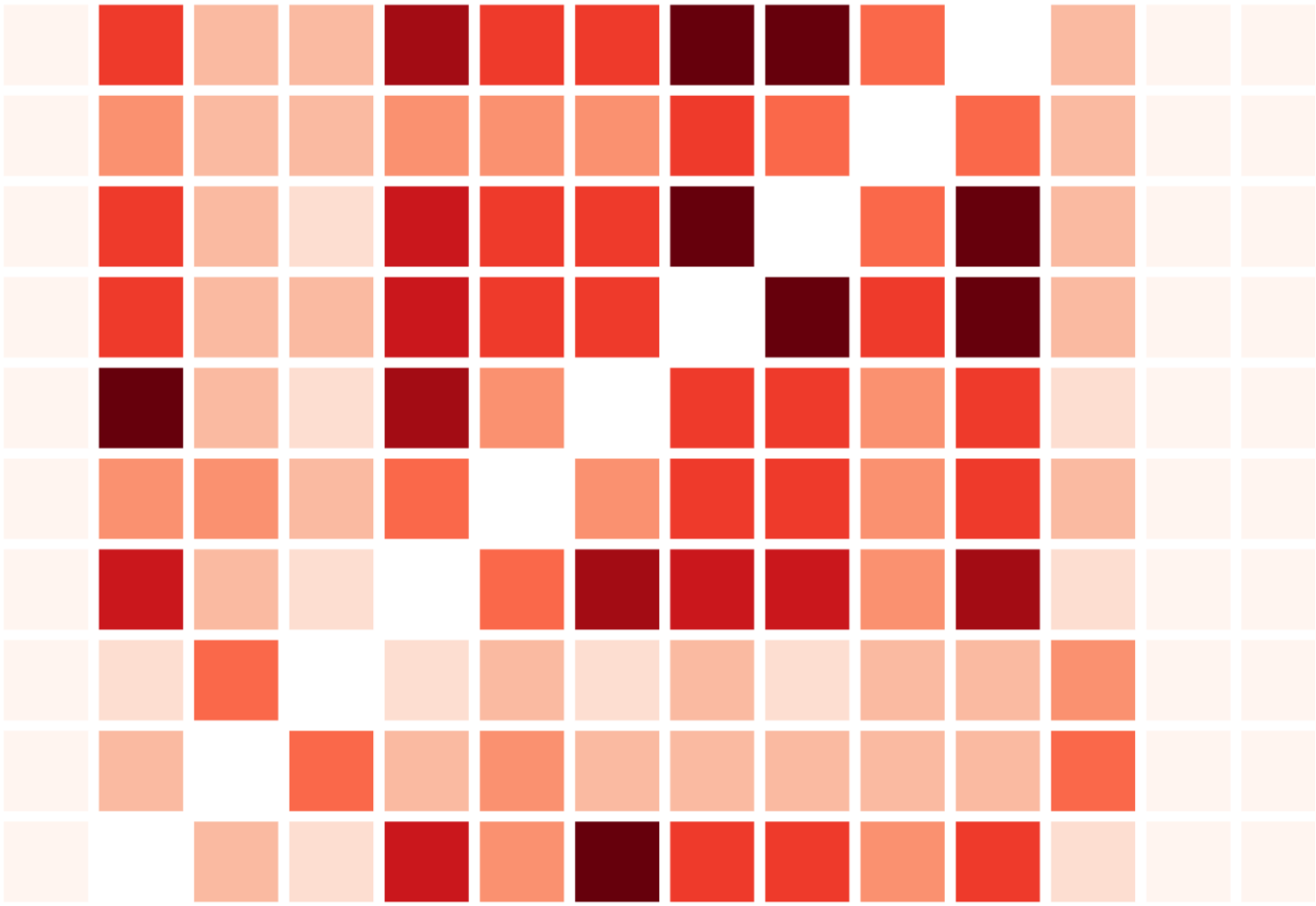
Geert van Geest, Msc thesis

Association of novel candidate genes with tuber quality traits in *Solanum tuberosum*

New insights in bruising, tuber shape and tuber starch content



Supervisors: Dr. Christiane Gebhardt, Dr. Yuling Bai
Examiner: Prof. Dr. Richard Visser
Performed at the Max Planck Institute for Plant Breeding Research in Köln

Association of novel candidate genes with tuber quality traits in *Solanum tuberosum*

New insights in bruising, tuber shape and tuber starch content

MSc thesis

Student:	Geert van Geest
Supervisor (MPIPZ):	Dr. Christiane Gebhardt
Supervisor (WU):	Dr. Yuling Bai
Examiner (WU):	Prof. Dr. Richard Visser
Credits:	36 ECTS
Course code:	PBR-80436
Performed at:	Max Planck Institute for Plant Breeding Research (MPIPZ), Köln, Germany
Student number:	870625254070
Study program:	MSc Plant Sciences, specialization: Plant Breeding and Genetic Resources at Wageningen University (WU), Wageningen, The Netherlands
Period:	January 2012 – July 2012
Submitted:	5 July 2012

CONTENTS

SUMMARY

Introduction

Throughout the supply chain the occurrence of black to brown spots in potato (*Solanum tuberosum*) tuber flesh called 'bruising' is a main reason for discarding potato tubers. There exists genetic variation for bruising susceptibility, so breeding against it is possible. However, tests for bruising susceptibility need a large amount of potato tubers which are only available late in the breeding program. Additionally, it is known that bruising susceptibility correlates positively with starch content; a favourable trait. These two reasons hamper breeding against bruising susceptibility. This study aims at finding molecular markers associated with bruising susceptibility by combining a bulked segregant analysis with a candidate gene approach.

Methods

Ten highly bruising resistant (cases) and ten highly bruising susceptible (controls) cultivars were genotyped with the SolCAP SNP array containing approximately 8300 SNPs that are located in annotated genes on the potato genome. Eleven of these SNPs were selected as potentially associated SNPs, and genotyped in a large population in order to find robust SNP trait associations. Selection was based on significant differences of allele frequency between the cases and controls, and putative function of the candidate gene they are in.

Results & discussion

SNPs strongly associated with bruising susceptibility ($p<3.5e^{-4}$) were found at seven loci spread over six chromosomes. In two loci, 20 SNPs and two INDELs were found that explained more than 20% of bruising susceptibility ($p<1e^{-7}$). These SNPs were also strongly associated with starch content ($p<1e^{-8}$) in the same polarity as bruising susceptibility. The finding of such strong associations proves the feasibility of combining a BSA with a candidate gene approach. In one locus on chromosome 10, SNPs were strongly associated with tuber shape, explaining up to 47% of the tuber shape in the population. These SNPs were very proximate to markers linked to the *Ro* locus – a locus previously described as causative for tuber shape. By sequencing parts of neighbouring loci, one putative boarder of the haplotype block containing the *Ro* locus was located. This haplotype block has a size of at least 420 kb, and contains several candidate genes that could have a function in shaping tubers.

Conclusions

The molecular markers provided in this study have potential to strongly improve selection efficiency in a potato breeding program for bruising susceptibility and starch content independent of plant maturity. In addition, the mapping of the haplotype block containing the *Ro* locus provides an initial step for the characterization of genes affecting tuber shape. This knowledge will greatly enhance our fundamental understanding on how plant organs are shaped.

LIST OF ABBREVIATIONS

Abbreviation	**Meaning**
BI	Bruising Index
BIPx	Bacterial-induced peroxidase
BSA	Bulked Segregant Analysis
Ce	Carbohydrate esterase
ChoK	Choline/ethanolamine kinase
CuBP	Copper ion binding protein
Gt	Glycosyltransferase (locus)
GT	Glycosyltransferases (Gene family)
GWAS	Genome wide association study
INDEL	Insertion/deletion
LD	Linkage disequilibrium
Mco	Mulicopper oxidase
nsLTP	Non-specific lipid transfer protein
Pest1	Pectin methylesterase1
PM	Plant Maturity
POD	Peroxidase
PPO	Polyphenol oxidase
QTL	Quantitative trait locus
SCB	Starch Corrected Bruising
SG	Specific Gravity
SKIP10	Ski-interacting protein10
SNP	Single nucleotide polymorphism
StHSL1	(*Solanum tuberosum*) High-sugar induible like1
TS	Tuber Shape
TY	Tuber Yield
Vps35	Vacuolar protein sorting35
Zds	Zeta-carotene desaturase

1. INTRODUCTION

Despite an emerging worldwide food crisis, large parts of the harvested food crops do not reach the consumer's mouth. For tuber and root crops the losses in the chain from harvest to consumption are tremendous: it was estimated to be 68% in Europe and 82% in North America (Gustavsson et al., 2011). Potato (*Solanum tuberosum*), the fourth most important food crop in the world, is the largest crop in this group. A main reason for discarding potatoes is the occurrence of enzymatic browning of the tuber flesh after mechanical damage. This enzymatic browning, also referred to as 'bruising', reduces nutritional value, texture and flavour. Therefore, severely bruised tubers are useless for further processing and consumption. Measures in growing and processing potatoes can reduce the severity of bruising (Storey, 2007). There is also a strong genetic component affecting bruising susceptibility. Improving the resistance to bruising is a very efficient way of reducing problems with bruising wherever in the chain, since once there are well performing bruising resistant cultivars available, there is less need for repetitive investment in resources, like improved, but more costly, processing measures or sophisticated growing techniques.

Bruising susceptibility can only be assessed late in a potato breeding program, since there is a substantial amount of tubers needed for assessing the trait. It takes several years of propagation before enough tubers can be harvested from one seed as starting material. In addition, bruising susceptibility is positively correlated to specific gravity (Urbany et al., 2011), a measure for starch content, which is a favourable trait. These two reasons hamper breeding against bruising susceptibility. Molecular markers explaining variation in bruising susceptibility could provide very useful information to enable selection early in the breeding program without the need of phenotyping large amounts of potato tubers, and to select efficiently against bruising susceptibility independent of starch content. In potato, these kind of molecular markers linked to resistance to Potato Virus Y (Kasai et al., 2000; Song et al., 2005), the nematode *Globodera pallida* (Sattarzadeh et al., 2006) and the nematode *Globodera rostochiensis* (Paal et al., 2004) have been found and used successfully in breeding programs (Ortega and Lopez-Vizcon, 2012).

Many agriculturally important traits are quantitative and are affected by many genes. Loci that contribute to these traits are called quantitative trait loci (QTL). QTL mapping with use of experimental mapping populations has been applied successfully to map markers linked to several QTL affecting agriculturally important traits in potato (e.g. Van Eck et al., 1994; Leonards-Schippers et al., 1994; Bradshaw et al., 2008). However, this method has its limitations. First, QTL mapping needs development of progenies segregating for the trait of interest. Second, the fact that potato is a non-inbreeding tetraploid makes analysis of inheritance rather complicated (Gebhardt, 2007). Third, by using a single mapping population markers are limited to this population, and not all alleles present in the entire germplasm are represented. An alternative to QTL mapping is association mapping, which has been used successfully in potato for several polygenic traits (Simko et al., 2004; Li et al., 2005; D'hoop et al., 2007; Urbany et al., 2011). The method is based on the linkage disequilibrium between a trait and a marker. For association mapping an entire breeding collection containing all relevant alleles can be used for analysis. Marker trait associations can be found with this method by calculating the associations of a certain trait with a large number of markers spread over the entire genome (genome wide association study or GWAS; reviewed by Xu et al. (2012)), or with markers in specific regions, like in candidate genes (candidate gene approach; reviewed by Pflieger et al. (2001) and specifically for potato by Gebhardt et al. (2007)).

For this project the genome wide and candidate gene approach were combined in order to find markers significantly associated with bruising susceptibility. The genome wide approach was applied by means of a bulked segregant analysis (BSA) with few genotypes (cases and controls for bruising susceptibility). These were genotyped with an Illumina SNP array (Hamilton et al., 2011) and SNPs with a significantly different allele frequency between the cases and controls were selected as potentially associated SNPs. Eleven of these SNPs were selected based on significance

and putative function of the candidate gene they are in. These SNPs were genotyped in a large population phenotyped for bruising susceptibility and five other agriculturally important traits. This data was used for an association analysis, resulting in SNPs in candidate genes highly significantly associated with bruising susceptibility, tuber shape and specific gravity.

2. BACKGROUND: PHYSIOLOGY AND GENETICS INVOLVED IN BRUISING SUSCEPTIBILITY

In order to be able to select for useful candidate genes that might explain variation in bruising susceptibility, the physiological and genetic background of bruising susceptibility provide indispensible information. This chapter treats the relevant literature involved in enzymatic browning and bruising. As a result a concept map is presented (Figure 1), which is used as a theoretical framework to select for candidate genes used in this study.

2.1 Oxidoreductases and phenols: enzymes and substrates causing discoloration

The brown to blue colour typical for bruising is caused by melanins which consist mainly of non-enzymatically polymerized quinones. Quinones are synthesized out of phenolic compounds and oxygen by two groups of enzymes (oxidoreductases): the polyphenol oxidases (PPO) and peroxidases (POD; Figure 1). In intact cells the enzymes and phenolic compounds are separated, since phenolic compounds are mainly located in the vacuoles, PPO in the plastids and POD in the apoplast and intracellular spaces (Pourcel et al., 2007). Upon mechanical damage cell membranes are disrupted and the PPO and POD convert their substrate into quinones eventually leading to the pigmentation. PPO, among these catechol oxidases and laccases, are mainly involved in two sequential reactions affecting biosynthesis of quinones: the hydroxylation of monophenols to o-diphenols (Hunt et al., 1993) and the oxidation of o-diphenols to o-quinones (Hunt et al., 1993; Pourcel et al., 2007). Peroxidases (POD) catalyze only the latter oxidation reaction (Pourcel et al., 2007). Laccase catalyzes a third reaction, comparable to the second, which comprehends the oxidation of p-diphenols into p-quinones, also leading to pigmentation (Flurkey, 2003). The PPO catechol oxidase and laccase contain copper at their active site, which is needed for enzymatic activity (Pourcel et al., 2007).The PPO *POT32* has two copper binding sites (Friedman, 1997) and allelic variation in these binding sites was suggested to cause variation in bruising resistance (Werij et al., 2007).

The phenolic substrates for PPO and POD originate from the phenylpropanoid pathway (Pourcel et al., 2007; Figure 1). A key enzyme in this pathway is 4-coumarate:CoA ligase (*4CL*; Hamberger and Hahlbrock, 2004). One product of the reaction catalyzed by *4CL* is coumaroyl CoA. This molecule can be further converted into o-diphenols, among which are chlorogenic acid and caffeic acid, that are substrates for PPO and POD (Friedman, 1997). Urbany et al. (2011) found associations with allelic variation in the gene *4CL* with bruising susceptibility, suggesting involvement of genes early in the biosynthesis pathway. On the other hand, phenylpropanoids are also precursors for lignin (Taiz and Zeiger, 2010; Figure 1), which is an important compound in the cell wall. Therefore, genes involved in the phenylpropanoid pathway may not only influence the colouring itself, but may also affect the susceptibility to rupture of membranes, causing the enzymes and substrates to mix due to possible effects on cell wall rigidity.

2.2 The role of superoxide radicals in melanin synthesis

Strong discoloration caused by melanins only appears upon mechanical impact. This means that certain processes must occur as a reaction on the impact, before PPO or POD and their substrates come together to form melanins. Johnson and Doherty (2003) found that upon mechanical impact there is a biphasic production of superoxide radicals in the potato tuber tissue. The amount of superoxide produced explains a large part of the bruising susceptibility of a certain tuber (Storey, 2007). The first peak, which reaches its maximum around two hours after impact, was suggested to be caused by disturbance of the cell organization or ion balance (Johnson and Doherty, 2003). For tubers of a potato cultivar susceptible to enzymatic browning the first peak was followed by a larger second peak, about 4 to 5 hours after impact. The second peak was suggested to be caused by a certain pectin derived factor (Figure 1) that is formed at the first peak of superoxide

production, suggesting a positive feedback mechanism between production of the superoxide radicals and the pectin derived factor (Figure 1). When cell extracts from a susceptible cultivar after mechanical impact were applied to intact tissue of a resistant cultivar, the tissue showed a comparable superoxide production as if it were susceptible, suggesting that this factor is not produced in resistant tubers, explaining their resistance. Superoxide radicals are highly reactive, and are most probably preferentially used as co-factor over oxygen in the synthesis of melanins by PPO (Valverde et al., 1996) and POD. Radical scavenging would then reduce the enzymatic browning (Figure 1). This has been reported: ascorbic acid and gluthathione, both reducing agents, inhibited browning in an apple extract (Henze, 1956), and this concept is widely applied in food processing (Ramirez et al., 2003). By taking into account the correlation between the amount of produced superoxide radicals and bruising susceptibility, Johnson and Doherty (2003) suggested that it is plausible that the presence of high amounts of superoxide radicals is causative for the conversion of phenols and diphenols into quinones, resulting in accumulation of melanins (Figure 1). Following this suggestion it may well be the case that not direct mechanical damage to the cell membranes causes the substrates and enzymes to come together, but disruption of cell membranes may be caused mainly indirectly by the superoxide radicals (Figure 1), which react with the unsaturated fatty acids in lipids that make up the cell membranes.

The ability to produce the pectin derived factor is most likely not the only factor explaining variability in bruising susceptibility, and there are probably more processes explaining variability between bruising susceptibility among genotypes in potato, since Johnson and Doherty (2003) used only a small amount of genotypes. Anti-sense expression of PPO genes can lead to a strongly reduced bruising susceptibility (Bachem et al., 1994; Coetzer et al., 2001; Rommens et al., 2006), and allelic variation in PPO (*POT32* and *POTPOLOXA*) correlated with PPO expression and bruising susceptibility (Werij et al., 2007; Urbany et al., 2011), strongly suggesting that allelic variation in PPO genes also underlies genetic variation in bruising susceptibility.

