGGTGTTGGCCGCT	581	61
Squalene monooxygenase (SMO)	SMO2	TGCAAGGAAGGATTTCAAGC	GCACATCGGCTTTGATTTTT	117	59
Biotin carboxylase carrier protein (BCCP)	BCCP	TGGGACGATCTGTTGCCTTC	AAAGTCATGAAATAGAGGTCTCGT	840	59
	AspS1	CCAAAGTCCAGAAGGACACTCTAT	GCAAGCCTGACCGTTCCT	552	60
	AspS2	GTACTCTAGCTTCTAAGCCCC	CGGTTGATGACTGATGTCCCC	830	59
Asparagine synthetase(AspS)	AspS3	CGCCTCAGAACCAAAGTGTCAT	CATCGTGGACCGGATTGGAGT	1053	60
Chloroplast protease (Chp)	Chp	GCTCCACTGAAACCAGGTGT	ATTTACTGCAGTGGGGGCTC	623	60
Avr9/Cf-9 rapidly elicited protein 20 (AC9)	AC9	GAACCACGGGCCGACATATT	ATCGAGATAAGAGCAGAACCGC	1218	62
Magnesium-protoporphyrin-IX-monomethyl-ester cyclase (Desaturase) (MPP)	MPP	GGAAGATGTTTTGTCAAGTGC	ACAAACCGCACAACTGCA	536	58
Serine protease inhibitor (Miraculin) (SPI)	SPI	CTCAAGGATACATTAAGGTAGCAATAC	GTTGCAGTCTGGGCGAGATTTG	587	58
4-coumarate-CoA ligase (4Cl)	4Cl	GATCCTTTTGCCTAATTGCCCTGA	CTAGCGCATCAGCACATGGATGTA	892	65
Lipoxygenase_1 (Lox1)	Lox1	GTGGAAGTGGTTCAGGCAAG	GACGTCGAATGATAACAGGGTT	838	63
	StAOS2-1*	CCTCTTCCTTCTCTTCACCAAC	GAAGAAAGAAGGAAGAACATCAA	485	59
Allene oxide synthase 2 (StAOS2)	StAOS2-2	TTGATGTTCTTCCTTCTTTCTTC	GCCAGCGGATTTTACTTCCGATC	573	60

* (Pajerowska-Mukhtar et al., 2009); Ta = annealing temperature

2.2.2.2. *Field assessment of tetraploid potato genotypes to late blight under natural infestation*

The 24 tetraploid potato genotypes selected for candidate genes expression analysis (Table 2.1) were planted in the field at the demonstration plot (WissenschaftsScheune, Max Planck Institute, Köln). Each genotype was replicated two times (two plants per plot) and the susceptible local genotype, Hansa, was planted around the border as infector row to increase the inoculum. Planting was done on May 2012 and plants were inspected for *Phytophthora* infection up to the end of August, 2012. Late blight infestation on each genotype was scored by percentage using CIP scale (Henfling, 1982).

2.2.2.3. *Multiplication and maintenance of P. infestans*

A mixture of complex *P. infestans* field isolates were used in the study. Using mixtures of *P. infestans* isolates insured that all major genes for late blight resistance potentially present in the plants were overcome. The isolate was propagated on rye agar (30 g/l PDA, 200 g/l rye, 1000 ml H_2O (Duchefa)) medium at 18 ^oC with 16 h light and 8 h dark in a growth chamber (Rubarth Apparate GmbH) and maintained by continuous sub culturing on tissue cultured potato leaves and on detached leaf assay.

A piece of *P. infestans* mycelium grown on a rye agar medium was cut from four to six weeks old cultures and put on apaxial side of tissue cultured potato leaves, which were placed on a plastic Petri-dish containing MS medium (4.4g/l MS salts with vitamins, 20g/l Sucrose, 4.4 g/l gelarite, pH 5.7 – 5.8 (Duchefa)). The Petri-dish was sealed with parafilm and incubated at 18 ^oC with 16 h light and 8 h dark in a growth chamber (Rubarth Apparate GmbH) for seven to ten days.

Pieces of *P. infestans* infected tissue cultured potato leaves were transferred into a new sterilized Petri-dish using sterilized forceps. *P. infestans* sporangia were collected in sterilized beaker by washing the infected leaves with sterilized tap water. The sporangia were counted using a hemocytometer under a microscope and the concentration was adjusted to 40 sporangia/µl. The sporangia solution was incubated at 4 ^oC in the dark for 1-2h to release zoospores.

Potato leaflets from the susceptible genotypes Grata and Granola were collected with sterile blade and placed on top of whatman paper, which was on top of a metal grid in a transparent box. The box was filled with sterile tap water up 2 to 3 cm height to keep the medium humid. Each leaflet was inoculated on apaxial side with 20 µl of sporangial solution at four spots. The lid of the box was partially sealed with parafilm paper to maintain high humidity. The box was placed in a growth chamber (Rubarth Apparate GmbH) at 18 ^{o}C with 16 h light and 8 h dark for four to five days.

2.2.2.4. *Infection of potato genotypes with complex isolates of P. infestans*

Tissue cultured plants from the 24 selected tetraploid potato genotypes (Table 2.1) were propagated in pots in the greenhouse for 5 to 6 weeks and then transferred to a growth chamber (Snijders Scientific B.V., Tilburg, Netherlands), which was set at 80 % humidity, 100 % light, 16 h light and 8 h dark at 22 ^{o}C and 18 ^{o}C day and night temperature respectively. For each genotype, four plants were grown for four different treatments. One plant was used to collect leaf samples before inoculation (T0), the second plant was sprayed with water as negative control. The remaining two plants were inoculated with the complex isolate of *P. infestans*.

P. infestans sporangia multiplied on Grata and Granola leaflets were collected as described in 2.2.2.3. The concentration was adjusted to 20 sporangia/µl and zoospores were released after 1-2h incubation in the dark.

On each plant, the 3rd, 4th and 5th leaves (counting from the top) were marked with rings and each compound leaf was sprayed with sporangia/zoospore solution two to three times by using 20 ml spraying bottle (cat. №10007245, neoLab, Heidelberg, Germany). Each plant was covered with a transparent plastic bag to maintain high humidity. At the second (T2) and third (T3) day post inoculation, the 4th and 5th leaves were collected using a sterile blade, rapped in aluminum foil, immediately frozen in liquid nitrogen and then kept at -80 ^{o}C until needed. The leaf samples were collected in the morning between 9 and 10 am. After collecting the leaf samples, the plants were kept in the chamber for one week to follow up *Phytophthora* growth, to monitor symptom development and to phenotype the 3rd leaf of each genotype.

The infection experiments were repeated five times and the three most successful infection trials were used for further candidate genes expression analysis.

2.2.2.5. *Phenotyping*

Growth of *P. infestans* on the 3rd leaf of each genotype was assessed in percentage using a modified CIP (Centro International de la Papa, international potato center) scale (Henfling, 1982) (Figure 2.1). Late blight infestation was scored on the 3rd, 4th and 5th day post inoculation and the score across time points was converted to area under the disease progress curve (AUDPC) using the following formula (Fry, 1978);

$$AUDPC = \sum_{i}^{n-1}\left[\left(\frac{y_i + y_{i+1}}{2}\right)(t_{i+1} - t_i)\right]$$

with; y_i = late blight infestation (in percent) at the *i*th observation

t_i = time (days) at the *i*th observation

n = total number of observation

a

TABLE 1 KEY FOR ASSESSING FOLIAR BLIGHT

Blight %		Description
5.0	1	Lesion per compound leaf
10.0	2	Lesions per compound leaf
25.0		Nearly every leaflet with blight lesions - plants still retaining their normal form - 75% plot leaf area remaining green
50.0		About half of the leaf area destroyed by blight
75.0		About three-quarters of the leaf area destroyed by blight
95.0		Stems green, only a few leaves remaining
100.0		All leaves dead, stems dead or dying

b

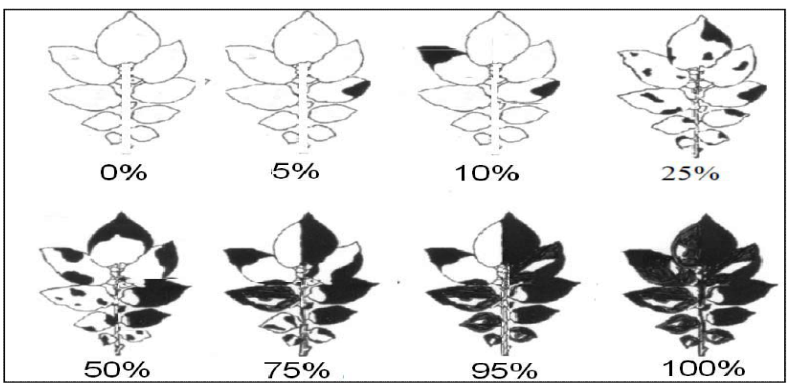


Fig. 2.1. Modified CIP scale used to evaluate resistance to late blight based on phenotype on the 3rd compound potato leaf as shown on the picture (b). The key corresponding to each picture is shown above (a) (Henfling, 1982).

2.2.2.6. *Selection of a gene for P . infestans growth quantification*

P. infestans tags generated by SuperSAGE analysis (Draffehn et al., 2013) were screened to select tags/genes that could be used to monitor the infection and disease progression of *P. infestans*. From the SuperSAGE data, five candidate tags with their transcript levels (hit count) corresponded to the resistant levels of three potato genotype (A1, B2, and A2 in order of increasing resistance) groups were selected (Table 2.5). DNA sequence corresponding to the tag was collected from *P. infestans* data base at http://www.broadinstitute.org/annotation/genome/*phytophthora_infestans*/. The selected tags/genes were further tested by qRT-PCR. Primers were designed for the selected candidate genes taking the 26 bp SuperSAGE tag as a reverse primer whenever possible (Table 2.6). The candidates were first tested on standard PCR on three types of cDNA template from *P. infestans* infected leaf samples collected at three time points of post inoculation; T0 = before inoculation; T1 = one day after inoculation; and T2 = two days after inoculation. qRT-PCR was carried out as described before on the three types of cDNA templates (T0, T1 and T2) from two potato genotypes with different resistance level.

Table 2.5. ***Phytophthora*** **tag sequence, their annotation and normalized hit count (hits per million, hpm) on three potato genotypes with different levels of maturity corrected resistance (MCR) to** ***P. infestans*****.**

Annotation	Tag_sequence	Potato population	MCR	Hit count (hpm)	*P-value
P. infestans cystatin-like cysteine protease inhibitor	CATGTAGTTATCCT TGGAAAAAGAAG	A1	26.2	10.41	4.11E-08
		B2	5.9	9.31	6.04E-06
		A2	-69.8	3.12	7.52E-05
P. infestans T30-4 60S ribosomal protein L23a	CATGGTCATCGCA CTCTCGCTGAAAG	A1	26.2	35.22	4.74E-26
		B2	5.9	34.56	6.28E-20
		A2	-69.8	24.44	9.07E-31
P. infestans T30-4 40S ribosomal protein S9-1	CATGCGAATTTGG GGACTTCAAAAAA	A1	26.2	39.43	1.30E-28
		B2	5.9	38.3	1.15E-20
		A2	-69.8	10.92	8.36E-12
P. infestans T30-4 40S ribosomal protein S3a	CATGCCTAAGCGA GTACTTGTTTCTA	A1	26.2	38.77	3.92E-28
		B2	5.9	32.57	7.98E-19
		A2	-69.8	14.56	9.30E-19
P. infestans T30-4 conserved hypothetical protein	CATGCCCGACCTC GACCCCAACGACA	A1	26.2	7.53	4.96E-06
		B2	5.9	7.31	7.68E-05
		A2	-69.8	4.16	4.09E-06

*The P-value was between before inoculation (T0) and two days post inoculation (T2)

Table 2.6. *Phytophthora* tags, primers used in qRT-PCR.

Annotation	Primer name	Forward (5‘ – 3‘)	Reverse (5‘ – 3‘)	Amplicon size	Ta
Cystatin-like cysteine protease inhibitor	Cysteine	F: CGTTTCAGGTTGCGGCGTCA	R: GAGGCATTTAGTTGGCGGGCGT	152	70
T30-4 60S ribosomal protein L23a	RL23a-1	F:CGCCTGACCGCTGACTACGA	R:GATCCCTCGCAGTACGGCTAGACC	110	66
	RL23a-2	F:CGCCTGACCGCTGACTACGA	T: GCGAGAGTGCGATGACGATG	164	64
T30-4 40S ribosomal protein S9-1	Rs9-1-1	F1:CGAGGAGTAAGCGTAGCGAACAA	R:ACGCCCCGAGAAACAAAGTC	157	62
	Rs9-1-2	F2:TCGCCCGTCGTCGTGCTG	T:GTTGAAGTCCCCAAATTCGCATG	247	64
T30-4 40S ribosomal protein S3a	RS3a-1	F1:GAACAAGGACGAGGACCAAGCC	R:GCCACTTGCGGATCAGCGAG	134	65
	RS3a-2	F2:ACGAGGGCGGTGCTGAGG	T:TAGAAACAAGTACTCGCTTAGGCATG	201	60
T30-4 conserved hypothetical protein	Hypo-1	F1:GCTCTGTGGGCGCCGTGAA	R:ACGGTCCGTCTTGGCCCAGAT	213	68
	Hypo-2	F2:GTTCTGGCCGCGTCTGCTAGCTGA	T:TTGGGGTCGGGTCGGGCTG	234	71

T = reverse primer designed from the tag sequence; Ta = annealing temperature

2.2.2.7. *RNA pooling*

Four different RNA pools, each with quantitatively resistant and susceptible contrasting genotype groups, were prepared by combining equal amounts of RNA from each genotype. The concentration of the pooled RNA was measured before performing the cDNA synthesis. The genotypes are showed in Table 2.7.

Table 2.7. Genotypes used in each contrasting groups of the four RNA pools

Pools_1		Pools_2		Pools_3		Pools_4	
R	S	R	S	R	S	R	S
SL001	SL028	SL318	SL330	SL311	SL191	SL311	SL322
SL029	SL078	SL311	SL368	SL028	SL368	SL412	SL330
SL049	SL116	SL411	SL192	SL411	SL173	SL312	SL368
SL262	SL125	SL412	SL116	SL412	SL322		
SL101	SL173	SL312	SL173	SL312	SL125		
SL311	SL191	SL028	SL078				
SL312	SL192						
SL318	SL207						
SL331	SL322						
SL411	SL330						
SL412	SL368						
SL427	SL410						

2.2.2.8. *Selection of novel candidate genes*

Candidate genes were selected based on differential tag counts analyzed in SuperSAGE (Draffehn et al., 2013), which generated 26 bp tags corresponding to each transcript in the sample and also quantified expression by tag counts. The SuperSAGE analysis was done on three potato genotype groups labeled as A1, B2 and A2 (Draffehn et al., 2013). A1 and B2 were quantitatively more susceptible while A2 was more resistant to complex isolates of *P. infestans*. Details for the genotype groups and SuperSAGE analysis were explained by Draffehn et al. (2013). From 9 SuperSAGE libraries that contained three genotype groups (A1, B2 and A2) each at three time points (T0, T1 and T2), uni-tags differentially expressed ($P < 10^{-4}$) in at least five of the six comparisons (B2 vs A2 and A1 vs A2 each at T0, T1 and T2) were selected. Uni genes corresponding to the tag were collected from DFCI (expressed sequence data bases of potato) at http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=potato using the transcript number (TC_number). The alignment of the uni gene sequence with the tag sequence was checked. The uni gene sequence was blasted against loci in the potato genome reference sequence version 4.03 at http://solanaceae.plantbiology.msu.edu/pgsc. From the few top hits in the potato genome browser, the one aligned with the tag sequence was taken as annotated candidate gene corresponding to the tag. Genomic DNA sequences and related information like map position, intron-exon regions and loci name of the candidate gene were collected from the genome browser. The position of a candidate locus was compared with known QTLs for pathogen resistance in the PoMaMo (Potato Maps and More) data base (Meyer et al., 2005) at https://www.gabipd.org/projects/Pomamo/ for their co-localization. The physical map position of the selected candidates and their co-localization with known QTLs (according to potato PGSC v4.03) was mapped using adobe illustrator CS5 (Version 15.0.2).

2.2.2.9. *Statistical methods*

2.2.2.10. Candidate genes linkage mapping

To analyze marker-trait linkage, a simple analysis of variance (ANOVA) was employed in R statistical software (www.r-project.org) using the *lm ( )* function. The model was;

$$\text{Phenotype} = \text{SNP dosage} + \text{error}$$

With SNP dosage varied from 2 to 5. P_value > 0.01 was used as a threshold to claim significant.

2.2.2.11. Candidate genes association mapping

Candidate genes association was analyzed using TASSEL2.1 (Bradbury et al., 2007) statistical software in mixed linear model (MLM) which takes into account population structure (Q) and kinship (K) in the analysis to avoid spurious association. The MLM model explained as follows;

$$y = X\beta + Zu + e$$

with;

- y = response vector for phenotypic values
- X = known design matrices
- β = vector containing fixed effects regarding population structure (Q)
- Z = known design matrices for markers
- u = unknown vector of random additive genetic effects from multiple background QTL for individuals or lines (K)
- e = unobserved vector of random residual

2.2.2.12. Linkage disequilibrium analysis

Linkage disequilibrium (LD) between two loci was analyzed by testing for independence of two loci with a X^2 (chi-square) test using the statistical software R. LD p-values were corrected for multiple testing using the R package 'qvalue', using the method described by Storey (2003).

2.3. Results

2.3.1. *P. infestans* SuperSAGE tags for *Phytophthora* quantification in qRT-PCR

One of the advantages of SuperSAGE is the simultaneous analysis of gene expression both from host and pathogen in infected samples. In addition, the SuperSAGE tags can be used as gene specific primers. In this study, these advantages were used to identify *P. infestans* genes that can be used for growth quantification of the pathogen in infected samples by qRT-PCR, which helps to monitor infection, disease progression (Draffehn et al., 2013) and also to evaluate potato genotypes for resistance to *P. infestans* during the biotrophic phase, when no disease symptoms are visible.

From the SuperSAGE data (Draffehn et al., 2013), five tags with hit counts corresponding to the resistant levels of the three potato genotype groups (A1, B2 and A2) were selected (Table 2.5). Primers were designed for the selected candidate genes taking the 26 bp SuperSAGE tag as a reverse primer whenever possible. The candidates were first tested by standard PCR on cDNA templates from two genotypes (BL_114, quantitatively more resistant and BL_067, quantitatively more susceptible) each at three time points; before infection (T0), one day after infection (T1) and two days after infection (T2). Out of the selected candidates, 40S ribosomal protein S9-1 (Rs9-1), 60S ribosomal protein L23a (RL23a_2) and conserved hypothetical protein (hypothetical_1) showed amplification (a clear band corresponded to the size of the amplicon) on samples from T1 and T2 but not on T0 (Fig. 2.2). These three candidates were further tested in qRT-PCR. qRT-PCR was carried out as mentioned above on the 3 types of cDNA templates (T0, T1 and T2). In qRT-PCR, 40S ribosomal protein S9-1 (Rs9-1) and 60S ribosomal protein L23a (RL23a_2) expression progressively increased from T0 to two days after inoculation (T2) (Fig. 2.3) corresponding to the resistance level of the genotypes. 60S ribosomal protein L23a (RL23a_2) was used subsequently for further analysis of *P. infestans* infection and growth progression and also to evaluate potato genotypes for resistance to *P. infestans* (Fig. 2.4).

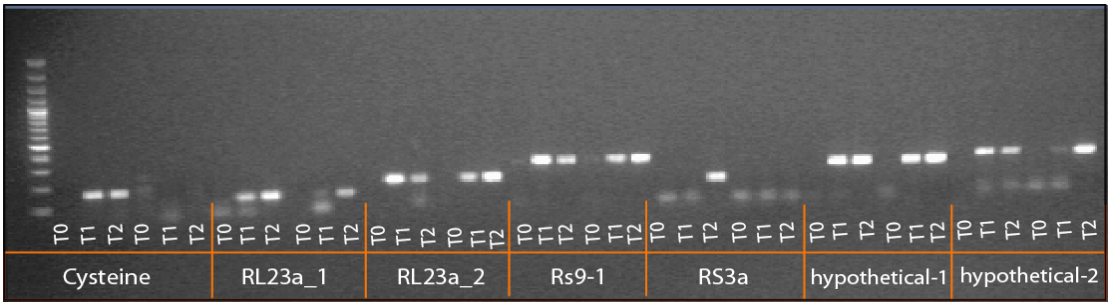


Fig. 2.2. PCR fragments for five *Phytophthora* SuperSAGE tags. Each tag (candidate gene) was tested on two genotypes (BL_114, the first three time points, and BL_067 the next three time points) each at three time points (T0, T1 and T2).

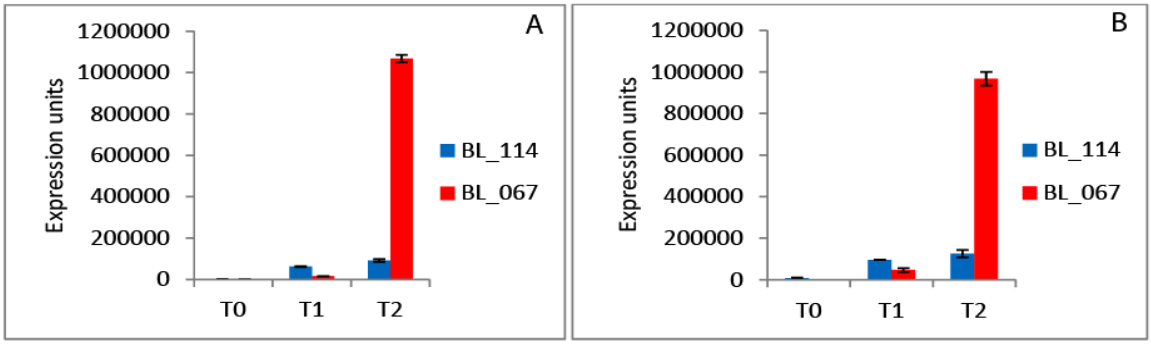


Fig. 2.3. Transcript quantification of the *P. infestans* tags 60S ribosomal protein L23a (RL23a_2) (A) and 40S ribosomal protein S9-1 (Rs9-1) (B) from infected leaves of two tetraploid potato genotypes, BL_114-resistant (blue bar) and BL_067-susceptible (red bar) each at three time points (T0, T1 and T2) of post inoculation.

2.3.2. Selection of tetraploid potato genotypes contrasting for MCR to *P. infestans*

Two tetraploid potato genotype groups contrasting for maturity corrected resistance (MCR) were selected (Fig. 2.5; Fig. 2.7) for the candidate genes differential expression analysis. The genotypes were selected based on previous (2010) field evaluation for MCR to *P. infestans* and plant maturity (PM). Each group consisted of 12 tetraploid potato genotypes with an average MCR value of -93 and 72 in quantitatively more resistant and susceptible groups, respectively. In MCR, negative values indicate resistance while positive values show susceptibility (Pajerowska-Mukhtar et al., 2009). PM ranged from 1 to 7, with an average value of 5. The genotypes were used to construct two pools (Pools_1), each with 12 genotypes, contrasting for MCR for studying differential gene expression. Another two contrasting pools, each with 3 genotypes, were selected based on the StAOS2 marker and the field MCR (Pools_4) (Fig 2.7). The genotypes MCR was further verified under controlled conditions by infecting with complex isolates of *P. infestans* as well as in the field under natural late blight infestation.

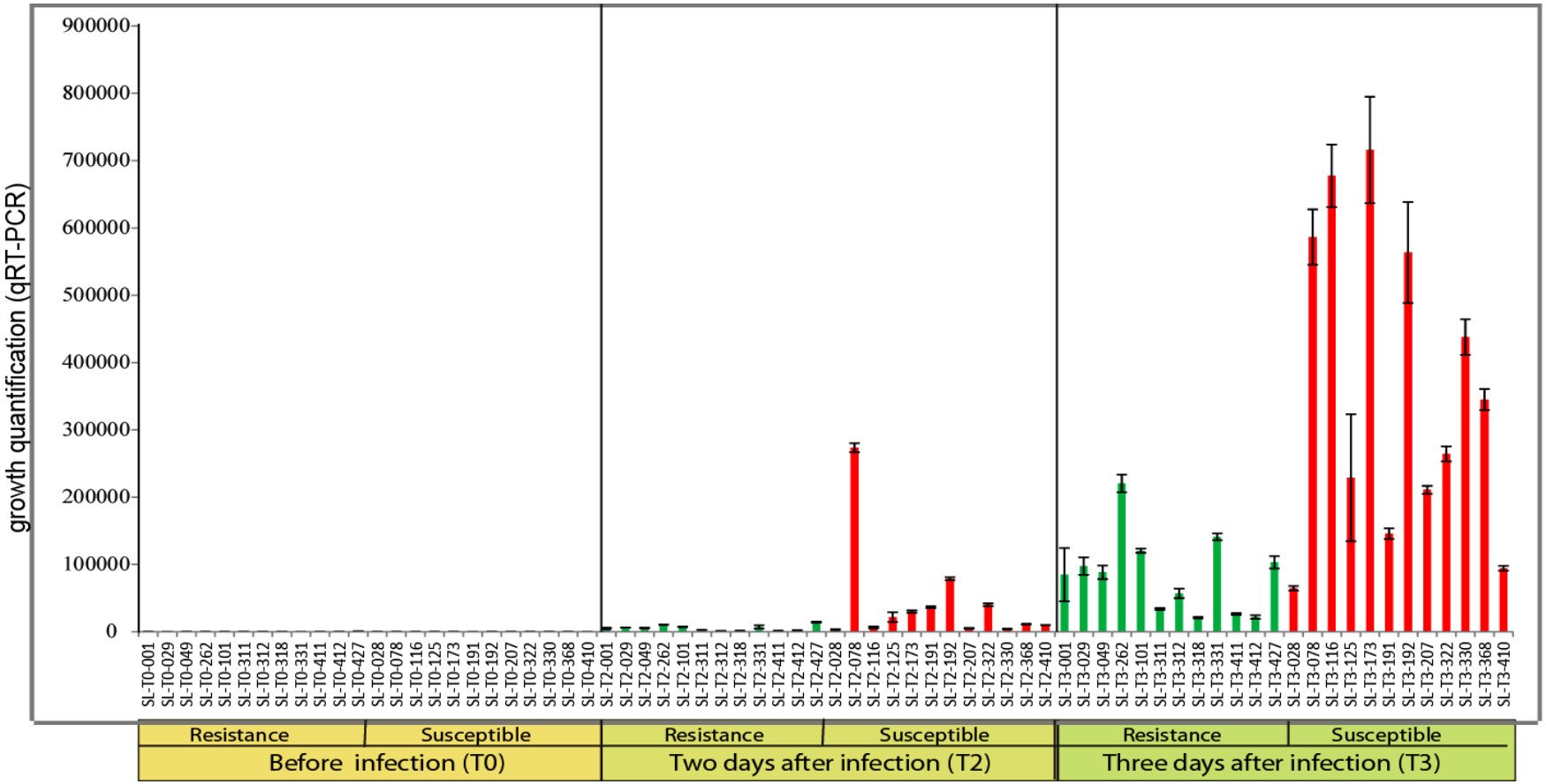


Fig. 2.4. ***P. infestans*** **growth quantification by qRT-PCR from infected leaf samples using the** ***Phytophthora*** **transcript, 60S ribosomal protein L23a (RL23a_2). The infected leaf samples were collected from the 24 tetraploid potato genotypes selected for candidate genes differential expression analysis. The quantitatively more resistant genotypes are shown by green bars while the susceptible genotypes are shown by red bars. Leaf samples were collected at three time points (T0, T2 and T3) post inoculation. The data was the average of three biological replicates.**

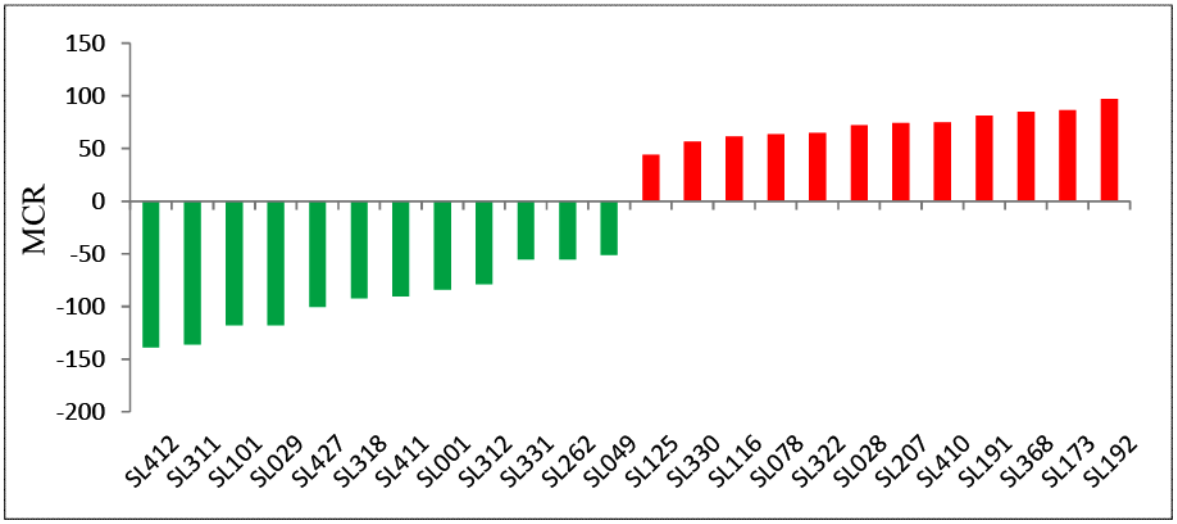


Fig. 2.5. Selected tetraploid potato genotypes contrasting for MCR, according to the 2010 field evaluation. The twelve quantitatively more resistant genotypes are shown as green bars while the twelve susceptible genotypes are shown as red bars.

2.3.3. Verification of the resistance level of the selected tetraploid potato genotypes

The 24 tetraploid potato genotypes were inoculated with complex isolates of *P. infestans* under controlled conditions in a growth chamber (Snijders Scientific B.V., Tilburg, Netherlands). The evaluation was done by qRT-PCR quantifying the growth of *P. infestans* on each genotype two (T2) and three (T3) days post inoculation using the *Phytophthora* 60S ribosomal gene, *RL23a*. The genotypes infected with the complex isolates of *P. infestans* (Fig. 2.6), showed different levels of *P. infestans* growth (Fig. 2.7) and late blight symptom severity (Fig. 2.8). Infection was detected two days post inoculation with some degree of *Phytophthora* growth difference between resistant and susceptible genotypes. Then, high growth progression with much more growth difference between resistant and susceptible genotypes was observed three days post inoculation (Fig. 2.4).

The highest *P. infestans* growth was observed on SL078 and the lowest on SL318. The average quantitative growth of *P. infestans* on each potato genotype is shown in Fig. 2.7, which was the average from three biological replications and two plants per genotype in each replication. Results were consistent with the 2010 field evaluation, except for SL028. This genotype showed less *P. infestans* growth and fell within the first 12 quantitatively more resistant genotypes in contrast to its previous (2010) assignment to the quantitatively more susceptible group. Based on this data, the 6 most resistant (SL318, SL412, SL411, SL311,

SL312 and SL028) and the 6 most susceptible (SL368, SL330, SL192, SL116, SL173 and SL078) genotypes were selected as contrasting group (Pools_2), for studying differential gene expression.

Fig. 2.6. Late blight symptoms on the 3rd compound leaf, one week post inoculation.

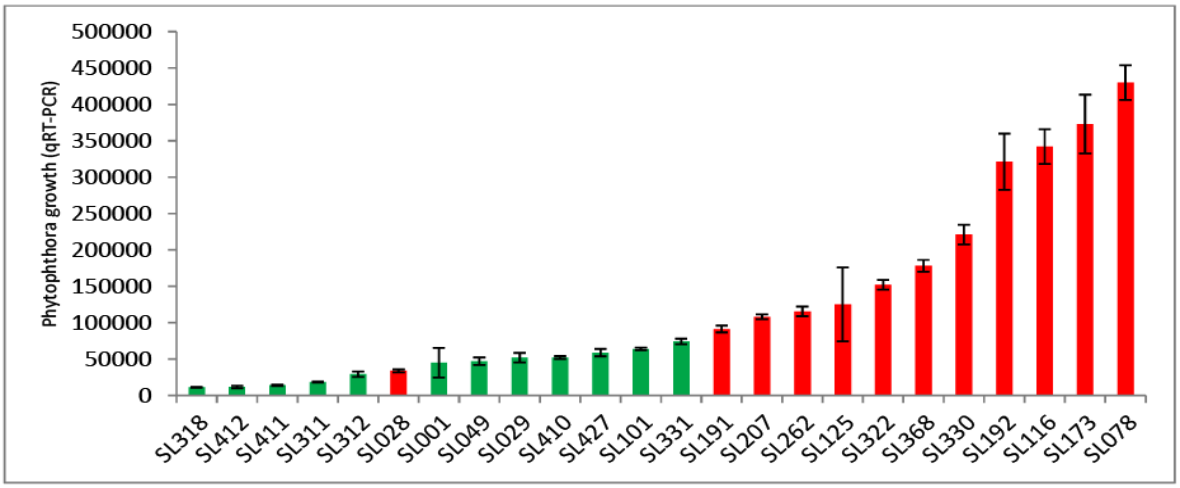


Fig. 2.7. *P. infestans* quantitative growth on 24 selected tetraploid potato genotypes. The quantification was done by qRT-PCR using the *Phytophthora* transcript RL23a. The twelve quantitatively more resistant genotypes are shown by green bars while the twelve susceptible genotypes are shown by red bars. The color was according to the 2010 field evaluation (Fig. 2.5).

In addition, each genotype was phenotyped for late blight symptom development on the 3rd compound leaf for three consecutive days from four days after inoculation. The disease score was taken in percentage using a modified CIP scale (Henfling, 1982) and converted to AUDPC. The average AUDPC of each genotype is shown in Fig. 2.8, which was the average of three biological replications and two plants per genotype. The lowest AUDPC was

recorded on SL311 while the highest was on SL125 (Fig. 2.8). Two genotypes (SL028 and SL116) that were classified within quantitatively more susceptible group in the 2010 field evaluation were grouped within the quantitatively more resistant group in this assessment. Contrasting genotypes were selected taking 5 most resistant (SL311, SL028, SL411, SL312, and SL412) and 5 most susceptible (SL191, SL368, SL173, SL322 and SL125) genotypes as contrasting groups (Pools_3), for studying differential gene expression.

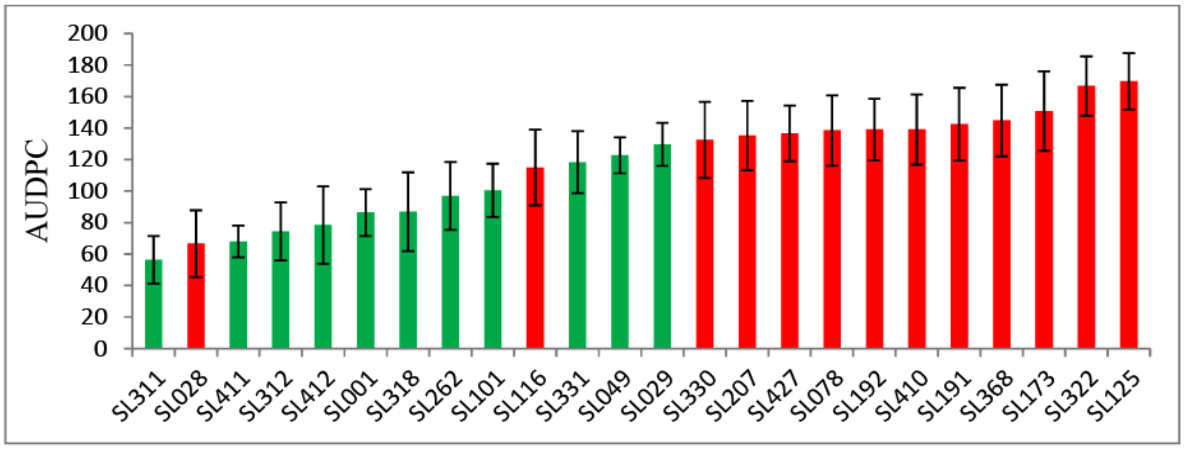


Fig. 2.8. AUDPC of selected tetraploid potato genotypes. The twelve quantitatively more resistant genotypes are shown by green bars while the twelve susceptible genotypes are shown by red bars. The color was according to the 2010 field evaluation (Fig. 2.5).

The selected 24 tetraploid potato genotypes were also planted in the field at the demonstration plot (WissenschaftsScheune, Max Planck Institute, Köln) for further verification of their resistance to late blight under natural infestation. Late blight infestation started in mid July 2012 on potato genotypes that were selected as quantitatively more susceptible, then later symptoms started on the susceptible variety, Hansa, was grown as infector row around the test plots. The genotypes that were selected as quantitatively more resistant showed less late blight symptoms (Fig. 2.9). The result was consistent with the 2010 field evaluation.

Fig. 2.9. Field response of the 24 tetraploid potato genotypes to late blight natural infestation. The 12 quantitatively more resistant genotypes are visible to the right and the 12 susceptible to the left on the picture. The susceptible variety, Hansa, was planted around the border as infector row to increase *Phytophthora* inoculums.

2.3.4. Selection of novel candidate genes for resistance to *P. infestans*

Transcriptome data generated by SuperSAGE (Draffehn et al., 2013) were used to select novel candidate genes. 165 SuperSAGE unitags (candidate genes) which were differentially expressed ($P < 10^{-4}$) between quantitatively more resistant (A2) and susceptible pools (B2 and A1) in at least five of six comparisons (three genotype pools each at three time points) were selected. The 165 candidates were represented by the same 26 bp unitag in all five or six comparisons (Supplementary table, Table 6.3).

The number of candidates was further reduced by selecting those that co-localized with known QTLs for pathogen resistance and a probable role in resistance based on literature. Hence, a final set of 23 novel candidate genes (Fig. 2.10) was selected and the expression of 22 was validated by qRT-PCR and pyrosequencing (Table 2.8). One candidate, SPI, was selected based on its differential expression both in SuperSAGE and RNA seq, therefore it was not tested in qRT-PCR or pyrosequencing. Thirteen candidates were more highly expressed in the quantitatively more resistant genotype pool A2 compared to pools A1 and B2, while two other candidates were more highly expressed in the susceptible pools A1 and B2. The 14 candidates were validated by qRT-PCR for the reproducibility of their differential expression. Eight further candidates were represented by two allelic SuperSAGE tags that showed contrasting expression between the resistant and susceptible pools, i.e., one

tag was expressed at higher level in A2 while the other was expressed at higher level in A1 and B2. These eight candidates were validated by allele specific pyrosequencing. The candidate genes were mapped in-silico using the potato genome reference sequence version 4.03, DFCI (expressed sequence data bases of potato) and PoMaMo data base (Meyer et al., 2005). The genomic position and annotation of the selected candidates are shown in Table 2.8.

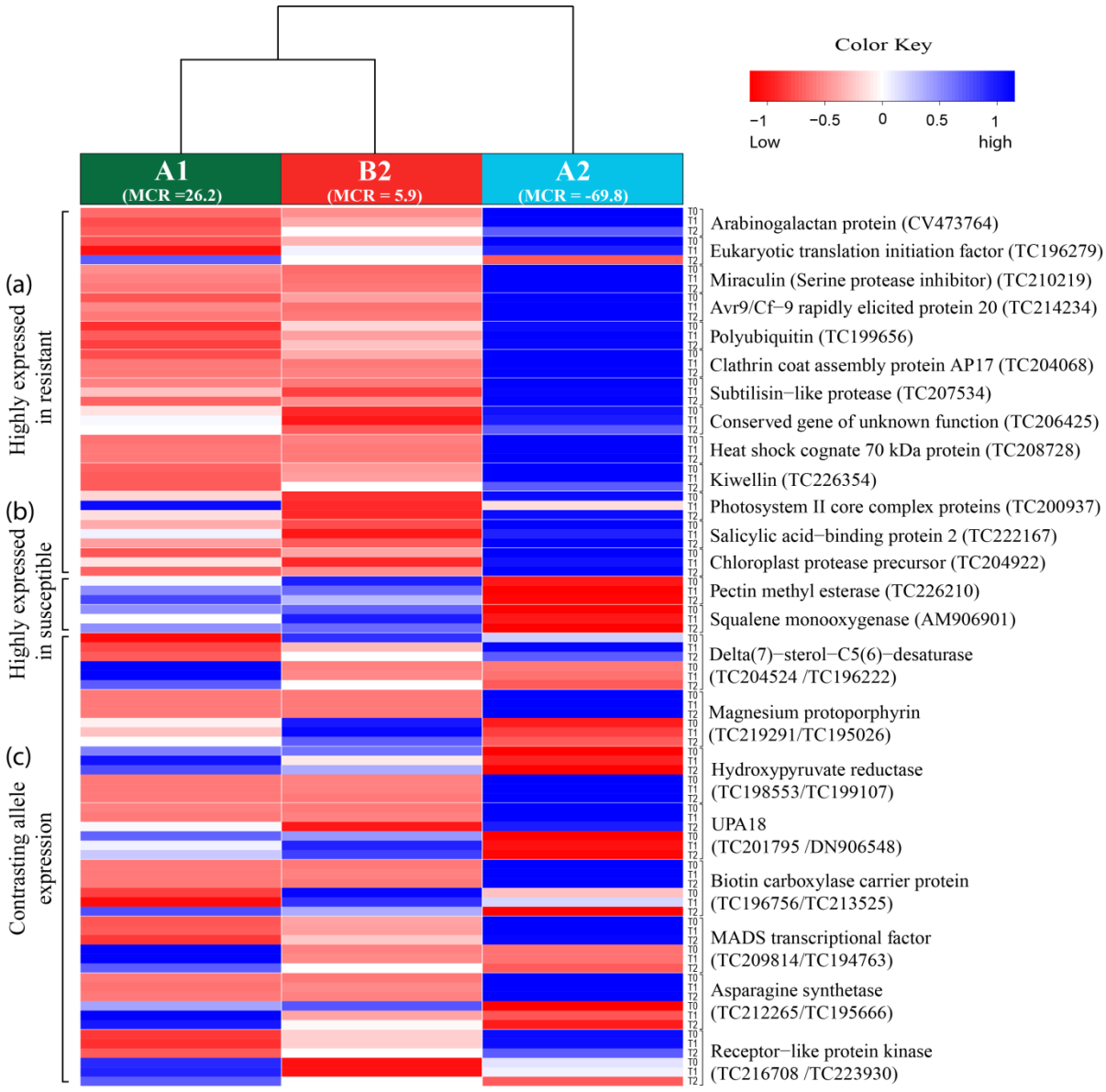


Fig. 2.10. Heatmap showing differential expression in SuperSAGE for the selected candidates. The three genotype pools (A1, B2 and A2) are in columns and the mean MCR for each pool is shown in parentheses. The candidates (transcript number in parentheses) are shown in rows to the right, each at three time points (T0, T1 and T2). The first 13 candidates (a) were highly expressed in quantitatively more resistant plants (A2), the next 2 candidates (b) were highly expressed in susceptible plants (A1 and B2) and the last 8 candidates (c) showed contrasting expression of alleles. The expression increases with the blue color. The raw data is shown in supplementary table (Table S1).

Table 2.8. Analyzed uni-tags (novel candidate genes), their annotation, the tag sequence, transcript (TC) number and position. For the eight candidates that showed contrasting allele expression, the alternative allele is shown in red.

Annotation	Acronym	The 26 bp TAG sequence	TC_number	Chr.	Locus
Arabinogalactan protein	Agp	GATCATTGAGATGAGGAGTTTGACGT	CV473764	10	PGSC0003DMG400008381
Eukaryotic translation initiation factor 4e type	EIF	CATGTAATATGAGCTAACAATTAGAT	TC196279	10	PGSC0003DMG400023664
Serine protease inhibitor (Miraculin)	SPI	CATGTGTGATGTGTAGCCTCTGTCGT	TC210219	3	PGSC0003DMG400010170
Avr9/Cf-9 rapidly elicited protein 20	AC9	GATCTTAAGGGTATGGTTAAAGAAGG	TC214234	10	PGSC0003DMG400025023
Polyubiquitin	Pubq	CATGATGGCTGTGCGCTTTGTTGTTT	TC199656	NA	Unknown
Clathrin coat assembly protein AP17	CCAP	CATGTTGATGTGGCCTCATTTAATAC	TC204068	4	PGSC0003DMG400009862
Subtilisin-like protease	SLP	CATGAGAGCTATGTTTTTAATTATGG	TC207534	1	PGSC0003DMG400006781
Conserved gene of unknown function	Cgu	CATGAACAGAGTGTGTATTTGTATAG	TC206425	12	PGSC0003DMG400004981
Heat shock cognate 70 kDa protein	HSP70	CATGTTTTTGGTTTCGTCAGTTAGTT	TC208728	9	PGSC0003DMG400008917
Kiwellin	Kiw	GATCCAGTGAGCAGTTCAGAGTCAGT	TC226354	11	PGSC0003DMG400008099
Photosystem II core complex proteins	PSP	GATCATCATAGTTTGTAATATTTGGA	TC200937	10	PGSC0003DMG400007201
Salicylic acid-binding protein	SABP	GATCCATCTCTACATCTTCTACTCTT	TC222167	2	PGSC0003DMG400013101
Chloroplast protease	Chp	GATCTAGAGTATAAACAATAGACTAT	TC204922	7	PGSC0003DMG400017311
Pectin methyl esterase	PME	CATGTTAATATTAACTATTGTGTTTA	TC226210	3	PGSC0003DMG400009178
Squalene monooxygenase	SMO	CATGATTGATTACACTATTAACTGGA	AM906901	4	PGSC0003DMG400004923
Delta (7)-sterol-C5 (6)-desaturase	DSD	CATGATTTGCAC/ATCAAGGATGTTCCT	TC204524 /TC196222	2	PGSC0003DMG400026401
Magnesium-protoporphyrin-IX-monomethyl-ester cyclase (Desaturase)	MPP	CATGTTCATTGTTGTAAGTTG/ATATAG	TC219291/TC195026	10	PGSC0003DMG400007188
Hydroxypyruvate reductase	HPR	CATGATTTTGCTATAATTAGGAC/TGGA	TC198553/TC199107	1	PGSC0003DMG400006186
UPA18	UPA18	GATCAGAAGCGTAGTGGGAAAATA/GGG	TC201795 /DN906548	9	PGSC0003DMG400006079
Biotin carboxylase carrier protein	BCCP	GATCGGTCAGGAACCATT/CGTTGAGGT	TC196756/TC213525	5	PGSC0003DMG401023454
MADS transcriptional factor	MADS	GATCAAGATGAT/CTCCTCAAATGCATC	TC209814/TC194763	4	PGSC0003DMG400009363
Asparagine synthetase	AspS	GATCAAAACCATATATAGGTTTTGAA	TC212265/TC195666	6	PGSC0003DMG400004170
Receptor-like protein kinase	RLPK	CATGAAACTAAGAT/GAG/ATATTTCTTGT	TC216708 /TC223930	3	PGSC0003DMG400015157

Chr. = chromosome position; NA = chromosome position not known

2.3.5. RNA pooling and pooling strategy

Four different RNA pools, each containing quantitatively more resistant and susceptible contrasting groups, were prepared based on the evaluation of the potato genotypes for their resistance to *P. infestans* under field and controlled conditions as well as based on evaluation methods, such as *Phytophthora* growth quantification by qRT-PCR and visual phenotyping using CIP scale and also using the *StAOS2* marker.

Pools_1 was constructed by pooling equal amounts of RNA from 12 potato genotypes each in quantitatively more resistant and susceptible contrasting group. The contrasting potato genotypes were selected based on MCR to late blight under field conditions in 2010. The quantitatively more resistant group had on average MCR of -93 and the susceptible group of 72 (Fig. 2.11a).

In pools_2, RNA was pooled from six potato genotypes each in quantitatively more resistant and susceptible contrasting group, which were selected based on *P .infestans* growth quantification by qRT-PCR after infection under controlled conditions. The average amount of *Phytophthora* was 19, 725 (expression unit calculated using C_t value) and 310, 895 in quantitatively more resistant and susceptible groups, respectively (Fig. 2.11b).

The resistant and susceptible grouping in Pools_3 was based on visual phenotyping after infection under controlled conditions. RNA was pooled from five potato genotypes in each contrasting group. The average AUDPC was 69 in the quantitatively more resistant group while the susceptible group had an average AUDPC of 155 (Fig. 2.11c).

The grouping in Pools_4 was based on *StAOS2* marker genotype and field MCR. Each contrasting pool was constructed using RNA from three potato genotypes. The quantitatively more resistant genotypes were homozygous for haplotype *StAOS2_*$A_{691}C_{692}$ while the susceptible genotypes had three copies of haplotype *StAOS2_*$G_{691}G_{692}$ and one copy of haplotype *StAOS2_*$A_{691}C_{692}$ and with mean MCR of -118 and 69, respectively (Fig. 2.11d).

Pools_1 had the highest significant difference in resistance between the contrasting groups while pools_4 had the lowest significant difference. The distribution of the data in each pool was shown by boxplots (Fig. 2.11).

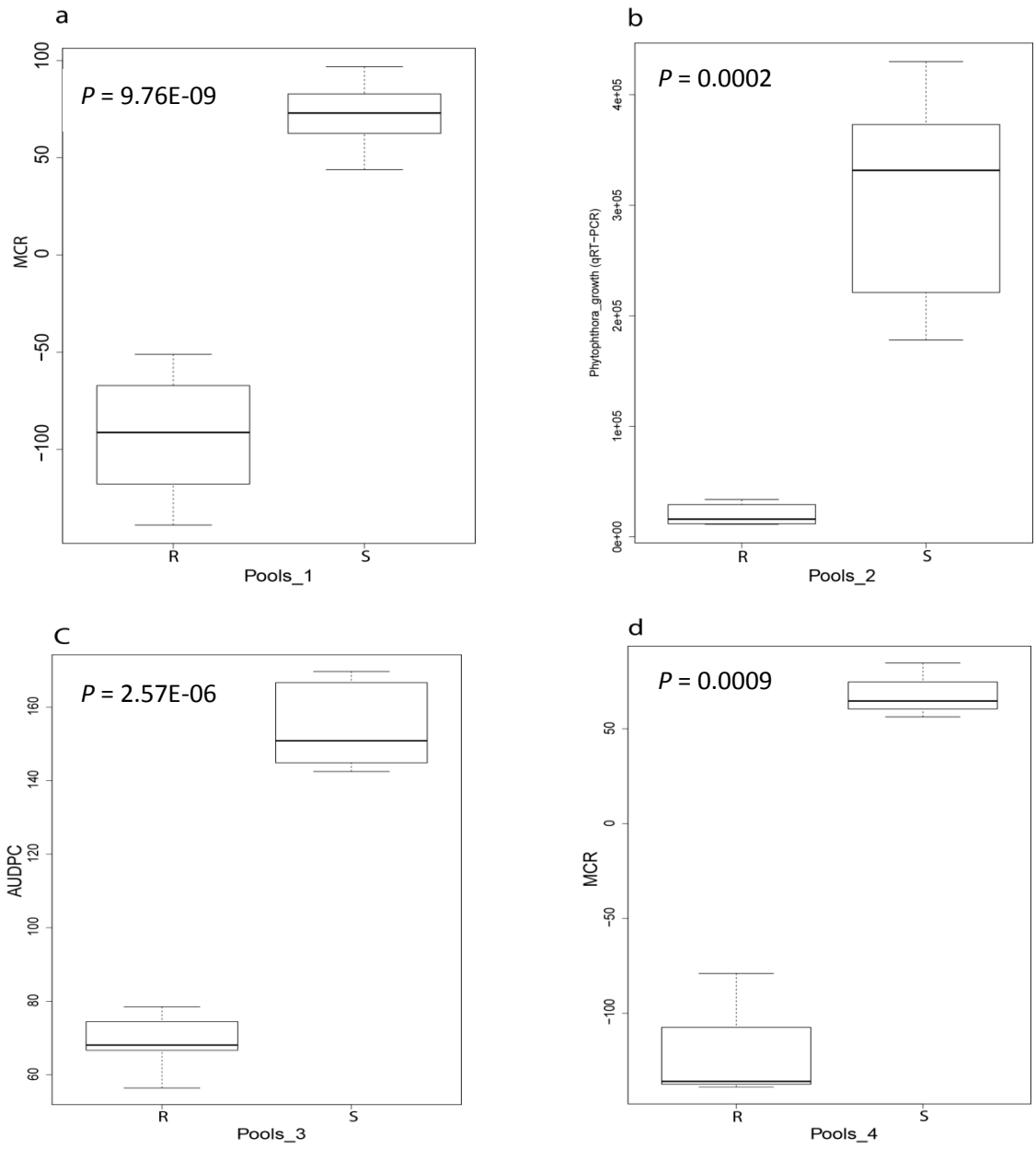


Fig. 2.11. Boxplots showing mean phenotypic data for *P. infestans* resistant and susceptible contrasting pools for each of the four different types of pools.

2.3.6. Reproducibility of differential expression

In total 22 candidates were analyzed for reproducibility of their SuperSAGE differential expression by qRT-PCR and pyrosequencing. Each candidate was analyzed on two contrasting resistant and susceptible genotype pools, at three time points (T0, T2 ad T3) and in three biological replications. In addition, candidates were tested on two to four different

types of RNA pools (Pools_1 to Pools_4). Therefore, each candidate was tested in a total of 36 to 72 replications.

2.3.6.1. *Differential expression analysis in qRT-PCR*

Fourteen candidates were tested for the reproducibility of differential expression by qRT-PCR. Six candidates were tested in all four pools while eight were tested only in pools _1 and pools_2. Pools_1 was based on MCR of genotypes under field condition while pools_2 was according to *Phytophthora* growth quantification from infection under controlled conditions.

The comparison for reproducibility was at T0 and T2, as these two time points were common in SuperSAGE and qRT-PCR expression analysis. The candidate's expression reproducibility ranged from 17 to 96 % with an average reproducibility of 60 %, taking the 48 (excluding T3 from the comparison) and or 24 RNA pools as 100%. Three candidate genes, named SMO, Pubq and AC9 were reproducible in more than 90 % of the tests. PME was reproducible in 75 % of the tests. Of the remaining ten candidates, seven were reproducible in 50 to 67 % of the tests while three were reproducible less than 50 % of the tests with the least reproducibility of 17 %. The lowest reproducibility was observed for salicylic acid binding protein (SABP). It was highly expressed in susceptible pools in contradiction to its expression in SuperSAGE.

Eleven candidates showed reproducible expression across time points. The expression of Agp, SMO, PSP and SABP was high before infection and then down regulated after infection (Fig. 2.12). Other five candidates were up regulated after infection. The expression of AC9 was high before infection (T0) and down regulated after infection in qRT-PCR test, which was contradictory to the expression observed in SuperSAGE analysis. Similarly, Cgu, which showed high expression before infection in SuperSAGE analysis, was up regulated after infection in qRT-PCR analysis.

In three independent replications, again SMO, Pubq and AC9, showed the best reproducibility.

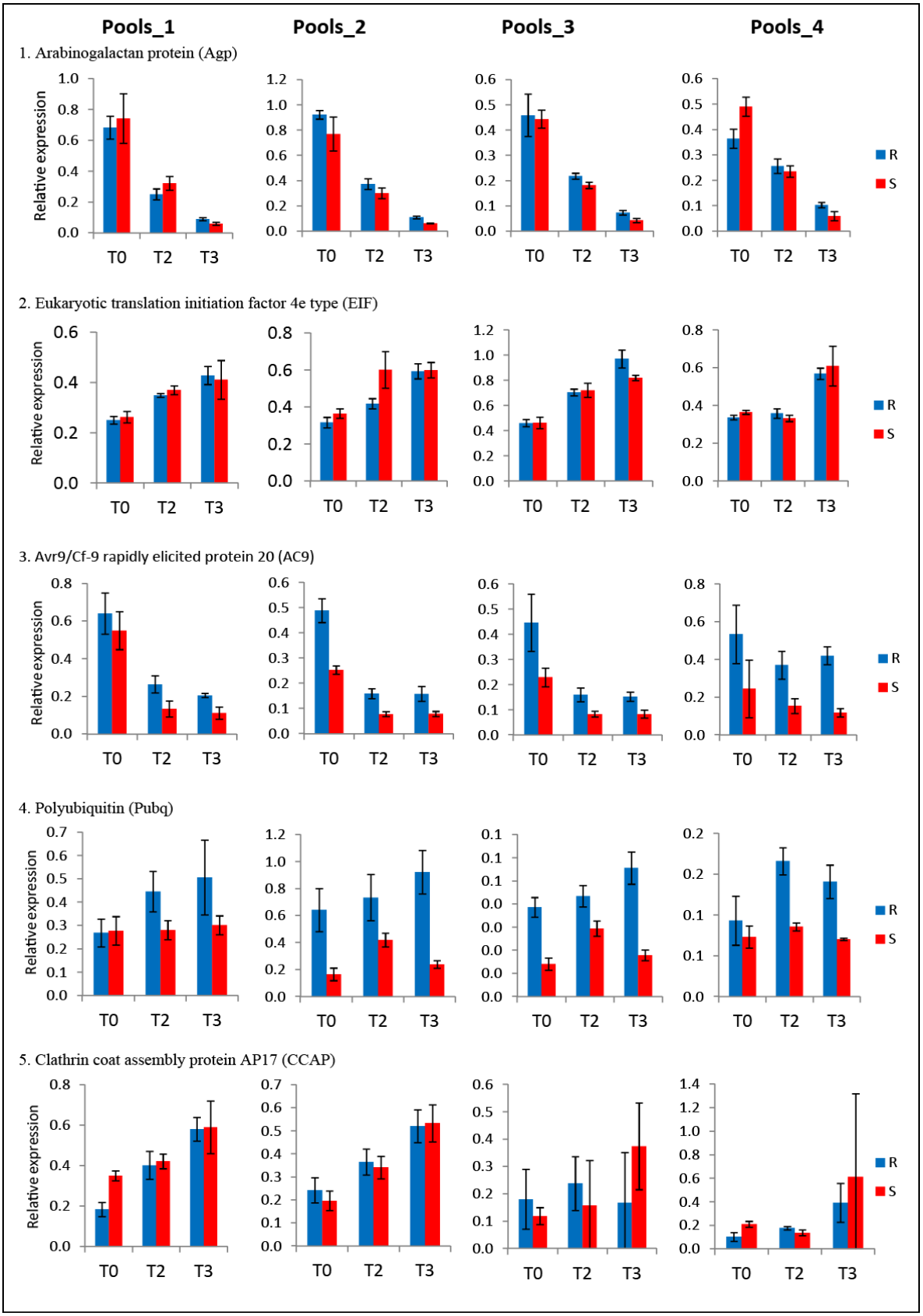


Fig. 2.12. Continued

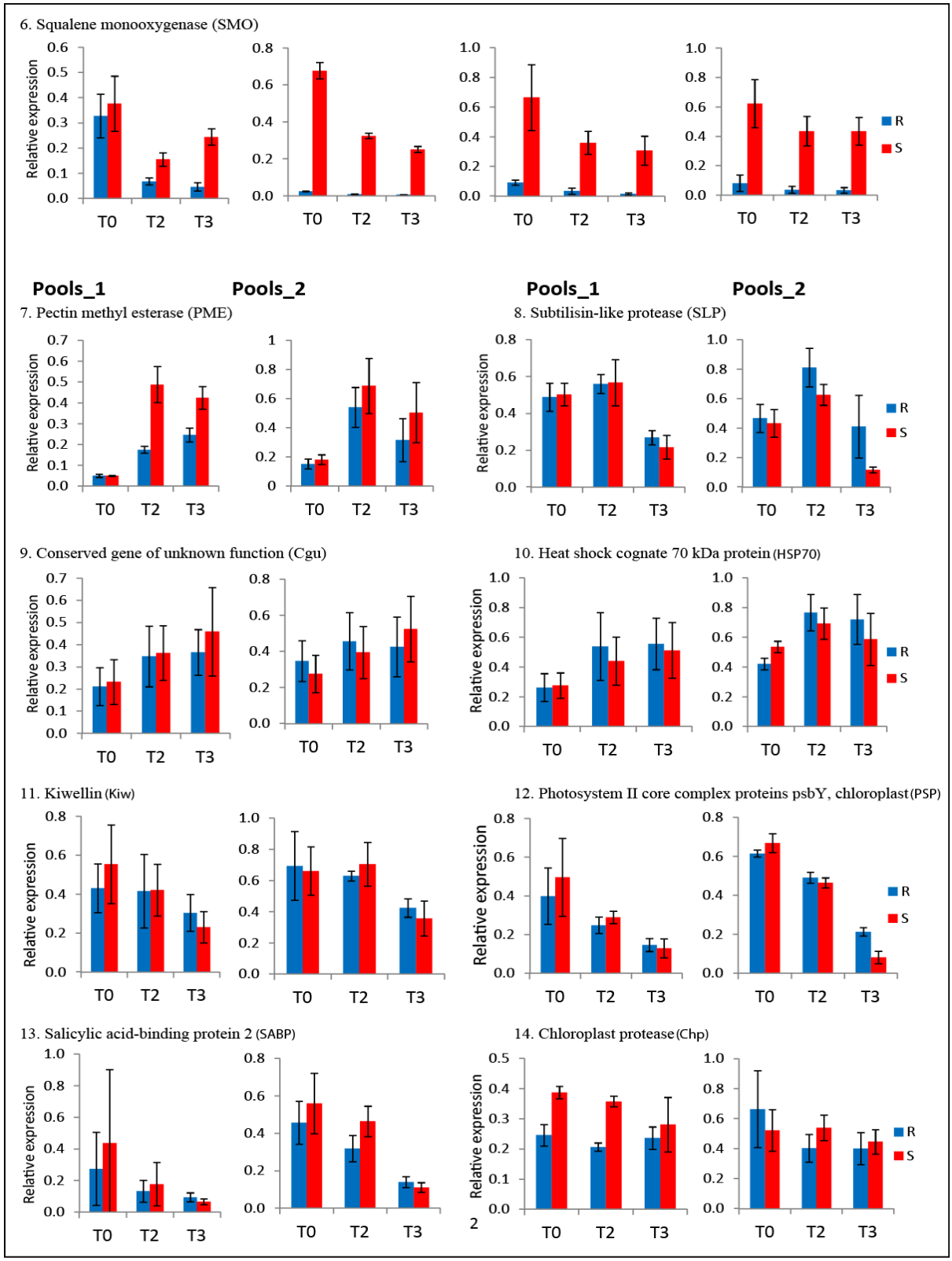


Fig. 2.12. Expression analysis by qRT-PCR. The resistant and susceptible pools are shown in blue and red bars, respectively, at three time points (T0, T2 and T3). The first six candidates were tested in 4 different types of RNA_pools, the last eight in Pools _1 and _2. The data were the average of three biological replicates.

2.3.6.2. *Allele specific differential expression analysis by pyrosequencing*

In pyrosequencing, most of the candidates showed reproducible differential allele expression in three biological replicates and in four different pooling strategies as well as in reference to their differential expression in SuperSAGE. The resistant allele (the allele highly expressed in the resistant pool in SuperSAGE analysis) differential expression between resistant and susceptible contrasting pools is shown by a heatmap (Fig. 2.13). The average of three biological replicates is shown. The heatmap clustered the resistant and susceptible groups separately, showing in most cases that the allele specific differential expression in SuperSAGE was reproducible by pyrosequencing. The reproducibility ranged from 64 % to 97 % with an average reproducibility of 89 %, taking the 72 tests (36-pairs) as 100 %. The differential allele expression of DSD, MPP, HPR and UPA18 were consistently reproducible in three independent biological replicates, in all 4 pooling strategies as well as across time points. BCCP was 100 % reproducible in pools_4, 78% in pools_1 and 67% in pools_3. However, its expression in pools_2 was contradictory to the SuperSAGE observation. The allele expression of RLPK was reproducible in pools_3 and pools_4 and was consistent across all time points. However, in pools_1 and pools _2, though it was consistent across all time points, the expression was contradictory to the SuperSAGE observation. In pools_1 and pools_2, the allele that was highly expressed in the resistant pool A2 in SuperSAGE showed less expression in pyrosequencing. Similarly, the differential expression of the AspS allele was reproducible in pools_1 and pools _2, but not in pools_3 and pools_4. To find out the reason for the above expression discrepancy, the alleles were genotyped for all individual genotypes used in the RNA pools (Pools_1 to _4) by pyrosequencing (2.3.6.2.1). MADS showed the least reproducible allele expression with 64 % reproducibility. However, it was reproducible in pools_3 and was consistent across all time points within this pool. At T3, the resistant allele was consistently higher expressed in all four resistant pools.

Across time points, the resistant allele of AspS was consistently induced from before infection (T0) to three days (T3) post infection in all the four different pools irrespective of the resistance type. This result was consistent with the SuperSAGE expression pattern. The allele of MPP was up regulated three days after inoculation in resistant pools. MADS was also up regulated two and three days post inoculation in resistant pools with reference to its

expression before infection. The resistant allele of UPA18 was up regulated two days post inoculation on resistant pools but it was down regulated on susceptible pools. DSD did not show expression difference across the three time points.

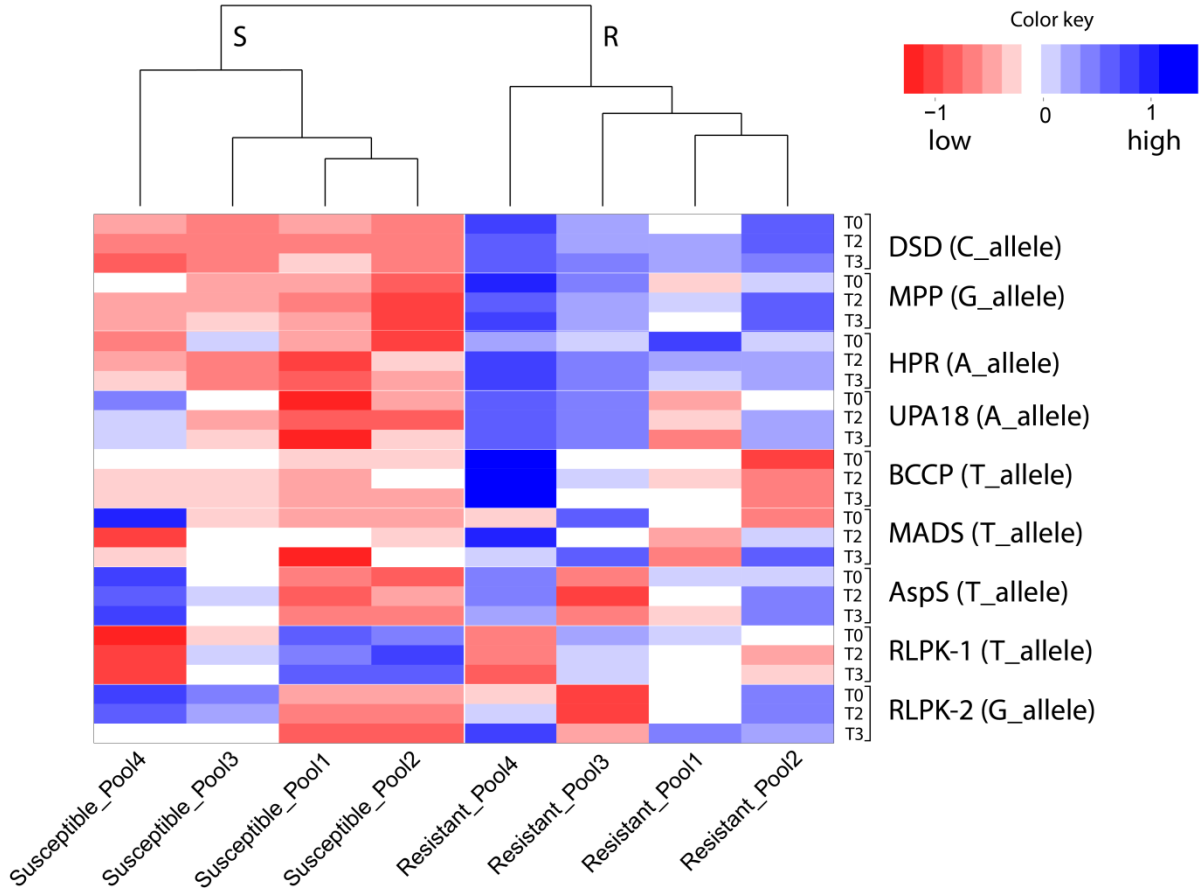


Fig. 2.13. Heatmap showing allele differential expression analysis by pyrosequencing. The expression of the resistant allele (the allele highly expressed in resistant in SuperSAGE analysis) is shown. The candidates are shown in rows to the right, each at three time points (T0, T2 and T3). The four different pools for each resistant and susceptible contrasting groups are in columns. The expression increases with the blue color. The data were the average of three biological replicates. The raw data is shown in supplementary table (Table S2).

2.3.6.2.1. Allele frequency analysis

The SNPs in the tag with contrasting SuperSAGE expression were genotyped by pyrosequencing on genomic DNA of the individual genotypes used in the RNA pools (Pools_1 to _4). The dosage of the SNPs were scored and the allele frequency was calculated for each quantitatively more resistant and susceptible contrasting groups in the four different

pools. The allele frequency was consistent with the allele expression in SuperSAGE and pyrosequencing (Fig. 2.14).

In DSD (C/A allele), MPP (G/A allele), HPR (A/G allele), UPA18 (A/G allele) and MADS (T/C allele), the allele frequencies in resistant and susceptible pools was consistent with the expression of the alleles in SuperSAGE and pyrosequencing. It was also consistent in all four pools. In AspS (T/G allele) and RLPK (T/G and G/A alleles), the allele frequencies in contrasting groups were consistent with the allele expression only in pools_1 and pools_2. In pools_3 and pools_4, the allele's frequency was still consistent with the allele's expression in pyrosequencing but it was contradictory to the SuperSAGE expression. In BCCP the allele genotyping by pyrosequencing did not work well for some of the genotypes, including the parents. The genotypes used for allele expression (used in RNA pooling) and allele frequency analysis in pyrosequencing were from two crosses with two resistant parents (Phy14 and Phy16) and one common susceptible parent (Phy20). When the allele was homozygous in one of the resistant parents, it caused the above discrepancy in allele expression and allele frequency when compared to the SuperSAGE expression. Different sets of genotypes were used in SuperSAGE analysis. The RLPK G/T SNP was homozygous in Phy14 while both Phy16 and the susceptible common parent Phy20 were homozygous for the AspS G/T SNP. It seems that the discrepancy for BCCP expression was also due to the same reason, but the pyrosequencing did not work well especially on the parental genotypes for this locus.

2.3.6.2.2. Amino acid substitution analysis

Of the 8 candidate genes with contrasting allele expression, the tag SNP in MPP and RLPK were noncoding SNPs as they are found in non-coding region. The tag SNPs in DSD, HPR, BCCP and AspS were synonymous while the tag SNPs in UPA18 and MADS were non-synonymous. The A/G SNP in UPA18 substituted methionine (M) by isoleucine (I) while the T/C SNP in MADS substituted aspartate (D) by glutamate (E) (Fig. 2.15).

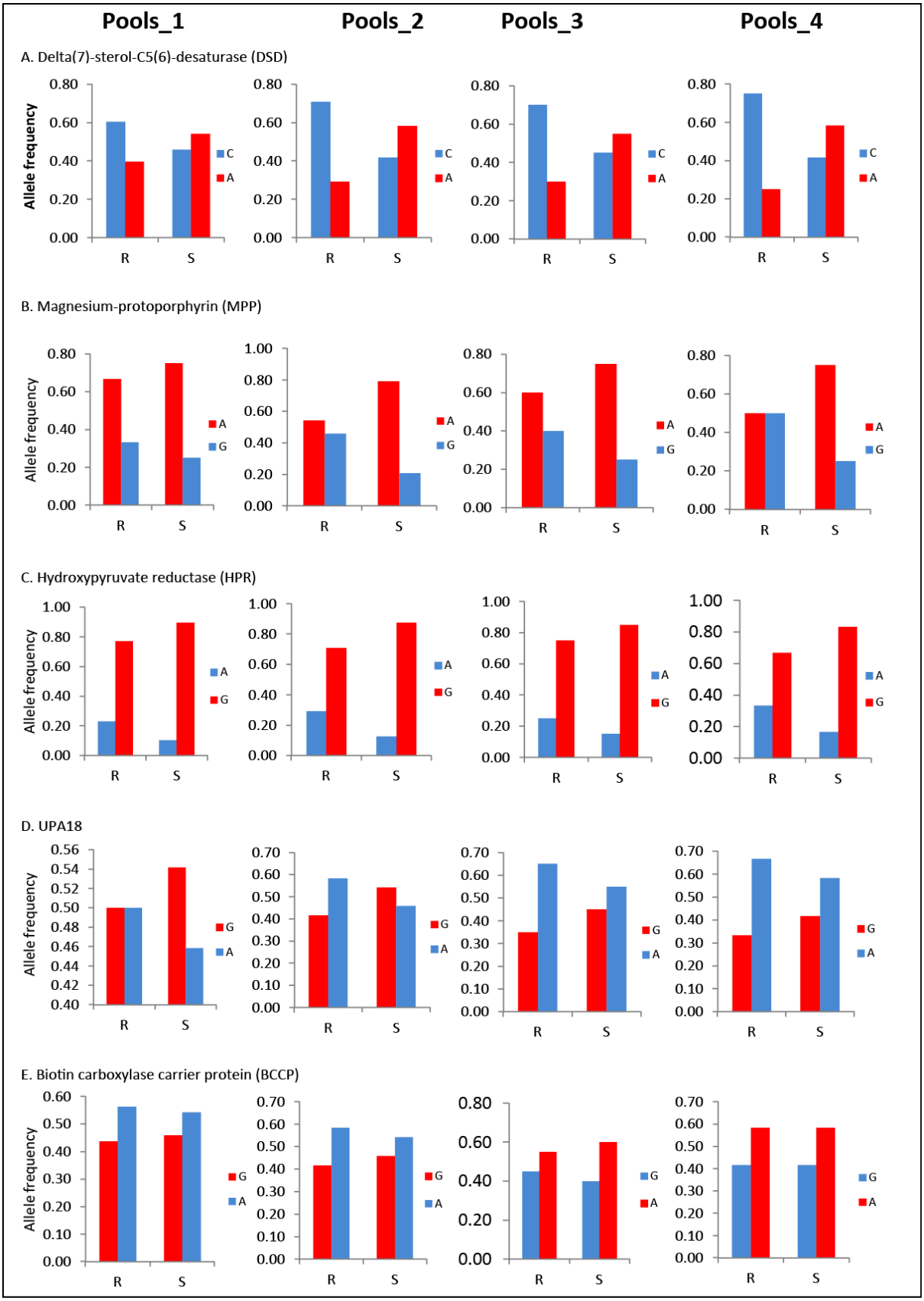


Fig. 2.14. Continued

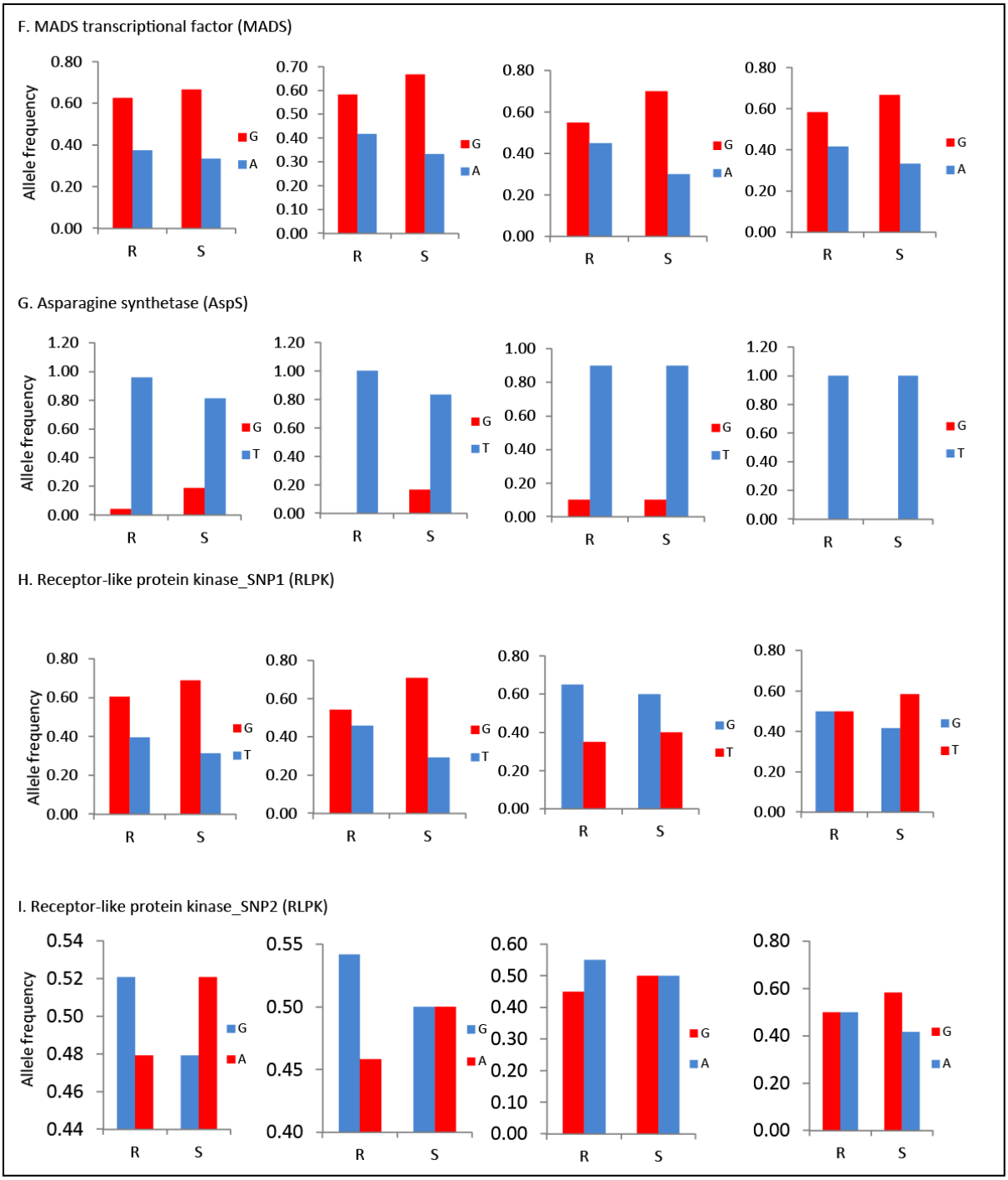


Fig. 2.14. Allele frequency of the nine allele-pairs with contrasting expression in SuperSAGE and pyrosequencing. The alleles were genotyped by pyrosequencing on genomic DNA of individual genotypes used in RNA pooling; pools_1 to pools_4. The frequency was calculated for each resistance (R) and susceptible (S) contrasting group in each pool. The frequency of the resistant and susceptible alleles is shown in blue and red, respectively.

a

```
                      10        20        30        40        50        60
                       |         |         |         |         |         |
UPA18_Allele_A MATGSSRPPQKTYDLEITIVSAKHLKNVNWHHGDLKPYVIFWVDPDQRRATQADDSGNTR
UPA18_Allele_G MATGSSRPPQKTYDLEITIVSAKHLKNVNWHHGDLKPYVIFWVDPDQRRATQADDSGNTR
                      70        80        90       100       110       120
                       |         |         |         |         |         |
UPA18_Allele_A PVWNERFVLHLPQSQSPPHDAVLTLEIFHSKPSETPKPLVGTLRVPLKELVNVDDYNKIR
UPA18_Allele_G PVWNERFVLHLPQSQSPPHDAVLTLEIFHSKPSETPKPLVGTLRVPLKELVNVDDYNKIR
                     130       140       150       160       170       180
                       |         |         |         |         |         |
UPA18_Allele_A TFELRRPSGRPHGKIRLKLAIRELSPTLDYQIPPPSSYYYSTAPPPPPPSYRTFPSSPYP
UPA18_Allele_G TFELRRPSGRPHGKIRLKLAIRELSPTLDYQIPPPSSYYYSTAPPPPPPSYRTFPSSPYP
                     190       200       210       220       230       240
                       |         |         |         |         |         |
UPA18_Allele_A SHPHPPPSHVVAPPPPASPSPPPPPPQQHSFPYSGFSDPYSTYFPGYYSQPPPPPRPFVE
UPA18_Allele_G SHPHPPPSHVVAPPPPASPSPPPPPPQQHSFPYSGFSDPYSTYFPGYYSQPPPPPRPFVE
                     250       260       270       280       290       300
                       |         |         |         |         |         |
UPA18_Allele_A RQSSYGSLGSRPSAPVDYYPPYDQKRSGKIGMGVGTGLAVGAVAGTLGGLTLGEGLKYEE
UPA18_Allele_G RQSSYGSLGSRPSAPVDYYPPYDQKRSGKMGMGVGTGLAVGAVAGTLGGLTLGEGLKYEE
                     310       320
                       |         |
UPA18_Allele_A AKIAERVENNTSARDDYSNYSEY
UPA18_Allele_G AKIAERVENNTSARDDYSNYSEY
```

b

```
                      10        20        30        40        50        60
                       |         |         |         |         |         |
MADS_Allele_A  MAREKIKIKKIDNITARQVTFSKRRRGLFKKAEELSVLCDADVALIIFSATGKLFDFAST
MADS_Allele_C  MAREKIKIKKIDNITARQVTFSKRRRGLFKKAEELSVLCDADVALIIFSATGKLFDFAST
                      70        80        90       100       110       120
                       |         |         |         |         |         |
MADS_Allele_A  SMKDILGKYKLQSASLEKVDQPSLDLQLENSLNMRLSKQVADKTRELRQMRGEELEGLSL
MADS_Allele_C  SMKDILGKYKLQSASLEKVDQPSLDLQLENSLNMRLSKQVADKTRELRQMRGEELEGLSL
                     130       140       150       160       170       180
                       |         |         |         |         |         |
MADS_Allele_A  EELQQIEKRLEAGFNRVLEIKGTRIMDEITNLQRKGAELMEENKQLKYKMEIMKKGKLPL
MADS_Allele_C  EELQQIEKRLEAGFNRVLEIKGTRIMDEITNLQRKGAELMEENKQLKYKMEIMKKGKLPL
                     190       200       210
                       |         |         |
MADS_Allele_A  LTDMMMEEGQSSESIITTNNPDQDDSSNASLKLGLPFS
MADS_Allele_C  LTDMMMEEGQSSESIITTNNPDQDESSNASLKLGLPFS
```

Fig. 2.15. Amino acid sequence alignment showing amino acid substitution by a SNP with contrasting expression between quantitatively more resistant and susceptible potato genotype pools, of UPA18 (a) and MADS (b).