2.3 New candidate genes and markers associated with bruising resistance

In Urbany et al. (2011) association mapping was used to find molecular markers associated with bruising resistance. Markers were derived from candidate genes, which were found by using comparative proteomics (Urbany et al., 2012) and literature research. For the association mapping experiment a population of 205 individuals was established and phenotyped (further referred to as ALL205 population). It consists of 85 varieties and 120 breeding clones, which were collected from six different breeders. In total 21 highly significant markers were obtained associated with bruising susceptibility using this method. In the population bruising susceptibility correlates strongly with specific gravity and tuber shape. Therefore, many markers associated with high bruising susceptibility also associated with these two other traits. 11 out of 18 markers found in Urbany et al. (2011) were associated with bruising susceptibility and specific gravity, and affected the trait in the same direction. The correlation with specific gravity corresponds with breeders' observations, and is seen as an adverse relation. It seems that a high specific gravity is causative for bruising susceptibility, since candidate genes originally associated with starch metabolism, like starch phosphorylase L-type, significantly associated with bruising susceptibility (Urbany et al., 2011; Urbany et al., 2012). The main hypotheses explaining the correlation between bruising susceptibility and specific gravity are: (i) physical linkage of genes affecting both traits and (ii) that a higher starch granule load would easier rupture cell membranes causing decompartmentation, leading to mixture of PPO and its substrates (Urbany et al., 2011; Figure 1). The first hypothesis doesn't seem likely, since genes on several different chromosomes affect both traits. It would be unlikely that there would be linkage between the traits on all these loci. The latter hypothesis seems more feasible, however, there is no evidence supporting the relation between the amount or size of starch granules and membrane rupture.

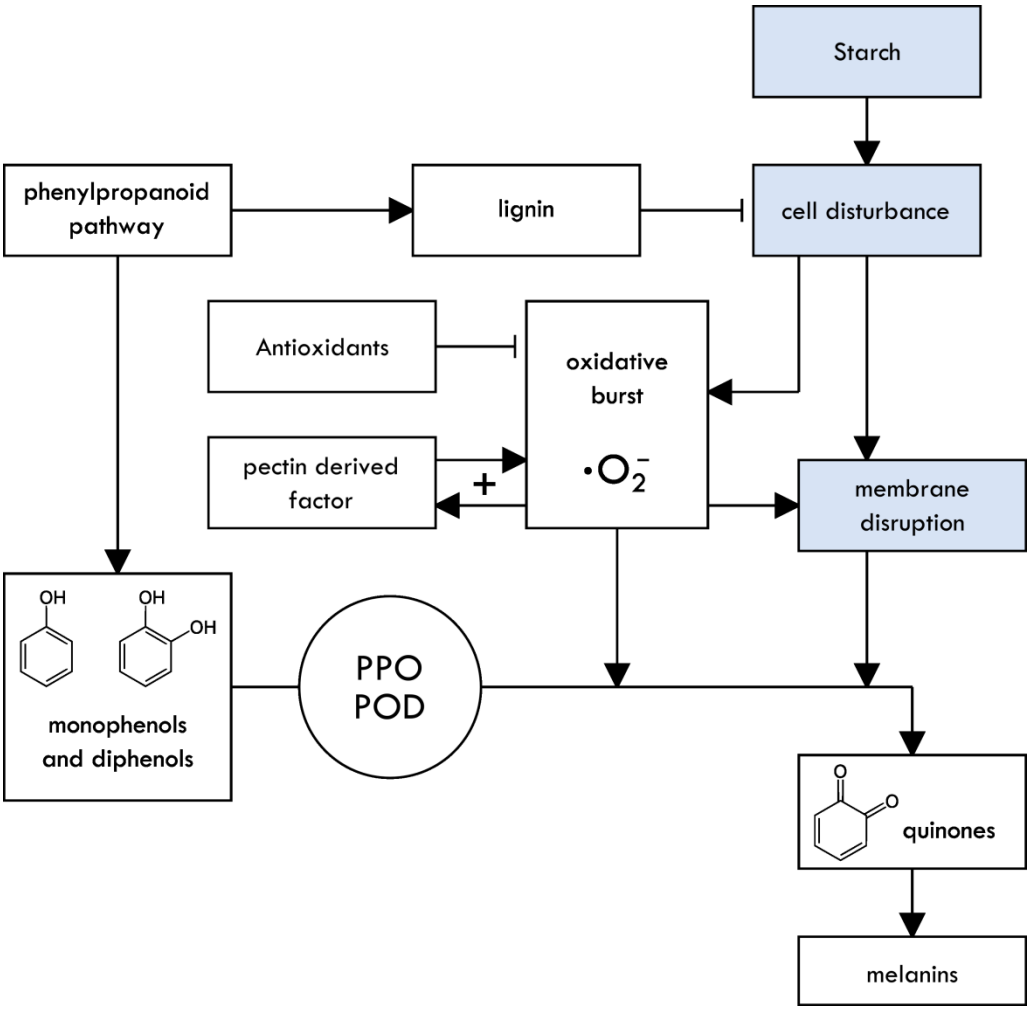


Figure 1. Schematic overview of processes involved in enzymatic browning. PPO: polyphenol oxidases, POD: peroxidases. The '+' sign indicates a positive feedback mechanism. Blue boxes represent the physical processes, the white boxes represent biochemical processes. Arrows indicate causative relations and lines with short vertical bars at their end indicate inhibitions.

3. BULKED SEGREGANT ANALYSIS

A bulked segregant analysis (BSA) was performed with 10 cultivars with low bruising susceptibility (cases) and 10 cultivars with high bruising susceptibility (controls) from cultivars in the ALL205 population described in Urbany et al. (2011; Table 1). These 20 cultivars were genotyped with an Illumina SNP array containing ~8300 SNPs developed by the SolCAP project (http://solcap.msu.edu/). The SNPs in this array were obtained from cDNA libraries from three North American varieties and from sequences from a European variety and two other North American varieties (Hamilton et al., 2011). All SNPs are in coding regions and many of them are in annotated genes, so associations with these SNPs provide interesting knowledge for developing diagnostic markers using a candidate gene approach. Genotyping these 20 cultivars with the SolCAP SNP array resulted in 677 SNPs that associated significantly ($p<0.05$) with bruising index according to a t-test corrected for multiple testing, in which the average allele frequency per bulk was compared (unpublished results).

Table 1. Cultivars and bruising index of genotypes used in the BSA

Cases		Controls	
Cultivar	**Bruising Index[1]**	**Cultivar**	**Bruising Index**
Remarka	11,2	Panda	60,2
Rafaela	11	Olga	66,2
Quarta	11,6	Maxilla	73,5
Marabel	11,4	Logo	73,3
Lolita	7,2	Lady Rosetta	67,8
Krone	7,7	Kuba	71,1
Gala	8,9	Kolibri	72,3
Elfe	9,4	Fitis	55,5
Carmona	6,6	Calla	56,1
Agila	10,4	Aspirant	73,7

[1]Bruising index: a measure for bruising susceptibility ranging from 0 (low susceptibility) to 100 (high susceptibility). For more details see materials and methods.

4. RESEARCH OBJECTIVE

The major objective of this project was to provide SNP markers in annotated genes that are associated with bruising susceptibility in the population described in Urbany et al. (2011). The starting point was the dataset of 677 significant SNPs (at $p<0.05$) from the already performed BSA. Candidate SNPs for association in a large population were selected based on their significance and/or biological function of their associated putative gene involved in processes described in Figure 1 and chapter 2. After that, they were evaluated in the entire population by pyrosequencing and Sanger sequencing, in order to find robust SNP-trait associations. Since bruising susceptibility strongly correlates with specific gravity and tuber shape in the ALL205 population, it is inevitable that the BSA also resulted in SNPs strongly associated with specific gravity and tuber shape. Therefore, part of this thesis will also deal with these traits in relation with bruising susceptibility. The overall aim of this project is providing a better understanding of the genetic background of these three traits and their interrelations, with an emphasis on bruising susceptibility.

5. MATERIALS AND METHODS

5.1 Population

The association mapping population used for this study consists of 188 genotypes (further referred to as ALL188 population), which were selected from the ALL population described in Urbany et al. (2011). 17 genetically similar cultivars were left out, based on population structure. The number of genotypes was reduced, in order to fit the population into two 96 well plates.

5.2 Phenotypic data

The phenotypic data evaluated for this study are bruising index (BI), specific gravity (SG), starch corrected bruising (SCB), tuber yield (TY) and tuber shape (TS). The BI was obtained by exposing tubers to mechanical stress (e.g. a washing machine), and after 4-5 days the bruising was scored per potato for three levels: light discoloration (L), average discoloration (M) and strong discoloration (S). The bruising index for the entire batch was calculated using the following formula: BI=[(0.3L + 0.5M+ S)/total tuber number] x 100, resulting in an index ranging from 0 to 100. The SG, expressed in estimated percentage starch content, was measured by dividing the tuber weight in air by the difference between the weight in air and the weight in water. This SG is an indirect measure for starch content, since starch makes up by far the greatest part of the tuber dry weight. The SCB was obtained from the residuals of the regression between BI and SG, resulting in a measure for bruising corrected for the SG. The TS was assessed by scoring from 1 (completely circular) to 9 (longitudinal). A detailed description of measurement and analysis of phenotypic data can be found in Urbany et al. (2011).

5.3 Polymerase chain reactions (PCR) and Sanger sequencing

The PCR were performed in a 25 µL reaction solution with 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 1.5 mM $MgCl_2$, 0.2 mM of each dNTP, one unit of homemade Taq polymerase, 0.4 µM of each primer and 20 ng of genomic DNA. Standard cycling conditions were 2 minutes initial denaturation at 94°, followed by 40 cycles of 45 seconds denaturation at 93°C, 45 seconds annealing at 57°C and 1 minute extension at 72°C. Reactions were finished by 10 minutes incubation at 72°. After PCR, the success of amplification was checked on an ethidium bromide stained 1.5% agarose gel. Primers were synthesized by SIGMA (Taufkirchen, Germany), and are displayed in Table 2. Prior to Sanger sequencing, leftover dNTPs and primers in the template were removed with ExoSap-IT (Affymetrix; according to manufacturer's protocol). Sanger sequencing of the amplicons was performed by the Max Planck Genome Centre (Cologne, Germany). SNP calling from sequencing data was performed with the software DAx (Van Mierlo Software Consultancy, Eindhoven, The Netherlands). The SNP dosage was determined according to the ratio of the area under the nucleotide curve, assuming the five different possible allele dosages that are possible in a tetraploid (AAAA, AAAB, AABB, ABBB, BBBB). The outcomes of the software analyses were manually checked and corrected when needed.

5.4 Pyrosequencing

For pyrosequencing the reverse primer was biotin labelled at the 5' end (Table 2). The PCR and amplicon quality check were carried out as described in subchapter 5.3, except for the number of cycles in the PCR, which were 50 in this case. After PCR, biotin labelled amplicons were captured onto sepharose beads (GE Healthcare) in a 80 µL solution containing 15 µL of PCR product, 40 µL of binding buffer (Qiagen) and 5 µL of sepharose beads solution. With use of a vacuum prep workstation (Biotage AB, Uppsala, Sweden), the beads were sucked out of the buffer solution and were washed with 70% ethanol for 3 seconds, denaturated with 0.2 M NaOH for 5 seconds, and washed again with 10 mM Tris-HCl (pH 7.6) for 8 seconds. After that, the beads were dropped into a 40 µL solution containing annealing buffer (Qiagen) and 0.25 µM sequencing primer. Primers were annealed to the sepharose captured amplicons for 2 minutes at 80 °C. The pyrosequencing

was performed using pyromark gold Q96 reagents kit (Qiagen) and a PSQ™ 96 MA pyrosequencing instrument (Biotage AB) according to manufacturers' protocols. The relative SNP dosage was estimated as a percentage by the software supplied by the manufacturer (PSQ™ 96MA 2.1.1; Biotage AB). The percentage of the dosage of the two nucleotides adds up to 100%. These values were used to classify the allele dosage per genotype into one of the 5 possibilities.

Table 2. PCR primers and amplicon length

Candidate gene[1]	Primers (5' to 3')[2]	Amplicon length (bp)
Ampliconsequencing		
Mco	f,s-AGAGATCGGGCCTGATTGAA	798
	r-TTTCCTTCCCTTGCTTTCCA	
Vps35	f,s-TTCGTTGCGGATTGCTAATG	909
	r-CGTTTGGTCCACTGGCCTAT	
Pest1	f-TGTTGCAGATTTGCTCTTCAGTC	852
	r,s-CTGCAACTGGAGCTGGTTTTATT	
StHSL1	f,s-TGCAGGAGAATCTGGAAGCA	820
	r-CTTGAATCCGCAGAGGAAGG	
Ala	f,s-CACAAAAGCCACCACGACAT	802
	r-CATGAGGTTGCAAGGTCGAA	
Ghf	f-TCGGAAACATCCTGTTGTGC	926
	r,s-GCGATTTTCCACCGTATGCT	
ATSIK	f,s-GAGGTGAACACCCTGGCTTC	1067
	r-TCCTTTGAATACCGGCCATC	
AspAT	f,s-GCTGTGCCCAATCTTTTTCC	695
	r-TGTTCCCAGCTCGAAGATGA	
Pyrosequencing		
CuBP	f-GGCGACGCTTCTCCTCTTT	416
	r,b-AGGAACTATGGAGCACACAAAGG	
	s-TCCTTATCTGCTTTGACACTCTCCC	
Ce	f-TCCACCAGAATGTCAGTCCAA	272
	r,b-ACGCGATCCGAGTCCTATTC	
	s-GGACTTTCATGGGAAGTAGCAA	
SKIP10	f-GGAATCAGGGTTCAAGTTAGAAGTG	414
	r,b-GAGTTCCTAATGGGCATGCAG	
	s-GAATTACCAACAACTCAGATCGTC	
ChoK	f-GTTCGTGTCGATTGCAGCTT	387
	r,b-GGGCTCTCTGGGCTTTGAT	
	s-ATAACTTGGGTAAAGAGATGACTTT	
Gt	f-AGCCTTTGGGAAACAATGAATG	363
	r,b-CATGGTGGATTTGCTCTTTGC	
	s-GTGTAGAAACCACTGAACCATTAAA	
Zds	f-TGCAATGCCTAACTGGGTTG	200
	r,b-CATGGAAGGGGCAACTCTTT	
	s-CACCTTTAGAGATCTCGTTTAACTC	
BIPx	f-CGTGTCAATGTTGGTGTCGTT	219
	r,b-GTACCATTGGGCCGAAGAGA	
	s-GGAGAGTAAAGTTGAAAGACTGGAA	

[1]More information about candidate genes can be found in Table 4
[2]r: reverse primer, f: forward primer, s: sequencing primer, b: biotin labelled at 5' end

5.5 Location of unanchored superscaffolds

For determination location of unanchored superscaffolds, coding sesquences (CDS) of the superscaffolds were BLASTed against the tomato genome (http://solgenomics.net/tools/blast/; release 2.3). The location was estimated relative to the potato sequence when the superscaffold bordered known homologues regions between the potato and tomato physical map. When the superscaffold did not border homologues regions, the distance between two regions of interest in potato was estimated by the distance between the regions on the tomato physical map. Regions in which unanchored superscaffolds were located were checked for macroinversions between potato and tomato, which are published by The Tomato Genome Consortium (2012).

5.6 Data analysis

Only the SNPs that were heterozygous in more than 5% of the genotypes in the ALL188 population were used for analysis. Statistical analysis was performed with the software R (www.r-project.org). For the difference of average phenotypic values between the cases and controls of the BSA analysis a t-test was performed. Significance of associations was tested with analysis of variance (ANOVA) according to the following general linear model (model 1):

$$Adjusted\ entry\ mean = SNP\ dosage + error$$

The adjusted entry means of above treated phenotypic traits were obtained from Urbany et al (2011). The number of levels of SNP dosage varied from 2 to 5.The stringent threshold of significance for the association analyses was derived from the Bonferroni multiple testing correction (at p=0.05), using all p-values resulting from ANOVAs performed for a single trait. The significance of linkage disequilibrium between two loci was calculated by testing for independence between two loci with a X^2 (chi-square) test to a 5 by 5 (5 possibilities of different allele dosages on both loci) contingency table for each locus combination. LD p-values were corrected for multiple testing using the R package ‘qvalue’, using the method described by Storey (2003). Protein and DNA sequences were aligned with ClustalW (Thompson et al., 1994). Sequence similarity was assessed with EMBOSS Needle (http://www.ebi.ac.uk/Tools/psa/emboss_needle/). The neighbour-joining tree was constructed with the software MEGA5 (Tamura et al., 2011).

6. RESULTS

6.1 Phenotypic data

The phenotypic data of the ALL205 population is described in Urbany et al. (2011). The ALL188 population, which was derived from the ALL205 population and used in this study, was reanalyzed for correlations among traits (Figure 2A) and phenotypic distribution (Figure S 1). As expected, the strength of correlations and shapes histograms were virtually the same when data of the ALL205 population is compared to the ALL188 population. Also partial correlations were calculated for the phenotypic data of the ALL188 population (Figure 2B). These are the correlations between the variables while removing the effects of all other variables on this relationship. Since SCB is a derived trait, it was not used in the partial correlation analysis. The correlation between BI and PM did not remain after partial correlation analysis, whereas the correlation between TS and PM was added after partial correlation analysis. Other correlations remained and kept the same polarity. The correlation between SG and BI was still highest of all correlations in the partial correlation network, although it was reduced.