2.3.7. Linkage analysis for selected candidates

2.3.7.1. *Phenotypic data*

Phenotypic data of AUDPC, MCR and PM collected by the breeding company SAKA (Pflanzenzucht GmbH & Co. Ka) were used. The data were collected on a total of 114 heterozygous tetraploid potato genotypes composed of 111 half sib F1 genotypes and three parents. The genotypes were evaluated in 2010 under field conditions for quantitative late blight resistance. The histograms for each trait are shown in Fig. 2.16. As shown in Fig. 2.17a, AUDPC was negatively correlated with PM showing the association between late blight resistance and late plant maturity in potato, however there was no correlation between MCR and PM (Fig. 2.17b).

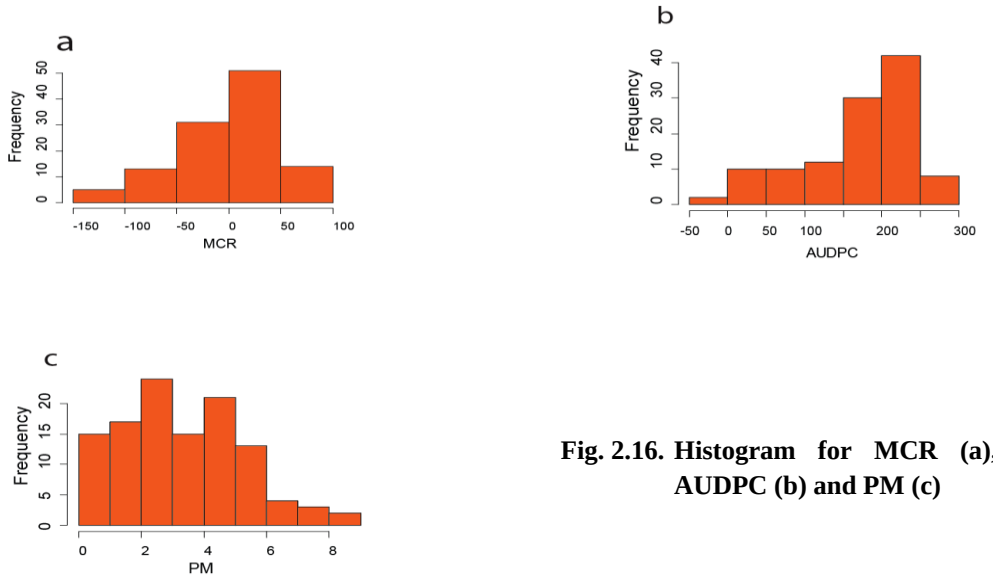


Fig. 2.16. Histogram for MCR (a), AUDPC (b) and PM (c)

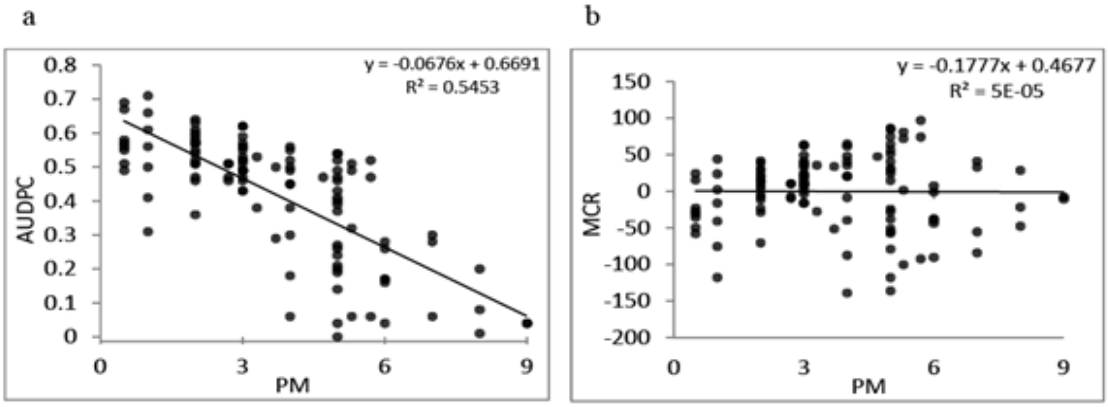


Fig. 2.17. Correlation between PM and AUDPC (a) and MCR (b). PM scored in 0 – 9 scale, 9 being very late maturing

2.3.7.2. *Marker data*

Ninety-two SNP markers from seven loci (candidate genes), with 5 to 28 SNP markers per locus, were collected by genotyping the 114 heterozygous potato clones (SAKA population) using Sanger sequencing. In addition, the 9 SNPs from 8 SuperSAGE tags that showed contrasting allele expression between quantitatively more resistant and susceptible genotype pools were genotyped on the segregating potato clones by pyrosequencing. Hence, a total of 101 SNPs were tested for linkage with the phenotypic traits (Fig. S1).

2.3.7.3. *Marker-trait linkage analysis*

A single marker-trait analysis was performed using general linear model (GLM) to detect QTLs for AUDPC, MCR and PM. Out of the seven candidates, SPI, AspS, MPP and DSD were selected based on differential expression in SuperSAGE analysis. 4Cl, lox1 and StAOS2 were selected based on their significant association with MCR in previous studies (Pajerowska-Mukhtar et al., 2009; Fritzemeier et al., 1987).

With a cutoff at P = 0.01, 8 SNPs from SPI were significantly linked with AUDPC and explained about 10 to 14 % the total variance. None of the SNPs were linked with MCR or PM. One SNP from DSD was significantly linked with both AUDPC and MCR. From AspS, two SNPs were significantly linked with AUDPC and four SNPs with MCR. Among them, one SNP was highly significantly linked with MCR with P-value of 3×10^{-8}. Another AspS SNP (SNP165) caused amino acid substitution (Fig. 2.19). Of the 9 SNPs that showed contrasting allele expression in SuperSAGE, one SNP from DSD was significantly linked with MCR. The range of total variation explained by the significant SNPs was 8 to 14 % in AUDPC and 13 to 30 % in MCR. The direction of the allele effect for linked SNPs is shown in Table 2.9.

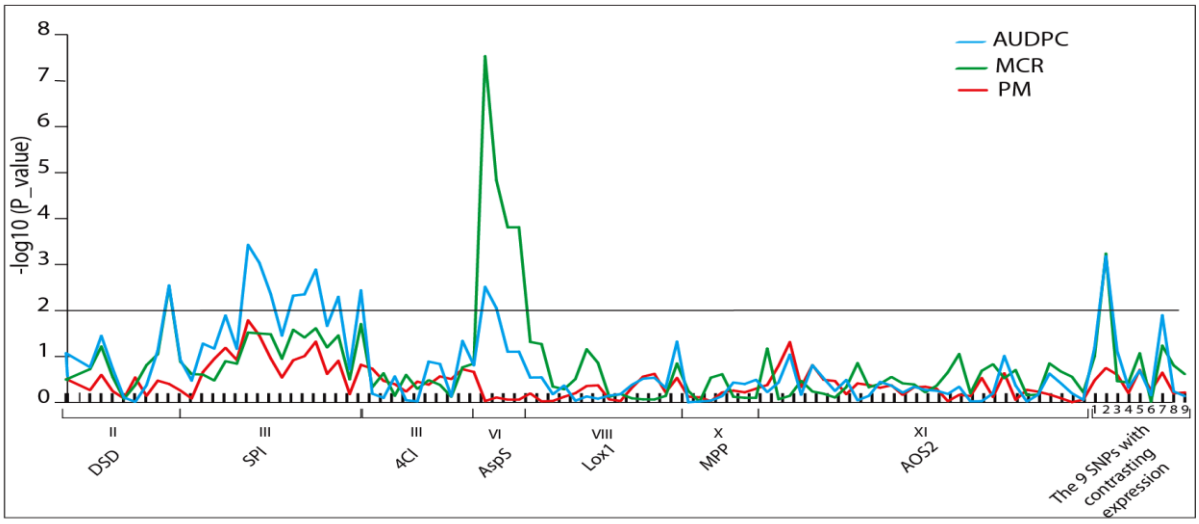


Fig. 2.18. Linkage plot for SNPs from selected candidate loci. SNPs are plotted in x-axis while the $-\log_{10}$ of P_values in y-axis. The three traits are shown by lines with different colors. The threshold level was set at P < 0.01. The SNPs with contrasting expression are shown at the right; 1 = HPR, 2 = DSD, 3 =RLPK1, 4 = RLPK2, 5 = MADS, 6 = BCCP, 7 = AspS, 8 = UPA18, 9 = MPP.

```
                        10         20         30         40         50         60
                         |          |          |          |          |          |
Asp_allele_C    MCGILALLGCSDDSQAKRVRVLELSRRLKHRGPDWSGIFQYGDFYLAHQRLAIIDPASGD
Asp_allele_T    MCGILALLGCSDDSQAKRVRVLELSRRLKHRGPDWSGIFQYGDFYLAHQRLAIIDPASGD
                        70         80         90        100        110        120
                         |          |          |          |          |          |
Asp_allele_C    QPLFNEDKKIVVTVNGEIYNHEKLRKLMPNHKFRTGSDCDVIAHLYEEYGENFVDMLDGV
Asp_allele_T    QPLFNEDKKIVVTVNGEIYNHEKLRKLMPNHKFRTGSDCDVIAHLYEEYGENFVDMLDGV
                       130        140        150        160        170        180
                         |          |          |          |          |          |
Asp_allele_C    FSFVLLDTRDNSFLAARDAIGITPLYIGWGLDGSVWISSELKGLNDDCEHFEVFPPGHLY
Asp_allele_T    FSFVLLDTRDNSFLAARDAIGITPLYIGWGLDGSVWISSELKGLNDDCEHFEVFPPGHLY
                       190        200        210        220        230        240
                         |          |          |          |          |          |
Asp_allele_C    SSKNGGLRRWYNPAWFSEAIPSTPYDTLVLRRAFENAVIKRLMTDVPFGVLLSGGLDSSL
Asp_allele_T    SSKNGGLRRWYNPAWFSEAIPSTPYDTLVLRRAFENAVIKRLMTDVPFGVLLSGGLDSSL
                       250        260        270        280        290        300
                         |          |          |          |          |          |
Asp_allele_C    VASVTTRYLAGTKAAKQWGAQLHSFCVGLEGSPDLKAAKEVADFLGTVHHEFHFTVQDGI
Asp_allele_T    VASVTTRYLAGTKAAKQWGAQLHSFCVGLEGSPDLKAAKEVADFLGTVHHEFHFTVQDGI
                       310        320        330        340        350        360
                         |          |          |          |          |          |
Asp_allele_C    DAIEDVIYHIETYDVTTIRASTPMFLMSRKIKSLGVKMVISGEGADEIFGGYLYFHKAPN
Asp_allele_T    DAIEDVIYHIETYDVTTIRASTPMFLMSRKIKSLGVKMVISGEGADEIFGGYLYFHKAPN
                       370        380        390        400        410        420
                         |          |          |          |          |          |
Asp_allele_C    KEEFHTETCRKIKALHQYDCLRANKATSAWGLEARVPFLDKEFIDVAMSIDPEWKMIKHD
Asp_allele_T    KEEFHTETCRKIKALHQYDCLRANKATSAWGLEARVPFLDKEFIDVAMSIDPEWKMIKHD
                       430        440        450        460        470        480
                         |          |          |          |          |          |
Asp_allele_C    QGRIEKWVLRKAFDDEEQPYLPKHILYRQKEQFSDGVGYSWIDGLKAHAEQHVTDRMMLN
Asp_allele_T    QGRIEKWVLRKAFDDEEQPYLPKHILYRQKEQFSDGVGYSWIDGLKAHAEQHVTDRMMLN
                       490        500        510        520        530        540
                         |          |          |          |          |          |
Asp_allele_C    AAHIFPHNTPTTKEGYYYRMIFERFFPQNSASLTVPGGPSIACSTAKAIEWDASWSNNLD
Asp_allele_T    AAHIFPHNTPTTKEGYYYRMIFERFFPQNSASLTVPGGPSIACSTAKAIEWDASWSNNLD
                       550        560        570        580        590
                         |          |          |          |          |
Asp_allele_C    PSGRAAIGVHNSAYDNHLSSVANGNLDTPIINNVPKMVGVGVAAELTIRS
Asp_allele_T    PSGRAAIGVHNSAYDNHLSSVANGNLDTPIINNVPRMVGVGVAAELTIRS
```

Fig. 2.19. Amino acid sequence alignment showing amino acid substitution by the significantly linked SNP165 (C/T) from AspS.

Table. 2.9. Significantly linked markers, allele frequencies and allele effects. Markers from SPI locus are not shown as the minor allele frequency was less than 5%.

Locus	Chr.	Marker	Allele	Allele Freq. (%)	MCR P_value (R^2)	AUDPC P_value (R^2)
Delta (7)-sterol-C5 (6)-desaturase (DSD)	II	DSD_SNP31	T/G	54/45	0.0030(0.13)↓/↑	0.0029(0.14)↓/↑
Asparagine synthetase (AspS)	VI	AspS_SNP112	C/T	39/61	0.0001 (0.17)↓/↑	NS
		AspS_SNP119	T/G	39/61	0.0001 (0.17)↓/↑	NS
		AspS_SNP165	G/A	36/64	1.47E-05(0.21) ↓/↑	NS
		AspS_SNP175	G/T	80/20	2.95E-08 (0.29)↓/↑	0.0031 (0.12)*

Chr. = Chromosome; Marker effect shown by arrow, ↑ show positive effect (higher resistance and earliness); ↓ show negative effect (susceptibility and late maturity); NS = not significant (P > 0.01); *the direction of the allele effect unclear

2.3.8. Candidate genes association mapping

2.3.8.1. *Phenotypic data*

Phenotypic data rAUDPC, MCR and PM collected on the CONQUEST2 association mapping population (Pajerowska-Mukhtar et al., 2009) were used in marker-trait association analysis. The population contained 184 tetraploid potato clones obtained from two breeding companies, Boehm-Nordkartoffel Agrarproduktion (BNA) and SAKA. The histograms of each trait are shown in Fig. 2.20.

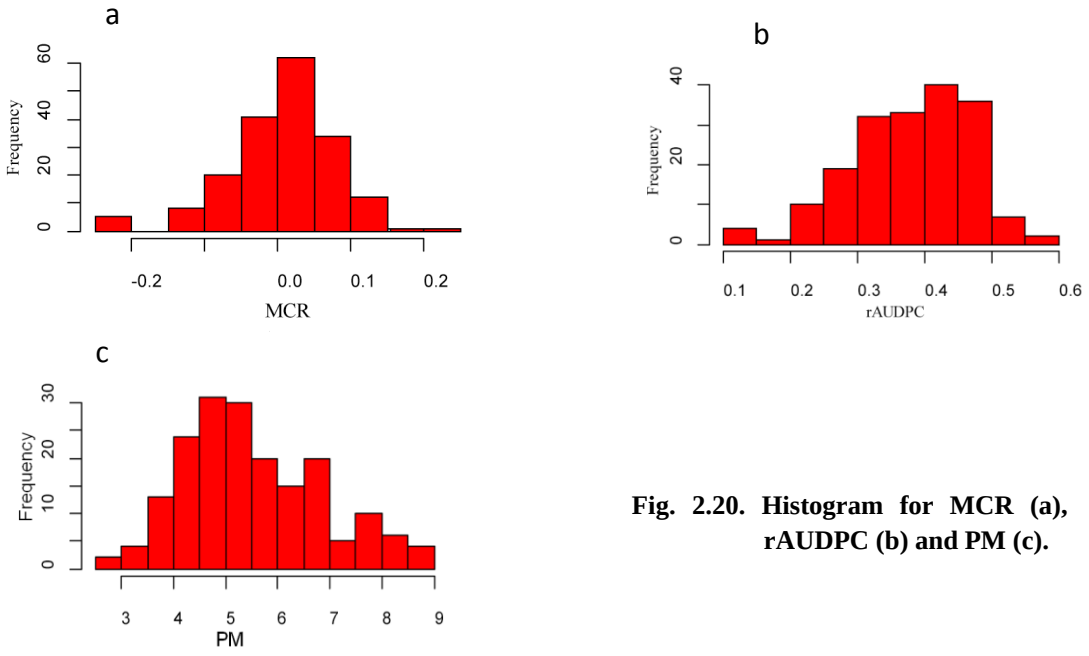


Fig. 2.20. Histogram for MCR (a), rAUDPC (b) and PM (c).

2.3.8.2. *Marker data*

Eleven candidate genes that showed reproducible differential expression between quantitatively more resistant and susceptible contrasting potato genotype pools were analyzed for their association with rAUDPC, MCR and PM. The candidates were genotyped in the CONQUEST2 population by Sanger sequencing. A total of 112 polymorphic bi-allelic and 2 tri-allelic SNP markers were scored in the amplicons of 9 candidates while no scorable sequence was found in repeated sequencing attempts with different primer pairs for two candidates; Pubq and UPA18. The dosages of the SNP alleles were scored. The number of SNP markers scored per locus ranged from 4 in MPP to 18 from DSD.

2.3.8.3. *Marker-trait association analysis*

The marker-trait association analysis was done using Tassel (2.1) statistical software (Bradbury et al., 2007) in a mixed linear model (MLM), which accounts for population structure (Q) and kinship (K) to avoid false positives. The Q and K matrices were calculated using SNP markers uniformly distributed on the 12 potato chromosomes (described and discussed in chapter III). Presence of two subpopulations (Q) in CONQUEST2 was determined based on Evano et al. (2005) method after running population structure analysis (Pritchard et al., 2000). Only one SNP from AspS had MAF (minor allele frequency) < 10 % and was excluded from the analysis.

2.3.8.4. Maturity corrected resistance (MCR)

Of 113 SNP markers analyzed, 8 SNPs from 6 loci were found to be significantly associated ($P < 0.01$) with MCR (Fig. 2.21; Table 2.10). The associated loci in order of their significance were BCCP (two SNPs), PME (one SNP), AC9 (one SNP), DSD (two SNPs), MPP (one SNP) and Chp (one SNP). No LD was detected between significantly associated SNP from different loci, whereas SNPs within a locus were in LD with each other (Fig. 2.23). The total variance explained by the SNPs ranged from 6 to 14 %. The two associated SNPs in BCCP locus caused amino acid substitution (Fig. 2.22).

2.3.8.5. Relative area under disease progress curve (rAUDPC)

For rAUDPC, four SNPs from Chp, two SNPs from BCCP and one SNP from each of PME, MPP and HPR were significantly associated (P < 0.01) (Fig. 2.21; Table 2.10). The total variance explained by the SNPs ranged from 8 to 12 %.

2.3.8.6. Plant maturity (PM)

Three SNPs from DSD and one SNP from AspS were significantly (P < 0.01) associated with plant maturity (Fig. 2.21; Table 2.10). The two SNPs from DSD were in LD and the best associated SNP explained 11 % of the total variance.

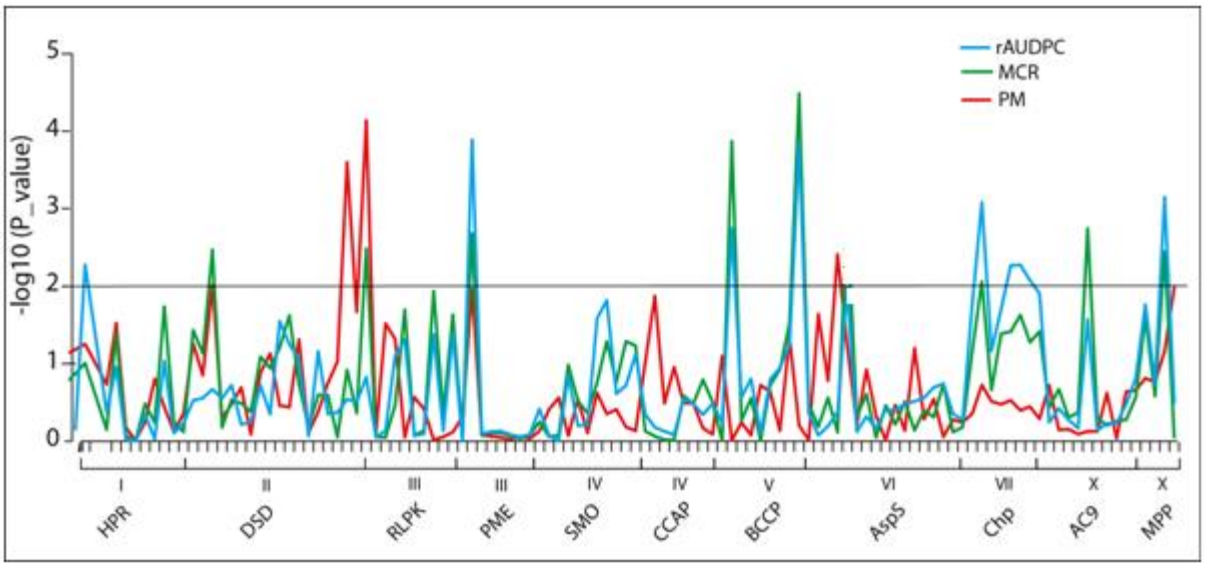


Fig. 2.21. Association plot of SNPs from eleven loci that showed reproducible differential expression. SNPs are plotted on the x-axis and the $-\log_{10}$ of P_values are plotted on the y-axis. The three traits are shown by lines with different colors. The threshold level was set at P-value < 0.01. The chromosome position is shown on the X-axis.

In all cases, the direction of the allele effect was the same for MCR and rAUDPC, suggesting that these two traits are controlled by the same genes. However, the alleles of the two associated SNPs in the DSD locus showed opposite effects. The alleles increasing MCR was causing late maturity or vise versa. It was difficult to determine the direction of allele effects from associated SNPs in Chp locus, as the allele with positive effect at simplex show negative effect at duplex or triplex and again show positive effect at quadraplex. The alleles

at this locus might have interaction effect. The allele effect for associated markers is shown in Table 2.10.

2.3.9. Candidate genes co-localization with known QTLs for pathogen resistance

Most of the candidates co-localized or were found in close proximity to anchor markers which co-localized with known QTLs for pathogen resistance. DSD on chromosome II and SPI on chromosome III co-localized with anchor markers RbcS-2 and StK1, respectively. On chromosome IV, MADS and CCAP co-localized with anchor markers TG22 and GP83. BCCP co-localized with anchor marker TG185, on chromosome V. The remaining candidates were in close proximity with different anchor markers as shown in Fig 2.24.

```
                    10        20        30        40        50        60
                     |         |         |         |         |         |
BCCP_C-C)     MTTSFGTTTTATPAASASPAGFPVKNAPKSLTLLPFCHHRHYNHLRSKLSITPKLRLSCK
BCCP_T-T)     MTTSFGTTTTATPAASASPAGFPVKNAPKSLTLLPFCHHRHYNHLRSKLSITPKLRLSCK
                    70        80        90       100       110       120
                     |         |         |         |         |         |
BCCP_C-C      GLQSSWNHSAVTRAQSNEVALGGSSNNSATPITQSEEVKSPSETSSATSVSEEAISNFIS
BCCP_T-T      GLQSSWNHSAVTRAQSNEVALGGSSNNSATPITQSEEVKSPSETSSATPVSEEAISNFIS
                   130       140       150       160       170       180
                     |         |         |         |         |         |
BCCP_C-C      QVSSLVKLVDSRDIVELQLKQLDCEILIRKKEALPQPPAPAPAQAPFIQSYHVPSVQSNA
BCCP_T-T      QVSSLVKLVDSRDIVELQLKQLDCEILIRKKEALPQPPAPAPAQAPFIQPYHVPSVQSNA
                   190       200       210       220       230       240
                     |         |         |         |         |         |
BCCP_C-C      PPPPAPAPAPIQTPAPSPAAAKSADSSLPPLKSPMAGTFYRSPAPSEPAFVKVGDKVQKG
BCCP_T-T      PPPPAPAPAPIQTPAPSPAAAKSADSSLPPLKSPMAGTFYRSPAPSEPAFVKVGDKVQKG

                   250       260       270       280       290       300
                     |         |         |         |         |         |
BCCP_C-C      QVLCIIEAMKLMNEIEADRSGTIVEVVAEDGKPVSVDTPLFVIKPNRSGKTQETICRGTW
BCCP_T-T      QVLCIIEAMKLMNEIEADRSGTIVEVVAEDGKPVSVDTPLFVIKPNRSGKTQETICRGTW
                   310       320
                     |         |
BCCP_C-C      QVFVYLGLMIFCSISVSVSMFATTGSVME
BCCP_T-T      QVFVYLGLMIFCSISVSVSMFATTGSVME
```

Fig. 2.22. Amino acid sequence alignment showing amino acid substitution by two significantly associated SNPs from BCCP locus.

2.3.10. Linkage disequilibrium (LD)

For linkage disequilibrium (LD) analysis, all SNPs from candidate genes in which one or more of the SNPs were found to be associated or linked with any of the three traits (rAUDPC, MCR or PM) were selected. Hence, a total of 73 SNPs from seven loci were arranged in order of their map position and the significance of LD between two loci was calculated by testing for independence between two loci with a X^2 (chi-square) test. As

shown by a LD_heatmap (Fig. 2.23), there was strong LD between SNPs within locus but the LD of SNPs between loci was generally weak.

The two (DSD_SNP81 and DSD_SNP248) significantly associated SNPs in DSD were in LD (P = 3.15E-25), interestingly both SNPs were in strong LD with a SNP (DSD_SNP31) that was significantly linked with MCR and AUDPC in linkage analysis. This SNP was differentially expressed both in SuperSAGE and pyrosequencing. Similarly, the two significant SNPs from BCCP were in strong LD (P = 3.22E-32). Three SNPs in ACP were in strong LD with most of the SNPs in AspS, including the significantly linked SNPs. The significantly associated/linked SNPs from the two loci were in strong LD with P-value of 1.7E-05. Therefore, these two loci were not independent. The LD blocks between these two loci shown by c, d and e (Fig. 2.23). Four other SNPs were showed weak LD (LD block a and b), but none of these SNPs were associated or linked to any of the traits. In general, many SNP within locus were in LD with each other, but most SNPs between loci were independent.

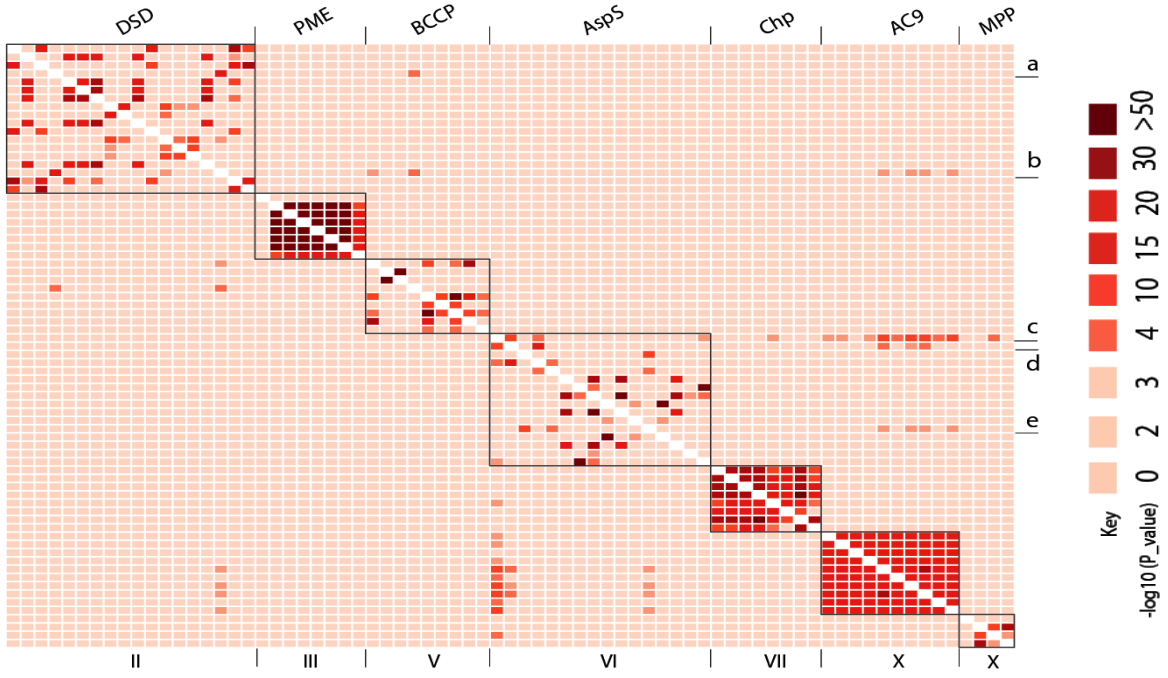


Fig. 2.23. LD_heatmap for loci with one or more of the SNPs significantly (P < 0.01) associated/linked with MCR or rAUDPC, AUDPC or PM. The SNPs were plotted on both axes and the seven loci are shown on top of the map. SNPs within locus are framed. The loci are arranged in order according to their map position and the chromosome position is shown below the map. LD blocks are shown at the right by a – e.

Table. 2.10. Significantly associated markers, allele frequencies and allele effects

Locus	Chr.	Marker	Allele	Allele Frequency (%)	MCR P_value (R^2)	rAUDPC P_value (R^2)	PM P_value (R^2)
Delta(7)-sterol-C5(6)-desaturase (DSD)	II	DSD_SNP81	A/T	75/25	0.0034 (0.06)↓/↑	NS	0.0097 (0.04)↑/↓
		DSD_SNP160	G/A	47/53	NS	NS	2.52E-04 (0.10)↓/↑
		DSD_SNP248	T/G	64/36	0.0036 (0.09)↓/↑	NS	7.29E-05 (0.12)↑/↓
Pectin methyl esterase (PME)	III	PME_SNP82	C/A	84/16	0.0021 (0.07)↓/↑	1.30E-04 (0.09)↓/↑	NS
Biotin carboxylase carrier protein (BCCP)	V	BCCP_SNP60	T/C	88/12	1.35E-04 (0.12)↓/↑	0.0018 (0.09)↓/↑	NS
		BCCP_SNP321	T/C	78/22	3.27E-05 (0.14)↓/↑	1.48E-04 (0.12)↓/↑	NS
Chloroplast protease (Chp)	VII	Chp_SNP86	C/T	51/49	0.0087 (0.08)[a]	8.31E-04 (0.11)a	NS
		Chp_SNP418	C/A	51/49	NS	0.0054 (0.09)a	NS
		Chp_SNP451	G/C	47/53	NS	0.0053 (0.09)a	NS
		Chp_SNP452	T/C	52/48	NS	0.0084 (0.09)a	NS
Avr9/Cf-9 rapidly elicited protein 20 (AC9)	X	AVR9_SNP370	G/C	82/18	0.0018 (0.08)↓/↑	NS	NS
Magnesium-protoporphyrin IX monomethyl ester cyclase (desaturase) (MPP)	X	MPP _SNP286	G/C	68/32	0.0036 (0.09)↓/↑	7.12E-04 (0.11)↓/↑	NS
Asparagine synthetase (AspS)	VI	Asp_SNP162	T/C	75/25	NS	NS	0.0039 (0.09)↓/↑

Chr. = Chromosome; Marker effect shown by arrow, ↑ show positive effect (higher resistance and earliness); ↓ show negative effect (susceptibility and late maturity); NS = not significant (P > 0.01); [a]the direction of the allele effect unclear.

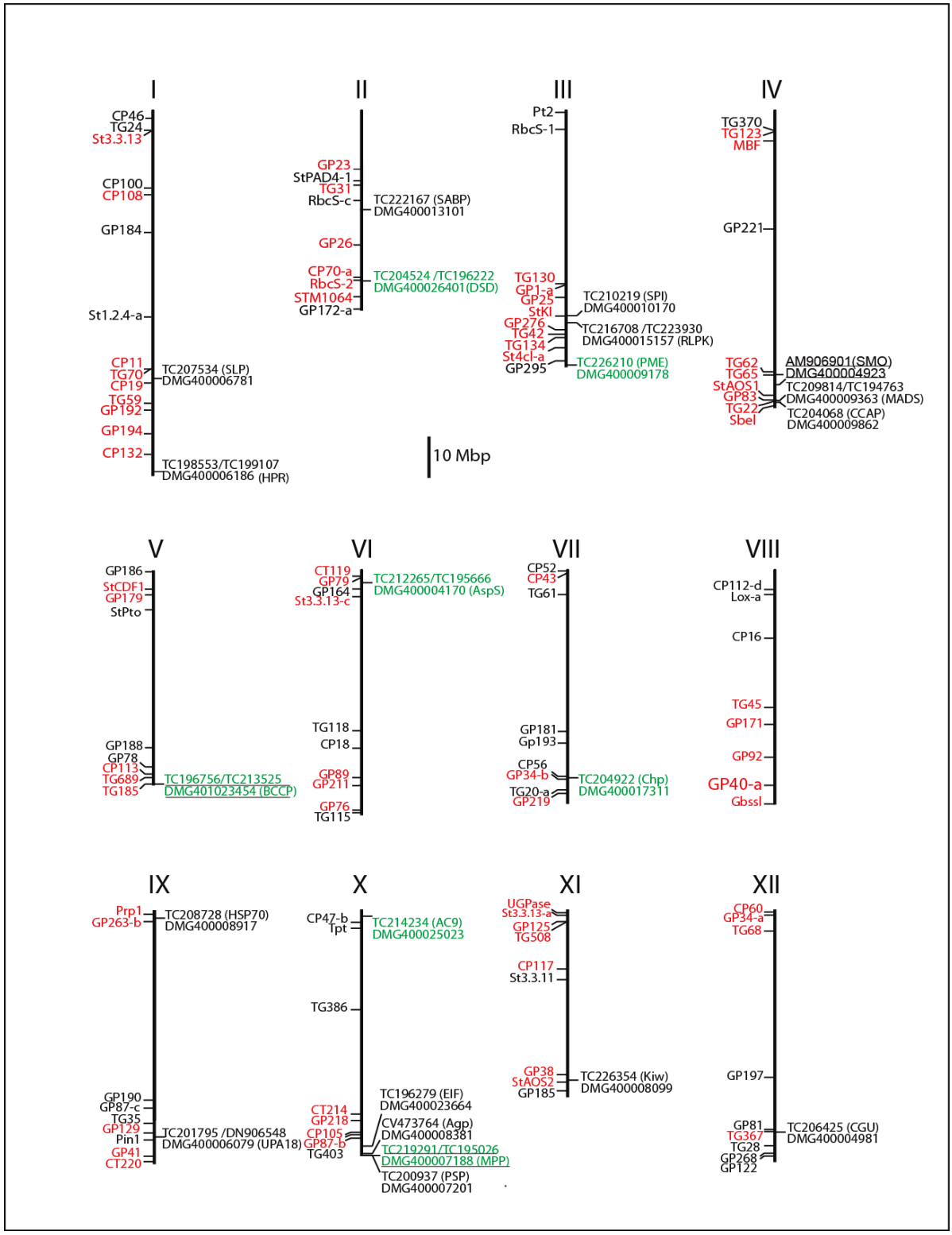


Fig. 2.24. Physical map of candidate genes. RFLP markers linked to resistance QTLs (Gebhardt, 2013) are shown in red and additional chromosome anchor markers in black, at the left. Candidate locus name, matching transcript number and annotation acronyms in parentheses are shown at the right. Candidates that showed reproducible differential expression are underlined. Candidates in which one or more SNP markers were associated/linked with any of the three traits (MCR, rAUDPC and PM) are shown in green. For Pubq, the position is not known and not shown on the map.