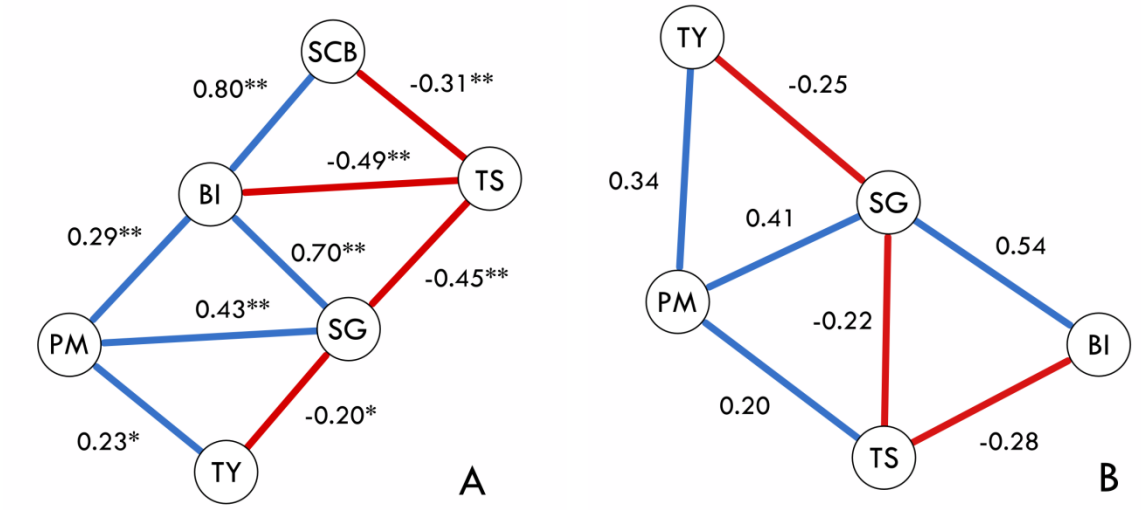


Figure 2. Correlation network (A) and partial correlation network (B) of analyzed phenotypic traits. Only correlations with a Pearson's correlation coefficient > 0.2 are displayed. Blue lines indicate positive correlations, red lines indicate negative correlations. * significant at p=0.01; ** significant at p=0.001.

The distribution of phenotypic values for the two groups (cases and controls) used for the BSA analysis are displayed in boxplots in Figure 3. All traits except TY differed significantly ($p<0.05$) between the cases and controls. The two groups were significantly very different for BI ($p<1e^{-5}$), SG and SCB, and less significantly different ($p<0.05$) for PM and TS. The differences between average trait values are in concordance with the correlation between bruising index and the other trait values. BI is positively correlated with SG, SCB and PM. This can also be seen from the boxplots where the cases show lower trait values than the controls for all four traits. BI is negatively correlated with TS, which can also be seen in the boxplots; for TS the average trait value is higher for the cases than the controls.

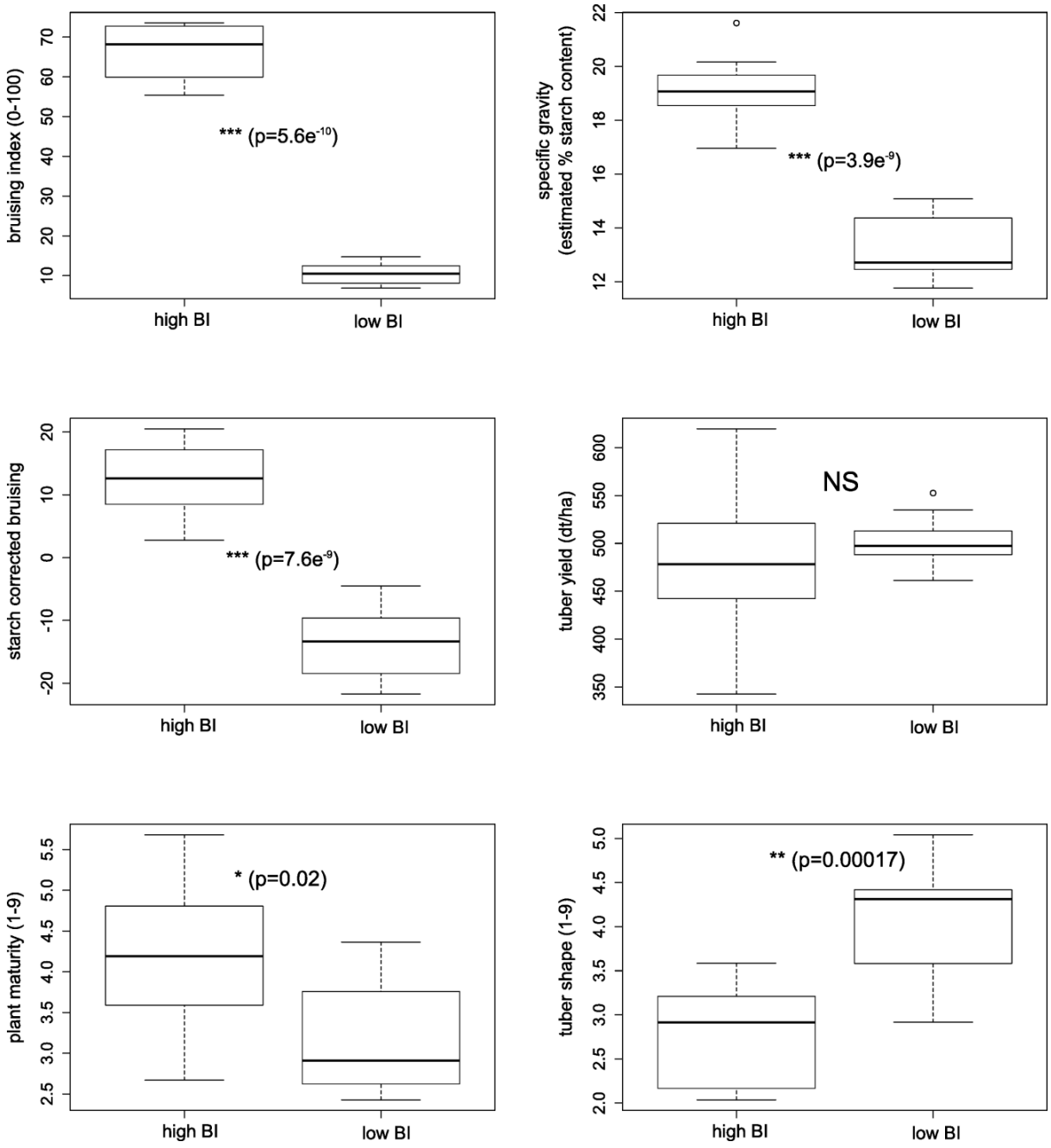


Figure 3. Boxplots of phenotypic values of cases (low BI) and controls (high BI) used in the BSA analysis. P-values for a t-test for the difference of the means between cases and controls are displayed in the graph. * significant at p=0.05, ** significant at p=0.001, * significant at p=1e^{-5}.**

6.2 Candidate genes

The BSA resulted in significant SNPs in gene accessions which were already found to be associated with bruising resistance in Urbany et al. (2011)(Table 3). Since their associations are already described, the genes in Table 3 were excluded as candidate genes. The same accessions of 'PPO isoform POTPOLOXA' and 'hydroxycinnamoyl CoA quinate transferase' (Table 3) that resulted from the BSA were previously described by Urbany et al. (2011). In addition, SNPs in homologs of genes which were found to be associated by Urbany et al. (2011) were found significant in the BSA analysis (Table 3). These homologs are located on the same chromosomes. The gene 'Polyphenol oxidase B, chloroplastic' is proximate to 6 other PPOs among which are *POT32* (~30 kbp away) and *POT33* (~50 kbp away).

Table 3. Candidate genes resulting from the BSA equal or similar to genes significantly associated in Urbany et al. (2011).

Annotation PGSC	Locus ID PGSC[1]	SolCAP SNP[2]	chromosome	q value BSA	Annotation in Urbany et al. (2011)	same accession
Catechol oxidase B, chloroplastic	400029575	c2_44293	VIII	0.0065	Polyphenol oxidase isoform POTPOLOXA	yes
hydroxycinnamoyl CoA quinate transferase	400011189	c1_15912	VII	0.0346	Hydroxycinnamoyl CoA quinate transferase	yes
Polyphenol oxidase B, chloroplastic	400018925	c2_50715	VIII	0.0020	-	no
4-coumarate:CoA ligase	400003155	c2_29643	III	0.0035	4-coumarate:CoA ligase	no

[1]Locus ID after PGSC0003DMG..
[2]SolCAP SNP ID after solcap_snp_..

In total 11 different candidate genes were selected based on their significance in the t-test from the BSA, their biological function and novelty (Table 4; Figure 4). Two candidate genes had a putative function in carbohydrate metabolism, two in membranes and vacuoles, two in copper metabolism, one in peroxidation, one in antioxidants and one in gene transcription. All candidate genes are on different superscaffolds except for *Vps35* and *Ce*, which were both on PGSC0003DMB000000034. They are about 2 Mb away from each other and in significant LD according to a X^2 test ($p=0.005$).

Table 4. Selected candidate genes

SolCAP SNP[1]	Annotation[2]	Abbreviation	Chromosome	q value[3]	Metabolism
c1_5286	Zeta-carotene desaturase	*Zds*	I	0.0080	Carotenoids
c2_46603	Bacterial-induced peroxidase	*BIPx*	III	0.00067	Peroxidation
c2_55285	Pectin methylesterase 1	*Pest1*	III	0.0041	Cell wall
c2_1484	Copper ion binding protein	*CuBP*	IX	0.00030	Copper
c2_45180	Multicopper oxidase	*Mco*	VII	0.0019	Copper oxidation
c2_44304	Choline/ethanolamine kinase	*ChoK*	VIII	0.0015	Membrane lipids
c1_8019	High-sugar inducible like1[4]	*StHSL1*	X	0.00007	Gene transcription
c2_55861	Transferase, transferring glycosyl groups	*Gt*	X[5]	0.0001	Carbohydrate
c2_7902	Vacuolar protein sorting	*Vps35*	XII	0.00028	Vacuolar assortment
c1_2350	Carbohydrate esterase	Ce	XII	0.00067	Carbohydrate
c2_45743	Ski-interacting protein 10	*SKIP10*	XII	0.00080	Stress

[1] SolCAP SNP ID after solcap_snp_..
[2]According to the PGSC Genome Browser (http://potatogenomics.plantbiology.msu.edu version DM1-3516R44; The Potato Genome Sequencing Consortium, 2011)
[3]Significance measure of a t-test of the BSA analysis corrected for multiple testing
[4]Annotation obtained by NCBI nBLAST of cDNA sequence (http://blast.ncbi.nlm.nih.gov)
[5]On an unanchored superscaffold, location inferred from homology search with the tomato physical map (Materials and methods)

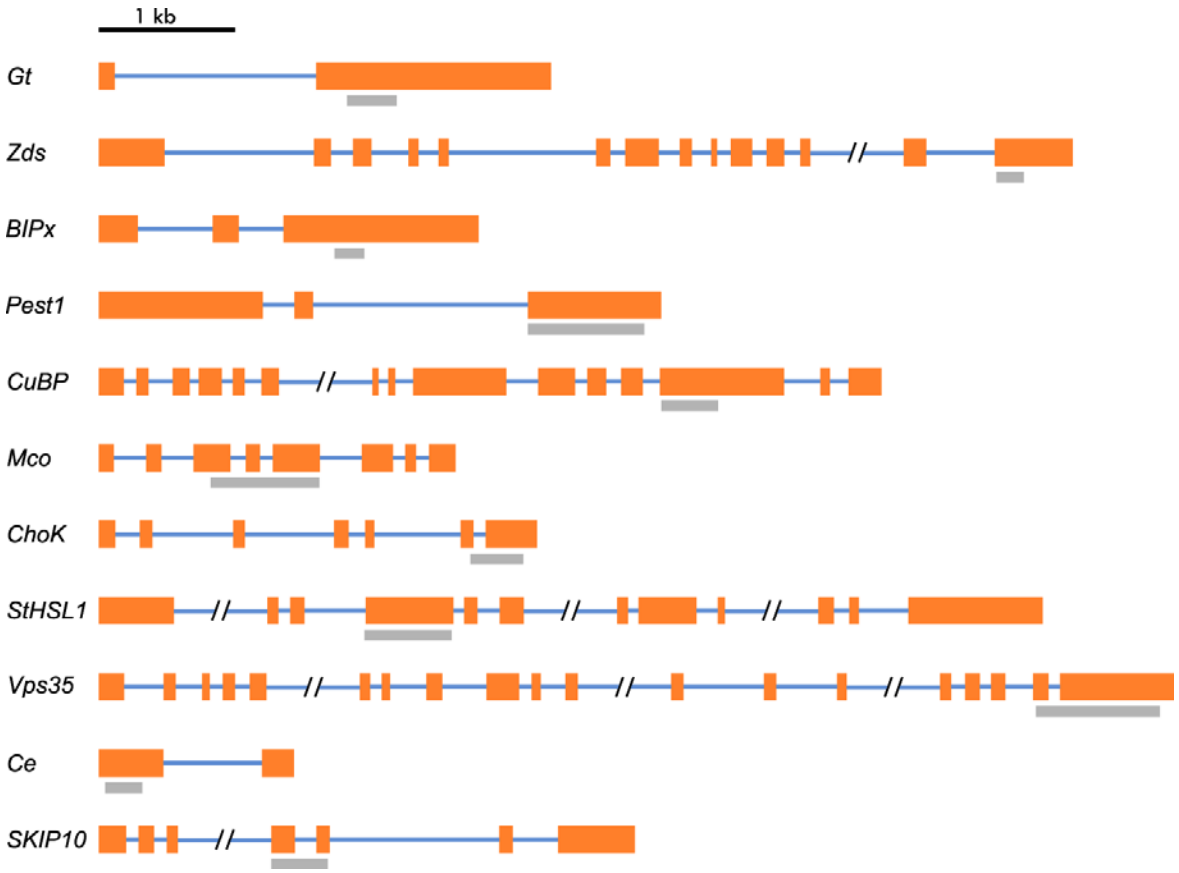


Figure 4. Gene models of selected candidate genes. Orange blocks represent exons, blue stripes represent introns and grey blocks indicate the amplicon used for Sanger- or pyrosequencing. Diagonal double stripes indicate that the intron is longer than displayed.

6.3 Association of the BSA derived candidate genes

In total 71 SNPs were found spread over 11 loci. These SNPs were all tested for association with BI, SG, SCB, TY, PM and TS using a linear model (Materials and Methods). This resulted in 33 SNPs associated at p=0.01 with BI, SG, SCB, TS, PM and/or TY (Table S 1). If SNPs were associated with more than one trait, the polarity in which it was associated was conform their correlations, meaning that in case of a positive correlation between two traits, polarities were in the same direction, and in case of negative correlation, polarities were in opposite direction (Table 5; Table 7). In this subchapter first the SNPs found with pyrosequencing are treated, after that SNPs found with ampliconsequencing are treated per candidate gene. Finally, the p-values found in the BSA analysis and in the association analysis of the same SNPs are compared.

6.3.1 Pyrosequencing

The quality of the pyrosequencing outcomes differed per SNP (Figure 5). While sequencing of some SNPs like solcap_snp_c2_1484 and solcap_snp_c2_45743 resulted in clear data in which all genotypes could be easily classified into one of the five possible SNP dosages, other experiments resulted in less clear cut results (e.g. in solcap_snp_c2_46603). For all SNPs the genotypes were classified into one of the five possible SNP dosages, except for solcap_snp_c2_46603, for which the quantitative dosage was used for analysis. This was because classifying it into the five dosages would probably have led to more mistakes than using the quantitative dosage.