2.4. Discussion

The comparative transcriptome profiling has been demonstrated as a powerful tool to identify candidate QTLs and diagnostic DNA markers for late blight resistance in potato (Gyetvai et al., 2012; Draffehn et al., 2013; Thümmler and Wenzel, 2005). A gene's differential expression between genotypes contrasting for a certain trait could be an indication for a role in governing the trait (Kloosterman et al., 2010) or it could also be due to the fact that the gene is located physically near the QTL and therefore the association with the QTL is indirect (Kloosterman et al., 2010). Thus, differential expression is a promising approach to select candidate genes linked or associated with QTLs governing the variation of the trait.

In this study we selected novel candidate genes differentially expressed between quantitative resistant versus susceptible potato genotype pools and validated the differential expression in independent *P. infestans* infection tests and by different transcript quantification methods. Validated novel candidate genes were further tested for linkage or association with the aim to identify genes for quantitative resistance to *P. infestans* not compromised by late maturity and to discover diagnostic SNP markers.

2.4.1. Tetraploid potato genotypes showed stable resistance to *P. infestans*

Correct identification of resistant and susceptible genotypes is a crucial step for the identification of differentially regulated candidate genes, hence accurate assessment of genotypes for their response to *P. infestans* infection was required. Evaluating the genotypes under different environments and controlled conditions as well as using different evaluation methods increases the accuracy. This is because the resistance response might vary in different environments (Bradshaw, 2007). Furthermore, each evaluation method has its own advantages and disadvantages (Simko et al., 2007). A major advantage of infection under controlled conditions is that the environment can be controlled to optimize disease pressure while minimizing the effects of confounding biotic and abiotic factors. However the controlled environment may not adequately mimic the complexities of a field infection (Simko et al., 2007).

Two groups of tetraploid potato genotypes contrasting for MCR and with early to mid maturity dates were selected with the aim to identify candidate genes differentially expressed

between the contrasting groups. The genotypes were first selected based on a previous field evaluation and were further evaluated in a field (in a new environment) under natural late blight infestation and under controlled conditions by infecting with complex isolates of *P. infestans*.

Overall, the result showed that most of the genotypes had a stable phenotype in different environments. This result is in line with previous reports (Wulff et al., 2007; Forbes et al., 2005), which showed that quantitative resistance to *P. infestans* in potato is stable under field conditions in multiple years. Our result also showed that the controlled condition mostly mimicked the field conditions. However, SL028 showed instable resistance response between field and controlled conditions. This genotype was susceptible according to the field evaluation, but reacted resistant under controlled infection tests. The resistant/susceptibility in this genotype might vary according to the environmental conditions.

2.4.2. The *Phytophthora* ribosomal genes are promising candidates to monitor infection and disease progression of *P. infestans* in potato

Resistant and susceptibility are measured by the growth of the pathogen on the plant. The measurement can be based on disease symptom expression or quantifying pathogen population on the plant (Simko et al., 2007). Disease symptom expression method is inexpensive; however, it is less accurate as different pathogens and abiotic factors cause similar symptoms. In addition, the method is subjective as the judgment varies from person to person.

Five *P. infestans* tags with their tag count in SuperSAGE analysis corresponding to the resistant levels of three potato genotype groups (A1, B2 and A2) were selected. Most were detected by standard PCR, showing the efficiency of TRIzol RNA isolation kit in extracting RNA simultaneously both from *Phytophthora* and potato in infected sample. However, only three of them showed amplification corresponding to the infection time points. In further qRT-PCR analysis the 60S ribosomal protein L23a and the 40S ribosomal protein S9-1 showed *Phytophthora* quantification correlated with the progression of *P. infestans* within two or three days after inoculation, showing the reproducibility of their tag count in SuperSAGE by qRT-PCR. In large scale application, 60S ribosomal protein L23a

quantification better discriminated the resistant and susceptible genotypes as compared to the phenotypic evaluation (comparison of Fig. 2.7 and Fig. 2.8).

These two genes, 60S ribosomal protein L23a and 40S ribosomal protein S9-1, may be useful in plant breeding programs for comparisons of cultivars with different levels of quantitative resistance. They can also be used as internal control to monitor the infection and disease progression during the initial biotrophic phase of the compatible interaction, when no disease symptom was visible (Draffehn et al., 2013).

2.4.3. Some candidate genes showed reproducible differential expression

Candidate genes were selected based on differential expression between quantitatively more resistant (A2) and susceptible (B2 and A1) genotype pools from transcriptome data generated by SuperSAGE analysis (Draffehn et al., 2013). The pedigree and genotype at the *StAOS2* locus of the three potato genotype groups was explained in Draffehn et al. (2013). A total of 165 uni-tag that were differentially expressed in at least five of six comparisons (three genotype pools each at T0, T1 and T2) selected. Due to time limit, only 22 prioritized candidates were further analyzed in this study. The 22 selected candidates were subjected to validation of differential expression by qRT-PCR and allele specific pyrosequencing.

In qRT-PCR, four candidates, namely Squalene Monooxygenase (SMO), Polyubiquitin (Pubq), Avr9/Cf-9 rapidly elicited protein 20 (AC9) and Pectin Methylesterase (PME), showed reproducible differential expression in four different pools and in three biological replicates as well as compared to the SuperSAGE expression analysis.

SMO and PME were highly expressed in susceptible genotype pools. SMO was down regulated upon infection while PME was up regulated upon infection. Their high expression in susceptible pools might reflect a role in susceptibility to *P. infestans*. There is evidence that plants possess susceptibility genes that attract pathogens, enable their entry into the host and facilitate nutrient provision (Lapin and Ackerveke, 2013; Hückelhoven et al., 2013). These host susceptibility genes might be a target for manipulation by pathogen effectors. In this line, down regulation of SMO upon infection might indicate that it was targeted by *P. infestans* effectors. *P. infestans* manipulate host defenses mainly using RXLR effectors (McLellan et al., 2013). However, it has also been hypothesized that plants possess the

susceptibility genes as negative regulators of immune response to avoid autoimmunity and unnecessary investment into defense in environments with little disease pressure (Hückelhoven et al., 2013). SMO is an important enzyme in sterol biosynthesis and catalyses the conversion of squalene into 2,3-oxidosqualene, the precursor of cyclic triterpenoids (Rasbery et al., 2007; Pose et al., 2009). Sterols are isoprenoid-derived lipids that play essential roles in plant growth and development and are integral components of the membrane lipid bilayer, playing a role in membrane permeability and fluidity. The homologus Arabidopsis mutant, *dry2/sqe1-5*, reveals a central role for sterols in drought tolerance and regulation of reactive oxygen species. In addition, the *dry2/ sqe1-5* mutant showed developmental defects, including altered root architecture, root hairs, diminished shoot size and chlorophyll content (Pose et al., 2009). There is evidence for a role of PME in susceptibility (Ma et al., 2013), which is discussed in the next section under association mapping (2.4.5.4). Another example for a susceptibility gene is UPA (UPregulated by AvrBs3), which is induced by the AvrBs3 effector of *Xanthomonas campestris* pv. *vesicatoria* (Xcv) upon infection of pepper (Kay et al. 2009; Kay et al. 2007). AvrBs3 binds to a conserved element in the pepper UPA20 promoter via its central repeat region and induces gene expression through its activation domain which then induces hypertrophy of plant mesophyll cells in susceptible pepper plants. The hypertrophy probably facilitates bacterial release from infected plant tissue at later infection stages (Kay et al., 2007). In our study, UPA18 showed reproducible differential allele expression in SuperSAGE and pyrosequencing analysis and the SNP with differential allele expression was nonsynonymous substituting isoleucine (I) by methionine (M). Unfortunately, no scorable sequence was found from this locus and was not further analyzed in association mapping.

Polyubiquitin (Pubq) was highly expressed in resistant genotype pools and was up regulated upon infection. Ubiquitin is a regulatory protein that acts through its post-translational attachment (ubiquitination) to other proteins, which alters the function, location or trafficking of the protein, or targets it for degradation by the 26S proteasome (Trujillo and Shirasu, 2010). There is also evidence that the ubiquitination regulatory system plays a role in plant defenses against pathogens (Trujillo and Shirasu, 2010; Marino et al., 2013). This candidate was the second best reproducible for its differential expression after SMO.

There was higher reproducibility of differential expression patterns before and after infection than between resistant and susceptible genotype pools, which is in line with the observations of Draffehn et al. (2013). Of 14 candidates tested in qRT-PCR, eleven showed a reproducible expression pattern before and after infection. Pubq and PME were up regulated upon infection while SMO and AC9 were down regulated upon infection.

Eight candidates detected as allele specific transcripts by SuperSAGE were tested by pyrosequencing for the reproducibility of differential allele specific expression. Most of the candidates showed reproducible differential allele expression compared to the SuperSAGE expression as well as in three biological replicates. Especially alleles of delta(7)-sterol-C5(6)-desaturase (DSD), magnesium-protoporphyrin-IX-monomethyl-ester cyclase (MPP), hydroxypyruvate reductase (HPR) and UPA18 were consistently differentially expressed even in four different pooling strategies. Allele specific expression of asparagine synthetase (AspS) and receptor-like protein kinase (RLPK) were also consistently reproducible at least in two of the pooling strategies. Biotin carboxylase carrier protein (BCCP) showed reproducible allele specific expression mainly in pools_4, which were constructed based on the *StAOS2* marker, which might indicate that the two genes are found in the same biosynthesis pathway. Both *allene oxide synthase* and *biotin carboxylase carrier* genes are found in fatty acid biosynthesis pathway (Pajerowska-Mukhtar et al., 2009; Nikolau et al., 2003).

As pyrosequencing gives the expression of one allele relative to the other, it was difficult to compare the candidate's expression before and after infection. However, the resistant allele of AspS was consistently up regulated upon infection while DSD allele expression seemed constitutive.

Overall, better reproducibility of differential expression was found with pyrosequencing than with qRT-PCR. In pyrosequencing, a region of about 100 to 300 bp is first amplified by standard PCR, then a small portion (20 to 30 bp) containing the SNP is sequenced from the single stranded template using a third sequencing primer. These further steps could have made pyrosequencing more specific than qRT-PCR. As stated by Draffehn et al. (2013), sensitivity and resolution of qRT-PCR might be insufficient for validation of quantitative differences between low expressed, allele specific transcripts detected by SuperSAGE.

However, the comparison is not accurate as we did not test the expression of the same candidates from the same sample. Pyrosequencing quantifies relative expression of alleles while qRT-PCR enables the analysis of differential expression of genes. In addition, prior knowledge of the SNP (alleles) information is required for pyrosequencing while it is not required when qRT-PCR is used.

2.4.4. Linkage mapping identified a QTL with minor allele frequency

The cultivated potato is a highly heterozygous autotetraploid species with a complex genetic inheritance as inheritance of traits is often masked by multiple alleles and tetrasomic segregation. Diploid potatoes suffer from self incompatibility, complicating more the mapping of genes in potato. Therefore, in potato there are no F2, recombinant inbreed lines (RILs), isogenic lines (NILs) mapping populations. Instead, the base for constructing molecular linkage maps in potato is an F1 or backcross progeny of partially heterozygous parents, which segregate for a sufficient number of DNA polymorphisms (Gebhardt, 2007).

In this study, we used hundred and eleven F1 heterozygous, tetraploid potato clones from two half sib families. A total of 101 SNP markers from seven loci were collected by genotyping in the F1 clones. A single marker-trait analysis was employed in general linear model (GLM) to detect QTLs linked with AUDPC, MCR and PM. 13 markers from three loci (SPI, AspS and DSD) for AUDPC; 5 markers from two loci (AspS and DSD) for MCR were found to be significantly linked with the respective traits. The range of variation explained by the significant markers was 8 to 14% in AUDPC and 13 to 30 % in MCR. None of the markers were linked with plant maturity.

SPI was selected based on its differential expression both in SuperSAGE as well as in RNA seq (Gebhardt, personal communication) analysis. This candidate was one of the Kunitz-type (KTI) protease inhibitors analyzed by Odeny et al. (2010). As shown in Fig. 2.24, this candidate co-localized with the *StKI* locus. Proteinase inhibitors (PIs) are defense-related proteins often present in seeds and induced in certain plant tissues by herbivory or wounding. Potato contains abundant levels of diverse PIs. Most potato Kunitz-type inhibitor genes map to the *StKI* locus on potato chromosome III, which is linked to a quantitative trait locus (QTL) for resistance to *P. infestans* (Odeny et al., 2010).

DSD was the second best linked candidate after AspS and was one of the candidates that showed contrasting differential allele expression. More interestingly, the SNP that showed contrasting allele expression was also significantly linked with MCR, which makes this candidate one of the most promising markers for MCR to *P. infestans*.

AspS was the best linked candidate in linkage analysis. It was one of the candidates that showed contrasting allele differential expression in SuperSAGE which was reproducible in pyrosequencing. It is located at the distal part of chromosome VI, co-localizing with QTLs for late blight resistance. However, the most significantly linked SNP was not detected in association mapping. A marker found to be linked in linkage analysis can lose its linkage with QTLs in diverse population due to repeated recombination events during meiosis. However, this is not so frequent in potato as potato is propagated vegetatively, which reduces the number of meiotic generations (Gebhardt et al., 2004). The allele linked with resistance was rare in CONQUEST2 population, which was present only in 6 (4 triplex and 2 simplex) out of the 184 genotypes. Of which, 5 genotypes with 4 triplex and one simplex resistant alleles were quantitatively more resistant with MCR ranging from -0.03 to -021. The result might show the advantage of linkage mapping over association mapping in detecting QTLs with rare alleles. A difference between association mapping in a multiparental population and genetic linkage mapping in a biparental population is that association mapping generally identifies the association of common alleles as rare alleles do not reach statistical significance, whereas a population originating from a biparental cross enables the identification of rare alleles in the population at large (Simko et al., 2007).

AspS is a key enzyme in the production of the nitrogen-rich amino acid asparagines and is required for plant nitrogen assimilation and defense responses to microbial pathogens. Hwang et al., (2011) reported that *Capsicum annuum* asparagine synthetase1 (CaAS1) was induced rapidly and strongly in pepper leaves by *X. campestris* pv. *vesicatoria* (Xcv) infection. They also showed that silencing of CaAS1in pepper plants resulted in enhanced susceptibility to Xcv infection. They further showed transgenic arabidopsis plants that overexpressed CaAS1 exhibited enhanced resistance to *Pseudomonas syringae* pv. *tomato* DC3000 and to the oomycete pathogen *Hyaloperonospora arabidopsidis*.

2.4.5. Differentially expressed candidates associated with MCR to *P. infestans*

Compared to the conventional linkage mapping, association mapping is a powerful high-resolution mapping tool for complex quantitative traits. Hence, most of the candidate genes that showed reproducible differential expression were analyzed for association with rAUDPC, MCR and PM using a panel of 184 tetraploid potato clones. Association mapping was carried out using a mixed linear model (MLM), which takes into account population structure and kinship to avoid false positives. Thirteen candidates were selected for association mapping based on their reproducible differential expression. A total of 114 SNP markers were scored from eleven loci. There were no scorable SNPs from the remaining two candidates, Pubq and UPA18. The sequence from these two loci had background noise due to presence of indels, which causes frame shifts in the sequence trace file. Almost all scored SNPs were bi-allelic; only two tri-allelic markers were scored from the *RLPK* locus. Significantly associated markers are discussed below;

2.4.5.1. *Biotin-carboxylase-carrier-protein (BCCP)*

In BCCP, two SNP markers were significantly associated with MCR and rAUDPC. The genetic variance explained by these two markers was 14 % and 12 % of MCR and 9 % and 12 % of rAUDPC. None of the markers was associated with PM, which make this candidate a promising marker for *P. infestans* resistance not compromised by late plant maturity. Both markers caused an amino acid substitution and both had the C/T SNPs with both the C nucleotides encode serine while both the T nucleotides encode proline. Amino acid substitution might show their functional relevance in resistance. As stated by Pajerowska-Mukhtar et al. (2008), nonsynonymous amino acid substitution could possibly change the enzyme's substrate affinity and other kinetic properties and could lead to variable quantitative allele expression that might influence quantitative defense responses in a concentration dependent manner.

The candidate is located on the south arm of chromosome V and co-localized with QTLs for *Phytophthora* resistance. BCCP is a subunit in a heteromeric complex of the plastid acetyl-CoA carboxylase, an enzyme that catalyzes the first committed step of fatty acid synthesis (Nikolau et al., 2003). Hence, being in fatty acid biosynthesis it might play role in Jasmonic

acid signaling pathway, which is the major signaling pathway for *P. infestans* resistance in potato (Pajerowska-Mukhtar et. al., 2009).

2.4.5.2. *Delta(7)-sterol-C5(6)-desaturase (DSD)*

Two SNP markers from DSD were significantly associated with MCR and explained 6 % and 9 % of the genetic variance. However, from this locus the same two markers and one additional marker were associated with PM explaining 4 %, 10 % and 11 % of the genetic variance, respectively. This might show that MCR still has some association with PM, though there is no correlation between the two traits (Pajerowska-Mukhtar et al., 2009). The association of both MCR and PM with the same marker can be due to either pleiotropic effect or physically linked marker control the two traits. As shown in the previous section, this locus had reproducible contrasting allele expression in SuperSAGE and pyrosequencing analysis and the SNP with contrasting differential expression was found to be significantly linked with MCR in linkage analysis. This means that the candidate was validated for its association with MCR by differential expression as well as statistically by linkage and association mapping. It also co-localized with *RbcS-2* locus at the long arm of chromosome II. RbcS locus is linked with the Rubisco- small Subunit , which plays a vital role in resistance to tobamoviruses (Zhao et al., 2012). Thus, this candidate is one of the potential markers for quantitative resistance to *P. infestans* in potato.

DSD function in the sterol biosynthesis pathway, which is involved in many biological processes, by acting as signaling molecules in the cell cycle both in plants and animals. Sterols are crucial lipid components that regulate membrane permeability and fluidity and are the precursors of bioactive steroids (Silvestro et al., 2013). In Arabidopsis, mutants of two alleles of this gene (*dwf7* and *STE1*) cause a dwarfing phenotype (Choe et al., 1999).

2.4.5.3. *Desaturase or Magnesium-protoporphyrin IX monomethyl ester cyclase (MPP)*

This candidate was annotated as desaturase or Magnesium-protoporphyrin IX monomethyl ester [oxidative] cyclase, (chloroplast) in PGSC. The latter annotation seems more correct as the same annotation was found by blasting the sequence in tomato and arabidopsis genome browsers. Although different primer pairs were tested for amplicon sequencing, only 4 SNP markers were scored due unreadable sequences probably due to indels that caused frame shifts and overlapping of peaks in the sequence trace file. Presence of one Indel every 250 to

300 bps has been reported in potato genome (Xu et al., 2011). Nevertheless, one SNP marker was significantly associated both with MCR and rAUDPC and explained 9 % and 11 % of the genetic variation of the traits, respectively. This candidate is located on chromosome X and co-localized with QTLs for pathogen resistance. MPP is an enzyme involved in chlorophyll biosynthesis (Peter et al., 2010), and in absicic acid signaling (Wu et al., 2009).

2.4.5.4. *Pectin methyl esterase (PME)*

In PME, one SNP marker was significantly associated with MCR and rAUDPC and explained 7 % and 9 % of the genetic variation of MCR and rAUDPC, respectively. PME was expressed at higher level in quantitatively more susceptible than in resistant pools. No or very small differences were observed before infection. High differential expression was observed only after infection (at T2 and T3). This result is in line with the findings of Ma et al., (2013) who showed a significant increase in PME expression in susceptible but not in resistant banana cultivars when plants were wounded and attacked by *Fusarium oxysporum* f.sp. *cubense*. Lionetti et al (2007) reported that the over expression of PME-inhibitor in arabidopsis reduced the activity and expression level of PME and resulted in higher resistance to *Botrytis cinerea.* PMEs catalyzes the demethylation of pectin, which is one of the main components of plant cell wall and plays a key role in providing a barrier against environmental stresses and pathogen attacks. The action of PME makes pectin susceptible to hydrolysis by pathogen enzymes such as endopolygalacturonases (An et al., 2008).

2.4.5.5. *Avr9/Cf-9 rapidly elicited protein 20 (AC9)*

AC9 genes are well known resistance genes in tomato. As the name indicates, the Cf genes in tomato confer resistance to race 9 of the fungal pathogen *Cladosporium fulvum* (Cf-9) through recognition of secreted avirulence (Avr9) peptides (Rowland et al., 2004). The *Cf* genes are immediately activated (in 15 to 30 min) upon perception of the pathogen Avr protein and induce rapid production of ROS (reactive oxygen species) and activation of MAP kinases and calcium-dependent protein kinases (Rowland et al., 2004). In line with this report, AC9 was highly expressed after infection at T1 and T2 in SuperSAGE. In qRT-PCR, this candidate was one of the best candidates that showed reproducible differential expression. But this up regulation upon infection was not reproducible, probably due to environmental impact on induction or suppression of this gene upon infection. In association

analysis, one SNP marker was significantly associated with MCR and explained 8 % of the genetic variation.

2.4.5.6. *Chloroplast protease (Chp)*

Out of 8 SNP markers scored in the *Chp* locus, 4 were significantly associated with rAUDPC explaining 9 % to 11 % of the genetic variation. But only one marker was associated with MCR and 8 % of the genetic variation was explained by this marker. The homologue of this gene in tomato and Arabidopsis is annotated as ATP-dependent Zn protease (FtsH). FtsH protease is important in chloroplast biogenesis and thylakoid maintenance (Chi et al., 2012).

2.5. Conclusions and outlook

Eight SNPs from six differentially expressed candidate genes were associated with MCR. Five SNPs of those were also associated with rAUDPC. Only 2 of the SNPs were associated with PM. All the positive alleles had minor allele frequencies. Their allele frequency need to be increased in potato cultivars by breeding to improve resistance to *P. infestans*. The SNPs from different loci were independent, therefore six (taking one SNP per locus) independent SNPs diagnostic to MCR were identified. The genetic variation explained by these six SNPs ranged from 6 % to 14 %. Similarly, 4 independent SNPs from 4 loci diagnostic to rAUDPC were identified, with a range of 9 % to 12 % explained genetic variation. Therefore, these SNPs can be used as diagnostic markers for resistance to *P. infestans*. However, it is advisable to test them in multiple populations with different genetic backgrounds (Gebhardt, 2013). The associated candidate genes need to be further functionally characterized to determine whether they are the causal gene for MCR. The method for functional characterization in potato was already demonstrated for the *StAOS2* allele by dsRNAi mediated silencing and also by complementation in mutant backgrounds of *Arabidopsis thaliana* (Pajerowska-Mukhtar et al., 2009). Based on the results found in this study, it can be inferred that most of the candidate genes that showed reproducible differential expression were either directly or indirectly involved in conferring or contributing towards resistance responses against *P. infestans*. Therefore, comparative transcript profiling combined with association mapping can be used to identify QTLs and diagnostic DNA markers to be used by breeders in marker-assisted resistance breeding.

3. Genome wide association mapping for maturity corrected resistance to *Phytophthora infestans* in potato using SNP markers from high-density genotyping SolCAP SNP array

3.1. Introduction

Genome wide association (GWA) mapping is conducted by scanning high density DNA variants (DNA markers), such as single nucleotide polymorphisms (SNPs), more or else evenly distributed over the genome (Stich et al., 2013; Soto-Cerda and Cloutier, 2012). The genotyped DNA markers are analyzed statistically for association with phenotypic traits to find markers linked with a quantitative trait locus (QTL) contributing to the trait (Stich et al., 2008; Yu et al., 2006). Association mapping is based on linkage disequilibrium (LD), which is a non random association between loci, fore example between SNP markers and causal QTL (Flint-Garcia et al., 2003). Opposed to conventional linkage mapping, in which the LD is generated by crossing two parental lines, association mapping exploits the existing LD in a population that has undergone historical recombination events in the past (Soto-Cerda and Cloutier, 2012). Hence, association mapping uses existing collections of genotypes routinely developed in breeding programs (Malosetti et al., 2007; Gebhardt et al., 2004). This makes association mapping more applicable to a much wider germplasm base as compared to conventional linkage mapping. In addition, the high number of historical recombination events gives association mapping high resolution power (Soto-Cerda and Cloutier, 2012; Xu et al. 2011).

However, the absence of a design behind the populations has consequences for the analysis. In such types of populations, significant marker trait association may or may not be the consequence of physical linkage between markers and QTLs, leading to false positives or spurious associations (Stich et al., 2008; Yu et al., 2006). A major cause of false positives is the genetic correlation between individuals resulting from the heterogeneous genetic relatedness between them (Malosetti et al., 2007; Pritchard et al., 2000).

The cultivated potato is a highly heterozygous autotetraploid species with a complex genetic inheritance. Loci can have multiple alleles and show tetrasomic segregation. Diploid potatoes

suffer from self incompatibility, which complicates the use of conventional linkage mapping in potato (Gebhardt, 2007). Therefore, association mapping is more appropriate for gene mapping in potato (Gebhardt et al., 2014; Gebhardt, 2013).

Although some association mapping experiments have been performed so far in potato, most of them were based on the candidate gene approach (Urbany et al., 2011; Achenbach et al., 2009; Pajerowska- Mukhtar et al., 2009; Malosetti et al., 2007; Li et al., 2005; Gebhardt et al., 2004). The only genome wide association mapping on potato reported so far was by using amplified fragment length polymorphism (AFLP) markers (D'hoop et al., 2014; D'hoop et al., 2008). However, AFLP markers do not distribute genome wide, rather cluster at centromers and show local gaps. In addition, AFLP markers are almost exclusively dominant which introduces problems when using them for estimating population structure or for direct use in mapping (Stich et al., 2013). Moreover, AFLPs are anonymous and their position on the physical map cannot be easily inferred (Gebhardt et al., 2014; Stich et al., 2013). Now a days SNP markers are widely used because of their abundance, simplicity and genome wide distribution. Furthermore, SNP alleles are co-dominant and the allele dosage in polyploid species, like tetraploid potato, can be estimated (Stich et al., 2013).

In this study, the SolCAP potato genotyping array was used to generate genome wide SNP markers on the CONQUEST2 population. The SolCAP potato genotyping array was originally developed from five North American (Atlantic, Premier Russet, Snowden, Kennebec, Shepody) and one European (Bintje) potato variety and contains a total of 8303 SNP markers (Hamilton et al., 2011). The objectives of this study were to analyze 1) population structure 2) genome wide LD and LD decay, and 3) to identify QTLs and diagnostic SNP markers for *Phytophthora infestans* resistance not compromised by late plant maturity.

3.2. Materials and methods

3.2.1. Potato genotypes

The association mapping population 'CONQUEST2' that contained 184 tetraploid potato clones was used in this study. The genotypes were phenotyped for relative area under the disease progress curve (rAUDPC) and plant maturity (PM) in replicated *Phytophthora infestans* infection tests. Adjusted means were calculated as described by Pajerowska-Mukhtar et al. (2009). Maturity corrected resistance (MCR) was calculated by regressing rAUDPC against PM.

3.2.2. Genotyping

DNA was extracted from leaf samples of each of the 184 tetraploid potato genotypes as described in chapter I (Section 2.2.1.1). The genotypes were custom genotyped for 8303 SNPs using the Illumina SolCAP SNP potato genotyping array (Hamilton et al., 2011) at the Life & Brain Center (Department of Genomics, Bonn, Germany). For each bi-allelic SNP, one of the five possible genotypes (nulliplex = AAAA, simplex = AAAB, duplex = AABB triplex = ABBB and quadraplex = BBBB) was assigned using the R package fitTetra (Voorrips et al., 2011).

3.2.3. Linkage disequilibrium (LD) and LD decay

Linkage disequilibrium (LD) and LD decay analysis was done by Jinquan Li (quantitative crop genetics group, Max Planck Institute for Plant Breeding Research, Köln) according to the methods used by Stich et al. (2013). LD between two loci was analyzed by measuring r^2 (squared correlation coefficient). LD decay was investigated by plotting r^2 against physical distance (Stich et al., 2013; Achenbach et al., 2009) using R statistical software.

3.2.4. Kinship

From the SolCAP SNPs, a subset of 241 polymorphic SNPs that were uniformly distributed on the 12 potato chromosomes were selected to estimate the relative kinship in the CONQUEST2 population. The selected SNPs had a minor allele frequency (MAF) of ≥ 10 % and no missing data. Relative kinship between pairs of genotypes was analysed using

SPAGeDi software (Hardy and Vekemans, 2002) according to Ritland (1996). Negative kinship values between genotypes were automatically set to zero.

3.2.5. Population structure

A Bayesian clustering approach implemented in STRUCTURE (Pritchard et al., 2000) software was used to assess the presence of population structure in the CONQUEST2 population. The selected 241 SNPs were used for the analysis. Burn-in time as well as iteration number was set to 100, 000 with 10 repetitions, testing the probability of 20 subpopulations. An admixture model with correlated allele frequencies was used. The results of the run were uploaded to the online software, Structure Harvester (Earl and VonHoldt, 2012), at http://taylor0.biology.ucla.edu/structureHarvester/ and the most likely number of subpopulations was determined by the log likelihood (Rosenberg et al., 2001) combined with the Evanno method (Evanno et al., 2005).

In addition, population structure in CONQUEST2 was analyzed by principal component analysis (PCA). It was performed using the 241 SNP markers in R using the function *prcomp ( )*.

3.2.6. Statistical analysis

Marker-trait association mapping was done using TASSEL2.1 (Bradbury et al., 2007) in mixed linear model (MLM) which takes into account population structure (Q) and kinship (K), as described in chapter I (Section 2.2.12.3). P_value significance threshold level calculated by Joao Paulo (Wageningen University, The Netherlands) was used. It was calculated based on M_{eff} (effective number of independent tests) according to Li and Ji (2005). The result was compared with three other statistical models; simple analysis of variance (ANOVA), using only Kinship (K) to avoid false positive (K_model) and using only population structure (Q) to avoid false positive (Q_model).

3.3. Results

3.3.1. SolCAP SNP markers in the CONQUEST2 population

The genotyping of 184 tetraploid potato clones with 8303 SNPs of the SolCAP SNP array produced a total of 6286 bi-allelic informative SNPs, which was about 76 %. This was a fairly good result as the SNP array was originally prepared based on SNPs in one European and five North American potato varieties. A similar result was reported by Stich et al., (2013). The distribution of the SNPs across the 12 potato chromosomes is shown in Fig. 3.1. About 9 % of the SNPs were without chromosome position information and about 1.5 % were without both chromosome and position (within chromosome) information. The SNPs with unknown position are indicated by X in Fig. 3.1. Of the 6286 SNPs, 211 duplicated SNPs were discarded. About 2 % of the SNP markers had a minor allele frequency (MAF) of less than 0.05 and 6 % had missing value above 25 %. The remaining about 5600 SNP markers were used for further analysis.

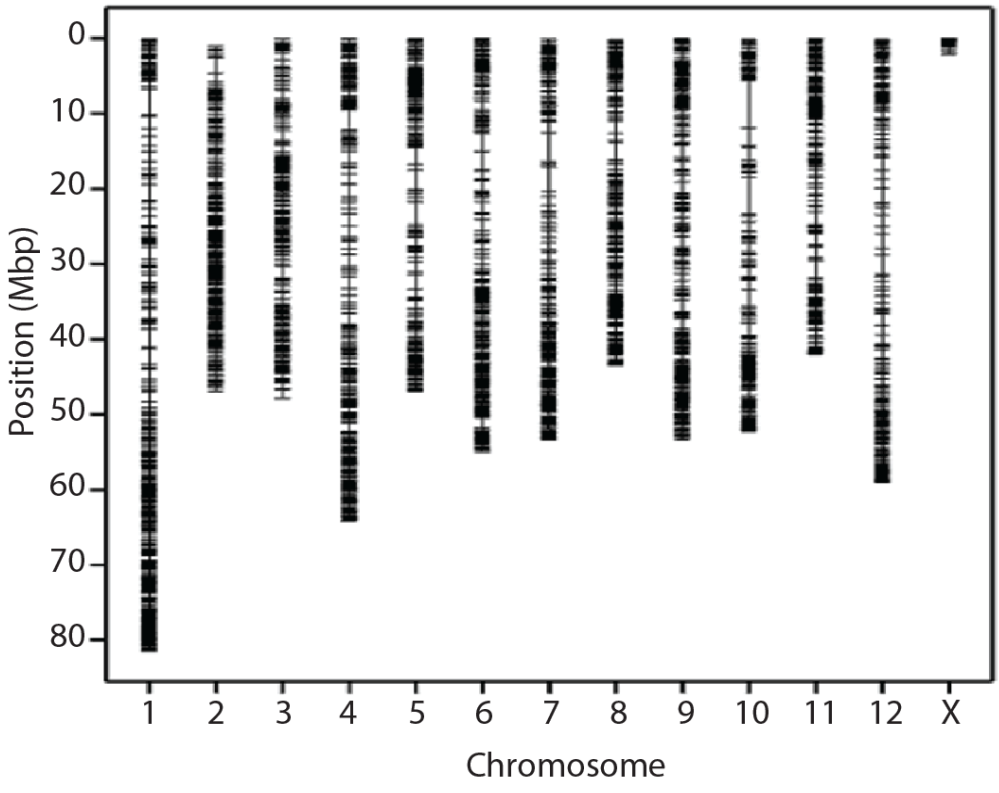


Fig. 3.1. Physical position of the SolCAP SNP markers across the 12 potato chromosomes. Markers without position information are indicated by X.

3.3.2. Relative Kinship/relatedness between genotypes in the CONQUEST2 population

Pair wise kinship coefficients were calculated according to Ritland (1996) using 241 SNPs uniformly distributed on the 12 potato chromosomes. The kinship coefficient reflects the degree of identity or relatedness between two genotypes. The pair wise kinship coefficient ranged from 0 to 0.3. About 53 % of the pair wise kinship coefficients were 0 and about 86 % were less than 0.03 (Fig. 3.2) showing that the 184 tetraploid potato genotypes were highly unrelated. Only about 1 % of the pair wise coefficients were above 0.1.

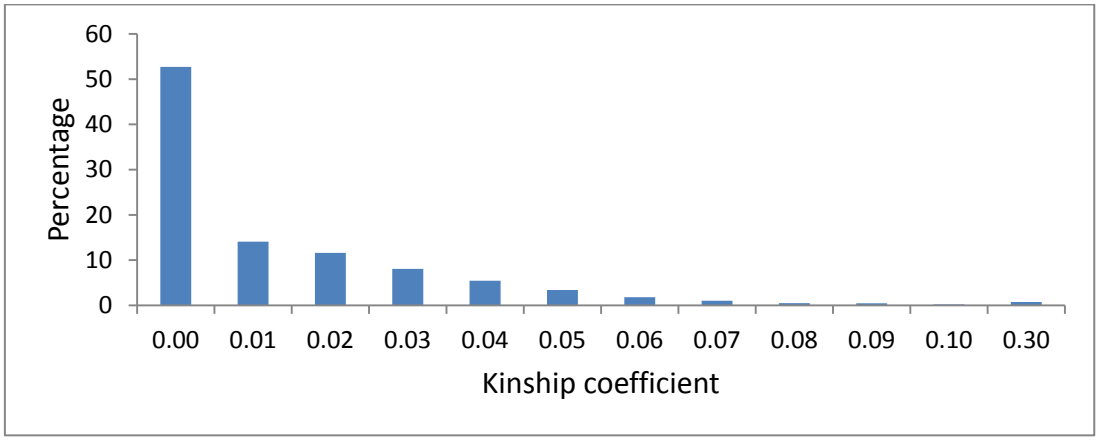


Fig. 3.2. Frequency distribution of pair wise kinship coefficient values in the CONQUEST2 population. The kinship coefficient was estimated using 241 polymorphic SNPs that were uniformly distributed on the 12 potato chromosomes.

3.3.3. Population structure in the CONQUEST2 population

Presence of sub populations in CONQUEST2 was analyzed with STRUCTURE (Pritchard et al., 2000) software using 241 polymorphic SNPs. The SNPs score was in dosage dependent manner. As showed in Fig 3.3b, the largest likelihood change (Rosenberg et al., 2001) was observed when K increased from one to two. In addition, the Δk (Evanno et al., 2005) showed a clear inflection point at K = 2 (Fig. 3.3a). Therefore, K = 2 (two subpopulation) was considered as the possible number of sub populations in CONQUEST2. The genotypes from the two breeding companies (Böhm-Nordkartoffel Agrarproduktion (BNA) and Pflanzenzucht GmbH & Co. Ka (SAKA)) were mixed and both were distributed in the two subpopulations inferred by STRUCTURE. The placement of each genotype into its corresponding inferred cluster (G1 or G2), for K =2 is shown by a plot in Fig. 3.3c.

In addition, population structure in CONQUEST2 was analyzed in principal component analysis using the 241 SNPs (Fig. S2). The first two components explained 6.30 % and 4.39 % of the variance. The genotypes from the two breeding companies (BNA and SAKA) were again intermixed.