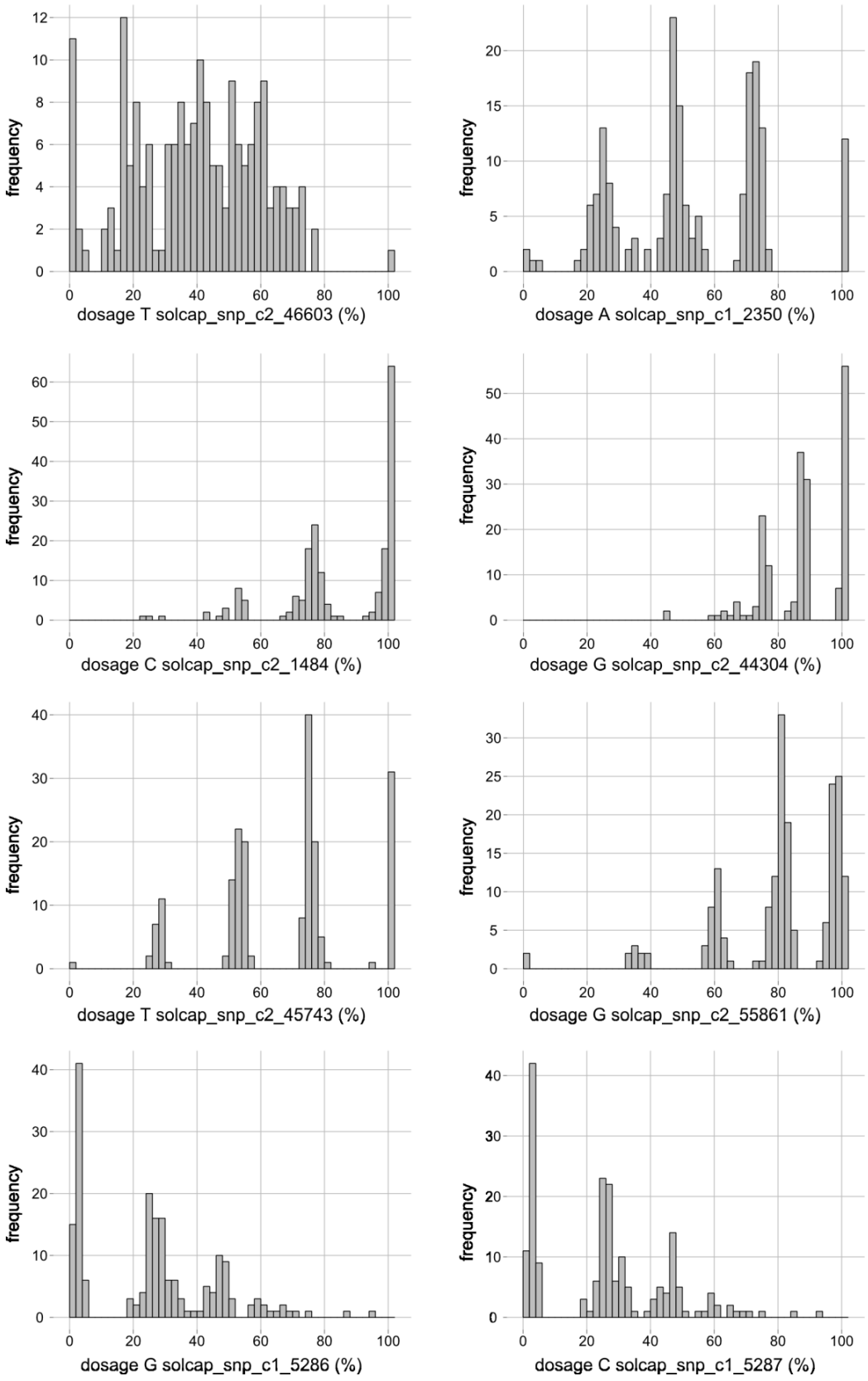


Figure 5. Histograms of allele relative dosages for eight SNPs in the genotypes of the ALL188 population. Only the relative dosage of the allele of the reference sequence (DM3-1 516R44 PGSC Genome Assembly) is shown. Relative dosages of complementary alleles always add up to 100%.

For all SNPs, the estimated SNP dosages genotyped by pyrosequencing showed no or very few differences with the outcomes of the Illumina array in the 20 BSA genotypes (except solcap_snp_c2_46603, for which the SNP dosage was not estimated). Six SNPs genotyped with pyrosequencing were significantly associated with BI in the ALL188 population (Table 5). Most of them were also associated with SG, except for *ChoK*, which was therefore associated with SCB. Although their dosages showed no differences with the Illumina array in the 20 BSA genotypes, *Ce* and *SKIP10* were not significantly associated with BI in the ALL188 population. None of the SNPs was associated with TY.

Table 5. Alleles, allele frequency, p-values, R^2 and polarity of associated SNPs (p=0.01) that were genotyped with pyrosequencing. Alleles are displayed with the allele with lowest frequency first. R^2 in % are in parentheses. The polarity of the effect of the allele is indicated by an arrow, and is the effect of the allele with lowest allele frequency.

SolCAP SNP[1]	Locus	Alleles	Allele frequency	BI[2] p (R^2)	SG p (R^2)	SCB p (R^2)	TS p (R^2)	PM p (R^2)
c2_46603[3]	*BIPx*	A/G	-	0.0034 (4.5) ↓	1.4e^{-6} (11.9) ↓	NS	NS	0.0012 (5.5) ↓
c1_2350	*Ce*	C/A	0.46	NS	NS	NS	NS	NS
c2_1484	*CuBP*	A/C	0.16	0.00067 (8.9) ↑	1.46e^{-6} (15.0) ↑	NS	NS	NS
c2_44304	*ChoK*	T/C	0.24	7.9e^{-5} (11.1) ↑	NS	0.0011 (8.4) ↑	0.0066 (6.5) ↓	NS
c2_45743	*SKIP10*	C/T	0.35	NS	NS	NS	NS	NS
c2_55861	*Gt*	A/G	0.23	9.4e^{-9} (21.1) ↑	2.9e^{-10} (24.2) ↑	NS	2.5e^{-7} (18.1) ↓	NS
c1_5286	*Zds*	C/A	0.26	0.00012 (11.9) ↓	0.0034 (8.2) ↓	NS	NS	NS
c1_5287	*Zds*	G/A	0.26	0.00012 (11.9) ↓	0.0034 (8.2) ↓	NS	NS	NS

[1]SolCAP SNP ID after solcap_snp_..
[2]NS: p>0.01
[3]Statistical analysis was performed with quantitative SNP dosage

6.3.2 Pectin methylesterase1

In total 23 SNPs were found in the *Pest1* locus (Figure 6). Of these SNPs 15 were associated with BI, SG, SCB and/or PM with p<0.01. Four SNPs were associated with only BI and SG at the more stringent threshold of p=3.5e^{-4}. There exists one large LD block within the gene consisting of 16 SNPs, and all 15 associated SNPs (at p=0.01) are within this block, all of them explaining variation in SG and most of them in BI. The most significantly associated SNP (solcap_snp_c2_55284) was associated with BI, SG and SCB. For this SNP, all SNP-trait associations had p-values below 0.001 and R^2 of 11.2, 12.4 and 8.7% respectively.

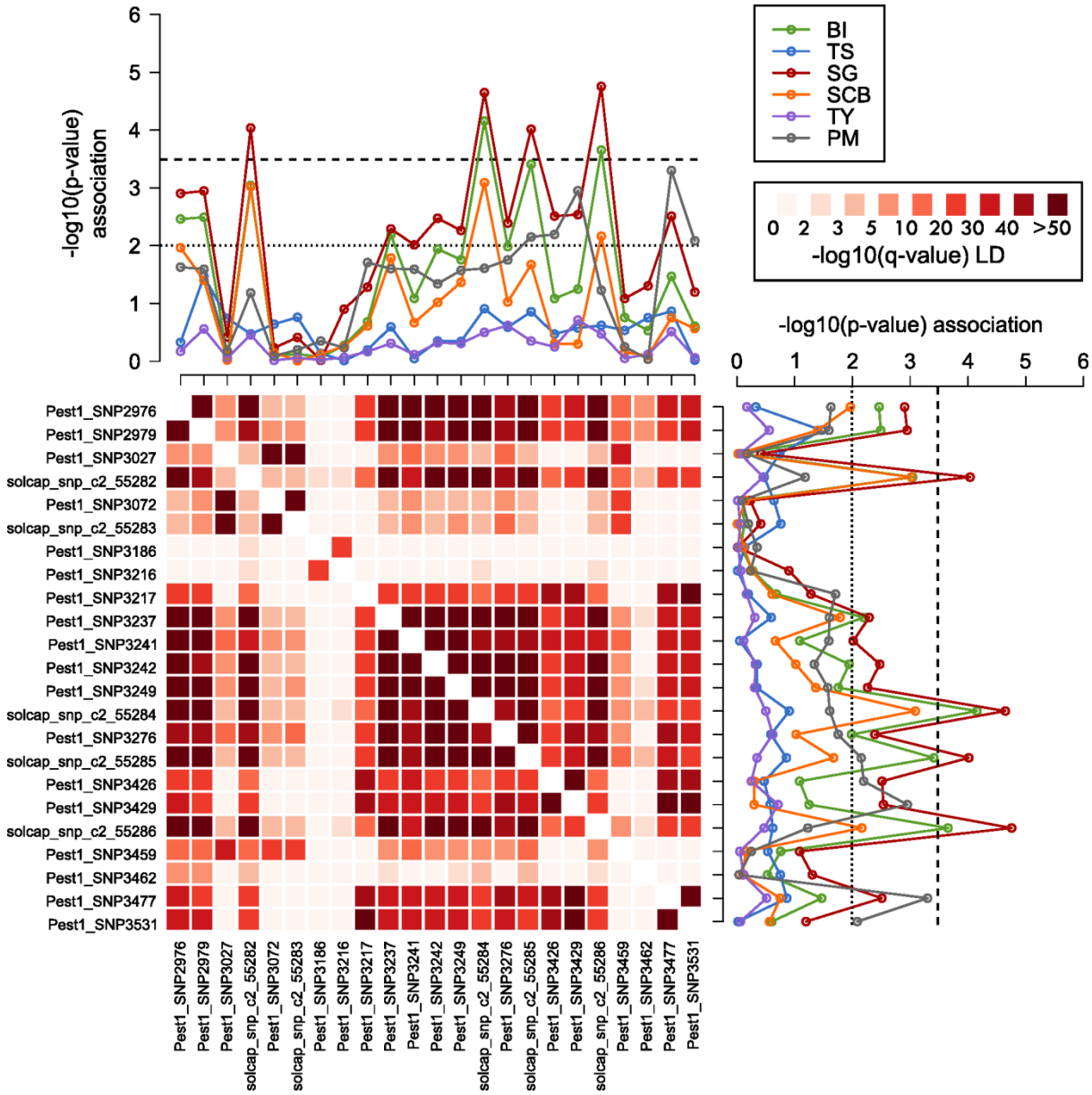


Figure 6. Combined LD heatmap and association significance plot of the SNPs in the *Pest1* locus. The q-values are p-values corrected for multiple testing. LD p-values were calculated with a X^2 test of a 5x5 contingency table (see Materials and methods). The dotted line in the lineplots indicates a significance threshold at p=0.01, the dashed line indicates the stringent significance threshold at p=3.5e-4 (see Materials and methods). Exact p-values and R^2 can be found in Table S 1.

6.3.3 Multicopper oxidase

In total 23 SNPs were found in the *Mco* locus, of which one was associated at a threshold of p=0.01 with PM, 2 with SG and TY, one with both SG and BI and one with only BI (Figure 7). One SNP remained to be associated with SG when the more stringent threshold of $p=3.5e^{-4}$ was applied. There existed no large continuous LD blocks in the amplicon, although strong LD can be observed between several SNPs (Figure 7). The 2 SNPs associated with PM (Mco_SNP1165), SG and TY (Mco_SNP1202) were not in LD with any other SNP. The 2 SNPs associated with SG and TY were in

strong LD with each other. Remarkably, the candidate SNP originating from the BSA did not show association with any of the traits.

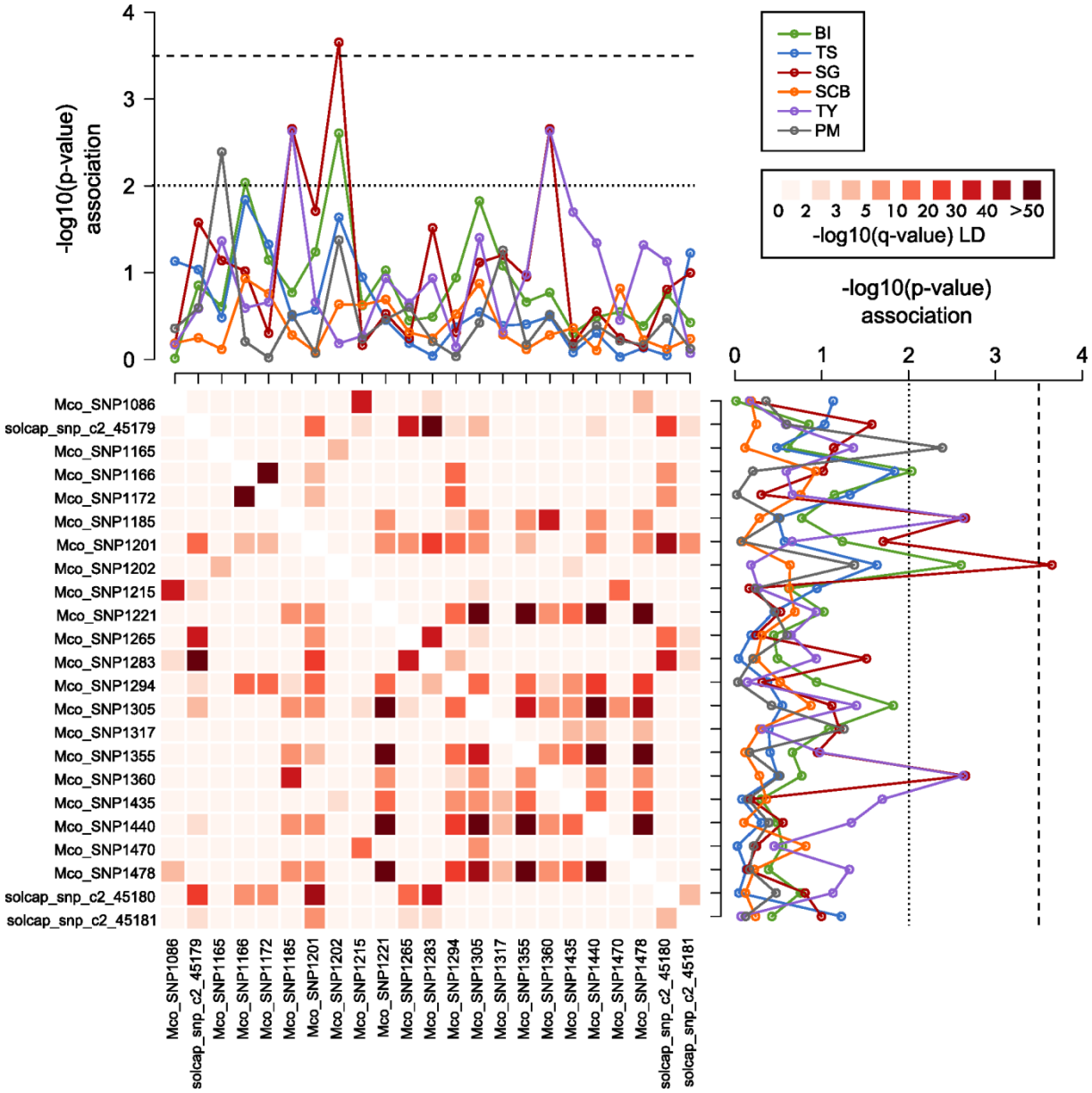


Figure 7. Combined LD heatmap and association significance plot of the SNPs in the *Mco* locus. The q-values are p-values corrected for multiple testing. LD p-values were calculated with a X^2 test of a 5x5 contingency table (see Materials and methods). The dotted line in the lineplots indicates a significance threshold at p=0.01, the dashed line indicates the stringent significance threshold at p=3.5e-4 (see Materials and methods). Exact p-values and R^2 can be found in Table S 1.

6.3.4 Vacuolar protein sorting35

In total 11 SNPs were found on the *Vps35* locus (Figure 8). One of those was associated at p=0.01 with SCB (R^2=6.9%) and one was associated with SCB (R^2=8.1%) and BI (R^2=8.2%).

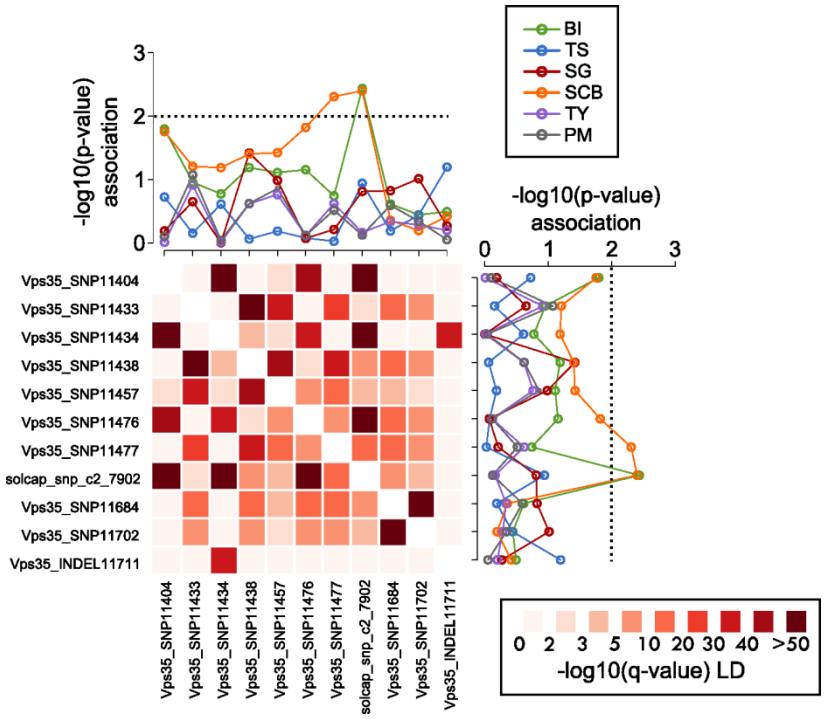


Figure 8. Combined LD heatmap and association significance plot of the SNPs in the *Vps35* locus. The q-values are p-values corrected for multiple testing. LD p-values were calculated with a X^2 test of a 5x5 contingency table (see Materials and methods). The dotted line in the lineplots indicates the significance threshold at p=0.01. Exact p-values and R^2 can be found in Table S 1.

6.3.5 Comparing significance of the BSA with association analyses

The two SNPs with the lowest p-value resulting from the t-test of the BSA analysis (in the genes *StHSL1* and *Gt*) also showed strongest association in the ALL188 population (Figure 9). Except for these two associations, a higher p-value resulting from the BSA analysis did not seem to be an indication for the p-value of the association in the ALL188 population.