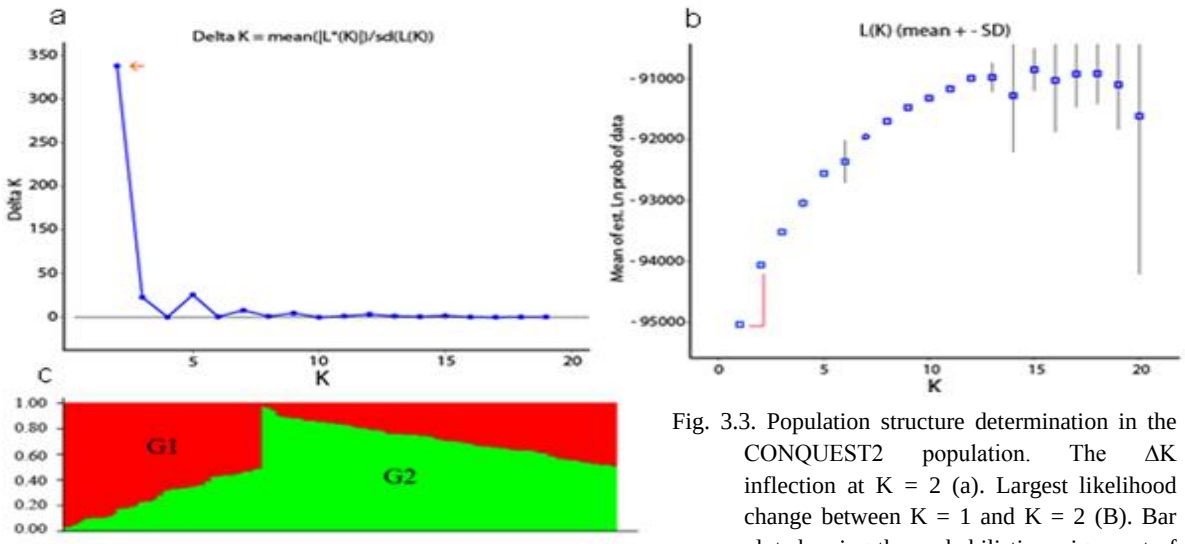


Fig. 3.3. Population structure determination in the CONQUEST2 population. The ΔK inflection at K = 2 (a). Largest likelihood change between K = 1 and K = 2 (B). Bar plot showing the probabilistic assignment of each genotype in two groups (G1 and G2)

3.3.4. Linkage disequilibrium (LD) and LD-decay

Linkage disequilibrium (LD), which is the non random association between two loci, was analyzed using 5600 SNPs for the CONQUEST2 population that contained 184 tetraploid potato clones. Pair wise LD was calculated using r^2 (square of the correlation coefficient between two loci), which acquires values between 0 (absolute linkage equilibrium) and 1 (absolute LD) (Hill and Robertson, 1968). The r^2 value of 0.1 and above was taken as a threshold to claim LD between two loci. The r^2 value was calculated between all pairs of SNPs (dosage form) within a chromosome. The proportion of pair wise r^2 values > 0.1 (loci in LD) and > 0.8 (loci in strong LD) was only 1.57 % and 0.01%, respectively. The remaining about 98 % pairs of loci were in linkage equilibrium. LD decay across the potato genome was analyzed by plotting r^2 values against the physical distance between the loci. The nonlinear regression curve was calculated to determine the average LD decay point. The trendline of the nonlinear regression curve reached the threshold level, $r^2 = 0.1$, between 270 and 280 bps (Fig. 3.4), showing a rapid LD decay in potato genome. This result was in agreement with Stich et al. (2013) who found that in potato LD decayed at 275 bp. However, in some cases SNPs very far apart up to 31Mbp were in strong LD (Table 3.1).

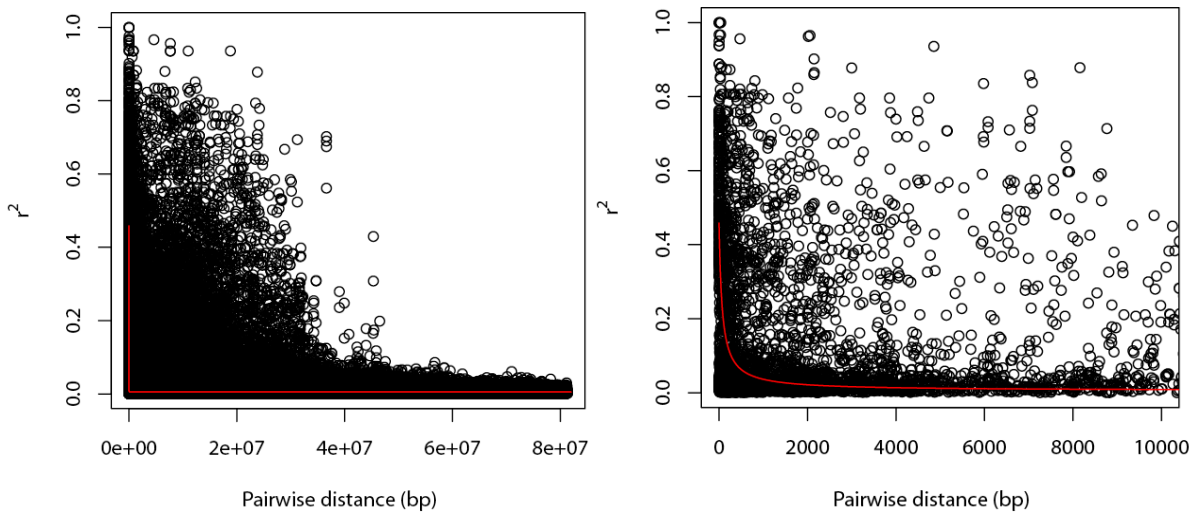


Fig 3.4. Linkage disequilibrium (LD) decay plots. r^2 (square of the correlation coefficient between two loci) plotted against the physical distance (bp) between all pairs of SNPs within the same chromosome; for the entire SNPs (left) and SNPs located within 10 000 bp (right). The SNPs were genotyped on 184 tetraploid potato clones. The trend line of the non linear regression of r^2 versus the physical distance between SNPs is indicated in red.

Table 3.1. An example data showing LD for distantly found markers within chromosome

Marker_1	Marker_2	Chr.	P_value	r^2	Distance between markers (bp)
solcap_snp_c1_9601	solcap_snp_c2_12403	VI	1.99E-145	0.370035	10603633
solcap_snp_c2_1511	solcap_snp_c2_1486	IX	1.99E-145	0.395828	16529059
solcap_snp_c2_681	solcap_snp_c2_1486	IX	1.99E-145	0.395828	11342017
solcap_snp_c2_1968	solcap_snp_c1_2795	VII	4.41E-145	0.349495	19154332
solcap_snp_c1_513	solcap_snp_c2_11400	VII	1.12E-140	0.342672	26758249
solcap_snp_c2_49759	solcap_snp_c1_15714	I	6.70E-139	0.398366	31886296

3.3.5. Marker-trait association analysis

Marker-trait association analysis was carried out using 5600 SolCAP SNP markers for three traits; Maturity corrected resistance (MCR), relative area under the disease progress curve (rAUDPC) and plant maturity (PM). Four different models were tested; simple analysis of variance (ANOVA), using only kinship (K) to avoid false positives (K_model), using only population structure (Q) (Q_model) and using a mixed linear model (MLM), which included kinship (K) and population structure (Q) (QK_model). The threshold level to claim significant association was determined based on M_{eff} according to Li and Ji (2005). The numbers of associated markers in each model and across models are shown in Table 3.2.

Larger numbers of associated markers were found by using simple ANOVA and the K_model. The Q_model and the QK_model gave similar results, suggesting that only population structure can be used to avoid false positives in association mapping using CONQUEST2 population. Comparing across models, three markers with MCR, two markers with rAUDPC and four markers with PM were found to be associated in all the four models (Table 3.2). The QK model was used to claim association as discussed below for each of the three traits.

Table 3.2. Number of associated markers obtained with different statistical models. In all the four models, association was claimed when P_value was above the threshold level calculated according to M_{eff} (effective number of independent tests).

	Number of markers associated with		
Model	MCR	rAUDPC	PM
ANOVA	30	24	132
K-model	21	15	70
Q-model	6	3	9
QK-model	6	2	9
*Common between ANOVA-K_model	11	10	56
Common between ANOVA -Q_model	3	2	4
Common between ANOVA -QK_model	3	2	4
Common between K_model-Q_model	5	2	5
Common between K_model-QK_model	5	2	5
Common between Q_model-QK_model	6	2	9
Common between ANOVA -K_model-Q_model	3	2	5
Common between ANOVA -K_model-QK_model	3	2	4
Common between ANOVA -Q_model-QK_model	3	2	4
Common between ANOVA -K_model-Q_model-QK_model	3	2	4

* Common = Markers found to be associated in different statistical models.

3.3.5.1. *Maturity corrected resistance (MCR)*

With a $-\log_{10}$(P_value) cut of 4.54 (threshold level after multiple testing), six markers were found to be significantly associated with MCR (using the QK_model) (Fig. 3.5). The total variance explained by the six markers ranged from 10 % to 14 %. Of these, three markers (Solcap_snp_c2_35100, Solcap_snp_c1_3326 and Solcap_snp_c2_47952) were associated in all the four statistical models (ANOVA , K_model, Q_model and the QK_model). The annotation of the markers according to PGSC v.4.03 is shown in Table 3.3.

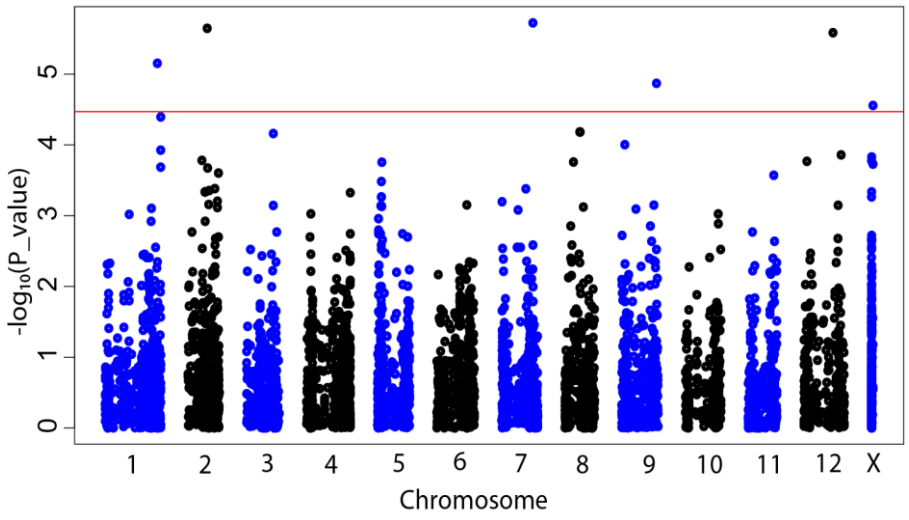


Fig. 3.5. Manhattan plot showing GWA scan for the trait MCR. 5600 SNPs are plotted in the x-axis and $-\log_{10}$ of P_value are plotted on the y-axis. The threshold level was calculated according to M_{eff} (effective number of independent tests). SNPs with no position information are indicated by X.

3.3.5.2. *Relative area under the disease progress curve (rAUDPC)*

Two SNP markers were found to be significantly ($-\log_{10}$P_value > 4.54) associated with rAUDPC (Fig. 3.6) with explained total variance of 13 % and 14 %. The associated markers were Solcap_snp_c2_50302 ($-\log_{10}$P_value > 6.05) and Solcap_snp_c2_50298 ($-\log_{10}$P_value > 5.11). The former was also associated with PM. Both SNPs were associated in all the four statistical models and both were found in chromosome V.

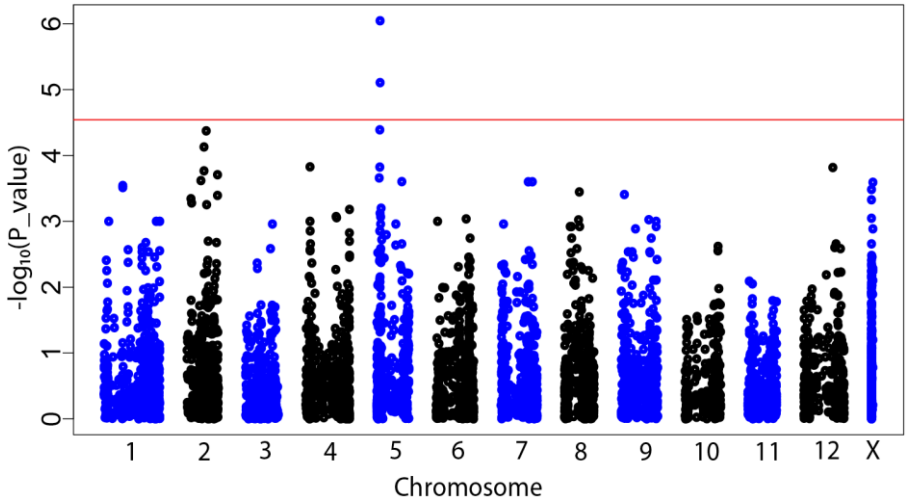


Fig. 3.6. Manhattan plot showing GWA scan for the trait rAUDPC. 5600 SNPs are plotted in the x-axis and $-\log_{10}$ of P_value are plotted on the y-axis. The threshold level was calculated according to M_{eff} (effective number of independent tests). SNPs with no position information are indicated by X.

3.3.5.3. *Plant maturity (PM)*

With PM, nine SNP markers were found to be significantly ($-\log_{10}$P_value > 4.54) associated (Fig. 3.7) with explained total variance ranging from 11 % to 26 % (Table 3.3). Four of these SNP markers were in Chromosome V and were associated in all the four statistical models. Another four were in chromosome IV, and three of them were from same locus. The

remaining one was in chromosome II. The best associated (-$\log_{10}$P_value > 14.04) marker was Solcap_snp_c2_23049, which was annotated as late blight resistance protein.

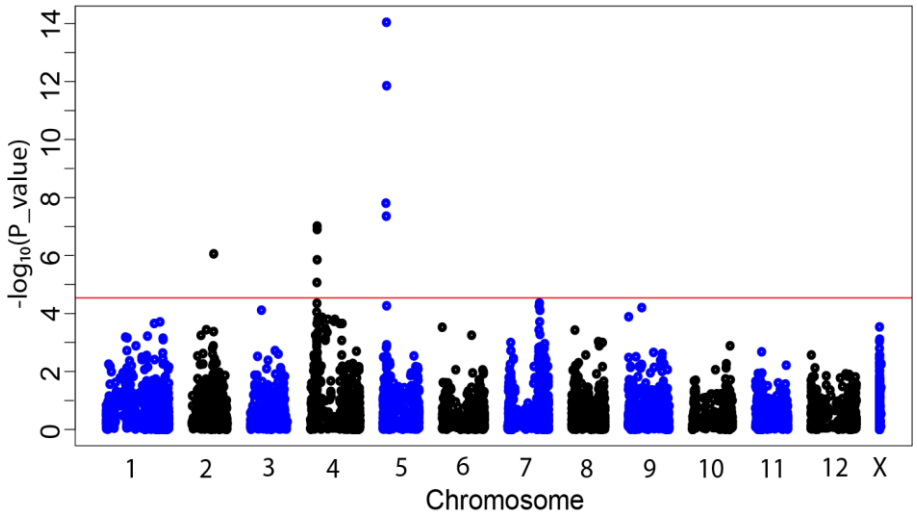


Fig. 3.7. Manhattan plot showing GWA scan for the trait PM. 5600 SNPs are plotted in the x-axis and $-\log_{10}$ of P_value are plotted on the y-axis. The threshold level was calculated according to M_{eff} (effective number of independent tests). SNPs with no position information are indicated by X.

3.3.6. Linkage disequilibrium (LD) in associated markers

Sixteen SolCAP SNP markers significantly associated with any of the traits (MCR, rAUDPC and PM) were selected and LD between markers was calculated by testing for independence between markers with a X^2 (chi-square) test (Fig. 3.8). The intra-chromosomal markers in chromosome IV and V were in strong LD. From inter chromosomal markers, Solcap_snp_c2_46802 with unknown position and Solcap_snp_c1_7412 in chromosome II were in LD. Another two markers from chromosome I (Solcap_snp_c2_35100) and VII (Solcap_snp_c2_4920) were also in LD. The physical map positions of the associated SNP markers and their co-localization with known QTLs are shown in Fig. 3.9.

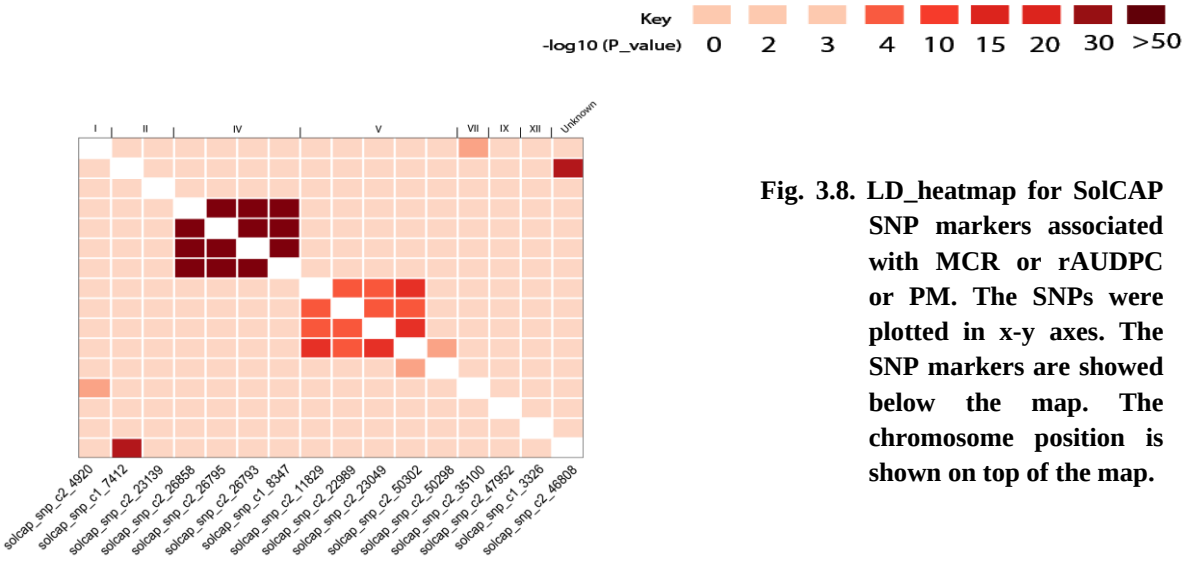


Fig. 3.8. LD_heatmap for SolCAP SNP markers associated with MCR or rAUDPC or PM. The SNPs were plotted in x-y axes. The SNP markers are showed below the map. The chromosome position is shown on top of the map.

Table 3.3. Significantly associated (according to the QK_model) SolCAP SNP markers, allele frequencies and allele effects.

SolCAP_SNP_Marker	Chr.	Allele	Allele frequency (%)	MCR, $-\log_{10}$(P_value) (R^2)	rAUDPC, $-\log_{10}$(P_value) (R^2)	PM, $-\log_{10}$(P_value) (R^2)	Annotation (PGSC v4.03)
solcap_snp_c2_4920	I	T/C	88/12	5.15 (0.11)↓↑	NS	NS	RNA splicing protein mrs2, mitochondrial
solcap_snp_c1_7412	II	A/C	13/87	5.72 (0.13)*	NS	NS	Cold regulated 314 thylakoid membrane 2
solcap_snp_c2_50298	V	T/A	71/29	NS	5.11 (0.13) ↓↑	NS	methyltransferase
solcap_snp_c2_50302	V	A/C	19/81	NS	6.05 (0.13) ↓↑	11.86 (0.21) ↓↑	methyltransferase
solcap_snp_c2_22989	V	A/G	82/18	NS	NS	7.36 (0.13) ↓↑	Bacterial spot disease resistance protein 4
solcap_snp_c2_11829	V	T/C	20/80	NS	NS	7.80 (0.16) ↑↓	Conserved gene of unknown function
solcap_snp_c2_23049	V	A/G	30/70	NS	NS	14.04 (0.25) ↑↓	late blight resistance protein
solcap_snp_c2_35100	VII	T/G	8/92	5.72 (0.13) ↑↓	NS	NS	C3HL domain class transcription factor
solcap_snp_c2_47952	IX	A/G	82/18	4.87 (0.14) ↓↑	NS	NS	Rpi-vnt1
solcap_snp_c1_3326	XII	T/C	6/94	5.59 (0.10) ↑↓	NS	NS	RBP50
solcap_snp_c2_46808	NA	A/C	11/89	4.56 (0.11) ↑↓	NS	NS	NA
solcap_snp_c2_23139	II	A/G	54/46	NS	NS	6.06 (0.14) ↓↑	Conserved gene of unknown function
solcap_snp_c1_8347	IV	T/G	12/88	NS	NS	7.02 (0.14) ↑↓	Glycosyltransferase, CAZy family GT8
solcap_snp_c2_26793	IV	A/C	88/12	NS	NS	6.90 (0.12) ↓↑	Glycosyltransferase, CAZy family GT8
solcap_snp_c2_26795	IV	G/C	12/88	NS	NS	5.85 (0.13) ↑↓	Glycosyltransferase, CAZy family GT8
solcap_snp_c2_26858	IV	T/G	13/87	NS	NS	5.07 (0.11) ↑↓	Transcriptional factor B3

Chr. = Chromosome position; Marker effect shown by arrow, ↑ show positive effect (resistance and earliness); ↓ show negative effect (susceptibility and late maturity); NS = not significant. NA = chromosome position and annotation of the marker are not known; *The direction of the allele effect was not clear

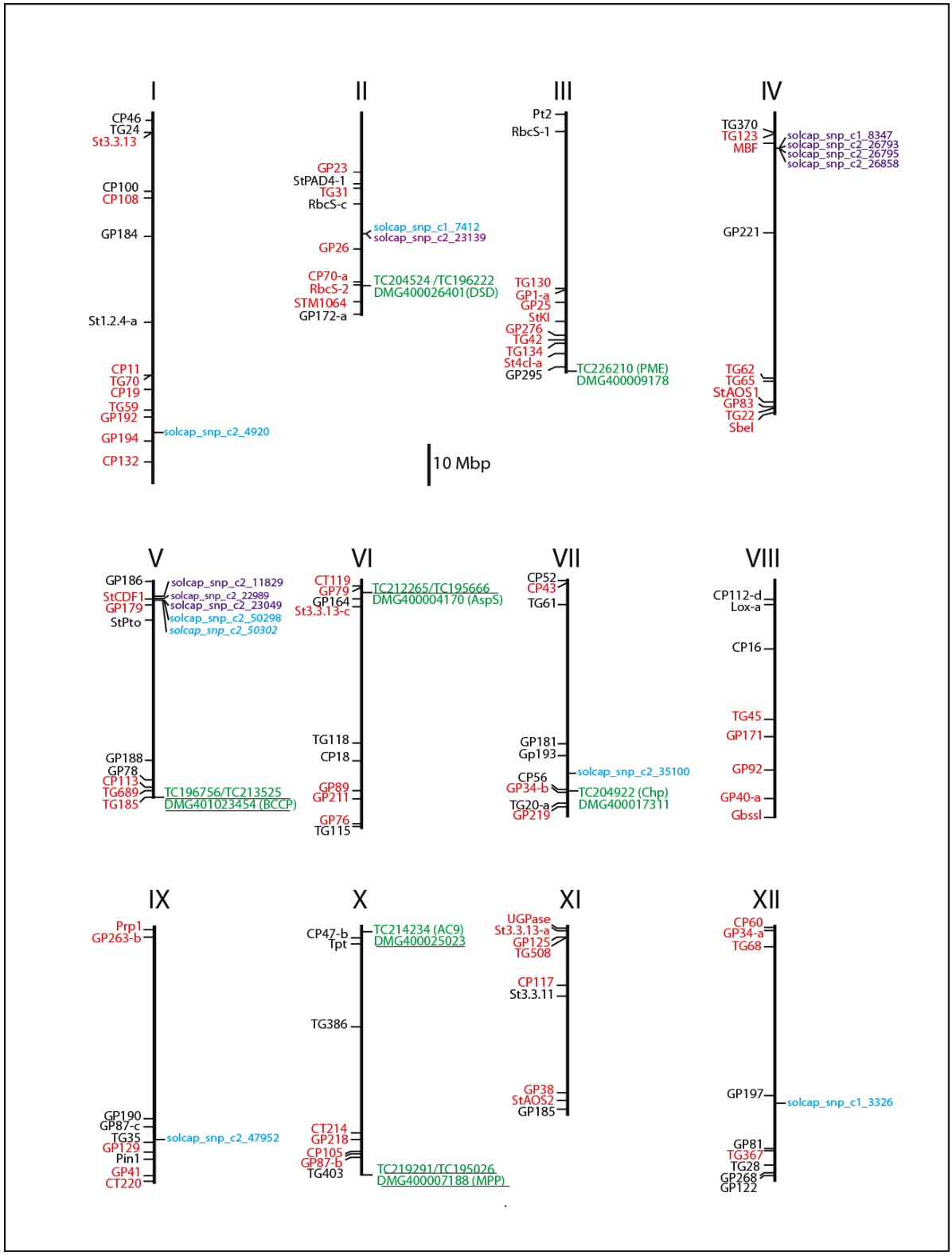


Fig. 3.9. Physical map position of associated SolCAP SNP markers. SNP markers associated with MCR and or rAUDPC are shown in blue; and those associated with PM in pink. RFLP markers linked to resistance QTLs (Gebhardt et al., 2013) are shown in red and additional chromosome anchor markers in black, at the left. Associated candidate genes are shown in green to the right.

3.4. Discussion

3.4.1. Population structure and relatedness

Kinship and population structure analysis suggested no or very weak relatedness and sub populations in the CONQUEST2 population, which contained 184 tetraploid potato clones. This finding is in agreement with earlier reports (Stich et al., 2013; D'hoop et al. 2008; Li et al., 2008; Malosetti et al. 2007; Simko et al. 2004).

The kinship or relatedness between two tetraploid genotypes is defined as the probability that the eight alleles at a particular locus chosen randomly are identical by descent, which means that the identical alleles from the two genotypes arise from the same allele in an earlier generation (Hardy and Vekemans, 2002). The pair wise kinship coefficient in CONQUEST2 population was estimated using 241 SNP markers uniformly distributed across the 12 potato chromosomes. More than 50 % of the pair wise kinship coefficients were negative or zero and more than 80 % were less than 0.03 suggesting the genotypes were unrelated. Negative relative kinship coefficients indicate that the genotypes are less related than random individuals. The negative values were converted to zero by SPAGeDi software, as kinship and probability of identity by descent do not allow negative values (Hardy and Vekemans, 2002).

Population structure is the presence of subgroups in a population, in which individuals are more closely related to each other than the average pair of individuals taken at random from the population (Xu, 2010). In a structured population, linkage disequilibrium (LD) arises due to unequal distribution of alleles within sub-populations (when allele frequencies differ at two loci across subpopulations), irrespective of the linkage status of the loci. This unequal distribution of alleles between sub-populations leads to erroneous, spurious associations between a phenotype and an unlinked candidate gene (Soto-Cerda and Cloutier, 2012; Pritchard et al., 2000).

Pajerowska-Mukhtar et al. (2009) analyzed population structure in the CONQUEST2 population using 31 simple sequence repeat (SSR) markers in STRUCTURE software. Taking the point at which the log likelihood reached the maximum, they used 15 sub populations for association mapping analysis. In this study, a similar result was found from

STRUCTURE software by using the 241 SNP markers in CONQUEST2 population. However, we used the relative rate of change in the likelihood (Rosenberg et al., 2001) and the Δk inflection point (Evanno et al., 2005) to determine the optimal number of sub populations. Both criteria suggested presence of two sub populations. However, these two sub populations were not clearly separated groups. They were rather mixed as shown by a bar plot from STRUCTURE analysis and PCA. Our analysis seems to be logical in reference to previous reports that showed absence of population structure in tetraploid potato (Stich et al., 2013; D'hoop et al. 2008; Li et al., 2008; Malosetti et al. 2007; Simko et al. 2004). Principal component analysis also did not show presence of sub groups in the CONQUEST2 population. The absence of population stratification in tetraploid potato may be mainly attributed to its vegetative propagation. Though the genotypes in the CONQUEST2 population were from two different breeding companies, they were distributed in both sub populations showing that both breeding companies used the same germplasm pools.

3.4.2. Linkage disequilibrium (LD) and LD-decay

Linkage disequilibrium (LD) is defined as the non-random association of alleles at different loci in a population (Soto-Cerda and Cloutier, 2012; Palmer and Cardon, 2005). It is the correlation between polymorphisms, like SNPs, that is caused by their shared history of mutation and recombination. Levels of LD increase by linkage, selection, and admixture (Soto-Cerda and Cloutier, 2012). However, high LD is mostly due to linkage as recombination between two physically close SNPs is rare (Flint-Gracia, 2003). In contrast, SNPs found distantly in a chromosome or on separate chromosomes experience high recombination rate, different selection pressures and independent segregation, so these SNPs have a much lower level of LD. However, LD decays due to successive recombination events that break up the alleles correlations (Palmer and Cardon, 2005). The rate of LD decay or the extent of LD determines the number and density of markers needed to perform an association analysis. In addition, the rate of LD decay determines whether genome wide association mapping versus candidate genes association approaches can be used (Soto-Cerda and Cloutier, 2012; Xu, 2010).

In this study, we estimated LD in the CONQUEST2 population using about 5600 SolCAP SNP markers. We observed that the proportion of pair wise r^2 values > 0.1 (loci in significant

LD) and > 0.8 (loci in strong LD) was only 1.57 % and 0.01%, respectively. These percentages were lower than earlier report by Stich et al. (2013), who observed 19.8 % (r^2 > 0.1) and 0.7 % (r^2 > 0.8) on 36 tetraploid potato clones using a similar number of SolCAP SNP markers. This might be explained by the number of different genotypes used in the two studies. LD decay across the potato genome was analyzed by plotting loci pairs r^2 against physical distance between the loci. The trendline of the nonlinear regression curve reached the threshold level, r^2 = 0.10, between 270 and 280 bps, showing a rapid LD decay in potato genome. This result was in agreement with the recent report by Stich et al. (2013) who showed an LD decay at 275 bp. We used the same methodology and a similar number of SolCAP SNP markers as Stich et al. (2013). As stated earlier, the difference between the two studies was number and type of genotypes.

However, the fast LD decay observed in our study is in contrast to previous reports that estimated LD decay of 5 to 10 cM (D'hoop et al., 2010; Simko et al., 2006), which corresponded to about 4 Mbp to 8 Mbp. In the potato genome, 1cM corresponds to approximately 0.8 Mbp (Stich et al., 2013; Gebhardt et al., 1991). Simko et al. (2006) estimated LD decay at about 10 cM using 47 potato accessions composed of different ploidy levels (monoploid to tetraploid). However, Stich et al. (2013) showed clear separation between tetraploid and diploid potato clones in principal coordinate analysis. LD can be overestimated in structured populations and can lead to a slower LD decay (Soto-Cerda and Cloutier, 2012; Xu, 2010). D'hoop et al. (2010) observed LD decay at about 5 cM using 720 amplified fragment length polymorphism (AFLP) and 53 microsatellite markers on 221 tetraploid potato cultivars. However, AFLP markers are almost exclusively dominant which introduces a number of problems when using them in association mapping (Stich et al., 2013).

Fast LD decay is expected in potato because of its out crossing mating system. Out crossing crops, in contrast to self pollinating crops, usually show fast LD decay in short distance (Xu, 2010). However, the vegetative propagation of potato reduces the number of meiotic generations considerably. Therefore, the average LD decay in potato expected to be somewhere between the recent very fast and the previous slower decays. But still, the gap between recent and earlier estimations of LD decay was very large. The earlier slower LD

decay estimation might be due to presence of large haplotype blocks (Stich et al., 2013; D'hoop et al., 2010) and population structure (Stich et al., 2013; D'hoop et al. 2010) in the populations analyzed. In contrast, the fast LD decay observed in this study and by Stich et al. (2013) might be explained by the fact that most of the SolCAP SNPs were dense in the distal arms of the 12 chromosomes (Fig. 3.1), which are regions where recombination events are more frequent. This might also be the reason why only about 1.6 % of the SNPs were in significant LD.

The fast LD decay might be the reason why the *StAOS2* marker or any allele or SNPs linked with this marker on chromosome XI were not detected by the SolCAP SNPs analyzed in this study. *StAOS2* is a major QTL for *P. infestans* resistance in potato explaining about 30 % of the genetic variation (Pajerowska-Mukhtar et al., 2009).

The fast LD decay observed in this study implies that a large number of markers is required for detecting QTLs in potato. According to Stich et al. (2013), about 3 million genome wide equally distributed SNPs are required to detect all possible marker-trait associations. This will be very expensive, time consuming and also increases false positives (spurious associations) due to multiple testing problems. This will lead to the conclusion that candidate genes association mapping approaches are more appropriate than genome wide association mapping in potato. On the other hand, the fast LD decay might also show that markers found to be associated with different traits in this study as well as in other studies were either in tight linkage with the QTL or they were within the QTL contributing to the trait.

In general, our result showed an average LD decay at about 275 bp. This was on average, in specific cases we have found some SNPs that were physically close to each other but were completely independent ($r^2 < 0.1$), whereas SNPs that were very far apart were in high LD. Thus, when LD is low, screening nearly all of the SNPs in a given region could still miss the relevant locus. When LD is high, evidence for association can be found for most of the loci examined, which would reveal little about the precise localization of the functional variant (Soto-Cerda and Cloutier, 2012).

3.4.3. Marker-trait association

Genome-wide association (GWA) mapping is a method that surveys or scans genetic variation in the whole genome to find signals of association for various complex traits (Xu, 2010). Association mapping does not need to develop specific crosses, instead it uses existing collections of genotypes allowing targeting a broader and more relevant genetic spectrum for plant breeders than conventional QTL mapping does. However, in such types of populations, significant marker trait association may or may not be the consequence of physical linkage between markers and QTLs, leading to false positives or spurious associations. A major cause of false positives is the genetic correlation between individuals resulting from the heterogeneous genetic relatedness between them and presence of population structure. To deal with non-functional, spurious associations between a phenotype and an unlinked candidate gene caused by the genetic correlation and population structure, several methods have been proposed (Stich et al., 2008; Malosetti et al., 2007; Yu et al., 2006). One such approach is the QK mixed-model association mapping approach that promises to correct for LD caused by population structure and family relatedness (Yu et al., 2006).

In this study, we calculated kinship coefficients (Ritland, 1996) and population structure (Pritchard et al., 2000) in the CONQUEST2 population. Then, GWA mapping was employed using a mixed linear model (MLM) in Tassel 2.1. using 5600 SolCAP SNP markers for three traits; MCR, rAUDPC and PM.

Six markers were found to be significantly associated with MCR. The range of variation explained by the markers ranged from 10 % to 14 %. Five of the associated markers were distributed on five different chromosomes (I, II, VII, IX, XII). However one marker was without location and position information showing that the potato genome sequencing is not yet fully done. The marker with unknown position was in strong LD with the marker on chromosome II, which might suggest its position. However, markers on different chromosomes can also be in LD due to selection, genetic drift and population structure (Soto-Cerda and Cloutier, 2012). The associated markers from chromosome I and VII were also in LD. The remaining associated markers were independent. The marker on chromosome IX was annotated as *Rpi-vnt1*. A homologoue of this gene, *Rpi-vnt1.1,* was reported as an *R*-gene encoding a coiled-coil nucleotide-binding leucine-rich repeat protein and conferred resistance

to *P. infestans* in potato and tomato (Foster et al., 2009). Therefore, our candidate can be a defeated *R*-gene, strengthening the idea that defeated *R*-genes still play role in field resistance (Gebhardt, 2013). It has also been reported that genetic positions of QTLs often correspond to regions of *R* gene clusters (Gebhardt and Valkonen 2001).

Two markers from the same locus on chromosome V were associated with rAUDPC. The two markers were in LD and each explained about 13 % of the genetic variation. One of them was also associated with PM showing the correlation between the two traits. This correlation might be explained by pleiotropic effect, in which a QTL contributes to more than one trait. The marker might also be in LD with different QTLs contributing to resistance and PM.

With PM, a cluster of 4 markers on chromosome IV and another cluster of 4 markers on chromosome V were in significant association. One additional associated marker was on chromosome II. Thus, a total of nine markers were associated with PM. One on chromosome V was annotated as late blight resistance protein, further strengthening the report on correlation between PM and late blight resistance in potato. The four associated markers on chromosome V were in strong LD with each other and co-localized with StCDF1. StCDF1 is a major QTL for plant maturity and initiation of tuber development in potato and belongs to the family of DOF (DNA-binding with one finger) transcription factors1 (Kloosterman et al., 2013). There is about 38 kbp difference between the best associated marker in this study and StCDF1.

3.5. Conclusions and outlook

The result of this study demonstrated that the SolCAP potato genotyping array is a useful tool for the identification of diagnostic SNP markers associated with agronomic traits in potato. However, the number of SNPs on the array were too low and not representative in reference to the observed very fast LD decay. Based on our observation of fast LD decay in the CONQUEST2 population, the associated markers found are expected to be either within or at least in tight physical linkage with the QTLs contributing to resistance to *P. infestans* and PM. Most of the markers are regarded as novel as they were not reported before. Functional characterization of the genes containing to the associated SolCAP SNP markers is required to determine whether they are causal or linked genes.

4. References

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. *Theor. Appl. Genet.* 118, 619–29. doi:10.1007/s00122-008-0925-x.

An, S. H., Sohn, K. H., Choi, H. W., Hwang, I. S., Lee, S. C., and Hwang, B. K. (2008). Pepper pectin methylesterase inhibitor protein CaPMEI1 is required for antifungal activity, basal disease resistance and abiotic stress tolerance. *Planta* 228, 61–78. doi:10.1007/s00425-008-0719-z.

Ballvora, A., Ercolano, M. R., Weiss, J., Meksem, K., Bormann, C. A., Oberhagemann, P., Salamini, F., and Gebhardt, C. (2002). The *R1* gene for potato resistance to late blight (*Phytophthora infestans*) belongs to the leucine zipper/NBS/LRR class of plant resistance genes. *Plant J.* 30, 361–71.

Bormann, C. A., Rickert, A. M., Ruiz, R. A. C., Paal, J., Lübeck, J., Strahwald, J., Buhr, K., and Gebhardt, C. (2004). Tagging quantitative trait loci for maturity-corrected late blight resistance in tetraploid potato with PCR-based candidate gene markers. *Mol. Plant. Microbe. Interact.* 17, 1126–38. doi:10.1094/MPMI.2004.17.10.1126.

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23, 2633–5. doi:10.1093/bioinformatics/btm308.

Bradshaw, J. E. (2007). "Potato breeding strategy", in Potato Biology and Biotechnology Advances and Perspectives, eds D. Vreugdenhil, J. Bradshaw, C. Gebhardt, F. Govers, D. K. L. Mackerron, M. A. Taylor and H. A. Ross (Amsterdam, Boston, Heidelberg, London, New York, Oxford, Paris, San Diego, San Francisco, Singapore, Sydney, Tokyo: Elsevier), 157-174. doi: 10.1016/B978-044451018-1/ 50049-X

Bradshaw, J. E., and Bonierbale, M. (2010). "Potatoes", in Root and Tuber Crops. , ed. J. E. Bradshaw New York, NY: Springer New York, 1-51. doi:10.1007/978-0-387-92765-7.

Chi, W., Sun, X., and Zhang, L. (2012). The roles of chloroplast proteases in the biogenesis and maintenance of photosystem II. *Biochim. Biophys. Acta* 1817, 239–46. doi:10.1016/j.bbabio.2011.05.014.

Choe, S., Noguchi, T., Fujioka, S., Takatsuto, S., Tissier, C. P., Gregory, B. D., Ross, a S., Tanaka, a, Yoshida, S., Tax, F. E., et al. (1999). The Arabidopsis dwf7/ste1 mutant is defective in the delta7 sterol C-5 desaturation step leading to brassinosteroid biosynthesis. *Plant Cell* 11, 207–21.

Collard, B. C. Y., and Mackill, D. J. (2008). Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 363, 557–72. doi:10.1098/rstb.2007.2170.

Czechowski, T., Stitt, M., Altmann, T., and Udvardi, M. K. (2005). Genome-wide identification and testing of superior reference genes for transcript normalization. 139, 5–17. doi:10.1104/pp.105.063743.1.

D'hoop, B. B., Keizer, P. L. C., Paulo, M. J., Visser, R. G. F., van Eeuwijk, F. a, and van Eck, H. J. (2014). Identification of agronomically important QTL in tetraploid potato cultivars using a marker-trait association analysis. *Theor. Appl. Genet.* 127, 731–48. doi:10.1007/s00122-013-2254-y.