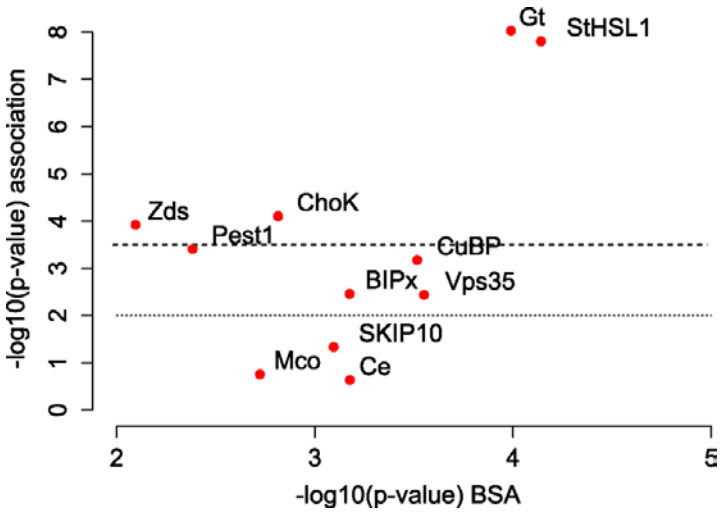


Figure 9. –log10 p-values of a t-test of the difference in allele frequency between cases and controls in the BSA analysis (-log10(p-value) BSA analysis) of selected SNPs, versus –log10 p-values of associations between the same SNP and BI in the ALL188 population (-log10(pvalue) association analysis). The dotted line in the lineplots indicates a significance threshold at p=0.01, the dashed line indicates the stringent significance threshold at p=3.5e-4 (see Materials and methods). Exact p-values and R^2 can be found in Table S 1.

6.4 Homology analysis of *StHSL1*

There was no annotation in the PGSC genome browser for the coding sequence (CDS) containing solcap_snp_8019; the top hit in the BSA analysis (Table 4). Therefore the CDS surrounding this SNP was blasted against the *Arabidopsis thaliana* genome (TAIR 10; http://www.arabidopsis.org/) and strongest homology was found with At4G32010, HSI2-like1 (*HSL1*). Using the CDS and protein sequence information of At4G32010 and CDS of potato, the intron-exon structure of its presumed homolog in potato (*StHSL1*) was estimated. This resulted in the finding of much larger exons in potato: up to 6.2 kb (Figure 10). The overall similarity of the CDS of At4G32010 and predicted CDS of *StHSL1* was 55.3% with some highly conserved regions, among which the specific domains (Figure 10; Figure 11).

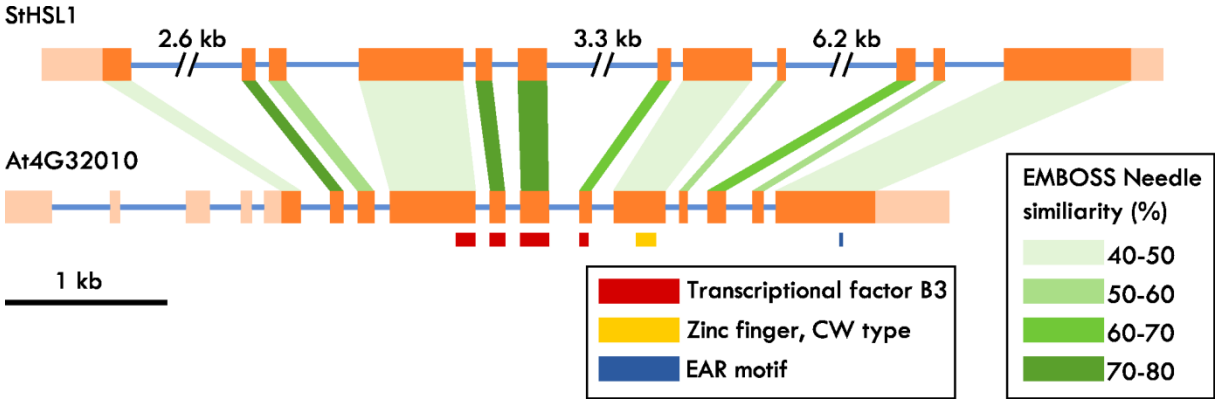


Figure 10. Gene structure, specific domains and similarities per exon of *StHSL1* and At4G32010. Orange bars indicate exons, light orange bars UTRs, and blue lines indicate introns. Green planes link the homologues exons. Similarity was calculated with EMBOSS Needle (http://www.ebi.ac.uk/Tools/psa/emboss_needle/).

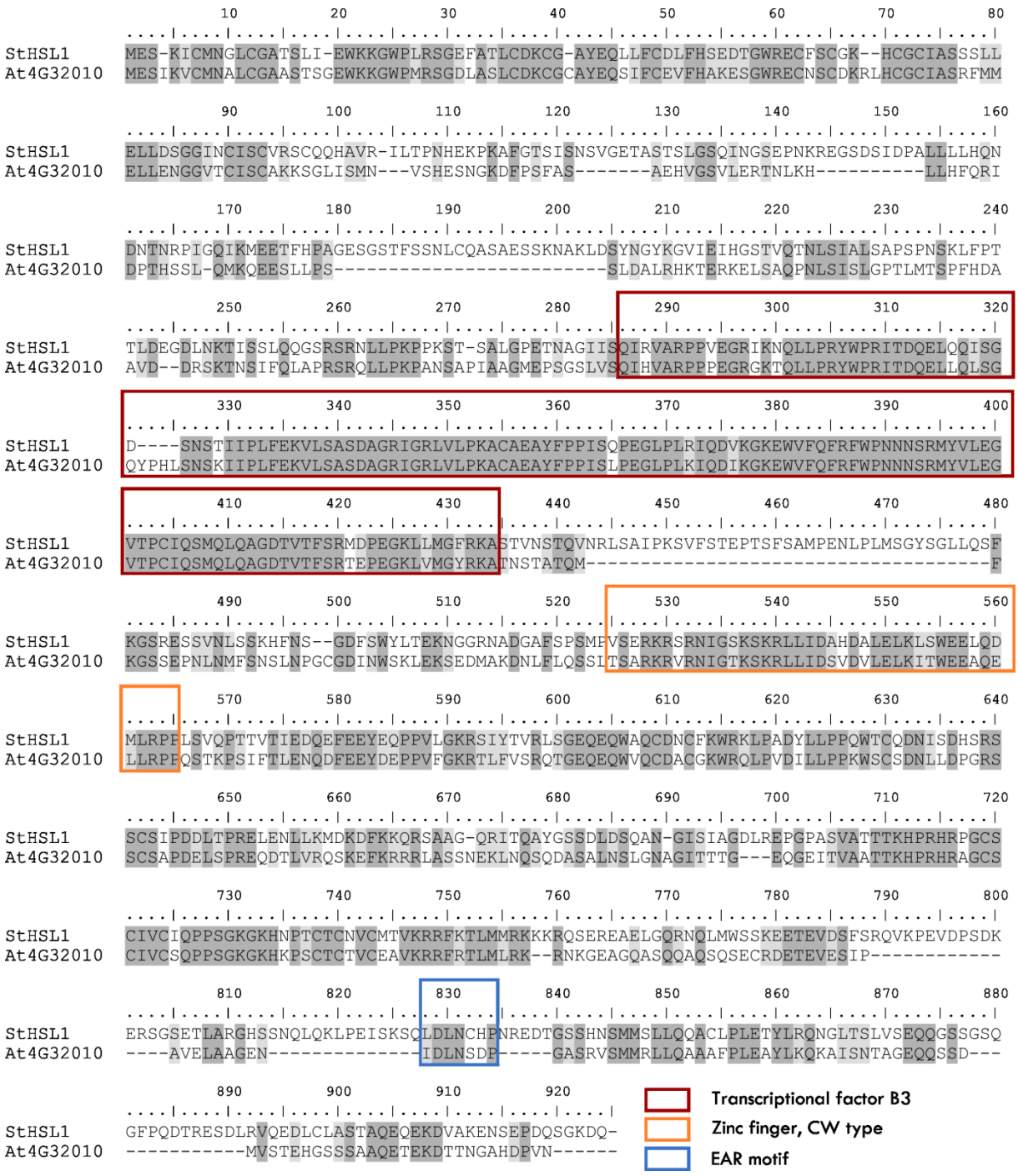


Figure 11. Protein sequence alignment between *StHSL1* and At4G32010.

StHSL1 was found to be very similar to the feature Solyc10g075030.1.1 on chromosome 10 in the tomato genome. The gene model of this tomato homolog is given the name 'AP2 domain-containing transcription factor' in the ITAG2.3 release of the tomato genome. Since it shows such strong similarity with *StHSL1* (and therefore also with At4G32010), it will from here further be referred to as *SlHSL1*. Two loci similar to *StHSL1* were found on the potato genome: PGSC0003DMG400030467 (*PGSC_30467*) on chromosome 6, and PGSC0003DMG400022362 (*PGSC_22362*) on chromosome 2. Both are annotated as 'transcription factor' on the potato genome. The gene models don't seem to contain the B3 binding domain, but there is unannotated CDS surrounding these gene models and they show some similarity in the Zinc finger region and EAR motif. In *A. thaliana* the four top hits from blasting the CDS of *StHSL1* to the TAIR10 *A. thaliana* genome were all B3 domain binding containing proteins: *HSL1*, *HSI2*, *VAL3* and *ABI3*. All

genes described above were used to construct a neighbour-joining tree based on sequence similarity (Figure 12).

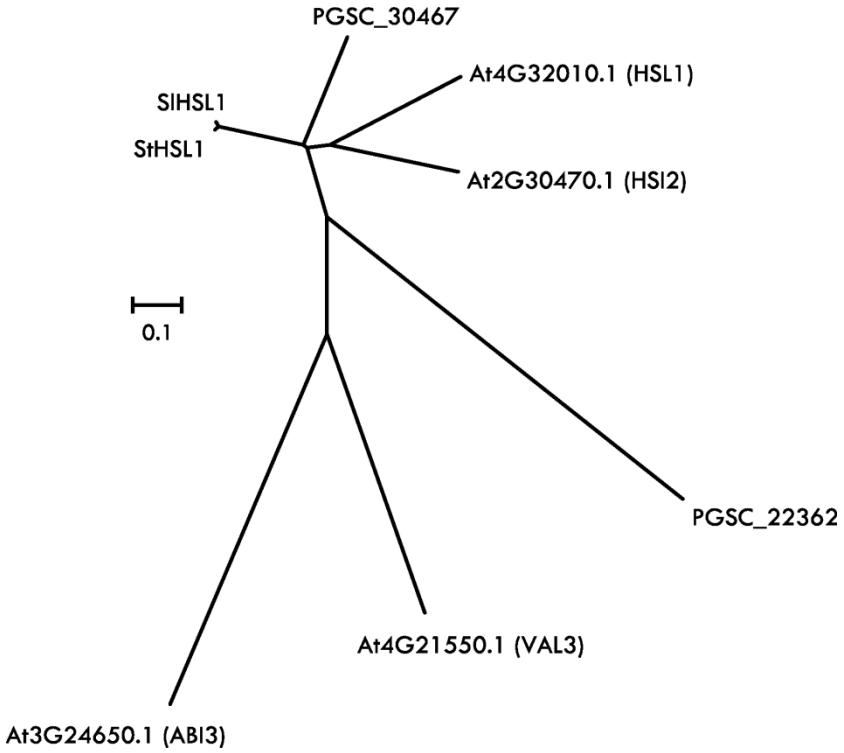


Figure 12. Phylogenetic tree of CDS similar to *StHSL1*. Branch lengths are proportional to the amount of nucleotide differences.

6.5 Association analysis of *StHSL1* neighbouring region

The top hit from the BSA analysis, solcap_snp_c1_8019 in the gene *StHSL1* (Table 4), was strongly associated with BI and SG, but most strongly with TS (explaining 39.8% of the variation; Table S 1). The closest marker on the PoMaMo map to *StHSL1* is TG63, and this marker is indeed linked to the *Ro* locus, which explains variation in tuber shape (Van Eck et al., 1994). However, *StHSL1* and TG63 are still 4.3 Mb away from each other. The marker CT217 (Zhang, 2009), which is in the gene 'calcium-binding protein CAST ', is also closely linked to the *Ro* locus and lays only 415 kb away from *StHSL1*. In order to locate the *Ro* locus on the physical map, parts of five loci in the neighbouring region were sequenced. The SNPs and INDELs found in these sequences are displayed in Table 6 and Figure 13. An overview of p-values and R^2 of all SNPs and INDELs are displayed in Table S 1.

Table 6. Sequenced genes neighbouring *StHSL1*.

Annotation	Abbreviation	Locus ID PGSC[1]	Superscaffold[2]
Protein kinase atsik	*ATSIK*	400031235	385
Phospholipid-transporting ATPase	*Ala*	400031222	385
Hydrolase, hydrolyzing 0-glycosyl compounds	*Ghf*	400031221	385
Calcium-binding protein CAST	*CAST*	400027692	385
Aspartate aminotransferase	*AspAT*	400006678	546

[1]Locus ID after PGSC0003DMG..
[2]Superscaffold ID after PGSC0003DMB000000...

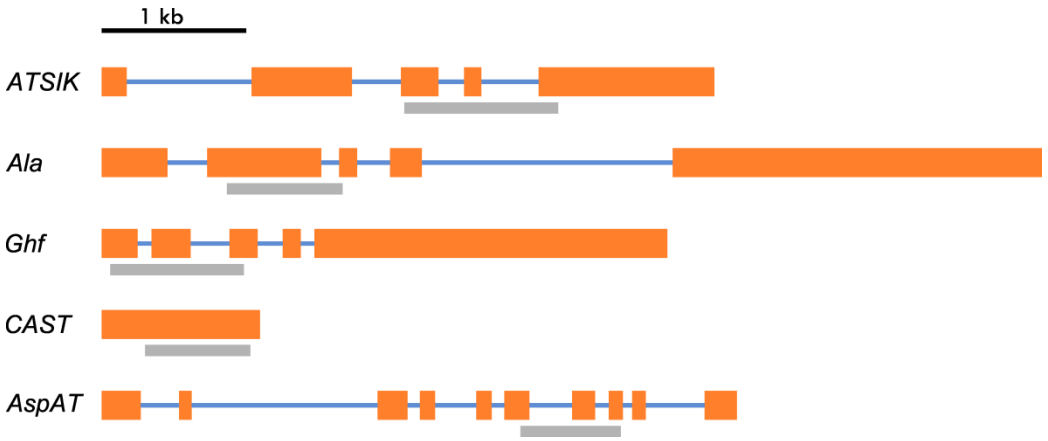


Figure 13. Gene models of sequenced genes neighbouring *StHSL1*. Orange blocks represent exons, blue stripes represent introns and grey blocks indicate the amplicon used for Sanger sequencing.

It was unknown which superscaffolds neighboured the superscaffold containing the *StHSL1* locus. This knowledge was needed in order to be able to sequence loci located in a larger region proximate to the *StHSL1* locus, since this locus is close to the superscaffold end (Figure 14). In order to find these neighbouring superscaffolds, loci on the homologous region on the tomato genome were BLASTed to the potato genome, resulting in two homologous superscaffolds: PGSC0003DMB000000773 and PGSC0003DMB000000546. These superscaffolds were unanchored in the potato genome, and located to the right end of the superscaffold PGSC0003DMB000000385 that contains *StHSL1* (Figure 14).

Four out of five sequenced loci in the region neighbouring *StHSL1* showed strong associations ($p<1e^{-15}$) with TS (Figure 14). Only SNPs in the locus *CAST*, which is approximately 400 kb away from *StHSL1* showed a reduced association with TS, and a reduction of LD with the stronger associated SNPs (Figure 14). No reduction in association with TS was found on the other side of *StHSL1*, SNPs in the locus *AspAT* showed even the strongest associations with TS ($p=1.5e^{-23}$; Table S 1).