D'hoop, B. B., Paulo, M. J., Kowitwanich, K., Sengers, M., Visser, R. G. F., van Eck, H. J., and van Eeuwijk, F. a (2010). Population structure and linkage disequilibrium unraveled in tetraploid potato. *Theor. Appl. Genet.* 121, 1151–70. doi:10.1007/s00122-010-1379-5.

D'hoop, B. B., Paulo, M. J., Mank, R. a., Eck, H. J., and Eeuwijk, F. a. (2008). Association mapping of quality traits in potato (*Solanum tuberosum* L.). *Euphytica* 161, 47–60. doi:10.1007/s10681-007-9565-5.

Dangl, J.L. and Jones, J.D. (2001). Plant pathogens and integrated defence responses to infection. *Nature*, 411, 826-833.

Draffehn, A. M., Li, L., Krezdorn, N., Ding, J., Lübeck, J., Strahwald, J., Muktar, M. S., Walkemeier, B., Rotter, B., and Gebhardt, C. (2013). Comparative transcript profiling by SuperSAGE identifies novel candidate genes for controlling potato quantitative resistance to late blight not compromised by late maturity. *Front. Plant Sci.* 4, 423. doi:10.3389/fpls.2013.00423.

Earl, Dent A. and vonHoldt, Bridgett M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4 (2), 359-361 doi: 10.1007/s12686-011-9548-7

Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14, 2611–20. doi:10.1111/j.1365-294X.2005.02553.x.

Expósito-Rodríguez, M., Borges, A. a, Borges-Pérez, A., and Pérez, J. A. (2008). Selection of internal control genes for quantitative real-time RT-PCR studies during tomato development process. *BMC Plant Biol.* 8, 131. doi:10.1186/1471-2229-8-131.

Flint-Garcia, S. A, Thornsberry, J. M., and Buckler, E. S. (2003). Structure of linkage disequilibrium in plants. *Annu. Rev. Plant Biol.* 54, 357–74. doi:10.1146/annurev.arplant.54.031902.134907.

Forbes, G. a., Chacon, M. G., Kirk, H. G., Huarte, M. a., Van Damme, M., Distel, S., Mackay, G. R., Stewart, H. E., Lowe, R., Duncan, J. M., et al. (2005). Stability of resistance to *Phytophthora infestans* in potato: an international evaluation. *Plant Pathol.* 54, 364–372. doi:10.1111/j.1365-3059.2005.01187.x.

Foster, S. J., Park, T., Pel, M., Brigneti, G., Jadwiga, Ś., Jagger, L.,van der Vossen, E., Jones, J. D. G. (2009). *Rpi-vnt1.1*, a *Tm-2*^{2} homolog from *Solanum venturii*, confers resistance to potato late blight. *Ammerican Phyto. soci*, *22*(5), 589–600.

Fritzemeier, K. H., Cretin, C., Kombrink, E., Rohwer, J., Taylor, J., Hahlbrock, K. (1987). Transient induction of phenylalanine ammonia-lyase and 4-coumarate:CoA ligase mRNAs in potato leaves infected with virulent or avirulent races of *Phytophthora infestans*. *Plant Physiol*. 85, 34-41.

Fry, W. E. (1978). Quantification of general resistance of potato cultivars and fungicide effects for integrated control of potato late blight. *Phy.path.* 68, 1650–1655.

Fry, W. E., and Goodwin, S. B. (1997). Re-emergence of potato and tomato late blight in the United States. *Plant Dis*. 81:1349-1357.

Gebhardt et al., (2014). "Dissection of potato complex traits by linkage and association genetics as basis for developing molecular diagnostics in breeding programs", in Genomics of Plant Genetic Resourses, Volume2 Crop Productivity, Food Security and Nutritional Quality, eds R. Tuberosa, A. Graner and E. Frison (Springer Dordrecht Heidelberg NewYork London), 47–85. doi: DOI 10.1007/978-94-007-7575-6_1.

Gebhardt, C. (2007). "Molecular markers , maps, and population genetics", in Potato Biology and Biotechnology Advances and Perspectives, eds D. Vreugdenhil, J. Bradshaw, C. Gebhardt, F. Govers, D. K. L. Mackerron, M. A. Taylor and H. A. Ross (Amsterdam, Boston, Heidelberg, London, New York, Oxford, Paris, San Diego, San Francisco, Singapore, Sydney, Tokyo: Elsevier), 77–89. doi: 10.1016/B978-044451018-1/ 50049-X

Gebhardt, C. (2013). Bridging the gap between genome analysis and precision breeding in potato. *Trends Genet.* 29, 248–56. doi:10.1016/j.tig.2012.11.006.

Gebhardt, C. and Valkonen , J. P. (2001). Organization of genes controlling disease resistance in the potato genome. *Ann. Rev., Phytopathol. 39*, 79–102. doi:10.1146/annurev.phyto.39.1.79

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. *Mol. Breed.* 13, 93–102. doi:10.1023/B:MOLB.0000012878.89855.df.

Gebhardt, C., Li, L., Pajerowska-Mukhtar, K., Achenbach, U., Sattarzadeh, A., Bormann, C., Ilarionova, E., and Ballvora, A. (2007). Candidate gene approach to identify genes underlying quantitative traits and develop diagnostic markers in potato. *Crop Sci.* 47, S–106. doi:10.2135/cropsci2007.04.0005IPBS.

Gleason, C.A., LIU, Q.L. and Williamson, V.M. (2008). Silencing a candidate nematode effector gene corresponding to the tomato resistance gene *Mi-1* leads to acquisition of virulence. *Molecular Plant-Microbe Interactions* 21, 576-585.

Guenthner, J. F., Michael, K. C. and Nolte, P. (2001). The economic impact of potato late blight on US growers. *Potato Res.*, 44, 121–125.

Gyetvai, G., Sønderkær, M., Göbel, U., Basekow, R., Ballvora, A., Imhoff, M., Kersten, B., Nielsen, K.-L., and Gebhardt, C. (2012). The transcriptome of compatible and incompatible interactions of potato (*Solanum tuberosum*) with *Phytophthora infestans* revealed by DeepSAGE analysis. *PLoS One* 7, e31526. doi:10.1371/journal.pone.0031526.

Haas, B. J., Kamoun, S., Zody, M. C., Jiang, R. H. Y., Handsaker, R. E., Cano, L. M., Grabherr, M., Kodira, C. D., Raffaele, S., Torto-Alalibo, T., et al. (2009). Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature* 461, 393–8. doi:10.1038/nature08358.

Hamilton, J. P., 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, 302. doi:10.1186/1471-2164-12-302.

Hardy, O. J. and Vekemans, X. (2002). SPAGeDi : a versatile computer program to analyze spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*. 2, 618-620.

Haverkort, A., Koesveld, F. Van, Schepers, H., Wijnands, J., Wustman, R. and Zhang, X. (2012). Potato prospects for Ethiopia : on the road to value addition. *Applied Plant Research*, 1-65.

Henfling, J. (1982). Field screening procedures to evaluate resistance to late blight. International Potato Center. *Technology evaluation series No*. 1982-5. 15 p.

Hill, W. G. and Robertson, A. (1968). Linkage disequilibrium in finite populations. *Theor. Appl. Genet*, 38, 226–231.

Hückelhoven, R., Eichmann, R., Weis, C., Hoefle, C., and Proels, R. K. (2013). Genetic loss of susceptibility: a costly route to disease resistance? *Plant Pathol.* 62, 56–62. doi:10.1111/ppa.12103.

Hwang, I. S., An, S. H., and Hwang, B. K. (2011). Pepper asparagine synthetase 1 (CaAS1) is required for plant nitrogen assimilation and defense responses to microbial pathogens. *Plant J.* 67, 749–62. doi:10.1111/j.1365-313X.2011.04622.x.

Jannink, J., & Jansen, R. (2001). Mapping Epistatic Quantitative Trait Loci With One-Dimensional Genome Searches. *Genetics*, 157, 445-454.

Jones, J.D.G. and Dangl, J.L. (2006). The plant immune system. *Nature*, 444, 323-329.

Judelson, H. S., and Blanco, F. A. (2005). The spores of *Phytophthora*: weapons of the plant destroyer. *Nat. Rev. Microbiol.* 3, 47–58. doi:10.1038/nrmicro1064.

Kamoun, S., and Smart, C. D. (2005). Late blight of potato and tomato in the genomics era. *Plant Dis.* 89, 692–699. doi:10.1094/PD-89-0692.

Kay, S., Hahn, S., Marois, E., Hause, G., and Bonas, U. (2007). A bacterial effector acts as a plant transcription factor and induces a cell size regulator. *Science* 318, 648–51. doi:10.1126/science.1144956.

Kay, S., Hahn, S., Marois, E., Wieduwild, R., and Bonas, U. (2009). Detailed analysis of the DNA recognition motifs of the *Xanthomonas* type III effectors AvrBs3 and AvrBs3Δrep16. *Plant J.* 59, 859–71. doi:10.1111/j.1365-313X.2009.03922.x.

Kloosterman, B., Abelenda, J. a, Gomez, M. D. M. C., Oortwijn, M., de Boer, J. M., Kowitwanich, K., Horvath, B. M., van Eck, H. J., Smaczniak, C., Prat, S., et al. (2013). Naturally occurring allele diversity allows potato cultivation in northern latitudes. *Nature* 495, 246–50. doi:10.1038/nature11912.

Kloosterman, B., Oortwijn, M., uitdeWilligen, J., America, T., de Vos, R., Visser, R. G. F., and Bachem, C. W. B. (2010). From QTL to candidate gene: genetical genomics of simple and complex traits in potato using a pooling strategy. *BMC Genomics* 11, 158. doi:10.1186/1471-2164-11-158.

Kou, Y., and Wang, S. (2010). Broad-spectrum and durability: understanding of quantitative disease resistance. *Curr. Opin. Plant Biol.* 13, 181–5. doi:10.1016/j.pbi.2009.12.010.

Lapin, D., and Van den Ackerveken, G. (2013). Susceptibility to plant disease: more than a failure of host immunity. *Trends Plant Sci.* 18, 546–54. doi:10.1016/j.tplants.2013.05.005.

Li, J., and Ji, L. (2005). Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity (Edinb).* 95, 221–7. doi:10.1038/sj.hdy.6800717.

Li, L., Paulo, M.-J., Strahwald, J., Lübeck, J., Hofferbert, H.-R., Tacke, E., Junghans, H., Wunder, J., Draffehn, A., van Eeuwijk, F., et al. (2008). Natural DNA variation at candidate loci is associated with potato chip color, tuber starch content, yield and starch yield. *Theor. Appl. Genet.* 116, 1167–81. doi:10.1007/s00122-008-0746-y.

Li, L., Strahwald, J., Hofferbert, H.-R., Lübeck, J., Tacke, E., Junghans, H., Wunder, J., and Gebhardt, C. (2005). DNA variation at the invertase locus *invGE/GF* is associated with tuber quality traits in populations of potato breeding clones. *Genetics* 170, 813–21. doi:10.1534/genetics.104.040006.

Lindqvist-kreuze, H., Carbajulca, D., and Gonzalez-escobedo, G. (2010). Comparison of transcript profiles in late blight-challenged *Solanum cajamarquense* and B3C1 potato clones. 11, 513–530. doi:10.1111/J.1364-3703.2010.00622.X.

Lionetti, V., Raiola, A., Camardella, L., Giovane, A., Obel, N., Pauly, M., Favaron, F., Cervone, F., and Bellincampi, D. (2007). Overexpression of pectin methylesterase inhibitors in arabidopsis restricts fungal infection by *Botrytis cinerea*. *Plant Physiol.* 143, 1871–80. doi:10.1104/pp.106.090803.

Ma, L., Jiang, S., Lin, G., Cai, J., Ye, X., Chen, H., Li, M., Li, H., Takác, T., Samaj, J., et al. (2013). Wound-induced pectin methylesterases enhance banana (Musa spp. AAA) susceptibility to *Fusarium oxysporum* f. sp. *cubense*. *J. Exp. Bot.* 64, 2219–29. doi:10.1093/jxb/ert088.

Malosetti, M., van der Linden, C. G., Vosman, B., and van Eeuwijk, F. A. (2007). A mixed-model approach to association mapping using pedigree information with an illustration of resistance to *Phytophthora infestans* in potato. *Genetics* 175, 879–89. doi:10.1534/genetics.105.054932.

Marino, D., Froidure, S., Canonne, J., Ben Khaled, S., Khafif, M., Pouzet, C., Jauneau, A., Roby, D., and Rivas, S. (2013). Arabidopsis ubiquitin ligase MIEL1 mediates degradation of the transcription factor MYB30 weakening plant defence. *Nat. Commun.* 4, 1476. doi:10.1038/ncomms2479.

Matsumura, H., Reich, S., Ito, A., Saitoh, H., Kamoun, S., Winter, P., Kahl, G., Reuter, M., Kruger, D. H., and Terauchi, R. (2003). Gene expression analysis of plant host-pathogen interactions by SuperSAGE. *Proc. Natl. Acad. Sci. U. S. A.* 100, 15718–23. doi:10.1073/pnas.2536670100.

McLellan, H., Boevink, P. C., Armstrong, M. R., Pritchard, L., Gomez, S., Morales, J., Whisson, S. C., Beynon, J. L., and Birch, P. R. J. (2013). An RxLR effector from *Phytophthora infestans* prevents re-localisation of two plant NAC transcription factors from the endoplasmic reticulum to the nucleus. *PLoS Pathog.* 9, e1003670. doi:10.1371/journal.ppat.1003670.

Menéndez, C. M., Ritter, E., Schäfer-Pregl, R., Walkemeier, B., Kalde, A., Salamini, F., and Gebhardt, C. (2002). Cold sweetening in diploid potato: mapping quantitative trait loci and candidate genes. *Genetics* 162, 1423–34.

Meyer, S., Nagel, A., and Gebhardt, C. (2005). PoMaMo--a comprehensive database for potato genome data. *Nucleic Acids Res.* 33, D666–70. doi:10.1093/nar/gki018.

Milbourne, D., Pande, B., and Bryan, G. J. (2007). "Potato", in Genome Mapping and Molecular Breeding in Plants, Volume 3 Pulses, Sugar and Tuber Crops, eds C. Kole. Springer-Verlag Berlin Heidelberg, 205-236.

Naess, S. K., Bradeen, J. M., Wielgus, S. M., Haberlach, G. T., McGrath, J. M., and Helgeson, J. P. (2001). Analysis of the introgression of *Solanum bulbocastanum* DNA into potato breeding lines. *Mol. Genet. Genom*. 265:694-704.

Nikolau, B. J., Ohlrogge, J. B., and Wurtele, E. S. (2003). Plant biotin-containing carboxylases. *Arch. Biochem. Biophys.* 414, 211–222. doi:10.1016/S0003-9861(03)00156-5.

Nowicki, M., Foolad, M. R., Nowakowska, M. and Kozik, E. U. (2012). Potato and tomato late blight caused by *Phytophthora infestans*: an overview of pathology and resistance breeding. *Plant Dis.* 96, 96, 4-17. doi: 10.1094/PDIS-05-11-0458

Pajerowska-Mukhtar, K. M., Mukhtar, M. S., Guex, N., Halim, V. a, Rosahl, S., Somssich, I. E., and Gebhardt, C. (2008). Natural variation of potato *allene oxide synthase 2* causes differential levels of jasmonates and pathogen resistance in *Arabidopsis*. *Planta* 228, 293–306. doi:10.1007/s00425-008-0737-x.

Pajerowska-Mukhtar, K., Stich, B., Achenbach, U., Ballvora, A., Lübeck, J., Strahwald, J., Tacke, E., Hofferbert, H.-R., Ilarionova, E., Bellin, D., et al. (2009). Single nucleotide polymorphisms in the *allene oxide synthase 2* gene are associated with field resistance to late blight in populations of tetraploid potato cultivars. *Genetics* 181, 1115–27. doi:10.1534/genetics.108.094268.

Palmer, L. J., and Cardon, L. R. (2005). Shaking the tree : mapping complex disease genes with linkage disequilibrium. *Genetic Epidemiology* 4, 1223–1234.

Park, T. H., Vleeshouwers, V. G. A. A., Jacobsen, E., Visser, R. G. F., and Vossen, E. A. G. v. d. (2009). Molecular breeding for resistance to *Phytophthora infestans* (Mont.) de Bary in potato (*Solanum tuberosum* L.): A perspective of cisgenesis. *Plant Breed*. 128,109-117.

Pereira, A. S., Gomes, C. B., Castro, C. M. and Silva, G. O. (2012). "Breeding for Fungus Resistance", in Plant breeding for biotic resistance, eds R. Fritsche-Neto and A. Borem (Springer Heidelberg New York Dordrecht London), 13-36. doi: 10.1007/978-3-642-33087-2.

Peter, E., Rothbart, M., Oelze, M.-L., Shalygo, N., Dietz, K.-J., and Grimm, B. (2010). Mg protoporphyrin monomethylester cyclase deficiency and effects on tetrapyrrole metabolism in different light conditions. *Plant Cell Physiol.* 51, 1229–41. doi:10.1093/pcp/pcq071.

Pluthero, F. G. (1993). Rapid purification of high-activity Taq DNA polymerase. *Nucleic Acids Res.* 21, 4850–1.

Posé, D., Castanedo, I., Borsani, O., Nieto, B., Rosado, A., Taconnat, L., Ferrer, A., Dolan, L., Valpuesta, V., and Botella, M. a (2009). Identification of the Arabidopsis *dry2/sqe1-5* mutant reveals a central role for sterols in drought tolerance and regulation of reactive oxygen species. *Plant J.* 59, 63–76. doi:10.1111/j.1365-313X.2009.03849.x.

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–59.

Rasbery, J. M., Shan, H., LeClair, R. J., Norman, M., Matsuda, S. P. T., and Bartel, B. (2007). *Arabidopsis thaliana squalene epoxidase 1* is essential for root and seed development. *J. Biol. Chem.* 282, 17002–13. doi:10.1074/jbc.M611831200.

Rietman, H., Bijsterbosch, G., Cano, L. M., Lee, H.-R., Vossen, J. H., Jacobsen, E., … Vleeshouwers, V. G. a a. (2012). Qualitative and quantitative late blight resistance in the potato cultivar Sarpo Mira is determined by the perception of five distinct RXLR effectors. *Molecular plant-microbe interactions : MPMI*, *25*(7), 910–9. doi:10.1094/MPMI-01-12-0010-R

Rigden, D.J., Mello, L.V. and Bertioli, D.J. (2000). Structural modeling of a plant disease resistance gene product domain. *Proteins: Structure, Function, and Genetics*, 41, 133-43.

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

Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlén, M., and Nyrén, P. (1996). Real-Time DNA Sequencing Using Detection of Pyrophosphate Release. *Analytical Biochemistry* 242, 84-89.

Ros, B., Thümmler, F., and Wenzel, G. (2005). Comparative analysis of *Phytophthora infestans* induced gene expression in potato cultivars with different levels of resistance. *Plant Biol. (Stuttg).* 7, 686–93. doi:10.1055/s-2005-872946.

Rowland, O., Ludwig, A. A., Merrick, C. J., Baillieul, F., Tracy, F. E., Yoshioka, H., Durrant, W. E., Fritz-laylin, L., Nekrasov, V., Sjo, K., et al. (2005). Functional analysis of *Avr9 / Cf-9 rapidly elicited* genes identifies a protein kinase , ACIK1 , That Is Essential for Full Cf-9 – Dependent Disease Resistance in Tomato. *The plant cell*, 17, 295–310. doi:10.1105/tpc.104.026013.2.

Silvestro, D., Andersen, T. G., Schaller, H., and Jensen, P. E. (2013). Plant sterol metabolism. Δ(7)-Sterol-C5-desaturase (STE1/DWARF7), Δ(5,7)-sterol-Δ(7)-reductase (DWARF5) and Δ(24)-sterol-Δ(24)-reductase (DIMINUTO/DWARF1) show multiple subcellular localizations in *Arabidopsis thaliana* (Heynh) L. *PLoS One* 8, e56429. doi:10.1371/journal.pone.0056429.

Simko, I., Haynes, K. G., and Jones, R. W. (2006). Assessment of linkage disequilibrium in potato genome with single nucleotide polymorphism markers. *Genetics* 173, 2237–45. doi:10.1534/genetics.106.060905.

Simko, I., Jansky, S., Stephenson, S., and Spooner, D. (2007). "Genetics of resistance to pests and disease", in Potato Biology and Biotechnology Advances and Perspectives, eds D. Vreugdenhil, J. Bradshaw, C. Gebhardt, F. Govers, D. K. L. Mackerron, M. A. Taylor and H. A. Ross (Amsterdam, Boston, Heidelberg, London, New York, Oxford, Paris, San Diego, San Francisco, Singapore, Sydney, Tokyo: Elsevier), 117–155. doi: 10.1016/B978-044451018-1/ 50049-X

Soto-Cerda, B. J. and Cloutier, S. (2012). "Association mapping in plant genomes", in Genetic Diversity in Plants, eds M. Caliskan, (InTech, online). Available from: http://www.intechopen.com/books/genetic-diversity-in-plants/association-mapping-in-plant-genomes.

St Clair, D. A (2010). Quantitative disease resistance and quantitative resistance loci in breeding. *Annu. Rev. Phytopathol.* 48, 247–68. doi:10.1146/annurev-phyto-080508-081904.

Stich, B., Möhring, J., Piepho, H-P., Heckenberger, M., Buckler, E.S., Melchinger, A.E., (2008). Comparison of Mixed-Model Approaches for Association Mapping. *Genetics.* 178, 1745-1754.

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 Breed.* 132, 718–724. doi:10.1111/pbr.12102.

Storey, J. D. (2003).The positive false discovery rate: a Bayesian interpretation and the *q*-value. *Annals of statistics*. 31, 2013–2035.

Tiwari, J. K., Siddappa, S., Singh, B. P., Kaushik, S. K., Chakrabarti, S. K., Bhardwaj, V., and Chandel, P. (2013). Molecular markers for late blight resistance breeding of potato: an update. *Plant Breed.* 132, 237–245. doi:10.1111/pbr.12053.

Trujillo, M., and Shirasu, K. (2010). Ubiquitination in plant immunity. *Curr. Opin. Plant Biol.* 13, 402–8. doi:10.1016/j.pbi.2010.04.002.

Uitdewilligen, J. G. a M. L., Wolters, A.-M. a, D'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. doi:10.1371/journal.pone.0062355.

Urbany, C., Stich, B., Schmidt, L., Simon, L., Berding, H., Junghans, H., Niehoff, K.-H., Braun, A., Tacke, E., Hofferbert, H.-R., et al. (2011). Association genetics in *Solanum tuberosum* provides new insights into potato tuber bruising and enzymatic tissue discoloration. *BMC Genomics* 12, 7. doi:10.1186/1471-2164-12-7.

Visker, M. H. P. W., Keizer, L. C. P., Van Eck, H. J., Jacobsen, E., Colon, L. T., and Struik, P. C. (2003). Can the QTL for late blight resistance on potato chromosome 5 be attributed to foliage maturity type? *Theor. Appl. Genet.* 106, 317–25. doi:10.1007/s00122-002-1021-2.

Voorrips, R. E., Gort, G., and Vosman, B. (2011). Genotype calling in tetraploid species from bi-allelic marker data using mixture models. *BMC Bioinformatics* 12, 172. doi:10.1186/1471-2105-12-172.

Wu, F.-Q., Xin, Q., Cao, Z., Liu, Z.-Q., Du, S.-Y., Mei, C., Zhao, C.-X., Wang, X.-F., Shang, Y., Jiang, T., et al. (2009). The magnesium-chelatase H subunit binds abscisic acid and

functions in abscisic acid signaling: new evidence in Arabidopsis. *Plant Physiol.* 150, 1940–54. doi:10.1104/pp.109.140731.

Xu, X., Pan, S., Cheng, S., Zhang, B., Mu, D., Ni, P., Zhang, G., Yang, S., Li, R., Wang, J., et al. (2011). Genome sequence and analysis of the tuber crop potato. *Nature* 475, 189–95. doi:10.1038/nature10158.

Xu, Y. (2010). Molecular Plant Breeding. CAB International, Nosworthy Way, Wallingford, Oxfordshire, OX10 8DE, UK

Yoshida, K., Schuenemann, V. J., Cano, L. M., Pais, M., Mishra, B., Sharma, R., Lanz, C., Martin, F. N., Kamoun, S., Krause, J., et al. (2013). The rise and fall of the *Phytophthora infestans* lineage that triggered the Irish potato famine. *Elife* 2, e00731. doi:10.7554/eLife.00731.

Yu, J., Pressoir, G., Briggs,W. H., Vroh, I. Bi, Yamasaki, M., Doebley, J. F., Mc-Mullen, M. D., Gaut,B. S., Nielsen, D. M., Holland, J. B., Kresovich, S. and Buckler, E. S. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* 38, 203–208.

Zhao, J., Liu, Q., Zhang, H., Jia, Q., Hong, Y., and Liu, Y. (2013). The rubisco small subunit is involved in tobamovirus movement and *Tm-2*^{2}-mediated extreme resistance. *Plant Physiol.* 161, 374–83. doi:10.1104/pp.112.209213.

Appendices

5. Appendix

Table . S1. Candidates expression (Tag hits per million in SuperSAGE) comparison between genotype pools (A1, B2 and A2) at three time points (T0, T1 and T2). The A2 pool was quantitatively more resistant while the B2 and A1 pools were quantitatively susceptible.

Comparison	Transcript number	Annotation	Hits per million in B2-T0, B2-T1, B2-T2, A1-T0, A1-T1, A1-T2	Hits per million in A2-T0, A2-T1, A2-T2	p-value	Log fold_change
B2-T0 vs A2-T0	TC208593	Arabinogalactan protein	23.2706	260.6525	3.22E-135	-3.485545514
B2-T1 vs A2-T1			17.0465	77.8889	4.77E-21	-2.19194418
A1-T0 vs A2-T0			0.5889	260.732	3.61E-195	-8.790379493
A1-T1 vs A2-T1			0	77.5325	1.64E-88	-10.59865691
A1-T2 vs A2-T2			0	20.1786	2.46E-21	-8.656682115
B2-T0 vs A2-T0	TC196279	Eukaryotic translation initiation factor 4e type	89.5458	137.9956	3.77E-08	-0.62392377
B2-T1 vs A2-T1			171.4619	234.2814	3.02E-06	-0.450354142
A1-T0 vs A2-T0			75.5974	138.0376	2.81E-13	-0.868653179
A1-T1 vs A2-T1			93.7492	233.2093	1.42E-51	-1.314747454
A1-T2 vs A2-T2			129.2239	84.1396	5.91E-07	0.619015115
B2-T0 vs A2-T0	TC210219	Serine protease inhibitor (Miraculin)	2.4874	29.9096	6.44E-17	-3.587906871
B2-T1 vs A2-T1			0.5912	15.7689	8.84E-09	-4.737178436
B2-T2 vs A2-T2			0	21.9298	2.97E-11	-8.776746521
A1-T0 vs A2-T0			0	29.9187	3.16E-23	-9.224904114
A1-T1 vs A2-T1			0	15.6968	1.64E-18	-8.294324948
A1-T2 vs A2-T2			0	21.814	5.49E-23	-8.769109004

Table S1. Continued

Table . S1

B2-T0 vs A2-T0	TC214234	Avr9/Cf-9 rapidly elicited protein 20	68.6082	252.6885	2.11E-69	-1.880906756
B2-T1 vs A2-T1			625.2647	872.4893	8.83E-22	-0.480670393
B2-T2 vs A2-T2			295.0861	585.1985	4.54E-37	-0.987790046
A1-T0 vs A2-T0			191.3858	252.7656	5.29E-07	-0.401316346
A1-T1 vs A2-T1			442.6272	868.4968	9.10E-116	-0.972428524
A1-T2 vs A2-T2			323.6996	582.1087	1.13E-47	-0.846633002
B2-T0 vs A2-T0	TC199656	Polyubiquitin	11.9394	32.7581	1.47E-07	-1.456116998
B2-T1 vs A2-T1			7.6862	47.5884	4.56E-16	-2.630263232
B2-T2 vs A2-T2			11.9603	41.2488	1.55E-07	-1.786095457
A1-T0 vs A2-T0			0	32.7681	2.18E-25	-9.356148647
A1-T1 vs A2-T1			0	47.3706	2.29E-54	-9.887849462
A1-T2 vs A2-T2			0	41.0311	2.12E-42	-9.680572329
A1-T0 vs A2-T0	TC204068	Clathrin coat assembly protein AP17	0	37.8337	3.15E-29	-9.563528498
A1-T1 vs A2-T1			0	56.9008	3.73E-65	-10.15230594
A1-T2 vs A2-T2			0	49.3411	8.54E-51	-9.946647189
B2-T0 vs A2-T0			7.9596	37.8222	5.47E-14	-2.24845935
B2-T1 vs A2-T1			0	57.1624	6.19E-35	-10.15892287
B2-T2 vs A2-T2			0.6645	49.603	5.77E-23	-6.222095318
B2-T0 vs A2-T0	TC207534	Subtilisin-like protease	132.8263	266.0212	3.52E-30	-1.002000619
B2-T1 vs A2-T1			193.3381	474.7577	7.00E-60	-1.296065509
B2-T2 vs A2-T2			54.486	120.6137	8.20E-11	-1.146436747
A1-T0 vs A2-T0			129.3113	266.1024	5.42E-32	-1.041132564
A1-T1 vs A2-T1			85.9368	472.5852	4.51E-241	-2.459227438
A1-T2 vs A2-T2			61.5141	119.9769	1.60E-13	-0.96376741

Table S1. Continued

Table . S1

B2-T0 vs A2-T0	TC206425	Conserved gene of unknown function	5.4722	85.456	1.45E-49	-3.96497652
B2-T1 vs A2-T1			3.5475	83.6317	9.54E-42	-4.559180134
B2-T2 vs A2-T2			1.9934	53.7801	2.26E-22	-4.753777737
A1-T0 vs A2-T0			37.7987	85.482	2.57E-13	-1.177284451
A1-T1 vs A2-T1			45.5725	83.249	1.22E-10	-0.869268841
B2-T0 vs A2-T0	TC208728	Heat shock cognate 70 kDa protein	0.8024	27.2101	1.07E-18	-5.083611049
B2-T1 vs A2-T1			0.6819	83.2375	2.43E-48	-6.931616157
B2-T2 vs A2-T2			0	27.3416	7.28E-14	-9.094953987
A1-T0 vs A2-T0			0	27.2184	3.52E-21	-9.08843629
A1-T1 vs A2-T1			0	82.8566	1.55E-94	-10.6944727
A1-T2 vs A2-T2			0	27.1972	2.00E-28	-9.087316469
B2-T0 vs A2-T0	TC226354	kiwellin	2.4073	21.403	2.70E-11	-3.152323799
B2-T1 vs A2-T1			4.0912	21.7287	2.39E-07	-2.409019537
A1-T0 vs A2-T0			0	21.4096	8.91E-17	-8.742111541
A1-T1 vs A2-T1			0.9236	21.6292	1.48E-20	-4.549596867
A1-T2 vs A2-T2			0.673	14.476	2.71E-12	-4.42696755
B2-T0 vs A2-T0	TC200937	Photosystem II core complex proteins	2273.7008	2857.3881	3.61E-45	-0.32965463
B2-T1 vs A2-T1			1313.2605	1877.3564	7.91E-51	-0.515549458
B2-T2 vs A2-T2			1107.2448	1586.2539	5.23E-33	-0.518649552
A1-T0 vs A2-T0			2480.3601	2858.2596	2.78E-19	-0.20458738
A1-T1 vs A2-T1			2803.3055	1868.7656	5.09E-153	0.58504336
A1-T2 vs A2-T2			1292.7794	1577.8786	1.18E-18	-0.287510071

Table S1. Continued

Table S1. Continued

B2-T0 vs A2-T0	TC222167	Salicylic acid-binding protein 2	132.0006	207.3938	2.21E-12	-0.65182765
B2-T1 vs A2-T1			83.1868	300.1898	1.34E-62	-1.851448522
B2-T2 vs A2-T2			53.2122	118.6273	8.69E-11	-1.156605472
A1-T0 vs A2-T0			148.3976	207.457	7.39E-08	-0.483344484
A1-T1 vs A2-T1			201.3411	298.8161	3.83E-17	-0.569616364
A1-T2 vs A2-T2			62.9229	118.0009	3.71E-12	-0.907141333
B2-T0 vs A2-T0	TC204922	Chloroplast protease precursor	395.6007	693.5247	3.72E-54	-0.809902435
B2-T1 vs A2-T1			297.9724	815.6605	5.29E-120	-1.452790188
B2-T2 vs A2-T2			283.2612	367.7887	1.80E-05	-0.376744151
A1-T0 vs A2-T0			327.417	693.7362	7.17E-86	-1.083258182
A1-T1 vs A2-T1			509.3561	811.9281	2.48E-58	-0.672677411
A1-T2 vs A2-T2			273.8997	365.8468	1.41E-09	-0.41759202
B2-T0 vs A2-T0	TC226210	Pectin methyl esterase	15.4218	0	3.51E-20	8.268825501
B2-T1 vs A2-T1			55.5773	0	4.22E-47	10.11835229
B2-T2 vs A2-T2			71.762	0	2.76E-39	10.48707689
A1-T0 vs A2-T0			28.8464	0	7.70E-37	9.172246317
A1-T1 vs A2-T1			60.1557	0	1.26E-66	10.23255824
A1-T2 vs A2-T2			53.5482	0	7.04E-38	10.06469538
B2-T0 vs A2-T0	AM906901	Squalene monooxygenase	172.6244	0	2.85E-215	11.75342104
B2-T1 vs A2-T1			84.5485	51.5306	1.13E-05	0.714347985
B2-T2 vs A2-T2			91.6959	0	5.56E-50	10.84071385
A1-T0 vs A2-T0			191.9776	0	2.80E-239	11.90672236
A1-T1 vs A2-T1			116.4052	51.2948	1.62E-22	1.182270351
A1-T2 vs A2-T2			99.5732	0	6.14E-70	10.95961333
B2-T0 vs A2-T0	TC204524	Delta(7)-sterol-C5(6)-desaturase	140.2885	92.419	1.85E-08	0.602134252

B2-T1 vs A2-T1			39.6136	159.6605	1.19E-37	-2.010939248
A1-T0 vs A2-T0			1.4921	92.4472	6.01E-65	-5.953258426
A1-T1 vs A2-T1			0	158.9299	1.18E-180	-11.63417495
A1-T2 vs A2-T2			0	26.4884	1.04E-27	-9.049216923
B2-T0 vs A2-T0	TC196222	Delta(7)-sterol-C5(6)-desaturase	2.9849	1.4243	0.1349922	1.067444958
B2-T1 vs A2-T1			3.5475	0	0.0007274	6.148725944
A1-T0 vs A2-T0			78.5815	1.4247	2.94E-85	5.785459379
A1-T1 vs A2-T1			58.5932	0	6.50E-65	10.19459039
A1-T2 vs A2-T2			11.5062	0	1.36E-08	7.846271858
B2-T0 vs A2-T0	TC219291	Magnesium protoporphyrin	0	962.0126	0	-14.23184004
B2-T1 vs A2-T1			1.1825	396.4762	9.26E-234	-8.389255132
B2-T2 vs A2-T2			0	92.4183	2.98E-45	-10.85203465
A1-T0 vs A2-T0			0.9947	962.306	0	-9.918014671
A1-T1 vs A2-T1			0	394.6619	0	-12.94640164
A1-T2 vs A2-T2			0	91.9303	8.07E-94	-10.84439713
B2-T0 vs A2-T0	TC195026	Magnesium protoporphyrin	269.6324	0	0	12.39677823
B2-T1 vs A2-T1			195.7031	18.3032	6.69E-99	3.41849608
A1-T0 vs A2-T0			47.8414	0	1.88E-26	9.90211439
A1-T1 vs A2-T1			123.176	18.2195	2.41E-71	2.757167729
A1-T2 vs A2-T2			69.7012	0	3.94E-49	10.44504016
B2-T0 vs A2-T0	TC198553	Hydroxypyruvate reductase	175.1118	0	2.33E-218	11.77406081
B2-T1 vs A2-T1			60.8985	4.2238	2.06E-35	3.849786418
B2-T2 vs A2-T2			61.1306	0	1.40E-33	10.25575134
A1-T0 vs A2-T0			168.6021	0	2.86E-210	11.71940679
A1-T1 vs A2-T1			136.4571	4.2045	3.41E-126	5.020371574
A1-T2 vs A2-T2			78.1096	0	5.48E-55	10.60935651

Table . S1. Continued

B2-T0 vs A2-T0	TC199107	Hydroxypyruvate reductase	4.9748	155.7198	2.44E-101	-4.968178952
B2-T1 vs A2-T1			2.365	108.9746	4.97E-59	-5.52601327
B2-T2 vs A2-T2			0.6645	85.1084	7.32E-40	-7.000967864
A1-T0 vs A2-T0			2.9841	155.7672	3.23E-107	-5.705948372
A1-T1 vs A2-T1			0	108.476	1.54E-123	-11.08315978
A1-T2 vs A2-T2			0	84.659	1.79E-86	-10.72551973
B2-T0 vs A2-T0	TC201795	UPA18	0.8024	87.4372	1.52E-63	-6.767718196
B2-T1 vs A2-T1			2.7274	109.9805	9.60E-59	-5.333557999
B2-T2 vs A2-T2			0.5375	97.4596	4.24E-46	-7.502402289
A1-T0 vs A2-T0			1.7666	87.4639	2.20E-60	-5.629608679
A1-T1 vs A2-T1			0	109.4772	1.13E-124	-11.09641454
A1-T2 vs A2-T2			51.1459	96.945	1.63E-10	-0.922548476
B2-T0 vs A2-T0	DN906548	UPA18	84.6569	16.7574	5.51E-40	2.336828662
B2-T1 vs A2-T1			158.8732	0	6.14E-133	11.63365961
B2-T2 vs A2-T2			99.437	0	3.90E-54	10.95763941
A1-T0 vs A2-T0			93.6318	16.7625	1.65E-47	2.481759445
A1-T1 vs A2-T1			87.9713	0	4.19E-97	10.7808889
A1-T2 vs A2-T2			69.6526	0	4.26E-49	10.44403369
B2-T0 vs A2-T0	TC196756	Biotin carboxylase carrier protein	1.2037	53.0928	1.23E-36	-5.463024638
B2-T1 vs A2-T1			4.773	43.4573	1.68E-17	-3.186627116
B2-T2 vs A2-T2			0	25.1366	8.42E-13	-8.973647691
A1-T0 vs A2-T0			1.1778	53.109	9.91E-37	-5.494840123
A1-T1 vs A2-T1			0	43.2585	1.03E-49	-9.756838576
A1-T2 vs A2-T2			0	25.0039	3.29E-26	-8.966010173

B2-T0 vs A2-T0	TC213525	Biotin carboxylase carrier protein	30.0913	11.1163	2.98E-08	1.436670456
B2-T1 vs A2-T1			60.0036	39.7802	0.0017036	0.59299963
B2-T2 vs A2-T2			34.3998	0	3.07E-19	9.426257947
A1-T0 vs A2-T0			38.866	11.1197	6.82E-14	1.805392901
A1-T1 vs A2-T1			42.023	39.5981	0.6084145	0.085747823
A1-T2 vs A2-T2			43.0702	0	1.40E-30	9.750546729
B2-T0 vs A2-T0	TC209814	MADS transcriptional factor	8.8268	50.2722	2.30E-20	-2.509801409
B2-T1 vs A2-T1			10.9097	72.8746	6.64E-25	-2.73979855
B2-T2 vs A2-T2			11.8249	37.0435	2.67E-06	-1.647385534
A1-T0 vs A2-T0			0	50.2876	1.14E-38	-9.974058269
A1-T1 vs A2-T1			0	72.5411	7.35E-83	-10.50265509
A1-T2 vs A2-T2			0	36.8479	3.57E-38	-9.525437582
B2-T0 vs A2-T0	TC194763	MADS transcriptional factor	1.6049	0	0.0049203	5.004385284
B2-T1 vs A2-T1			4.773	2.0057	0.0825333	1.250778196
A1-T0 vs A2-T0			40.0438	0	9.80E-51	9.645435099
A1-T1 vs A2-T1			61.1874	1.9965	7.39E-57	4.937656994
A1-T2 vs A2-T2			67.6337	0	1.09E-47	10.40159842
B2-T0 vs A2-T0	TC212265	Asparagine synthetase	0.8024	59.2317	1.95E-42	-6.205839308
B2-T1 vs A2-T1			10.2279	190.8779	1.41E-89	-4.222070565
B2-T2 vs A2-T2			6.45	291.4968	2.85E-124	-5.49804369
A1-T0 vs A2-T0			0.5889	59.2497	3.53E-43	-6.652692292
A1-T1 vs A2-T1			0	190.0045	7.88E-216	-11.8918177
A1-T2 vs A2-T2			0	289.9577	7.19E-294	-12.50162662

Table S1. Continued

Table S1.

B2-T0 vs A2-T0	TC212265	Asparagine synthetase	99.1008	0	4.99E-124	10.95275252
B2-T1 vs A2-T1			87.9598	0	5.20E-74	10.78070072
B2-T2 vs A2-T2			282.1862	0	3.44E-152	12.46243156
A1-T0 vs A2-T0			81.8542	0	1.28E-102	10.67691333
A1-T1 vs A2-T1			516.5138	0	0	13.33459125
A1-T2 vs A2-T2			588.8506	0	0	13.52368594
B2-T0 vs A2-T0	TC216708	Receptor-like protein kinase	106.46	241.967	2.20E-35	-1.184498247
B2-T1 vs A2-T1			48.4823	109.8194	3.45E-13	-1.179601823
A1-T0 vs A2-T0			39.2908	242.0408	3.35E-97	-2.622988311
A1-T1 vs A2-T1			16.9269	109.3169	3.94E-63	-2.691123332
A1-T2 vs A2-T2			20.1359	96.0854	4.53E-37	-2.254546261
B2-T0 vs A2-T0	TC223930	Receptor-like protein kinase	26.8637	76.5939	8.31E-17	-1.511568277
B2-T1 vs A2-T1			7.095	23.935	8.12E-06	-1.754251949
A1-T0 vs A2-T0			109.9146	76.6172	1.28E-05	0.520642954
A1-T1 vs A2-T1			37.2393	23.8255	0.0009722	0.644319569
A1-T2 vs A2-T2			74.1267	22.3334	1.35E-17	1.73079309

Table S2. Allele specific differential expression analysis by pyrosequencing. The expression of the resistant allele (the allele highly expressed in resistant in SuperSAGE) is shown. Each candidate was tested on 4 different pools, each with resistant (R) and susceptible (S) pools at three time points (T0, T2 and T3).

		Pools_1		Pools_2		Pools_3		Pools_4	
Candidate	Time points	R	S	R	S	R	S	R	S
Delta(7)-sterol-C5(6)-desaturase (DSD)	T0	57.50	49.90	68.73	45.30	62.73	47.10	73.80	48.43
	T2	65.60	43.93	71.47	44.23	65.13	42.40	71.27	45.97
	T3	63.63	50.37	68.87	44.50	66.57	42.67	71.30	41.70
Magnesium-protoporphyrin-IX-monomethyl-ester cyclase (Desaturase) (MPP)	T0	24.37	22.17	30.20	16.87	37.60	20.83	45.63	26.67
	T2	31.33	21.43	39.00	16.20	34.80	24.13	39.23	22.67
	T3	31.60	22.87	45.87	11.00	38.03	26.53	50.97	23.77
Hydroxypyruvate_ Reductase (HPR)	T0	21.63	12.60	17.23	9.50	17.43	16.93	18.50	11.63
	T2	17.80	10.23	18.70	14.60	18.90	12.97	21.33	13.47
	T3	19.33	9.37	20.57	12.80	23.83	11.87	27.13	15.70
UPA18	T0	62.20	55.83	65.17	62.10	70.33	64.73	72.13	70.03
	T2	61.20	54.07	69.37	55.50	72.27	60.40	74.37	65.83
	T3	56.33	49.80	68.53	60.87	70.53	59.93	71.80	64.47
Biotin_carboxylase_ carrier_protein (BCCP)	T0	64.47	63.40	55.47	63.53	64.37	64.67	80.00	64.67
	T2	61.30	60.13	56.30	62.60	64.37	61.17	77.90	61.60
	T3	61.03	56.80	53.47	57.23	61.80	59.30	78.73	59.83
MADS_transcription_ Factor (MADS)	T0	50.23	49.23	48.23	49.17	52.83	49.47	49.83	54.00
	T2	50.20	51.63	52.20	50.37	51.67	51.23	55.73	47.90
	T3	49.60	47.63	54.67	51.57	54.10	51.40	52.50	51.17
Asparagine_ Synthetase (AspS)	T0	79.17	58.53	78.43	55.07	59.10	71.97	89.23	97.60
	T2	77.73	66.80	86.40	71.07	63.50	79.77	88.87	90.37
	T3	78.60	69.37	90.30	69.87	69.00	81.20	88.40	97.70
Receptor_like_ protein_kinase_SNP1 (RLPK_1)	T0	36.00	39.63	34.93	37.90	36.97	33.57	30.23	26.67
	T2	37.30	44.40	32.20	48.40	40.50	38.77	30.50	26.23
	T3	32.87	37.27	31.30	38.30	33.17	32.90	26.97	26.17
Receptor_like_ protein_kinase_SNP2 (RLPK_2)	T0	80.23	78.60	82.30	78.50	76.73	82.23	79.37	83.93
	T2	78.90	76.10	82.17	75.87	73.87	81.43	80.43	83.73
	T3	83.43	80.33	82.97	80.30	81.13	82.33	84.30	82.10

Fig. S1. Amplicon sequences of 15 candidate genes analyzed in association and linkage mapping. SNPs are in parenthesis, significantly associated or linked SNPs highlighted by green, forward primer by gray, reverse primer by blue. The primer used for sequencing is underlined. Exon regions are in red while introns are in black.

1. Amplicon sequence of *HPR*

```
TATGATCAGGGAGTACTGACAGAGACTACAGCAGAGTTAGCGGCTTCACTTTCTTTAGCTGCAGCAAGAAGAATTGTGGAGGCAGATGAGTTCATGA
                                                          snp53 snp54
GAGCTGGCAAATATGAAGGATGGCTTCCACATTTGTATGAAAATCTTCGATTCTATTTATT(G/C)(T/C)GTCTTCTCAGGTCTATAACTGAATTC
                      snp102  snp105
TTTTGGTCTAAAAGTTCAATG(T/A)TT(T/G)CATTGCCAGGTTTGTGGGGAACTTACTTAAAGGACAAACGGTTGGCGTAATTGGGGCTGGTCGT

ATTGGATCTGCTTATGCTAGAATGATGGTAATCAGTTAGTATATATTATCCTCTAAGTTTGCTTTGAAACTAGGTTGCCTTGTGCTTTTTGCTCTAC

ATGTTACAAAACTTCATCTTTGCTGAAACTCATGAATAAGATATTGTGTGTAAATCTGCTATGTTAATTGGAGCAACTTTGGTGTTTTATCGCAACA

GGTTGAAGGCTTCAAAATGAACTTGATATACTACGACCTGTATCAATCTACTCGTCTAGAAAAATTCGTGACAGGCACGTTTTAATCAGTCTGCTAC

ATTAGCAAACTTCTCAAGCTCATTGCATGTACTTATGTTTTCTCGCCTTTGCAGCCTATGGTCAGTTTCTGAAAGCCAATGGTGAACAACCTGTAAC

ATGGAAAAGAGCTTCATCCATGGAGGAGGTCCTCCAAGAGGCTGATGTGGTAATTATCTACATTGAGAGTGCACAATTAGATTTCTTATTCACTTAT
                                                                                  snp285
CTGTATATACATATTTATTGGATTCTCATTCAAGTTCTTATCTTGATTCTTGATCAAAATGCTTTTCTAGATAAG(C/T)CTACATCCAATCCTTGA
        snp261 snp258
TAAAACAAC(G/A)TAC/T)CATCTTGTGAACAAAGAACGGCTAGCACTGATGAAGAAGGTATATCATTTGCAAGAATATATCTATTTTGCTTGCTG
                      snp157              snp142      snp134 snp133
ATTTACTCAAAATGTGATAT(C/A)TTGATATTACTTGG(C/T)CCAAACT(G/T)(T/G)TAGGAAGCAATACTTGTGAACTGTAGTAGGGGGCCC
    snp93
GTG(G/A)TTGATGAAGTAGCTCTTGTTGAGCATCTGAGGGAAAATCCCATGTTCCGTGTTGGCCTTGATGTCTTTGAGGTAATAACAATAGTAATC

TGTTCCTTTCTGTTGATCCCATTCTCCCGTTATGAAGGTGTAAAAGGTGAATCCCAAGGTCAATGAGATACTGTTGTGTCACG
```

2. Amplicon sequence of *PME*

```
ATATACATTTATTAATATTGATGAGTTGTCCAAAACAGAAATGACATAGTAACTAATTTCATGCAGCCGCGGTGGGGGAAAAGTTCTTGGCCCGGGA
                                                                  snp82          snp93
CATAACCTTCCAAAACACAGCAGGAGCCTCGAAGCATCAAGCCGTGGCACTTCGCGTGGGGTCTGATTTGTC(C/A)GCATTTTATA(A/G)ATGTG
   snp103                                                                 snp166    snp172
ACAT(A/G)TTAGCTTATCAGGACACCCTCTATGTCCACTCCAATCGTCAATTCTTTGTACAATGTTTAGT(T/C)GCTGG(T/G)ACCGTTGATTT
      snp190                                                          snp247
TATTTT(T/C)GGAAATGGTGCTGCTGTTTTACAAGACTGTGATATTCACGCCAGGCGTCCTGGTTC(G/A)GGCCAAAAGAACATGGTTACTGCCC
                                                                  snp344
AAGGAAGAACTGACCCCAACCAGAACACTGGTATTGTTATCCAGAAATGTAGAATTGGTGCAACGTCTGAT(T/C)TGAGACCGGTGCAGAAGAGTT

TCCCTACGTATCTAG
```

3. Amplicon sequence of *RLPK*

```
TCTAGTGGAAAGAATTCGATTCACATTTCGTACCATGCTTTATAATATAGCAGAGGAAACTTCTGCAAAGAACTTTCATACAGCTAAATATAAACTT

CTCATTTTGGAGAGGAAACAAAACACACAAGGACAGTCTCTCATATTGGAGATGTAACAGAATACACAAGGACCGTCATACACAAGAAATATTCTCT

TAGTTTCATGACAAACTTTTATTTACTGTAAGTTATCCACTACATACTTAATGCTTGATCTGCATCAACATGCACCTAACATCCTTACTATTAGGCC
                                                                  snp735
TCTCCTTAGGATTCTCCAGGGTACAAAAACAGGCAATCTTGAGAACCAAAAGCATTTGGTCCTCATTTCCATT(A/G)CCCATCAGCTTTGGATCAA
                  snp689
TTGCTCTGTTTGGATCCTCAGAAG(T/C)CATGACATTTCTCATCCATTTCACTAAACTCATCTCAGAAGTGTTCTGGAAGAACTCATCAGATGGAA
          snp600                  snp585                                  snp551
GCTTTCCCATAA(C/G)TAGCACGCCTAGCAGCACCCC(G/A)AAGCTGTATATATCACACTTATCAGTGAACTT(C/T)AGTGTCTGATGATATTC

TGGAGCGATGTATCCTATAGTTCCTGCCACATTTGAAGTTGTAATATGTGTATGAGCATCTGGGACAGCCTTTGCAAGGCCAAAATCCGCAATTCGA
                  snp413                                                  snp380
GCTTCCATGTCATCATCAAGGAG(G/A)ACATTGCCTGGCTTTAGATCTCTGTGAATTATACGCTGAGTATGATT(T/C)ATATGGAGATACTCGAG
                                  snp311                  snp290
TCCAGCAGCTATTCCCATTGCAATTCGGTGACGTGC(A/T/C)GACCAATCTAGTTCCCGTGT(C/T)CCCTCTCTGACTTGCTGGAGGGTATCCTG

TAAGCTCCCATTTTTCATGTACTCATAGACCAAGTAATGGCAGTCTGGTCTTGGCATATGTGCCAGTAGGGGAAGCAAATTCCGGTGTCTGATTTGA
                                  snp121
CCTACAATTTTGATTTCTGATTTAATCTGGCGCATTTTCTTA(C/G/T)TCATAGCCTTGCTATCTTCCTCAGTGAGTTCTGCAGCATCCCTTGGTG

GTTGTATGATCTTCTTTACGGCTATAATCTTTCCGTCACTTCCAGGTAACGCAGCTTTATAAACTTTTCCACATCCACCTTGCCCAATGAGTTCTAG

TGATTCCAATCCATCTTCGTTCTCCAGAAAGGCCAAGTGCTCGG
```

4. Amplicon sequence of *DSD*

CGCCCTAGCCCCTCTTCTTTTTCCTCTCCATCTTCTCACTCCATACTTCCTCCTCTTCTGGTGGACCATCGCCGCCGGAGTCTGGTAAAACCCGACC

CGACCCGGATATATTTCTCTTTCTCCGAGTATACCGTCGGAGATGGAGGATTACTTGAAGCTGTTCGTGGAGGAGACATCGTTTTACAACCGTATGG

TTTTGGGTACATTCTTGCCGGAATCATGGTGGGGACCACTTCCTCATATGCTTCAAGGATGGCTCCGTAACTACATTGGCGGTGTTTTACTTTACTT

CATCTCCGGTTTCCTCTGGTGCTTCTACATTTATCACTTGAAACGCAATGTCTATATTCCCAAAGGTATGATTTTTTCCCCCCTTTTTCTCCCCTTT

TTTTTTTTTGAATTTCGTTTGTGTATAGATGATTTCTTTGCATTATTTGAAAATTATAAGTTCATGGAATTAAGTTGACTGTTTTAGTAAAGCTTTT

GAAGGTTGGTTTCTGAATATTCATTCTTTAGATTTGAATATTTTGGCATTTTTTGGGTGTTACTGTATTTGAAGTATGTAAAGAGTAAGAAAAGGCC

AAGCATGCATATTCCCATTTTATTGATATGTATGAACAATTACTGTGGATCTGAACTATCAGGTATCAGTAACTCTGTCCACTAAGGTTAGAACACA

snp248

CGAGAAGAAATAACCTAGTGTTTTGTCTCTGTTGGTATTTGAACCTGAGACCTTATCATGGTCAACCCACTTGATTGACAGTAGGCCACATC(A/C)

snp181 snp160

TTGTGTATATTTGTATTAATTGATGGCCCTCTTAGTTCAAACTAGGGATGTCAATATCTCAATGAA(T/G)TTAGCAATTCCAGTTTTTCA(T/C)G

snp157 snp111 snp101 snp100

(C/T)TTCGATAAAGATAAGAGTAATTAATGGTTTGTATTTTCTCCTGATG(T/A)ACATTTTGA(T/C)(G/A)GCAGATGCCATACCATCAAAGG

AAGCAATGCTCTTGCAAATATCAGTTGCTATGAAAGCTATGCCGTGGTACTGTGCCCTTCCATCACTTTCTGAGTACATGATTGAAAACGGATGGAC

CAAATGTTTTGCGAGAATAAGTGATGTTGGATGGCCTACCTACATCATCAATGCGGCTATTTATCTCGTAATAGTGGAGTTTGGAATCTACTGGATG

CACAAGTTGTTGCATGACATAAAACCTCTGTACAAATATCTGCATGCTACACATCATATTTACAACAAGCAAAACACACTTTCCCCGTTTGCTGGTA

AGCTCTTAAGTCTAGTCATACCATATTGGTGCTTATTTCCGAAAACAACATCTAGCTCGTGGAGCCATTAGGTTTCTCTGAGATACTCTATTGATTG

TTTGTTGAATTAATGAGCGTTTCCTGTGTGATGACTAAAAACGTTGAAAACCTAGTCATGAGGACATTAAATTAATATGTACATTTTAACATCCATG

TCTTTAGTCATGCACCCTTTTGTTTCGTGGTCAGGAGTGTCGACTTAGGTATTTGGTTAGACTAGGGAATAACTGAACACATTGGAATTTTGTCTCT

TGCCTTTTAATTGTTTCTGTATGCCAAATTCAGTTTACTTGAAAAGAGTGGAGTAGCGTTTGATAACTTTTTTGTTCTTATTTTCTCCTTTCATTTT

GTTGGAAACTAGAACTTGTGATAGGATTGAGGAAAGAAAAATATTAAGCAAATTTCTCTGAATGTGGAAATTCGGAAAAATTGCTACAAGGCATTTC

CTATAATTCCATCACAAAGAGAAAAAGACTGTATGTTACAACTAATGACATTGTCCCTTCGATTCACACCCACAACATGCATTCATGTTTATTAAAT

CGACTCTCTCTGCATACCAGCCCCTAACCAACGCACAGCAGCATAAGCCCTCCCTATCTTCTCCTCCCTACCTTTTGTAACTTGTAACCATGCTATC

ATCTCTGACATTCAGAACTTTCTTTATAACTCTTTGCAATGCATATCCTGTTCTAGGAAATTGGTTGCTAAATCATCACAATATTGCAAGCTAGCTA

ATTCATGATGTCCAGTTAATTTATGGTTTGGTATGTGTCACCTATGTGTTGTAATTTTCTGACACTCTATCTAGCTAATTTGACCTCTGAGAGTGAG

snp357

AGTCGTAGCTTCTTTTCTGGTACATTTTCAAGCTAATTGCTAA(A/T)CATGTTGTGCTATTCAGGATTGGCATTCCACCCATTGGATGGAATACTG

snp290 snp284 snp245 snp236 snp235

CAGGCGGTGCCACACGT(T/G)GTAGC(T/C)CTATTCTTGGTGCCAATGCATTTCACTACACACATAGC(G/A)CTCATATT(C/T)(G/C)TGGA

snp218 snp191 snp163

AGCCTTATGGAC(G/A)GCTAATATTCATGACTGCATACATGG(T/G)AAGGTGTGGCCTGTAATGGGTGCCGG(C/T)TATCATACCATTCACCAT

snp83

ACAACATACCGCCATAATTATGGTCATTACACAATATGGATGGACTGGATGTTTGGAACTCT(T/A)CGTGATCCCGTTGAAGAGGATGCCAAGAAA

snp31

ATGTAATGTGCATGATTTGCA(A/C)TCAAGGATGTTCCTTGTTGTTCATTGTCCTAGTAATTTGTCTCGTGTTTCCATCGTTGGTGGATGTAAATT

TTG

5. Amplicon sequence of *CCAP*

TGAAAACAAGGTGAAAATTCTTCTCTGGCACCTCAGTTCTGACCCCTTGGGGCTTTTGCTAGTTTTTACAAATTAAAGAATCAAAGAGATTTCTCCA

ACTTTTTTCAATTGTTTCTCCAGCTATTTTGCCTGCAGATCTTCTGAACAAAATGGTAACTTTTTCCTCTTCCCTTTGATCTCACTGAAAATTATTT

GGTAGCAGAAATTACAATCTTTTTGTTTTCCATTTCATGAATTTGCTGTTTTCATTTCTCTTCTACTGTATTGATATTTGATCGTTTGTTTTCAACT

GGGTCATTTCGAAAATGGGTAGATCCGTTTTATATTGCTTCAGAACAGGCAAGGAAAGACTCGTTTGGCTAAGTATTATATACCTCTTGAGGAATCC

GAGAAGCACAAGGTGGAATATGAGGTATTTTAGTTTTTTTTTTGCTTTTTTCTTAGTTTGATTATGTGAATTTGGTGATTTTAGATTATGGGTTTTG

snp336 snp318

ATTATATGGTCAATTTGGATCATGTAGGTTCATCGGTTGGTGGT(G/C)AATAGAGATCCCAAATTCACCAACTT(T/C)GTTGAGGTAATGAAGTC

snp291 snp288 snp281 snp212

(T/A)GA(A/T)TTTTTT(A/T)ATTTCAGTTGATGGAAGCTAGCATTTTTCTGATTAAGCCCTTTTATATCTGTGGAATTGAGTTTAACT(A/T)G

snp188

ATCAGCTTGAAATGCAAAAGGC(A/G)AAAATGTGATTATCTGTTTTGTTTTCGATCTATAGAACACCATACTAAGCGAATTTGCTGCAATTGTGCT

snp80

AAGGCAAAGGGGAGACTCTTTCGCTAGTCTTTAAT(T/G)TTCTAAAAGTATTAATGTTTAACTATCTTATCGTATTTATCTTCCATTGCATCATAA

ATACTTGTTATCAAAGGAATAACTTTGTGAAATGATTGAAGGATAGGGAGTGGTGGGGGATTCAAGAACAGTTTATTGTTCAAAATTTAACTTATGT

GTAGAGTAACCTATGTTATGGCTGAGTATGAA

6. Amplicon sequence of *SMO*

CTACTCCATTGTCAAACTCAAAAGATGATTACTTCTTTGGTATCAGTAAGAGGCTTACAGTTATTGGCAAAACTAAGTGAACCGCAAGCCAATGTGG

GAGAACACAATAAGAAAAGGCAAAAAAGAAAACAAAGAGGAGGAAAAAAATAAAAGTACAGAGAAGAGACAAATGTATGATTGTACAACTTATATTT

snp205

ACTTTCCAGTT(A/C)ATAGTGTAATCAATCATGGGAACAAACTAAAAAGGAAAAAGATTACAAGGGATCACTAGCAACCAAAAATTATGCAACAGT

snp104 snp63 snp53

TGGACTAAACTCTTGAAACTC(A/T)CAACCAGTATACTGATGGAGGTG(G/A)TACAATG(G/A)CATAGCCTTATTCTGTTTATAGCAGTTACAA

CCTCACTTCAGGAGGGCTTCCTGTAATATGCAGGAACAGTTGCGGGGAAGAACATCTGCCTGACCCCTTCTGCCTTGATGATAGGGAAAATGATTGC

AGATGCACTCTGCATTTGAATCAACATGAAATTGTCAAAATACAAGGTCACATGGGAACACGTGTGTAAGATTGCAAATATTGAGCAAAAGCAAGAT

TGACACCTACCGATATTAGCCGGGCTCCAATCCATATACGTTTGGGTGATGGGAAAGGAAGCAGCAAGCGGCCAACACCAAAGATAGCCACAGCAAA

GAAATGACAAACGAGACTTAATGGACGAGGGTTTAAGCCAGAAAGCAAAGATACTGGTCCATCTGAGAATACTCCACCGAGGCTCAAATAATCAAAG

CATGCTTCACGCATCTCCTTCCTAGCTTGATCAGGTGAAGCACAAAACACCTTATACAAGGCACCAGCCAACGTATTGATGGTGGAGGCCACAGGCT

GCAGGTAAAATATGCGAGCAGAACAAGTCAATATCTATAAAAGTAATTATTTTTTTATTTTTTTTTGATAAACAATCTATAAAGGTAATAATATTCT

AGTTAAGTCTCAAAGTGCTGCAGGAGAATTGAGTCATCAGTCATCCTCTTTTGCTCTCTTTTTTATTTATTTATTTATGTAAAAGACAAGTTATAGG

AGAATTGAGTGAGTCAAGTCCTAAAACATTGTAATGAACCAAGCCTTTTTACATTATATCGTTTTTGGTAGATTGATAGCTGGTGAAATCATCAAAG

GAATGTCTGTTTTGTGGTTCCAGATCCTTCATTACTCACTAGTCAACATAGAGTTTTTACTTCACTGAGGATTCTTAAAGTAGAAAACGTGGTACTT

ACAGCAATCAAATATGTTACTTATGCAAGGAAGGATTTCAAGCGATTTTTACCTTACGCAAGGTGTAAAATGACTCCAGATATCTGCAAAGAGTAGA

TGCATCATTCAAATCACGAAGAGGTTTGAGAAGATCACGCAATACAACAATATCAGATAATGCTACAGTCATTCCTCCACCAGTCAAGGGGTGGCGC

snp190 snp205

ATATTGAATGCATCTCCCATGAGAAGAGCACCAGGAGTAGGATGAGGAG(C/A)AGCTGGCATGCTTC(G/T)GTTTGGCATTGTCCTGATGTTTCC

snp245 snp247 snp297 snp324

TTTATCAATTGC(G/A)G(C/T)AAAGGGACCTGGAACAACCACAC(G/A))AACATCTCATTGAACATCATGCAACT(T/A)AAGTATGTATACTA

snp353

CTATAATGACATGG(T/C)ATGTCATTCAAGTTTTCAACAATTCTTTACCTGGGGAGCAACTACACTTTTCAAATATTTGGCCATTTCGCCATTCGA

AATAGAAGGCACTTTTTGACCAGGTACATCAACCAGACAGCGGACCTCAGTACTGCTTATGGGATAGAACAAGATAGGTGATGGATCAGCCAAAATG

ACATGTCCATGATTTTCATGTGGAAGCTGGCAGTTCTCCAGGACCAGACCAACAAAACAAGAAGGGACTTCTACCTGTACAATATTACATGTCATTC

AGAGTGTTGAGAGAAACATAAGGAGGGGACAACAATGATGATGTTCATATAAAAATATTTGTTCATTCAGACAACTTCGGTACACTAATTCTTATTG

TCTCATACAGTAGTAAGTAAAAGCAAAAGATGATATACAATAAGTAAAAGCAAACGATTTCACAGATGAGCTTTTTAAACAAAAGATCCTTCTCAAC

AAACAATTAATCACCCAGTTCCATCCTTCTTATCTTAACAAAAATCAAGAGAAATATTAATAAGCGCCCTTACCTTTGGGTCGCAAAGGGTACGTCG

TAGATTCGAGAAACAACCATCACATACTACAGTTAAAGGTGCATATGCTTTCAACTCTTCACCAGATTTAGTCTTGTACTGAACACCTCTAATGATC

CCGTTTTCTTCAAGCAGAGACGTAACAGTGCCTTGCTCTAGTTTGACACTGTTAAAAATCAAAGCCGATGTGCACTTAGTCAATCAAAGCAAAGAT

AACTCAACAGGAAACTATTTCAAATTTTCAATCAATCA

7. Amplicon sequence of *BCCP*

```
AAATGGTTTTCATTGGAGGAATTCTTTCAAATATTGTAACCTTTGTGAAAACCTTTTTGGGACGATCTGTTGCCTTCAAGGTTGCTCTTGGTGGATC
                                                                                snp60 snp61 snp63
TTCAAATAATTCAGCTACTCCTATCACCCAGTCAGAAGAGGTGAAATCACCAAGTGAGACTTCATCTGCAACT(T/C)(C/T)G(G/A)TATCAGAG
          snp82                                                          snp130 snp131
GAAGCAATA(T/G)CTAACTTTATTAGCCAAGTTTCAAGCCTTGTCAAGTATGCATTCCCA(T/A)(T/G)TTTTTTTCACATCTCTGAAGAGACTT

TTCATGAATAATCATGACAGACTGTCCTTTGTGCAGGCTAGTTGATTCTAGGGATATTGTGGAGCTGCAGTTAAAACAACTTGATTGTGAAATTCTA
                      snp275                                                    snp321              snp339
ATCCGTAAGAAGGAGGCACT(A/G)CCACAACCACCTGCTCCTGCTCCTGCACAAGCTCCATTTATACAA(T/C)CATATCATGTACCATCT(G/A)

TACAATCTAATGCACCTCCTCCACCTGCACCTGCACCTGCTCCAATTCAAACCCCTGCACCTTCTCCAGCTGCAGCAAAGTCAGCTGATTCATCTCT

TCCACCGCTCAAATCCCCAATGGCAGGGACATTCTATAGAAGTCCAGCACCCAGTGAACCAGCATTTGTGAAGGTATTGCATGTCCATCTTTTATGT

AGTAAATTCAAATATGGAAACCTCAAAGATTAAAAAAATTGAAGTTTATGGTTTAATGCGCTCTCTCAATGCATTTGGTTTCCATGATATGCATTAC

AACAGGTAGGAGACAAAGTGCAGAAGGGTCAAGTTCTTTGCATTATTGAAGCGATGAAGTTGATGAACGAAATAGAGGTAACCTTTTCCTTTGGCTA

TTGCCATTGCCTGCTATATTGTTTTCTTGTACTTATACGAGACCTCTATTTCATGACTTTTAA
```

8. Amplicon sequence of *Chp*

```
                                                                                snp552
AATGGCTGCTTCATTTAAGAGGTTAGCAAGATCTGCTCCACTGAAACCAGGTGTCCTCATGGCTAT(G/A)ACTTCAAGAGACACATCTGAATCAAA
                                                                          snp452 snp451
CTTCTTGTTGCCGGCATGAACCTTTAAGATCTCTGTTCTTCCCTTGATGTCTGGAACATCCACGGATACCT(A/G)(G/C)ATAAGAGCTGTAAATT
                snp418 snp415
TTAGTTTCTTGCAAAA(G/T)TC(T/A)CAGTTAAAAAACGGAGGCAAGGACAGATGCTTACTTGTCTATCAAATCGTCCTGGCCTCAGCAAAGCAG

AGTCAAGAATATCTGCACGATTGGTTGCCGCAACAACTATTATACCAGTATTTCCTTCGAAACCATCCATTTCAGTCAATAGTTGGTTCAGGGTCTG
                    snp230
TTCCCTTTCATCATTTCC(T/C)CCTCCAATTCCAGTCCCTCTTTGCCGCCCAACAGCATCAATTTCATCAACAAATACAATGCAAGGAGCATTTTC
                                                                          snp86
CTTGGCCTTCTTGAAAAGATCACGGACTCGAGAGGCTCCAACACCAACAAACATCTCAACGAATTCTGA(G/A)CCTGAAATTGAGAAAAATGGAAC
   snp59
ACC(T/C)GCTTCACCAGCAATTGCCTTTGCTAGCAAGGTTTTCCCAGTACCAGGAGGACCAACAAGAAGAACACCTTTTGGAATACGAGCCCCCAC

TGCAGTAAATCTCTCAGGTTTCTTCAAAAATT
```

9. Amplicon sequence of *AC9*

```
TTCAACATCAGAACCACGGGCCGACATATTTGGTCTGTCTTCTGGAATCTGCTGGACACATAACAGACTTACATGTATGCACCGCGATATTTCCAAC

ATAGAGCTTGTGTCCTCTAACGAGGAATCAACTAACGCGAGGGGTGTCCCTTCCGTCCACATTCTCCAAGCCTGAAAAGGGGAATAAAGAAAGAAAG

GCACATTACATAACCAAACATTCAAACTTGGCTTCAGATGACAATTTAGCACTCCAATTTTGACTATGCACATCTAGACAGTTCAACTCGTCTCCAC

TGTGTCAGTTGAACACTCCAACTTCGTCTTTACTGTAACCTTGTTCTTTACTTCCACGGATAGTGACTAGTGAGATATTAACTCATTGACTCTAAAC

TCATTCCAATTATACCTATATTCCACACGCTGAGAAATAAATGTCATTGACTTGCAATGCTTGTGATTTCTCATAAGAAGAAGCATCAAAACTCAAA
            snp697   snp692
AGAAAGAGAAAGA(T/G)TTTG(C/T)GTACAACAACTACGCCTCAATCCCAAACAAGTGGTCAGCTATATGAATCCCCAGTTCTTCATTAAAGCTC
                                                                          snp552
ATTTTCATATCCTCATCGTGTGCATTAACGAAAAAATCTAATAAAAGTTTCAGCTTAGAATTTATCAAA(C/G)AGTGAGTTTCAGACATTGAATGG
   snp525
TTA(A/G)ATCATAAGGAAAAAAAAGTGAAAAGACCCTTGGTAGAGACAATAGAGAGAACTGATGAAATAAGCTGTTTCTCCATAGATACAACTTTC
snp435                                                                    snp370
(C/T)ATATATCCCTATCCATCAATAACGGAAACCCTTTACACAGCTCGTCTATCAAACCAACCCCAC(G/C)TAATTTCTCATACAAAATAGGTAA
snp346                    snp321
(C/T)AAATACTCATCCGTAGAGACAGTT(A/G)AGATAACTGATTAAACCTATTGTTTCTCCATTAGTACAACTTTCTGTATATGTTTTTGCAATG
                  snp238 snp237
TATTACATGTTTACATAGA(C/T)(A/G)AAAGTATAAAACGACAAAAGTTACAACGAAAAGTTGAATTCTTGGTAATGAGATTCTTTCTAACAAAA

ACTGAGACTCTCAAATGTGATGACCCCTTCATCCCTATCCATCAACAACTGAAACCCTTTACACAGCTCGTCAATCGAACCATCCCCACTTAGTTTT

TCTCATCCATATTAGGCAACAAATAGTTGAACATGATTACAATAATTCAGTGTATAACTATAAGAATTTTATCATAAAATGTGCGGTTCTGCTCTTA

TCTCGATGAGATAATAGCCACAAGATTTTGCTGTCACTTA
```