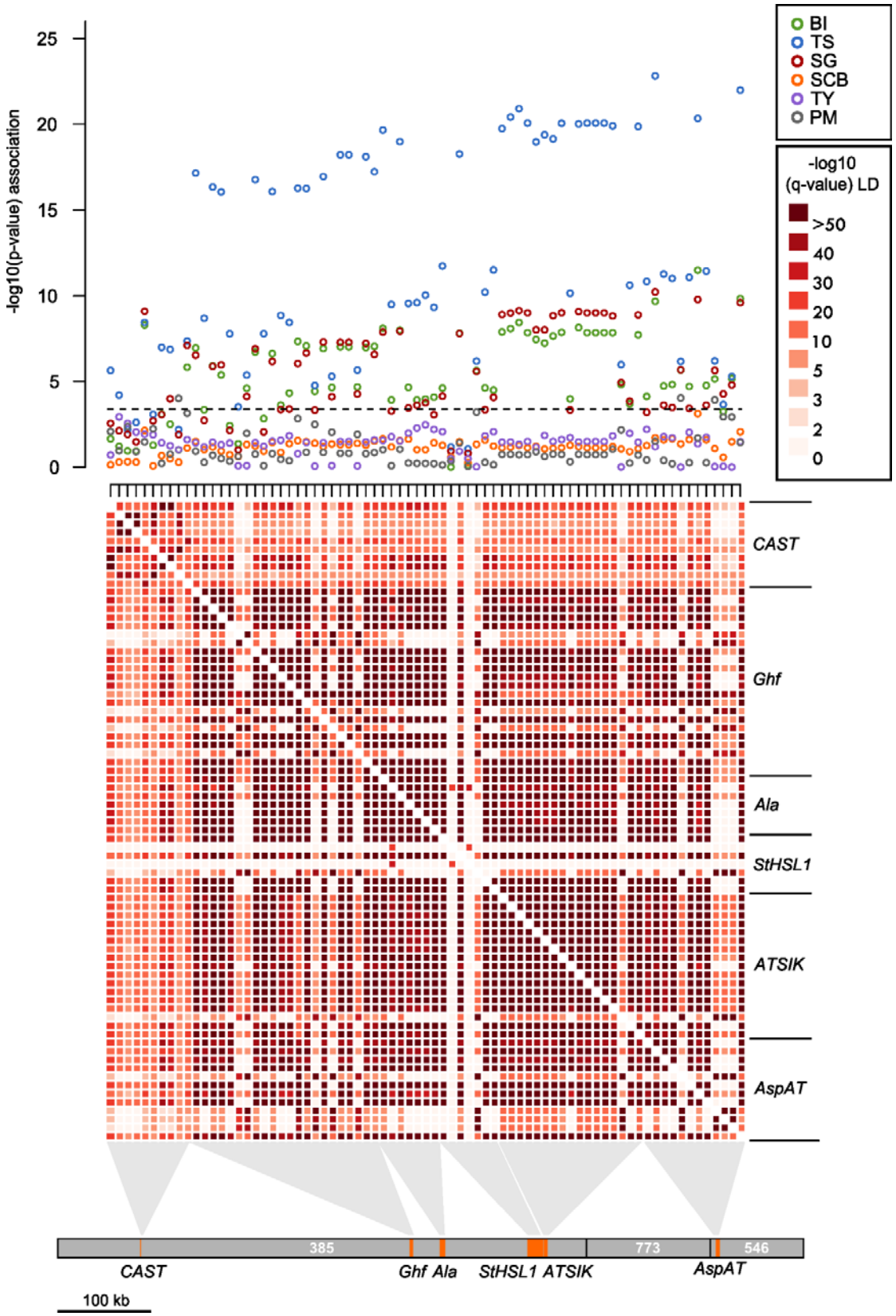


Figure 14. Combined LD heatmap and manhattan plot of the SNPs in the region neighbouring *StHSL1* on superscaffolds PGSC0003DMB000000385 and PGSC0003DMB000000546. Superscaffold numbers are written in white. The q-values are p-values corrected for multiple testing. LD p-values were calculated with a X^2 test of a 5x5 contingency table (see Materials and methods). The dashed line indicates the stringent significance threshold at p=3.5e-4 (see Materials and methods). Exact p-values and R^2 can be found in Table S 1.

6.6 LD analysis between associated loci

A selection was made of SNPs that were associated with one or more traits at $p<3.5e^{-4}$, and that located on different loci. When several highly significant SNPs were found on a locus, the SNP with strongest association with BI was selected. This resulted in the selection of 7 SNPs that spread over 6 chromosomes (Figure 15). P-values, R^2 and allele frequencies of selected SNPs presented in Figure 15 that were genotyped with pyrosequencing (in the loci *Zds*, *CuBP*, *ChoK* and *Gt*) can be found in Table 5. The same information about selected SNPs genotyped with Sanger sequencing is displayed in Table 7. Overall LD was weak between SNPs (Figure 15) except between the SNPs on chromosome 10 ($p<1e^{-40}$), between these two SNPs and solcap_snp_c2_44304 in *ChoK* ($p<1e^{-10}$) and between the two SNPs on chromosome 10 and solcap_snp_c1_55284 in *Pest1* ($p<1e^{-3}$). Of the SNPs that explained a high amount of variation of BI (>10%), only solcap_snp_c1_5286 in *Zds* was not in LD with the strongly associated SNPs on chromosome 10.

Table 7. Alleles, allele frequency, p-values, R^2 and polarity of highly associated SNPs that were genotyped with Sanger sequencing. All SNP trait associations with p<0.01 are displayed. Alleles are displayed with the allele with lowest frequency first. R^2 in % are in parentheses. The polarity of the effect of the allele is indicated by an arrow and is the effect of the allele with lowest allele frequency.

SNP ID	Locus	Alleles	Allele frequency	BI p (R^2)	SG p (R^2)	SCB[1] p (R^2)	TS p (R^2)
c2_55284[2]	*Pest1*	C/T	0.26	$7.0e^{-5}$ (11.2) ↑	$2.2e^{-5}$ (12.4) ↑	NS	NS
Mco_SNP1202	*Mco*	G/A	0.012	0.0025 (4.9) ↑	0.00022 (7.2) ↑	NS	NS
AspAT_SNP2921	*AspAT*	C/T	0.29	$3.3e^{-12}$ (28.0) ↑	$1.6e^{-10}$ (24.7) ↑	$7.5e^{-4}$ (10.0) ↑	$4.6e^{-21}$ (42.8) ↓

[1]NS: p>0.01
[2]SolCAP SNP ID after solcap_snp_..

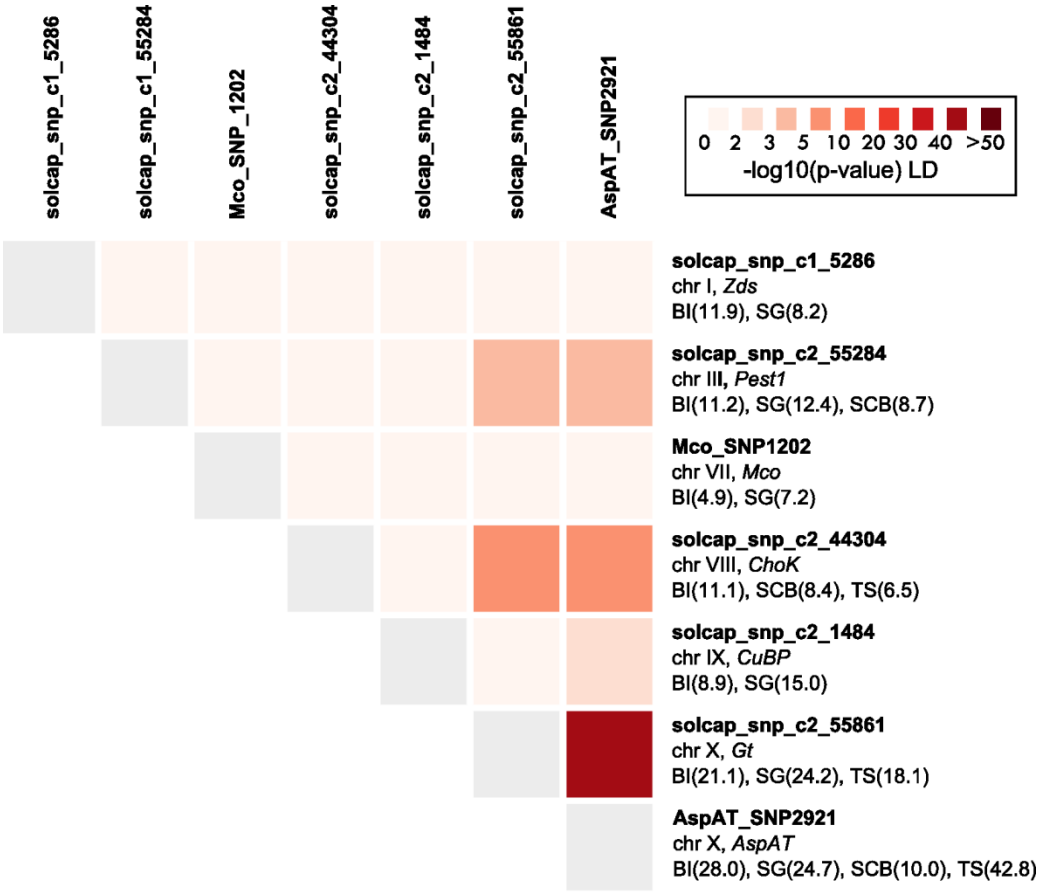


Figure 15. LD heatmap of highly associated SNPs ($p<3.5e^{-4}$ for one or more traits) in 7 different loci, spread over 6 chromosomes. On the vertical axis are displayed: SNP name, chromosome number, locus and associated traits ($p<0.01$) with R^2 in % in parentheses.

7. DISCUSSION

This chapter discusses the results described in the previous chapter and puts them in a larger framework with help of available literature. The strong correlations between agriculturally important traits in potato and its implications for the associations of SNP markers are discussed in the first subchapter. After that, the strengths and weaknesses of the BSA combined with the candidate gene approach are treated. The subchapter concludes with a suggestion for improvement for applying a BSA in order to find other useful SNP-trait associations. The location and genes that may be causative in the *Ro* locus for TS are discussed in the next subchapter. The findings in this study can have practical applications in potato breeding. These are discussed in subchapter 7.9. The chapter is finalized with the description of an additional hypothesis for the physiological background behind the strong correlation between SG and BI.

7.1 The correlation between BI, SG and TS

All traits measured in the ALL188 population were directly or indirectly correlated with each other, resulting in a single correlation network (Figure 2). The fact that the six measured traits are all economically important in potato stresses the need for association analysis in which the trait of interest is corrected for all other correlated traits in order to find marker trait associations solely for the trait of interest (Van Eck, 2007). This has been accomplished successfully for late blight resistance corrected for plant maturity (Visker et al., 2004; Pajerowska-Mukhtar et al., 2009) and partially for bruising susceptibility corrected for specific gravity (Urbany et al., 2011). However, the finding of markers in candidate genes associated with multiple traits may also provide interesting knowledge about the genetic background for the correlation between the traits, since associations of more than one trait with a certain candidate gene might give important clues for the gene's putative function.

In the partial correlation analysis, the correlation between PM and BI disappeared (Figure 2B) compared to the correlation analysis (Figure 2A). This is probably because PM and BI are correlated in the standard correlation analysis because of 'real' causative relations between SG and BI and SG and PM, meaning that BI and PM were correlated indirectly through SG. The correlation between SG and PM have a direct causative background: the potato plant probably needs time in order to accumulate a high amount of photosynthetic products in the form of starch into the tubers (Van Eck, 2007). Also the correlation between SG and BI seems to have a causative background, where SG causes a higher BI (see further in this chapter). This means that the partial correlation analysis was successful in eliminating correlations not established by direct relationships between traits.

In the ALL188 population BI is correlated with SG, SCB, TS and PM (Figure 2), and the cases and controls were significantly different for these traits (Figure 3). It was therefore expected that the candidate genes originating from the BSA would result in SNP trait associations with, next to BI, also with SG, SCB, TS and/or PM. This was indeed the case (Table 5; Figure 6; Figure 14; Table S 1). All SNPs strongly associated with BI were also associated with SG, and less frequent with other traits correlated with BI. This was expected, since the correlation between BI and SG is stronger than the correlation between BI and TS and BI and PM. Remarkably, only few SNPs were associated with SCB, although the cases and controls were very different for SCB. SCB is the most interesting trait in this context since it indicates bruising susceptibility independent of SG. The SNPs with strongest associations with SCB were in the genes *ChoK* ($p=0.0011$, $R^2=8.4$, Table 5) and *AspAT* (AspAT_SNP2921, $p=7.5e^{-4}$, $R^2=10.0$, Table S 1). The low amount of associations with SCB could be explained by the fact that SCB is a derived trait, and there might be only few genes affecting this trait specifically.

7.2 Combining BSA with the candidate gene approach

The fact that SNPs in the SolCAP Illumina array were found using transcriptome sequencing (Hamilton et al., 2011), directly allows allocation of a SNP from the array to an annotated locus on

the potato genome (The Potato Genome Sequencing Consortium, 2011). Because of this we were able to combine a BSA, in which potentially associated SNPs were found irrespective of their annotation, with a candidate gene approach, in which a more educated guess could be made about the potency of the SNP to be associated. The combination of methods greatly increases efficiency for finding SNP trait associations compared to using one of the methods separately. Using solely a candidate gene approach would lead to many possible candidate genes, since many genes and physiological processes are potentially related to bruising resistance, including genes in the phenylpropanoid pathway, starch metabolism, cell wall metabolism, membrane metabolism, antioxidant activity, stress and oxidoreduction of phenols (Figure 1). On the other hand, the BSA resulted in more than 600 potentially associated SNPs, too many to test all for association in a large population. Whether the method of preselecting SNPs using a BSA analysis would work was already partially confirmed before performing the association analysis in the ALL188 population. Four SNPs resulted from the BSA analysis that were in loci previously found to be associated with bruising index in Urbany et al (2011; Table 3).

Most selected SNPs from the BSA analysis that were potentially associated with BI (Table 4), were confirmed to be associated with BI in the ALL188 population. Eight out of eleven tested SNPs showed a p-value lower than 0.01 when tested for association in the ALL188 population (Figure 9). This is a much better score when compared to the approach were candidate genes where selected purely based on function (Urbany et al., 2011). When for these SNPs the Illumina outcomes were compared to the Sanger sequencing results in the 20 BSA genotypes, one of the SNPs that were not associated, solcap_snp_c2_45180 in the gene *Mco*, showed different allele dosages (Table S 2). This difference might be due to preferential pairing of certain alleles with PCR primers in the Sanger sequencing experiments, or with preferential binding of alleles in the Illumina array. The other two SNPs not significantly associated in the ALL188 population were in the genes *Ce* and *SKIP10* (Table 5). Both were genotyped with pyrosequencing, and showed generally the same allele dosages in the 20 BSA genotypes when comparing the Illumina array data with the pyrosequencing results (Table S 2). This shows that the significance of the t-test from the BSA was based on coincidence, and not on SNP scoring errors.

The two SNPs with lowest p-value in the BSA, which were in the loci *StHSL1* and *Gt*, showed strongest association with BI in the association analysis ($p=1.6e^{-8}$ and $p=9.4e^{-9}$ respectively; Figure 9). Higher p-values in the BSA did not seem to be an indication for the strength of association in the association analysis (Figure 9). The SNP in *Pest1* for example, with one of the highest p-value of the selected SNPs in the BSA analysis ($p=0.00414$; Table 4) showed strong associations in the association analysis ($p=0.00039$), and SNPs near the selected SolCAP SNP showed even stronger associations ($p=6.9e^{-5}$). *Pest1* was selected mainly for its gene function; variation in expression has profound effects on tuber rigidity in potato (Ross et al., 2011). This shows that it made sense to select not only for significance in the BSA analysis but also for putative gene function.

All SNPs found that were strongly associated with BI were also associated with SG (subchapter 7.1) and affected both traits in the same direction. Although this confirms the genetic correlation between BI and SG, this is also an artefact of the BSA. The cases and controls in the BSA were not only significantly very different for BI but also for SG (Figure 3). In order to be able to find SNPs more successfully that are only associated with BI, cases and controls could be selected based on physiological properties that are known to be causative for bruising susceptibility but do not affect SG. Examples of these are potential PPO activity , superoxide production after impact (Johnson and Doherty, 2003) or amount of polyphenol substrate available in the tuber, like tyrosine and chlorogenic acid, (Werij et al., 2007). When these processes are regulated by a small amount of genes, this strategy could also allow fine mapping of genes underlying the specific physiological processes.

7.3 A limitation of the statistical model

The statistical model used in this study does not account for population structure and kinship. A model which is better suited to find significant associations is described by Achenbach et al. (2009), which does correct for population structure and is based on effects of allele substitutions. Since it lacks correction for population structure, our model might have resulted in the finding of associations solely because a trait is more common in one interrelated group of genotypes than in another. If that is the case, the polymorphism found might simply be a polymorphism typical for the interrelated group and not for the trait of interest. However, the population structure in the ALL205 population is weak (Urbany et al., 2011), meaning that there are no distinctive interrelated groups present in the population. This suggests that correcting for population structure may have only a small effect on the strength of the associations found in this study

Adding population structure to a statistical model mainly has an effect on associations of traits which are typical for certain related groups of genotypes. This is for example the case for flowering time in Arabidopsis. Flowering time is typical for the geographic region from which a group of genotypes originates, so genotypes with the same flowering time are more likely to be closely related than genotypes with a different flowering time. Correcting for population structure in a GWAS for this trait strongly reduced the amount of significant associations, but it did not for traits that were not typical for closely related groups (Atwell et al., 2010). In addition, Atwell et al. (2010) also showed that correcting for population structure mainly reduced the associations at moderate p-values (around $p=1e^{-4}$), and the higher p-values remained. Suggesting that correcting for population structure in our data may reduce the amount of associations but the strong associations will remain.

7.4 Copper metabolism genes: no association with SCB

There were two genes involved in copper metabolism among the top 50 significant candidate genes resulting from the BSA analysis (Table 4). PPO and POD need copper as a co-factor (Friedman, 1997; Pourcel et al., 2007), and free copper ions can inhibit POD activity (Zancani et al., 1995). Therefore, it was proposed that these genes might have an effect on the activity of PPO and POD or that the genes might have a PPO/POD-like activity, suggesting that allelic variation in these genes would affect only bruising susceptibility and would have no effect on SG. However, SNPs in both genes were associated with both BI and SG, suggesting that allelic variation of these genes does not affect bruising susceptibility alone, as would be expected from their putative function in PPO and/or POD metabolism. Their association is probably caused by linkage to a gene affecting starch metabolism.