10. Amplicon sequence of *MPP*

GGAAGATGTTTTGTCAAGTGCATGTTGGCCCTCTGAGTCTTGACTCGTCAAAGAAGAAAACCTTCATCAATTGAAGAATGCCTAATAAATGAGACCA

GTGAAGATGAAAACTTTTAATGATGTACTCACAAGAAATTCATTCTAGATAATACAAGGGAGTGTCCTCTATGCAGGAAACTGTTTCTTGAATACAC

snp308 snp286 snp262 snp243

TATA(T/C)AACTTACAACAATGAACATG(G/C)ACCTCTATGTGAAAACAAAACAA(T/C)CGAGATTCTGTTTCTCGT(G/A)AAAAGGAAATTA

GTAGGTAAGTTGTGGTTCAAACTCTGCAAAATCAACTGAACCTGACTCAATAGGTTTCATGAGATACGCAGCCAATAACTCAGAAGCCAAAGCTGCA

ATCAGAGGGATCTTTTTGAAATTCTTCACCAGTGGAATCTCATCACTCTCACTAACAGCAATTAGTTTGGTGTTAATCTCCACCATCCTGTCCAACT

TCCTCTTGAATTCTGGGTTCTCAACATCAAGGACAGCAGGGAAAATCCTTGCAGTTGTGCGGTTTGTCTGCACAAGATAAGATTTTTAGAAAGGTTA

GTGAAGCAAATAAATT

11. Amplicon sequence of *SPI*

ACATTGGTAATGTTGTATATTCAATCAAATACCATATATATATATATAGTTGATTATATTATATTTCAACAAGTCGATTTTTATCTTCTTTTTTTTT

TTTTTGGTTACAGGTGTTGCAGTCTGGGCGAGATTTGGACAGAGGTTGGCCAGTGATTTTCAAACCAAAGGCAGCGAAACAAGTAGAAATAACTGAA

TCGAGCGATGTTAACATTGAGTTCTATATTGACAATCCAAGTGGGTTTTGTAACAACACAGTGTGGCAAGTAGAAGGTTTCCCAGGACATGCTATGC

snp369 snp368 snp364 snp350 snp349 snp342 snp310

CAATGTACTTGG(A/C)(C/)AC(C/A)AATGGGG(A/T)(G/A)GCTGGACATGTTAAGAATGTGGCCTCATGGTTTCAAATTA(A/T)GAAA(A/

snp306 snp298 snp297 snp266 snp264

G)CTG(G/A)AAGTTAC(A/T)(T/G)GTATAAGTTGATGTTTTGTCCTTATGGAGA(G/A)C(C/A)CATCTGCACTGACATTGGCATCGACTATA

snp230

CTGC(C/A)GGAAGGCGATTGGCCATAGGAACAGGCAATACCTTCAATCTTGTCTTCATTAAGGATACTAACATTGGAATTAAATCGATTGTATAAT

snp83 snp72

TTGGTTAATTTCGACCAAATAATATGTATCGATCCTTTTCACTGTATTATTTTCATCA(T/C)GTTTGTACGT(A/G)CGTACCTACCTACTTGTA(

snp53 snp51

T/C)T(C/G)TCTAAATAAAGCTACGCATGTGTGATGTGTAGCCTCTATCGTTGTAAAAGTATTGCTACCTTAATGTATCCTTGAGTTTGTTTTTTT

12. **Amplicon sequence of *AspS***

CCAAAGTCCAGAAGGACACTCTATATAGTGTCAATAGAGTTGAGGGTAAAAGGAAAAACAGTACAATTATCCACAAACAGCAACTAGATATCAACTT

AGAGCCCCCACAACTTGGAACATTAGTGGTGCTCAAGAACTTCAAAACCTCTATATGGTTTTGATCTTTTATTCTAGTTAAAACATTGACCCAATGC

TCTTTTATGCAGCAGGATATCTAAAAGACGTGCTGATCCTTCCTATAATAACAACAACAACAATAATACTGCTAATGTTTCAAACATCAAGTATGGA

snp175 snp162

CAGCAAGAGTAGTGCTAGTTAGCTCCTTATTGTGAGCTCTGCAGCCACGCC(C/A)ACGCCTACCATC(T/C)TTGGCACATTATTGATGATCGGGG

snp119 snp112

TGTCCAAATTCCCATTAG(C/A)AACACT(A/G)GATAGATGATTGTCATAAGCAGAGTTATGTACACCGATAGCAGCCCTACCGGAAGGATCAAGG

TTGTTCGACCAAGAAGCATCCCACTCAATTGCTTTTGCCGTGCTGCAAGCTATACTCGGTCCTCCAGGAACGGTCAGGCTTGCTGAGTTCTGCAAAT

TCAACAACCGATATGGTATTAGTGTTTCGAGACACCTTACTACATTTGGGAACTCTACAACTTTAAGCTTTTCATTTTACCCTGGATGAAATATTTT

CGTATATGGTACTAGTGTTTCGATTTTAAGAGCTCGCCTGCTTAATATTACCTGTGGGAAGAACCTCTCGAAAATCATTCTGTAATAGTATCCTTCC

TTTGTAGTCGGAGTGTTATGTGGGAAGATATGAGCAGCATTAAGCATCATCCTATCAGTCACCTAAGAGACACAAATCGTATACAAGATCAGCTGAC

AGAGTGAAAAAAACAACGAAAACTGAACGATAATGCTTGGAATAGTTTTGGATTTGTGTAGTAAGCTCACATGTTGTTCAGCATGTGCTTTGAGGCC

ATCGATCCAACTATAGCCTACGCCATCGCTGAATTGTTCTTTCTGTCTGTACAGAATATGCTGCAAAGGCACAAAAAACAGGACAGACATGTTAGTT

TTAATCCATAACTACTGGCAAATCAGCACGAAAAATATGACTTCCAACCTTTGGAAGGTACGGTTGCTCCTCATCATCAAACGCCTTCCTAAGAACC

CACTTCTCAATCCTTCCTTGATCATGCTTAATCTAAAATGGACAAAACAAGTTACAGTCAAACCCAAGTACAACACTGTTCCGGTGTGATATCTTTT

GGTTGCTATAGCAAAATGTTGTTATAGAGAACATATAATATAACTTAACATGAATAATCTGATACACAAAAGAACATGGTTGTTTTAGTGAAATGTT

GGTATAGAGGATGATTGTTATAAAAAGGTCTGACTGTACAGAGAAAGCATATCGAGTTATACGTCTATTCTCATCAATGATTCGAGCTTAGTTTCTT

ACCATCTTCCATTCGGGATCGATACTCATGGCAACATCGATGAACTCTTTATCCAGAAATGGTACTCTAGCTTCTAAGCCCCACGCGGATGTAGCCT

snp17 snp57

TGTTTGCTCTTAAACAGTCATACTGGTGAAG(C/T)GCTTTTATCTGCGCGTATAAATGAAGAACAAGATCAGCA(T/G)TTTTTAAAACTCGTCCT

snp82 snp83 snp115

GTCTCAA(T/C)(G/T)AGGTTCGAAACAAGAAGGCAATATTCCAAGAA(C/T)GATTATTACCTTGCGACATGTTTCCGTGTGGAACTCTTCCTTG

TTGGGAGCCTTGTGGAAGTACAAGTAACCACCAAAAATTTCGTCAGCGCCTTCCCCTGATATGACCATCTTCACTCCTAGTGATTTAATCTTACGCG

ACATAAGGAACATAGGAGTGCTGGCTCTTATTGTTGTTACATCATACGTCTCGATATGATATATAACATCTTCAATAGCATCAATACCGTCCTGCAA

ATATGAAACGCGTTGTTTGACAACATGAAAACAGAAGGTAATCCCGAGATCAAATAAAAAGTAACACCGCGATTTACAAACCTGAACAGTAAAGTGA

AACTCATGGTGAACGGTTCCTAAAAAGTCAGCAACTTCTTTTGCAGCCTTGAGATCTGGTGAGCCCTGTTCAATCGTAAAACACAAAAATTTATAAA

TGCACAACTGAAAAATGCAACGGGCACGTATAGGTTGAAACCATCAAGTTTAAATTCTTGATCCGCTTCTGGAGTGACTCACCTCGAGACCAACACA

GAAGGAATGAAGTTGTGCTCCCCATTGCTTAGCAGCTTTTGTTCCAGCCAAGTATCGAGTAGTGACAGAAGCAACCAAAGACGAATCAAGTCCCCCC

GAGAGCAGAACGCCAAAGGGGACATCAGTCATCAACCGTTTGATAACAGCCTACAAAATAACAACCACTCATTAAGATCAACAGACTTCTCGGGTAC

TAAGACCCAAAGTCAAAAGGAAAACAGCCTAGATTAATTCATCCCAAACAAAACTAATACCAAATAATAACAAATACTCACATTTTCGAAGGCACGC

CTCAGAACCAAAGTGTCATAAGGAGTGGAAGGAATTGCTTCAGAGAACCAAGCGGGATTGTACCATCTCCTAAGCCCTCCGTTCTTGCTAGAGTACA

AGTGCCCCGGAGGGAAAACTTCAAAATGTTCACAATCATCATTCAAGCCCTTCAGCTCAGATGATATCCACACAGAGCCTGTGACACACAATAACAT

ATATCACTTCTAGTTTGTTATGTTGTTCAGACCCTCCAAAAATGAACTCGTGTCACATTTTTGAAGAGTACGAGCAACATAATTAGTTTCTATTTGG

AAAATTTGTATAAAAAATCCTACCATCAAGTCCCCAACCAATATAGAGGGGTGTAATTCCGATGGCATCACGAGCAGCAAGAAAGCTATTATCGCGA

GTATCCAATAATACAAAAGAGAACACTCCATCCAGCATGTCAACAAAATTTTCTCCATATTCTTCATACTGATCAACACATATATTAAAACTCAAGT

AAACCAAAAATTACTTATAAGTTAATTTAACCAGCTTATAAGCGTATCTAAACGCCCCGATGTTTCTATTATAATTCACTAGCCTAAGTTGGTTGAT

AAATAAAGTAAGACTATATACTCACAAGATGAGCAATAACATCACAATCACTTCCAGTCCTAAACTTGTGATTAGGCATAAGTTTTCGAAGTTTTTC

ATGATTGTAGATCTCTCCATTAACCTGCAAAATTCGACACAAAATTAATCATCAAACAAATTTTTCATCCTTATGTAAGTCTAAAATTATGTTGCTT

snp138

CCAAACTCTTCAAAAATATCGACTAGTGCCTGCCAGATCCTCCAAAAGTAGCACATTTTCGAAAGACATGACACATACATACACGACAAGT(C/T)G

snp118 snp116 snp110 snp75

ACAACATTTTTGAAGAGT(C/T)C(G/A)AACAA(C/T)AGTAGAGTAGAGGTGTACTCACAGTAACAACAAT(C/T)TTTTTGTCTTCATTAAACA

snp48 snp30

GAGGTTG(A/G)TCACCAGAAGCAGGGTC(G/T)ATAATTGCTAGACGTTGATGTGCCAAGTAAAAATCACCATATTGAAATATTCCACTCCAATCC

GGTCCACGATGCTTCAACCTATTGTTACAAAAATTAACT

13. Amplicon sequence of *4CL*

```
ACTATAACGTGTAATTAAAATTAATTTTTTTTGGTTTCAGTGTATCTCTTAAAAATATTGTTTTTGGGAAACTAGCGTATCAGCACATGGATGGTAT

TTCAGGATAAGAACATGAAGATTAAAAGATCAATAACTTCGGCTAAATTTAAATTCTAACTCGGCGCAAAAGAAAATAATATTTACCTGACCAAGTT

TGGCGTTAGGGAATTTTGCTCGGACGGCATCTTCAAGTTCCTTTCCTAATGGTGCAGCACCAGACATGACTGTTCTTACTGATGAAAGATCGTAGTT
  snp599                                                                                    snp531
AT(G/C)AACTAATGGACTCTTAGCAATGGCGAGAACAATAGGCGGAACAAATGGTCCAATTGTCACTTTATGCTTC(G/T)GTATCAACTCTAAAA

ACTGAGCAATGTCGAATTTCTGCATAATCAAAATTGCTGCTCCGACTCTCAATGCACAGAGCAAAACAGAATTGAGTGAGTAAATATGGAACAAAGG
     snp409
CAACAC(G/A)CACATCAACACATCATCGCTGTGCATATACAAATTAGCATTTTCTCCATCAACTTGTTGTGCAACGCTTGTTACTAATCCTTTGTG
                       snp301                                    snp271
TGTCAACATCACCCCTTTCGG(C/A)AGCCCGGTAGTCCCGGAGGAATACGGCAG(C/A)GCTACGACATCGTCCGGCTGGATTTTCACATCAGGAA
                                                        snp180
TTTCGTGTTCGTCGGATTGAATCAATTCGGAGAAATGAACACAACCTTCCGGT(G/A)CTGAATCGACGCATATTACCTTCAAATCATTTTCAATTG
    snp139                                  snp101
CGTA(G/A)TCCTTCACTTTCCCCGCAAAACAAGCTTGCGTGATAA(C/T)AATCTTAGCACTTGAGGCTTTGGCTTGCTTTACTACCTCTGCTGGT
        snp46
GTAAACAG(A/G)GGATTAGCCATTGTTGAAATTGCTCCTAGATACGATGCGCCAATAAACGCAAACACAAATTCAGGGCAATTAGGCAAGAGGATC
```

14. Amplicon sequence of *Lox1*

```
CTTGCTGAAATCAAATTTTTCTGGTTTTCCAGACATACCTGCCTTGCAACACACCGGAGCCCCTCCGACCATACAGGGAACAGGAGCTTCTTAGTCT
                                                                                        snp61
TCGTGGAAGTGGTTCAGGCAAGCTCAAGGAGTGGGACAGAGTGTATGATTATGCATTCTACAATGATCTTGGATTTCCAGATAAGGG(G/A)CCAGA
                                                        snp121
CTATGTTCGACCTGTGCTTGGTGGGTCGAAGGAATATCCATATCCTCGTAGAGG(C/T)AGGACAAGCCGTCGAGCAACTAAAACAGGTTAGATCAC
                snp176                  snp194                  snp216
ATATAAAATTGAAGTC(G/A)TCATATTACAAGGTTAT(G/C)ACTTATGATTTGCATTGCTGT(A/T)GAAACTTCATTAGAATGAAATGAATTCA

AGACCAATGAAGTACTATGGATAATAAGCAGCTGTTATTCCTATTGGCTTTGCAGATCCTATGTCAGAGAGCCGGTTGCCACCTTTAGGTTTGGACA
                                                snp394                  snp412                  snp428
TTTATGTTCCACGAGATGAGCGTTTCACTCCTGTGAAGCTTTCAGATTTCCT(A/T)GCATATGCTGTTAAATC(A/T)TTGGGTCAGGTCCTC(A/
              snp441                                    snp477
G)TTCCTGAGATTG(T/C)TGCTTTATTTGACAAAACTATCAACGAGTTTGATA(G/A)CTTTGAGGACGTACTCAAGCTTTATGAAGGTGGGATTA
   snp520                                                                               snp589
AGCT(G/A)CCTGACCATCATTTAAAAAAACTCAGGCAATGCATCCCATGGGAGATGCTCAAAGAACTAGTCCGCTC(C/T)GATGGTGAACCATTT

CTCAAATTTCCGATGCCTGATGTAATCAAAGGTACCAAAGGGCTCGTTTAATAGATACTTCCTGAGATGTTAGATCTTGCATCATAAATTGATAACT
          snp706
AATTATAAGCA(T/C)CTTTTTAAAATTACTGCAGTTGATAGGTCTGCCTGGAGGACAGATGAAGAGTTTGGACGGGAAATGCTAGCTGGAGTCAAC

CCTGTTATCATTCGACGTCTTCAGGTAACAAACAAGATTCACTGTACTAGTTGGGTTTAATGATTTTTGCTCTCATGCATGTTTCTAACCCATTAA
```

15. Amplicon sequence of *StAOS2*

```
TATACATCAATCGTTCTATAGTCCCTACCACTACATTAAAGGAAGAAATTTCATCAAACACCTTTGTATCACTAACATTACCCATCCTTTTTTTGGT

TCAAAACAATGGCATTAACTTCATCTTTTTCTCTTCCTCTTCCTTCTCTTCACCAACAATTTCCATCAAAATACTCCACATTTCGTCCTATTATTGT
          snp99          snp110                                  snp138   snp143
TTCTTTATC(G/C)GAAAAACCAA(C/T)AATCGTGGTAACCCAACCTACAAAATT(A/C)CCTA(C/T)CAGGACAATACCCGGCGACTATGGG
  snp171                                                  snp219 snp222  snp226   snp231
TT(G/A)CCGGGTATTGGTCCATGGAAAGATAGGCTTGATTACTTTTACAATCA(A/G)GG(G/C)AAA(G/A)ACGA(A/T)TTTTTCGAATCAAG
      snp252                                  snp279                                  snp304          snp318 snp321
AGTAGT(G/A)AAATACAAATCAACTATATTCAGAAC(G/T)AACATGCCACCGGGACCATTCATT(T/G)CTTCTAACCCGAA(G/T)GT(T/A)A
              snp336                                  snp369                                  snp396
TTGTTTTGCTCGA(T/C)GGCAAGAGTTTCCCAGTCCTTTTCGATGTTTC(G/T)AAAGTCGAAAAAAAGGACCTCTTCAC(C/T)GGAACTTACAT
                        snp434 snp435
GCCGTCGACTGAACTCACCGGTGGTT(A/) (C/)CGTGTTCTTTCTTATCTTGACCCATCTGAACCAAACCATGAAAAATTGAAAAAATTGATGTTC

TTCCTTCTTTCTTCTCGACGTGATCACGTTATACCCAAATTCCATGAAACTTATACAGAGTTTTTTGAAACCCTAGATAAGGAAATGGCGGAAAAAG
                                                                                  snp678 snp681
GTACAGCTGGTTTAAACTCCGGCAATGATCAAGCTGCGTTTAATTTCTTAGCTAGATCGTTGTTCGGAGTTAACCCAGTTGA(A/G)AC(T/A)AAA
     snp691 snp692        snp702                                  snp727                  snp744
CTCGGA(A/G)(C/G)TGATGGTCC(G/A)ACATTGATCGGAAAATGGGTTTTG(C/T)TTCAGCTTCATCCTGT(A/G)CTCACTCTCGGTCTTCC
              snp774
GAAGTTTCTAGA(C/T)GACTTAATCCTCCATACTTTCCGGTTACCTCCGTTTCTGGTGAAGAAAGATTACCAGAGACTTTACGATTTCTTTTACAC
          snp866          snp875  snp879                          snp900
CAACTCCGCCA(G/A)TTTATTCG(C/T)CGA(A/G)GCTGAAAAACTCGGCATTTC(A/T)AAAGAAGAAGCTTGTCATAATCTTCTCTTCGCTAC
                    snp954                  snp982
TTGCTTCAATTCCTTCGG(C/T)GGGATGAAGATTTTCTTCCCGAATATG(C/A)TGAAATCGATAGCGAAAGCAGGGGTGGAGGTCCATACCCGTT

TAGCAAACGAGATCCGATCGGAAGTAAAATCCGCTGGCGGGAAGATCACGATGTCGGCGATGGAGAAAATGCCGTTAATGAAATCAGTAGTTTATGA

AGCTTTACGAGTTGATCCTCCG
```

Table S3. P_value and R^2 of SNPs used in linkage analysis, general linear model (GLM)

		rAUDPC		MCR		PM	
Locus	SNP_locus	p_value	R^2	p_value	R^2	p_value	R^2
Lipoxygenase_1 (lox1)	Lox1_SNP61	NS	NS	NS	NS	NS	NS
	Lox1_SNP_412	NS	NS	NS	NS	NS	NS
	Lox1_SNP_121	NS	NS	NS	NS	NS	NS
	Lox1_SNP_477	NS	NS	NS	NS	NS	NS
	Lox1_SNP_412	NS	NS	NS	NS	NS	NS
	Lox1_SNP_194	NS	NS	NS	NS	NS	NS
	Lox1_SNP_589	NS	NS	NS	NS	NS	NS
	Lox1_SNP_176	NS	NS	NS	NS	NS	NS
	Lox1_SNP_394	NS	NS	NS	NS	NS	NS
	Lox1_SNP_216	NS	NS	NS	NS	NS	NS
	Lox1_SNP_428	NS	NS	NS	NS	NS	NS
	Lox1_SNP_520	NS	NS	NS	NS	NS	NS
	Lox1_SNP441	NS	NS	NS	NS	NS	NS
	Lox1_SNP706	NS	NS	NS	NS	NS	NS
Delta (7)-sterol-C5 (6)-desaturase	DSD_SNP31	0.0029	13.75	0.003019	13.65	NS	NS
	DSD_SNP83	NS	NS	NS	NS	NS	NS
	DSD_SNP191	NS	NS	NS	NS	NS	NS
	DSD_SNP218	NS	NS	NS	NS	NS	NS
	DSD_SNP235	NS	NS	NS	NS	NS	NS
	DSD_SNP236	NS	NS	NS	NS	NS	NS
	DSD_SNP245	NS	NS	NS	NS	NS	NS
	DSD_SNP284	NS	NS	NS	NS	NS	NS
	DSD_SNP290	NS	NS	NS	NS	NS	NS
	DSD_SNP357	NS	NS	NS	NS	NS	NS
Serine protease inhibitor (Miracullin)	SPI_SNP51	0.0036	10.62	NS	NS	NS	NS
	SPI_SNP53	NS	NS	NS	NS	NS	NS
	SPI_SNP72	0.005	10.04	NS	NS	NS	NS
	SPI_SNP83	NS	NS	NS	NS	NS	NS
	SPI_SNP230	0.0013	12.69	NS	NS	NS	NS
	SPI_SNP264	0.0044	10.36	NS	NS	NS	NS
	SPI_SNP266	0.0048	10.24	NS	NS	NS	NS
	SPI_SNP297	NS	NS	NS	NS	NS	NS
	SPI_SNP298	0.0043	10.02	NS	NS	NS	NS

	SPI_SNP306	0.0009	12.58	NS	NS	NS	NS
	SPI_SNP310	0.0004	14.21	NS	NS	NS	NS
	SPI_SNP342	NS	NS	NS	NS	NS	NS
	SPI_SNP349	NS	NS	NS	NS	NS	NS
	SPI_SNP356	NS	NS	NS	NS	NS	NS
	SPI_SNP357	NS	NS	NS	NS	NS	NS
	SPI_SNP364	NS	NS	NS	NS	NS	NS
	SPI_SNP369	NS	NS	NS	NS	NS	NS
4-coumarate-CoA ligase	4CL_SNP540	NS	NS	NS	NS	NS	NS
	4CL_SNP531	NS	NS	NS	NS	NS	NS
	4CL_SNP46	NS	NS	NS	NS	NS	NS
	4CL_SNP101	NS	NS	NS	NS	NS	NS
	4CL_SNP139	NS	NS	NS	NS	NS	NS
	4CL_SNP180	NS	NS	NS	NS	NS	NS
	4CL_SNP271	NS	NS	NS	NS	NS	NS
	4CL_SNP301	NS	NS	NS	NS	NS	NS
	4CL_SNP409	NS	NS	NS	NS	NS	NS
	4CL_SNP599	NS	NS	NS	NS	NS	NS
Asparagine Synthetase	AspS_SNP112	NS	NS	0.000155	16.98	NS	NS
	AspS_SNP119	NS	NS	0.000155	16.98	NS	NS
	AspS_SNP165	0.0089	10.14	1.47E-05	20.66	NS	NS
	AspS_SNP175	0.0031	12.02	3E-08	29.49	NS	NS
Magnesiumprotoporphyrin (Disaturase)	MPP_SNP243	NS	NS	NS	NS	NS	NS
	MPP_SNP262	NS	NS	NS	NS	NS	NS
	MPP_SNP286	NS	NS	NS	NS	NS	NS
	MPP_SNP308	NS	NS	NS	NS	NS	NS
	MPP_SNP374	NS	NS	NS	NS	NS	NS
	MPP_SNP404	NS	NS	NS	NS	NS	NS
	MPP_SNP405	NS	NS	NS	NS	NS	NS
Allene oxide synthase 2	StAOS2_SNP_99	NS	NS	NS	NS	NS	NS
	StAOS2_SNP_110	NS	NS	NS	NS	NS	NS
	StAOS2_SNP_138	NS	NS	NS	NS	NS	NS
	StAOS2_SNP_143	NS	NS	NS	NS	NS	NS
	StAOS2_SNP_171	NS	NS	NS	NS	NS	NS
	StAOS2_SNP_219	NS	NS	NS	NS	NS	NS
	StAOS2_SNP_222	NS	NS	NS	NS	NS	NS

	StAOS2_SNP_226	NS	NS	NS	NS	NS	NS
	StAOS2_SNP_231	NS	NS	NS	NS	NS	NS
	StAOS2_SNP_252	NS	NS	NS	NS	NS	NS
	StAOS2_SNP_304	NS	NS	NS	NS	NS	NS
	StAOS2_SNP_318	NS	NS	NS	NS	NS	NS
	StAOS2_SNP_321	NS	NS	NS	NS	NS	NS
	StAOS2_SNP_336	NS	NS	NS	NS	NS	NS
	StAOS2_SNP_369	NS	NS	NS	NS	NS	NS
	StAOS2_snp678	NS	NS	NS	NS	NS	NS
	StAOS2_snp681	NS	NS	NS	NS	NS	NS
	StAOS2_snp691	NS	NS	NS	NS	NS	NS
	StAOS2_snp692	NS	NS	NS	NS	NS	NS
	StAOS2_snp702	NS	NS	NS	NS	NS	NS
	StAOS2_snp727	NS	NS	NS	NS	NS	NS
	StAOS2_snp744	NS	NS	NS	NS	NS	NS
	StAOS2_snp774	NS	NS	NS	NS	NS	NS
	StAOS2_snp866	NS	NS	NS	NS	NS	NS
	StAOS2_snp875	NS	NS	NS	NS	NS	NS
	StAOS2_snp879	NS	NS	NS	NS	NS	NS
	StAOS2_snp900	NS	NS	NS	NS	NS	NS
	StAOS2_snp954	NS	NS	NS	NS	NS	NS
	StAOS2_snp982	NS	NS	NS	NS	NS	NS
SNPs with contrasting expression	DSD_SNP_pyro	0.0007	16.16	0.000583	16.37	NS	5.57
	MPP_pyro	NS	NS	NS	NS	NS	NS
	AspS_pyro	NS	NS	NS	NS	NS	NS
	UPA_pyro	NS	NS	NS	NS	NS	NS
	MADS_Pyro	NS	NS	NS	NS	NS	NS
	RLPK1_Pyro	NS	NS	NS	NS	NS	NS
	RLPK2_pyro	NS	NS	NS	NS	NS	NS
	BCCP_pyro	NS	NS	NS	NS	NS	NS
	HPR_pyro	NS	NS	NS	NS	NS	NS

Table S4. P_value and R^2 of SNPs used in association mapping, mixed linear model (MLM)

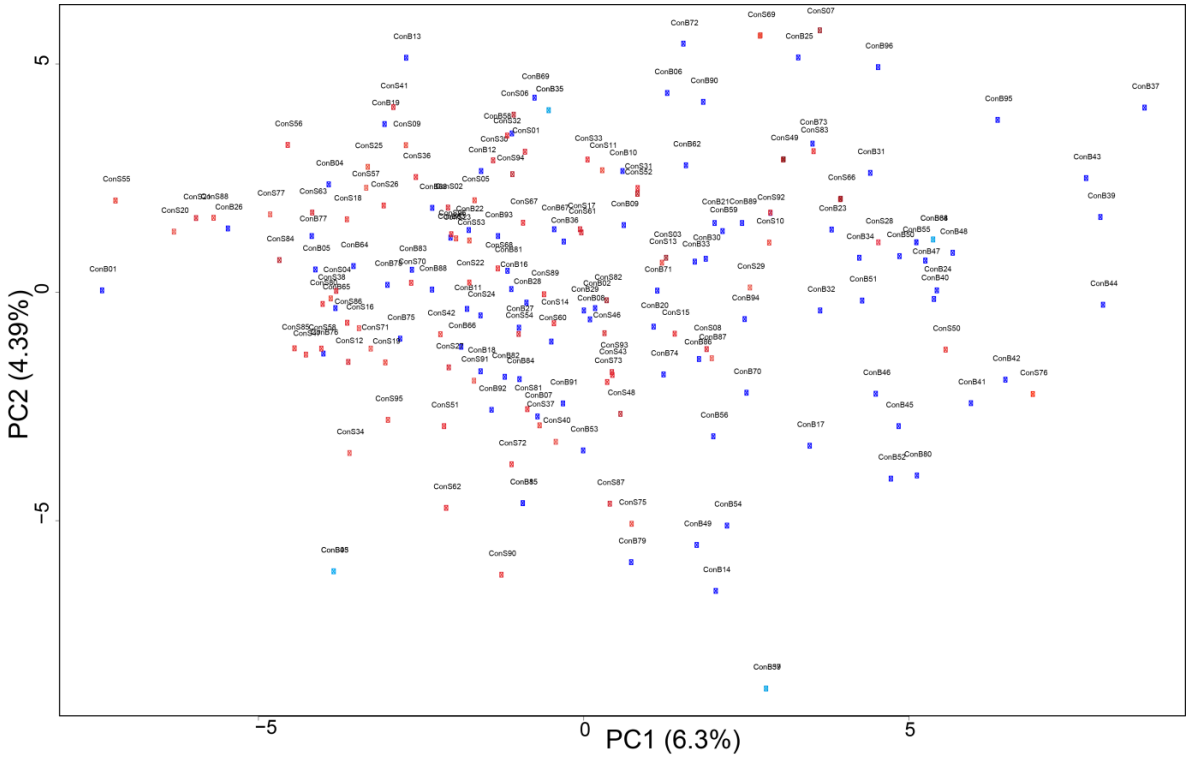
		rAUDPC		MCR		PM	
Locus	Marker	P_value	R^2	P_value	R^2	P_value	R^2
Hydroxypyruvate reductase	HPR_SNP53	NS	NS	NS	NS	NS	NS
	HPR_SNP54	0.0053	8.38	NS	NS	NS	NS
	HPR_SNP102	NS	NS	NS	NS	NS	NS
	HPR_SNP105	NS	NS	NS	NS	NS	NS
	HPR_SNP93	NS	NS	NS	NS	NS	NS
	HPR_SNP133	NS	NS	NS	NS	NS	NS
	HPR_SNP134	NS	NS	NS	NS	NS	NS
	HPR_SNP142	NS	NS	NS	NS	NS	NS
	HPR_SNP157	NS	NS	NS	NS	NS	NS
	HPR_SNP258	NS	NS	NS	NS	NS	NS
	HPR_SNP261	NS	NS	NS	NS	NS	NS
	HPR_SNP285	NS	NS	NS	NS	NS	NS
Delta (7)-sterol-C5 (6)-desaturase	DSD_SNP31	NS	NS	NS	NS	NS	NS
	DSD_SNP81	NS	NS	0.0034	6.08	0.0097	4.05
	DSD_SNP163	NS	NS	NS	NS	NS	NS
	DSD_SNP191	NS	NS	NS	NS	NS	NS
	DSD_SNP218	NS	NS	NS	NS	NS	NS
	DSD_SNP235	NS	NS	NS	NS	NS	NS
	DSD_SNP236	NS	NS	NS	NS	NS	NS
	DSD_SNP245	NS	NS	NS	NS	NS	NS
	DSD_SNP284	NS	NS	NS	NS	NS	NS
	DSD_SNP290	NS	NS	NS	NS	NS	NS
	DSD_SNP357	NS	NS	NS	NS	NS	NS
	DSD_SNP100	NS	NS	NS	NS	NS	NS
	DSD_SNP101	NS	NS	NS	NS	NS	NS
	DSD_SNP111	NS	NS	NS	NS	NS	NS
	DSD_SNP157	NS	NS	NS	NS	NS	NS
	DSD_SNP160	NS	NS	NS	NS	0.000252	9.9
	DSD_SNP181	NS	NS	NS	NS	NS	NS
	DSD_SNP248	NS	NS	0.0033	8.7	7.29E-05	11.39
Pectin methyl esterase	PME_SNP82	0.00013	9.35	0.0021	6.55	NS	NS
	PME_SNP93	NS	NS	NS	NS	NS	NS
	PME_SNP103	NS	NS	NS	NS	NS	NS
	PME_SNP166	NS	NS	NS	NS	NS	NS
	PME_SNP172	NS	NS	NS	NS	NS	NS

	PME_SNP190	NS	NS	NS	NS	NS	NS
	PME_SNP247	NS	NS	NS	NS	NS	NS
	PME_SNP344	NS	NS	NS	NS	NS	NS
Receptor-like protein kinase	RLPK_SNP-121	NS	NS	NS	NS	NS	NS
	RLPK_SNP-290	NS	NS	NS	NS	NS	NS
	RLPK_SNP-311	NS	NS	NS	NS	NS	NS
	RLPK_SNP-380	NS	NS	NS	NS	NS	NS
	RLPK_SNP-413	NS	NS	NS	NS	NS	NS
	RLPK_SNP-551	NS	NS	NS	NS	NS	NS
	RLPK_SNP-585	NS	NS	NS	NS	NS	NS
	RLPK_SNP-600	NS	NS	NS	NS	NS	NS
	RLPK_SNP-689	NS	NS	NS	NS	NS	NS
	RLPK_SNP-735	NS	NS	NS	NS	NS	NS
Squalene monooxygenase	SMO_SNP190	NS	NS	NS	NS	NS	NS
	SMO_SNP205	NS	NS	NS	NS	NS	NS
	SMO_SNP245	NS	NS	NS	NS	NS	NS
	SMO_SNP247	NS	NS	NS	NS	NS	NS
	SMO_SNP297	NS	NS	NS	NS	NS	NS
	SMO_SNP324	NS	NS	NS	NS	NS	NS
	SMO_SNP353	NS	NS	NS	NS	NS	NS
	SMO_SNP53	NS	NS	NS	NS	NS	NS
	SMO_SNP63	NS	NS	NS	NS	NS	NS
	SMO_SNP104	NS	NS	NS	NS	NS	NS
	SMO_SNP205	NS	NS	NS	NS	NS	NS
Clathrin coat assembly protein AP17	CCAP_SNP80	NS	NS	NS	NS	NS	NS
	CCAP_SNP188	NS	NS	NS	NS	NS	NS
	CCAP_SNP212	NS	NS	NS	NS	NS	NS
	CCAP_SNP281	NS	NS	NS	NS	NS	NS
	CCAP_SNP288	NS	NS	NS	NS	NS	NS
	CCAP_SNP291	NS	NS	NS	NS	NS	NS
	CCAP_SNP318	NS	NS	NS	NS	NS	NS
	CCAP_SNP336	NS	NS	NS	NS	NS	NS
Biotin carboxylase carrier protein	BCCP_SNP60	0.0018	9.33	0.000135	12.48	NS	NS
	BCCP_SNP61	NS	NS	NS	NS	NS	NS
	BCCP_SNP63	NS	NS	NS	NS	NS	NS
	BCCP_SNP82	NS	NS	NS	NS	NS	NS
	BCCP_SNP130	NS	NS	NS	NS	NS	NS
	BCCP_SNP131	NS	NS	NS	NS	NS	NS

	BCCP_SNP275	NS	NS	NS	NS	NS	NS
	BCCP_SNP321	0.000148	12.03	3.27E-05	13.55	NS	NS
	BCCP_SNP339	NS	NS	NS	NS	NS	NS
Asparagine synthetase	AspS_SNP112	NS	NS	NS	NS	NS	NS
	AspS_SNP119	NS	NS	NS	NS	NS	NS
	AspS_SNP162	NS	NS	NS	NS	0.0039	8.62
	AspS_SNP175	NS	NS	0.0012	8.18	NS	NS
	AspS_17	NS	NS	NS	NS	NS	NS
	AspS_57	NS	NS	NS	NS	NS	NS
	AspS_82	NS	NS	NS	NS	NS	NS
	AspS_83	NS	NS	NS	NS	NS	NS
	AspS_115_	NS	NS	NS	NS	NS	NS
	AspS_SNP-30	NS	NS	NS	NS	NS	NS
	AspS_SNP-48	NS	NS	NS	NS	NS	NS
	AspS_SNP-75	NS	NS	NS	NS	NS	NS
	AspS_SNP-110	NS	NS	NS	NS	NS	NS
	AspS_SNP-116	NS	NS	NS	NS	NS	NS
	AspS_SNP-118	NS	NS	NS	NS	NS	NS
	AspS_SNP-138	NS	NS	NS	NS	NS	NS
Chloroplast protease	Chp_SNP59	NS	NS	NS	NS	NS	NS
	Chp_SNP86	0.000831	11.42	0.0087	8.43	NS	NS
	Chp_SNP230	NS	NS	NS	NS	NS	NS
	Chp_SNP415	NS	NS	NS	NS	NS	NS
	Chp_SNP418	0.0054	9.38	NS	NS	NS	NS
	Chp_SNP451	0.0053	9.18	NS	NS	NS	NS
	Chp_SNP452	0.0084	8.69	NS	NS	NS	NS
	Chp_SNP550	NS	NS	NS	NS	NS	NS
Avr9/Cf-9 rapidly elicited protein 20	AC9_SNP237	NS	NS	NS	NS	NS	NS
	AC9_SNP238	NS	NS	NS	NS	NS	NS
	AC9_SNP321	NS	NS	NS	NS	NS	NS
	AC9_SNP346	NS	NS	NS	NS	NS	NS
	AC9_SNP370	NS	NS	0.0018	8.17	NS	NS
	AC9_SNP435	NS	NS	NS	NS	NS	NS
	AC9_SNP525	NS	NS	NS	NS	NS	NS
	AC9_SNP552	NS	NS	NS	NS	NS	NS
	AC9_SNP692	NS	NS	NS	NS	NS	NS
	AC9_SNP697	NS	NS	NS	NS	NS	NS

Magnesium-protoporphyrine (Disaturase)	MPP_SNP243	NS	NS	NS	NS	NS	NS
	MPP_SNP262	NS	NS	NS	NS	NS	NS
	MPP_SNP286	0.000712	10.6	0.0036	8.65	NS	NS
	MPP_SNP308	NS	NS	NS	NS	NS	NS

Fig. S2. First and second components from principal component analysis (PCA) of CONQUEST2 population. The population contain 184 tetraploid potato clones collected from two breeding companies, BNA and SAKA. BNA genotypes are indicated by blue while the SAKA genotypes by red. The analysis was done using 241 polymorphic SNPs that were uniformly distributed in the 12 potato chromosomes.



6. Supplementary data provided on CD

Table 6.1 SL_potato clones used in linkage mapping and their phenotypic data

Table 6.2 CONQUEST2 population and phenotypic data used in association mapping

Table 6.3 Selected 165 uni-tags (novel candidate genes) from SuperSAGE

Table 6.4 Candidate genes SNP markers on CONQUEST2 for candidate gene association mapping

Table 6.5 Population structure data for CONQUEST2 from STRUCTURE software

Table 6.6 Kinship matrix data for CONQUEST2 from SPAGeDi software

Table 6.7 Candidate genes SNP markers on SL_clones for linkage mapping

Table 6.8 Result_ chi-square for LD of significantly associated candidate genes

Table 6.9 CONQUEST2_SolCAP_SNP_marker_data_letterform

Table 6.10 SolCap_SNP_markers_map_positions

Table 6.11 Result_GWA mapping_MLM (QK_model)-using Tassel for SolCAP_SNPs

Table 6.12 Result_GWA mapping_GLM (Q_model)-using Tassel for SolCAP_SNPs

Table 6.13 Result_GWA mapping_K_model-using Tassel for SolCAP_SNPs

Table 6.14 Result_GWA mapping_ANOVA

Table 6.15 Result_LD_using _r-square and D'- for SoCAP_SNPs

Table 6.16 Result_LD_using _chi-square- for significantly associated SolCAP_SNPs

Table 6.17 Phenotypic data from controlled infection tests for the 24 genotypes used in expression analysis

Table 6.18 qRT-PCR data for Phytophthora growth quantification from controlled infection tests for the 24 genotypes used in expression analysis

Table 6.19 qRT-PCR data for Phytophthora growth quantification across time points, from controlled infection tests for the 24 genotypes used in expression analysis

Table 6.20 Pyrosequencing SNPs data for SuperSAGE tag SNPs with contrasting allele expression

7. Acknowledgements

First and most, I would like to thank my supervisor Christiane Gebhardt for giving me the opportunity to work in molecular genetics, which I love most as a science. Thank you for giving me the chance of working in your group and for the support, guidance and enthusiasm to develop this PhD project. I have appreciated your kind treatment. Thank you for all your critical comments, corrections and suggestions in my thesis document.

My salary was from the department of plant Breeding and Genetics headed by Maarten Koornneef, Max Planck institute for Plant Breeding Research, Köln.

I am grateful to Prof. Dr. Ulf-Ingo Flügge, Prof. Dr. Martin Hülskamp and Dr. Ales Pecinka for their willing to be my second thesis reviewer, chairing the examination board and protocol writer, respectively.

At the start of my PhD project, Astrid Draffehn introduced me with qRT-PCR and pyrosequencing and helped me much in infection works. Thank you Astrid for reading and commenting on part of this manuscript.

Birgit Walkemeier showed and guided me for the lab works; most importantly the infection work would be impossible without Birgit's help.

Elske Maria Schönhals has been willing to discuss with me on different topics.

Aman Mulki introduced me with TASSEL software. Thank you Aman for reading and commenting on part of this manuscript.

The support of the following people is gratefully acknowledged; Dr. Diro Terefe, Fabio Cericola, Dr. Joao Paulo, Geert van Geest. Dr. Jinquan Li did the LD decay analysis.

I would like to remember former Gebhardt's group memebers; Dr. Lena, Dr. Camila, Dr. Claude, Markus Kuckenberg, Dr. Jude, Dr. Rena, Dr. Janne, Dr. Li and Sandra

I am always grateful to my great mom, Meymuna Yesuf, who taught me the value of being patience. Seada is always in my heart. My sister, Sitina has been supporting me as usual. To my lovely wife, Hayat, thank you for your love, support and sacrifice.

Praise be to God!!!

8. Erklärung

Die vorliegende Arbeit wurde am Max-Planck Institut für Planzenzüchtungsforschung in Köln durchgeführt.

Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit - einschließlich Tabellen, Karten und Abbildungen -, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; nicht veröffentlicht worden ist sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde.

Die Bestimmungen der Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von PD Dr. Christiane Gebhardt betreut worden.

Köln, den 19.05.2014 Meki Shehabu Muktar

9. Lebenslauf

Name Meki Shehabu Muktar

Address Kolibriweg 14, 50829, Köln, Germany

Mobile: +491775194788

E-mail: muktar@mpipz.mpg.de

Geschlecht Männlich

Geburstdatum 16.10.1976, Dilla, Äthiopien

Staatsangehörigkeit Äthiopisch

Familienstand Verheiratet

Ausbildung

Seit Februar 2011:

Doktorand am Max-Planck-Institut für Pflanzenzüchtungsforschung, Köln
Finanzierung: Max-Planck-Institut Stipendium

September 2008 bis Juni 2010:

Studium "European Master of Science in Nematology", Abschluss: M.Sc., Universiteit Ghent (Belgien)
Finanzierung: ERASMUS MUNDUS Stipendium

September 2002 bis Januar 2005:

Studium "M.Sc. Plant Pathology", Abschluss: M.Sc., Haramaya University (Äthiopien)
Finanzierung: Ausbildungsförderung in Ramen eines landwirtschaftlichen Forschungsprojekts

September 1994 bis Juli 1998

Studium "B.Sc. Plant Science", Abschluss: B.Sc., Hawassa University (Äthiopien)

September 1990 to June 1994

Student in Dilla umfassende Hochschul, (Äthiopien)

Arbeitserfahrung:

Juni 2007 bis August 2008

Nutzpflanzenpathologe, Southern Agricultural Research Institute, Hawassa Agricultural Research Center, Hawassa (Äthiopien)

December 1998 to May 2007

Nutzpflanzenpathologe, Amhara Agricultural Research Institute, Debre Birhan Agricultural Research Center, Debre Berhan (Äthiopien)

Publikationen:

Draffehn et al. (2013). Front. Plant Sci. 4, 423.
Shehabu et al. (2008). International Journal of Pest Management, 54, 143 - 149
Shehabu et al., (2011). Archives of Phytopathology and Plant Protection, 5, 432-440
Shehabu et al. (2010). Tree and Forestry Science and Biotechnology, 4, 63-64
Shehabu et al. **(**2010). Tree and Forestry Science and Biotechnology, 4, 60-62
Shehabu, M. and Tekalign, A. (2004). Pest Management Journal of Ethiopia, 8, 111-18