7.5 The location of the *Ro* locus on the potato genome sequence

In the ALL188 population SNPs were found that explained a large amount of variation in TS (up to 46.7%; Table S 1). These are in very close proximity to the marker CT217 (415 kb), which is linked to the *Ro* locus; a qualitative locus explaining large amounts of variation of TS (De Jong and Burns, 1993; Van Eck et al., 1994). Variation in TS is probably explained by a single gene carrying multiple alleles that are in the classes *Ro* (round) and *ro* (oval), where *Ro* is dominant over *ro*. The SNPs found associated strongly with TS are most probably part of the haplotype block containing the alleles explaining variation in TS by the *Ro* locus, making mapping of this locus to a physical location on the potato chromosome possible. One boarder of the haplotype block containing the locus is probably in the superscaffold PGSC0003DMB000000385. It seems that it starts between the *CAST* and the *Ghf* locus, since p-values drop drastically in the *CAST* locus compared to the *Ghf* locus (Figure 14). The loci *Ala*, *StHSL1*, *ATSIK*, and *AspAT* all contain SNPs explaining high amounts of variability in TS, showing that the haplotype block containing the *Ro* locus spreads over at least 420 kb, and may stretch further along the superscaffold PGSC0003DMB000000546. Initially, the location of PGSC0003DMB000000546 was unknown, it was allocated to the position the potato physical map based on homology with the tomato physical map. The fact that SNPs in *AspAT* are highly associated with TS and are in strong LD with loci on superscaffold

PGSC0003DMB000000385 is an argument that the unanchored superscaffold PGSC0003DMB000000546 is located on the correct position in Figure 14, or that is at least close to PGSC0003DMB000000385.

The genotyped SNP in the locus *Gt* is in LD with SNPs in the haplotype block that harbours the putative *Ro* locus (e.g. $p<1e^{-40}$ between solcap_snp_c2_55861 (*Gt*) and AspAT_SNP2921; Figure 15). The *Gt* locus is on the unanchored superscaffold PGSC0003DMB000000553 of the potato genome. This superscaffold probably locates around 51 Mb on the potato physical map, and the haplotype block containing the *Ro* locus around 42 Mb. This suggests that in this case strong LD is maintained over approximately 9 Mb. Such large regions in which LD is maintained have been reported in potato: a decline to $r^2=0.10$ over 5 cM (5.3 Mb) by D'hoop et al. (2010) and 10 cM (approximately 10.6 Mb (D'hoop et al., 2010)) by Simko et al. (2006). There might be two reasons for these large distances: (i) the low amount of meiotic events separating the genotypes (Gebhardt et al., 2004; Simko et al., 2006; D'hoop et al., 2010) and (ii) the strong selection for loci in this region (Mackay and Powell, 2007), since they are associated with agricultural important traits. It should be noted that the unanchored superscaffold on which *Gt* is located may not be placed in the correct chromosomal location and is in reality closer to the *Ro* locus than assumed.

7.6 Annotation and possible functions of *StHSL1*

The CDS surrounding the SNP that originated from the BSA and showed strongest associations with BI (solcap_snp_c1_8019, Table S 1) was not annotated in the PGSC genome browser. BLASTing and comparing the CDS resulted in finding the transcription factor *HSL1* from *A. thaliana* as closest homolog. The potato homolog has a similar intron-exon structure, and the specific domains (B3 binding domain, Zinc finger and EAR motif) are highly conserved between potato and *A. thaliana* (Figure 10; Figure 11). In *A. thaliana*, *HSL1* has functions in blocking seed maturation and thereby initiating the transition from seed maturation to vegetative growth (Tsukagoshi et al., 2007). This transition is among others characterized by the change from sink to source of storage organs. Starch is the major storage compound in seeds and tubers. In potato, *StHSL1* may therefore play a major role regarding starch content in the tubers. Since starch content and bruising susceptibility are highly correlated, this might explain the association with bruising susceptibility. It should however be noted that despite of their similarities there are also a lot of differences between *HSL1* from *A. thaliana* and *StHSL1* (Figure 10; Figure 11; Figure 12). The similarities suggest that *StHSL1* is very likely to be a transcription factor but they might have function in different physiological processes.

Remarkably, SNPs in *StHSL1* and its surrounding region were most strongly correlated with TS. It could be that *StHSL1* regulates starch content, and SNPs in the gene also explain variation in TS just because of physical linkage with one or more gene(s) regulating TS. It could also be that the gene(s) regulating TS also affect SG and thereby bruising susceptibility. This would imply that the correlation between TS and bruising susceptibility should not be explained just by the notion that round tubers are more exposed to mechanical impact than oblong tubers (Urbany et al., 2011), but that genes affecting TS also affect SG and thereby bruising susceptibility. This is supported by the fact that TS is correlated with both SG and bruising susceptibility (Figure 2).

7.7 Candidates genes for the *Ro* locus

Next to *StHSL1*, there are also other genes that could be causative for TS in the described haplotype block based on their putative function. *Ghf*, one of the loci containing SNPs strongly associated with TS, is a glycosyl hydrolase. These enzymes can modify cell wall polysaccharides (Lee et al., 2007) and therefore may affect cell shape by altering the shape of the cell wall. It should be noted that the gene is not expressed in tubers according to the PGSC genome browser, but it might only be expressed during tuber growth, and could therefore have not been detected during RNA extraction for the RNAseq library of the PGSC sequence. Another gene on superscaffold PGSC0003DMB000000385 that might be causative for variation in TS is Auxin:hydrogen symporter

(*Aec*; Locus ID: PGSC0003DMG400027696). This gene is expressed in all the tissue from which data is available in the PGSC genome browser, including the tubers. Allelic variation in *Aec* might affect auxin transport. Auxin is involved in cell expansion (Taiz and Zeiger, 2010) and cell division (Petrásek et al., 2002), suggesting that altered auxin transport might have profound effects on TS.

Strongest associations with TS were found in the locus *AspAT*, suggesting that the functional gene(s) affecting TS are more likely to be found on superscaffold PGSC0003DMB000000773 or PGSC0003DMB000000546. Remarkably, there are 10 loci annotated as 'non-specific lipid transfer protein' (*nsLTP*) spread over the three superscaffolds depicted in Figure 14, and 6 of them are located on PGSC0003DMB000000773. These genes might be involved in variation in TS, since they can affect cell wall extension (Nieuwland et al., 2005). It should be noted that only few of the *nsLTP* genes are expressed in tubers. Other annotations that are present frequently in the region are different types of peroxidases (POD), of which 6 are located on PGSC0003DMB000000546. All of them are expressed in tubers. The presence of these POD in or near to the haplotype block might explain the correlation between TS and BI, since POD can convert o-diphenols into o-quinones (these make up the greatest part of melanins; see chapter 2). Other functions of POD (cell wall peroxidases) lay in modification of the cell walls (Passardi et al., 2004), thereby altering cell shape, suggesting involvement in variation in TS.

Although it is not mapped in close proximity to the haplotype block containing the *Ro* locus, *Gt* might also be causative for TS. Glycosyltransferases (GT) form a group of enzymes that play a central role in the synthesis of cell wall polysaccharides (Keegstra and Raikhel, 2001). In plants there are many members of the GT family, *A. thaliana* has hundreds of genes that are putative GT genes (Scheible and Pauly, 2004). Cell wall polysaccharides are very complex structures, and each GT has a specific role in their synthesis. Based on the annotation of the gene in which we found a SNP highly significantly associated with SG, BI and TS no conclusions can be drawn on its specific role in polysaccharide synthesis. However, because many GT are involved in synthesis in cell wall polysaccharides, it could be that our candidate gene for the *Ro* locus is involved in this process. Many *Gt* mutants have an altered cell wall rigidity (Scheible and Pauly, 2004). Allelic variation in the gene may therefore affect cell wall rigidity and cell shape.

7.8 LD patterns within and between loci

The locus *Mco* showed a more interrupted LD pattern than *Pest1* and the SNPs in the putative *Ro* locus. This coincided with less strong associations of the *Mco* locus with the measured agriculturally important traits than in *Pest1* and putative *Ro* locus. The less interrupted LD pattern in *Pest1* and the putative *Ro* locus is an indication for the presence of a smaller amount of different haplotype alleles than in *Mco*. Since *Pest1* and the putative *Ro* locus are strongly associated with agriculturally important traits, the occurrence of large uninterrupted LD blocks might be caused by selection for a few alleles that have a large effect on variation of the agriculturally important traits. The interrupted LD pattern in *Mco* might than be caused by absence of selection in this region allowing presence of a high number of haplotype alleles in the ALL188 population in this locus.

Figure 15 shows amongst others that LD was found between SNPs that were on different chromosomes: SNPs on chromosome 10 (AspAT_SNP2921 and solcap_snp_c2_55861) were in significant LD with SNPs on chromosome 3 (solcap_snp_c2_55284) and 8 (solcap_snp_c2_44304). This LD cannot be caused by a small number of crossing over events between the loci in the population's ancestry, since crossing-over between loci on different chromosomes should be independent of each other. The significant LD values between SNPs on different chromosomes in Figure 15 were in SNPs that were strongly associated with the same traits. It could be just by chance that they are in LD if they have comparable allele frequencies; in that case, genotypes with a certain trait value would have a high chance to have the same allele dosage of two SNPs in two independent loci that are strongly associated with that same trait. The SNPs in LD indeed have a very comparable allele dosage; the four SNPs on the different chromosomes have allele frequencies between 0.23 and 0.29 (Table 5; Table 7). It should be noted that the test for LD may have been

rather sensitive, an estimation of the r^2 (a different measure for indicating LD) might give results better comparable to results described in other publications. However, the very small p-values found in this study are still an unquestionable indication of LD between loci.

7.9 Practical implications: what's in it for the breeder?

The SNP markers provided by this study are strongly associated with several agriculturally important traits. SNPs with most significantly associated with BI are also associated with SG; these SNPs do therefore not allow selection against bruising susceptibility independent of SG. On the other hand, all SNP markers strongly associated with SG and BI did not show association with PM. The correlation between SG and PM is considered as a problem while selecting for SG (Van Eck, 2007). The markers found in this study allow selection for SG independent of PM, however, the increase in bruising susceptibility while selecting for SG should be taken for granted. Two different loci on chromosome 10 associated strongly with BI and SG independent of PM: in the putative *Ro* locus (a.o. AspAT_SNP2831, with $p=5.9e^{-11}$, $R^2=25.8\%$ for SG and AspAT_SNP2921 with $p=3.3e^{-12}$, $R^2=28\%$ for BI) and in *Gt* (solcap_snp_c2_55861, with $p=2.9E^{-10}$, $R^2=24.2\%$ for SG). Unfortunately, these SNPs are in strong LD with each other ($p<1e^{-40}$), suggesting that selection for both SNP markers will not result in a higher SG or lower BI of the selected material than selecting for only one of these SNPs. SNPs that may work additive to these SNPs are in *Zds* (solcap_snp_c1_5286), or in loci explaining smaller amounts of variation: *Mco* (Mco_SNP1202) and *CuBP* (solcap_snp_c2_55861). These SNPs are not in LD with each other nor with the highly associated SNPs on chromosome 10 (Figure 15). Therefore, selection for a combination of these SNP markers might allow efficient marker assisted selection for SG independent of plant maturity. It should be noted that validation of additive effects of SNPs should be carried out with stepwise regression.

Two SNPs were moderately associated with SCB (subchapter 7.1). The trait SCB is the most interesting in terms of breeding for bruising resistance, since it allows selection for bruising resistance independent of SG. However, their association should be confirmed with a model correcting for population structure, since their p-values are on the low side, and they explain a rather small percentage of the variation (10.0 and 8.4% respectively). In addition, the SNPs were in significant LD (Figure 15), so they might not work additive to each other (although this LD might be caused by chance; subchapter 7.8). SNPs associated with TS showed very strong associations. These SNPs explained up to 46.7% of the variation in TS (Table S 1). However, the need for molecular markers for TS is questionable, since the trait can be assessed as soon as the first potatoes are harvested and the measurement is non-destructive.

7.10 The correlation between specific gravity and bruising susceptibility: an additional hypothesis

It seems more plausible that a high SG is causative for a high BI than the other way around, since it does not seem likely that a high potential of synthesis of melanins would affect SG. Additionally, the partial correlation analysis shows that PM and BI are not directly correlated, suggesting that a high BI is caused by the high SG, which is typical for late maturing genotypes. The hypothesis that a higher starch granule load causes easier rupture of cell membranes (Urbany et al., 2011) fits in this suggestion. However, there can be more lines of thought how increased starch content, and therefore SG, could result in increased bruising susceptibility. An additional hypothesis could be that there is a competition between tuber starch content and cell wall rigidity. This is underpinned by the finding of strong associations of both BI and SG with SNPs in genes probably involved in cell wall rigidity: *Pest1* and *Gt*. SG is a measure for the percentage of dry matter in the tubers. If, due to allelic variation in genes involved cell wall rigidity, more sugars would be incorporated in the cell walls of the tuber cells than into starch stored in the tubers, it would not have a large effect on SG. However, when it is assumed that allelic variation in *Pest1* and *Gt* not only affects cell wall rigidity in tuber cells but also in green plant parts it could have a large effect on the competitive strength between tuber starch and cell walls, since in green plant parts the cell walls make up 45-65% of

the total dry weight (Robbins and Moen, 1975). Cell walls in green plant parts will therefore require a substantial amount of the total available sugars. When the plant incorporates more carbohydrates into all its cell walls, including the green parts, less will be left to be transported to the tubers and specific gravity may decrease.

8. CONCLUSIONS

The use of a combination of a BSA and a candidate gene approach has proved to be very successful in order to find strong associations with bruising susceptibility. It may therefore be a convenient method to use in order to map other traits in potato or to apply it in other crops. The method led to identification of diagnostic molecular markers that explain a large part of the variation (>20%) in bruising susceptibility, specific gravity independent of plant maturity and tuber shape. These molecular markers can be used in a breeding program to enable marker assisted selection that can greatly enhance efficiency and success of selection. More efficient and successful selection will improve tuber quality, thereby reducing waste throughout the potato supply chain.

This thesis stresses the strong correlation between specific gravity and bruising susceptibility, whereby SNP markers located in several locations across the potato genome are associated with both traits. In order to find markers that can be used to reduce bruising susceptibility in potato cultivars with high specific gravity, a better understanding about the physiological background and the relation between the two traits is indispensible.

The more precise location of the *Ro* locus provided in this thesis contributes to the finding of the functional gene or genes explaining large amounts of variation in tuber shape. When the nature and function of the gene or genes is unravelled it will add strongly to our fundamental knowledge on how plant organs are shaped.

9. ACKNOWLEDGEMENTS

This project was performed at the Max Planck institute for Plant Breeding Research in Köln at the Department for plant Breeding and Genetics headed by Maarten Koornneef. Special thanks go to Dr. Christiane Gebhardt for providing the opportunity for doing my MSc thesis at her workgroup, her guidance and interesting discussions. Further thanks go to Dr. Yuling Bai and Prof. Dr. Richard Visser for supervision from Wageningen University. Claude Urbany has provided the phenotypic data and DNA of the ALL188 population. He also helped me with important clues for which ways to go in this study, so thanks go out to him for both. Benjamin Stich and Jinquan Li have helped me with some R programming issues. Everyone from the Gebhardt group has helped greatly in achieving the goals reached in this thesis. Of this group special thanks go to Elske Maria Schönhals, who was always kind to respond to my numerous questions and introduced me with many techniques and procedures, Birgit Walkemeier, who helped greatly with many lab issues, Astrid Draffehn, who introduced me with pyrosequencing, Meki Shehabu Muktar, who has been a great collegue and answered many of my questions and Markus Kuckenberg for helping me in the lab.

10. REFERENCES

Achenbach U, Paulo J, Ilarionova E, Lübeck J, Strahwald J, Tacke E, Hofferbert H-R, Gebhardt C (2009) Using SNP markers to dissect linkage disequilibrium at a major quantitative trait locus for resistance to the potato cyst nematode Globodera pallida on potato chromosome V. Theoretical and Applied Genetics **118**: 619-629

Atwell S, Huang YS, Vilhjálmsson BJ, Willems G, Horton M, Li Y, Meng D, Platt A, Tarone AM, Hu TT, et al (2010) Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature **465**: 627-631

Bachem C, Speckmann G, van der Linde P, Verheggen F, Hunt MD, Steffens JC, Zabeau M (1994) Antisense expression of polyphenol oxidase genes inhibits enzymatic browning in potato tubers. Nature Biotechnology **12**: 1101-1105

Bradshaw JE, Hackett C a, Pande B, Waugh R, Bryan GJ (2008) QTL mapping of yield, agronomic and quality traits in tetraploid potato (Solanum tuberosum subsp. tuberosum). Theoretical and Applied Genetics **116**: 193-211

Coetzer C, Corsini D, Love S, Pavek J, Tumer N (2001) Control of enzymatic browning in potato (Solanum tuberosum L.) by sense and antisense RNA from tomato polyphenol oxidase. Journal of Agricultural and Food Chemistry **49**: 652-657

D'hoop BB, Paulo MJ, Kowitwanich K, Sengers M, Visser RGF, van Eck HJ, van Eeuwijk FA (2010) Population structure and linkage disequilibrium unravelled in tetraploid potato. Theoretical and Applied Genetics **121**: 1151-1170

D'hoop BB, Paulo MJ, Mank RA, Eck HJ, Eeuwijk FA (2007) Association mapping of quality traits in potato (Solanum tuberosum L.). Euphytica **161**: 47-60

Van Eck H, Jacobs J, Stam P, Ton J, Stiekema WJ, Jacobsen E (1994) Multiple alleles for tuber shape in diploid potato detected by qualitative and quantitative genetic analysis using RFLPs. Genetics **137**: 303-309

Van Eck HJ (2007) Genetics of Morphological and Tuber Traits. *In* D Vreugdenhil, J Bradshaw, C Gebhardt, F Govers, DKL Mackerron, MA Taylor, HA Ross, eds, Potato Biology and Biotechnology. Elsevier B.V., pp 91-115

Flurkey WH (2003) Laccase. *In* JR Whitaker, AG Voragen, DWS Wong, eds, Handbook of food enzymology. Marcel Dekker Inc., New York, NY, USA, pp 525-539

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

Gebhardt C (2007) Molecular Markers , Maps and Population Genetics. *In* D Vreugdenhil, J Bradshaw, C Gebhardt, F Govers, DKL Mackerron, MA Taylor, HA Ross, eds, Potato Biology and Biotechnology. Elsevier B.V., pp 77-89

Gebhardt C, Ballvora A, Walkemeier B, Oberhagemann P, Schüler K (2004) Assessing genetic potential in germplasm collections of crop plants by marker-trait association: a case study for potatoes with quantitative variation of resistance to late blight and maturity type. Molecular Breeding **13**: 93–102

Gebhardt C, Li L, Pajerowska-Mukthar K, Achenbach U, Sattarzadeh A, Bormann C, Ilarionova E, Ballvora A (2007) Candidate Gene Approach to Identify Genes Underlying Quantitative Traits and Develop Diagnostic Markers in Potato. Crop Science **47**: S-106

Gustavsson J, Cederberg C, Sonesson U, van Otterdijk R, Meybeck A (2011) Global food losses and food waste, Report of the Food and Agriculture Organization of the United Nations. International congress: Save Food!, Düsseldorf, Germany

Hamberger B, Hahlbrock K (2004) The 4-coumarate:CoA ligase gene family in Arabidopsis thaliana comprises one rare, sinapate-activating and three commonly occurring isoenzymes. Proceedings of the National Academy of Sciences of the United States of America **101**: 2209-2214

Hamilton JP, Hansey CN, Whitty BR, Stoffel K, Massa AN, Van Deynze A, De Jong WS, Douches DS, Buell CR (2011) Single nucleotide polymorphism discovery in elite North American potato germplasm. BMC Genomics **12**: 302

Henze RE (1956) Inhibition of enzymatic browning of chlorogenic acid solutions with cysteine and glutathione. Science **123**: 1174-1175

Hunt MD, Eannetta NT, Yu H, Newman SM, Steffens JC (1993) cDNA cloning and expression of potato polyphenol oxidase. Plant Molecular Biology **21**: 59-68

Johnson S, Doherty S (2003) Biphasic superoxide generation in potato tubers. A self-amplifying response to stress. Plant Physiology **131**: 1440-1449

De Jong H, Burns V (1993) Inheritance of tuber shape in cultivated diploid potatoes. American Journal of Potato Research **70**: 267-284

Kasai K, Morikawa Y, Sorri V a, Valkonen JP, Gebhardt C, Watanabe KN (2000) Development of SCAR markers to the PVY resistance gene Ryadg based on a common feature of plant disease resistance genes. Genome **43**: 1-8

Keegstra K, Raikhel N (2001) Plant glycosyltransferases. Current Opinion in Pant Biology **4**: 219-224

Lee E-J, Matsumura Y, Soga K, Hoson T, Koizumi N (2007) Glycosyl hydrolases of cell wall are induced by sugar starvation in Arabidopsis. Plant & Cell Physiology **48**: 405-413

Leonards-Schippers C, Gieffers W, Schäfer-Pregl R, Ritter E, Knapp SJ, Salamini F, Gebhardt C (1994) Quantitative resistance to Phytophthora infestans in potato: a case study for QTL mapping in an allogamous plant species. Genetics **77**: 67-77

Li L, Strahwald J, Hofferbert H-R, Lübeck J, Tacke E, Junghans H, Wunder J, 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**: 813-821

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

Nieuwland J, Feron R, Huisman B (2005) Lipid transfer proteins enhance cell wall extension in tobacco. The Plant Cell **17**: 2009-2019

Ortega F, Lopez-Vizcon C (2012) Application of Molecular Marker-Assisted Selection (MAS) for Disease Resistance in a Practical Potato Breeding Programme. Potato Research **55**: 1-13

Paal J, Henselewski H, Muth J, Meksem K, Menéndez CM, Salamini F, Ballvora A, Gebhardt C (2004) Molecular cloning of the potato Gro1-4 gene conferring resistance to pathotype Ro1 of the root cyst nematode Globodera rostochiensis, based on a candidate gene approach. The Plant Journal **38**: 285-297

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**: 1115-1127

Passardi F, Penel C, Dunand C (2004) Performing the paradoxical: how plant peroxidases modify the cell wall. Trends in Plant Science **9**: 534-540

Petrásek J, Elckner M, Morris D a, Zazímalová E (2002) Auxin efflux carrier activity and auxin accumulation regulate cell division and polarity in tobacco cells. Planta **216**: 302-308

Pflieger S, Lefebvre V, Causse M (2001) The candidate gene approach in plant genetics: a review. Molecular Breeding **7**: 275–291

Pourcel L, Routaboul J-M, Cheynier V, Lepiniec L, Debeaujon I (2007) Flavonoid oxidation in plants: from biochemical properties to physiological functions. Trends in Plant Science **12**: 29-36

Ramirez EC, Whitaker JR, Virador VM (2003) Polyphenol oxidase. *In* JR Whitaker, AG Voragen, DWS Wong, eds, Handbook of food enzymology. Marcel Dekker Inc., New York, NY, USA, pp 509-525

Robbins C, Moen A (1975) Composition and digestibility of several deciduous browses in the northeast. The Journal of Wildlife Management **39**: 337-341

Rommens CM, Ye J, Richael C, Swords K (2006) Improving potato storage and processing characteristics through all-native DNA transformation. Journal of Agricultural and Food Chemistry **54**: 9882-9887

Ross HA, Morris WL, Ducreux LJM, Hancock RD, Verrall SR, Morris J a, Tucker GA, Stewart D, Hedley PE, McDougall GJ, et al (2011) Pectin engineering to modify product quality in potato. Plant Biotechnology Journal **9**: 848-856

Sattarzadeh A, Achenbach U, Lübeck J, Strahwald J, Tacke E, Hofferbert H-R, Rothsteyn T, Gebhardt C (2006) Single nucleotide polymorphism (SNP) genotyping as basis for developing a PCR-based marker highly diagnostic for potato varieties with high resistance to Globodera pallida pathotype Pa2/3. Molecular Breeding **18**: 301-312

Scheible W-R, Pauly M (2004) Glycosyltransferases and cell wall biosynthesis: novel players and insights. Current Opinion in Plant Biology **7**: 285-295

Simko I, Costanzo S, Haynes KG, Christ BJ, Jones RW (2004) Linkage disequilibrium mapping of a Verticillium dahliae resistance quantitative trait locus in tetraploid potato (Solanum tuberosum) through a candidate gene approach. Theoretical and Applied Genetics **108**: 217-224

Simko I, Haynes KG, Jones RW (2006) Assessment of linkage disequilibrium in potato genome with single nucleotide polymorphism markers. Genetics **173**: 2237-2245

Song Y-S, Hepting L, Schweizer G, Hartl L, Wenzel G, Schwarzfischer A (2005) Mapping of extreme resistance to PVY (Ry (sto)) on chromosome XII using anther-culture-derived primary dihaploid potato lines. Theoretical and Applied Genetics **111**: 879-887

Storey J (2003) The positive false discovery rate: A Bayesian interpretation and the q-value. Annals of Statistics **31**: 2013-2035

Storey M (2007) The Harvested Crop. *In* D Vreugdenhil, J Bradshaw, C Gebhardt, F Govers, DKL Mackerron, MA Taylor, HA Ross, eds, Potato Biology and Biotechnology. Elsevier Ltd., Oxford, UK, pp 441-470

Taiz L, Zeiger E (2010) Plant Physiology, 5th ed. Sinauer associates Inc., Massachusetts U.S.A.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Bology and Evolution **28**: 2731-2739

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

The Tomato Genome Consortium (2012) The tomato genome sequence provides insights into fleshy fruit evolution. Nature **485**: 635-641

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research **22**: 4673-4680

Tsukagoshi H, Morikami A, Nakamura K (2007) Two B3 domain transcriptional repressors prevent sugar-inducible expression of seed maturation genes in Arabidopsis seedlings. Proceedings of the National Academy of Sciences of the United States of America **104**: 2543-2547

Urbany C, Colby T, Stich B, Schmidt L, Schmidt J, Gebhardt C (2012) Analysis of Natural Variation of the Potato Tuber Proteome Reveals Novel Candidate Genes for Tuber Bruising. Journal of Proteome Research **11**: 703-716

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

Valverde P, Manning P, McNEIL CJ, Thody AJ (1996) Activation of Tyrosinase Reduces the Cytotoxic Effects of the Superoxide Anion in B16 Mouse Melanoma Cells. Pigment Cell Research **9**: 77-81

Visker M, Raaij HV, Keizer L, Struik P (2004) Correlation between late blight resistance and foliage maturity type in potato. Euphytica **137**: 311-323

Werij JS, Kloosterman B, Celis-Gamboa C, de Vos CHR, America T, Visser RGF, Bachem CWB (2007) Unravelling enzymatic discoloration in potato through a combined approach of candidate genes, QTL, and expression analysis. Theoretical and Applied Genetics **115**: 245-252

Xu Y, Lu Y, Xie C, Gao S, Wan J, Prasanna BM (2012) Whole-genome strategies for marker-assisted plant breeding. Molecular Breeding 833-854

Zancani M, Nagy G, Vianello A (1995) Copper-inhibited NADH-dependent peroxidase activity of purified soya bean plasma membranes. Phytochemistry **40**: 367-371

Zhang Y (2009) Genetic Analysis Of Potato Tuber Flesh Pigmentation And Shape (Abstract Thesis). Cornell University, Ithaca, NY, USA

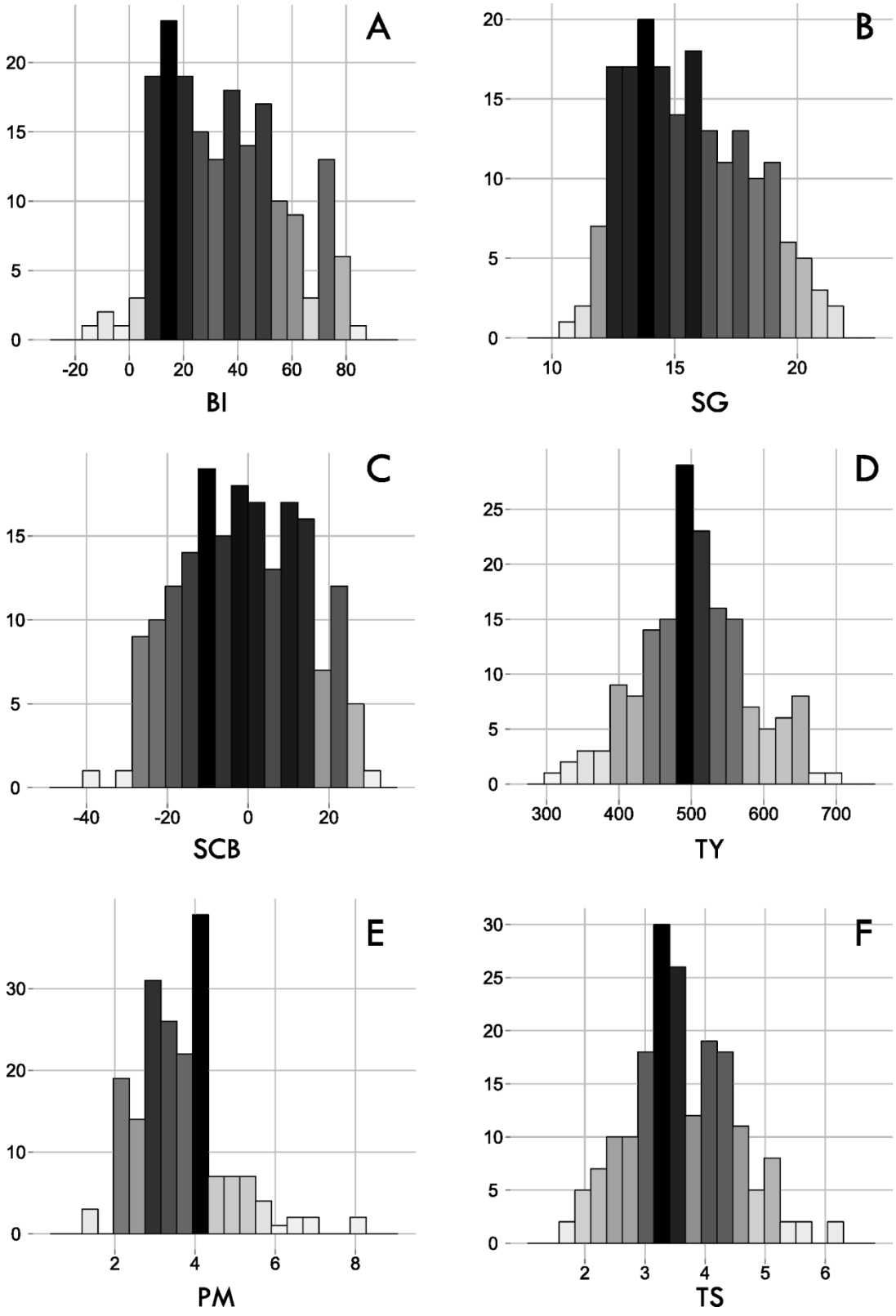


Figure S 1. Histograms of adjusted entry means. The numbers on the y-axis indicate the frequency. The white to black gradient of the bars indicates the counts per bar, where white represents low numbers and black high numbers. A represents the frequency distribution of bruising index (BI), B specific gravity (SG), C starch corrected bruising (SCB), D tuber yield (TY), E plant maturity (PM), and F tubershape (TS).

Table S 1. Position, p-value and R^2 of all SNPs at p<0.01. Per SNP, the p-value is displayed in the first row, the R^2 (%) in the second row in italics.

Gene	Position[1]	SNP ID	BI	SG	SCB	TS	PM	TY
Gt	2012	solcap_snp_c2_55861	9.4E-09	2.9E-10	NS	2.5E-07	NS	NS
			21.1	*24.2*	*NS*	*18.1*	*NS*	*NS*
Zds	7608	solcap_snp_c1_5287	0.00012	0.00348	NS	NS	NS	NS
			11.9	*8.2*	*NS*	*NS*	*NS*	*NS*
Zds	7611	solcap_snp_c1_5286	0.00012	0.00348	NS	NS	NS	NS
			11.9	*8.2*	*NS*	*NS*	*NS*	*NS*
BIPx	1730	solcap_snp_c2_46603	0.00348	1.4E-06	NS	NS	0.00122	NS
			4.5	*11.9*	*NS*	*NS*	*5.5*	*NS*
Pest1	2976	Pest1_SNP2976	0.00346	0.00125	NS	NS	NS	NS
			7.2	*8.2*	*NS*	*NS*	*NS*	*NS*
Pest1	2979	Pest1_SNP2979	0.00323	0.00113	NS	NS	NS	NS
			7.2	*8.4*	*NS*	*NS*	*NS*	*NS*
Pest1	3036	solcap_snp_c2_55282	0.00091	9.2E-05	0.00095	NS	NS	NS
			8.6	*11.0*	*8.5*	*NS*	*NS*	*NS*
Pest1	3237	Pest1_SNP3237	0.0064	0.00512	NS	NS	NS	NS
			6.5	*6.7*	*NS*	*NS*	*NS*	*NS*
Pest1	3241	Pest1_SNP3241	NS	0.00968	NS	NS	NS	NS
			NS	*6.0*	*NS*	*NS*	*NS*	*NS*
Pest1	3242	Pest1_SNP3242	NS	0.00336	NS	NS	NS	NS
			NS	*7.2*	*NS*	*NS*	*NS*	*NS*
Pest1	3249	Pest1_SNP3249	NS	0.00545	NS	NS	NS	NS
			NS	*6.7*	*NS*	*NS*	*NS*	*NS*
Pest1	3261	solcap_snp_c2_55284	7E-05	2.2E-05	0.00081	NS	NS	NS
			11.2	*12.4*	*8.7*	*NS*	*NS*	*NS*
Pest1	3276	Pest1_SNP3276	NS	0.00408	NS	NS	NS	NS
			NS	*7.0*	*NS*	*NS*	*NS*	*NS*
Pest1	3378	solcap_snp_c2_55285	0.00039	9.7E-05	NS	NS	0.00706	NS
			9.5	*10.9*	*NS*	*NS*	*6.4*	*NS*
Pest1	3426	Pest1_SNP3426	NS	0.00307	NS	NS	0.00638	NS
			NS	*6.1*	*NS*	*NS*	*5.4*	*NS*
Pest1	3429	Pest1_SNP3429	NS	0.0029	NS	NS	0.00112	NS
			NS	*6.2*	*NS*	*NS*	*7.2*	*NS*
Pest1	3438	solcap_snp_c2_55286	0.00022	1.8E-05	0.00686	NS	NS	NS
			10.1	*12.6*	*6.4*	*NS*	*NS*	*NS*
Pest1	3477	Pest1_SNP3477	NS	0.00309	NS	NS	0.0005	NS
			NS	*6.1*	*NS*	*NS*	*8.0*	*NS*
Pest1	3531	Pest1_SNP3531	NS	NS	NS	NS	0.00822	NS
			NS	*NS*	*NS*	*NS*	*5.1*	*NS*
CuBP	5548	solcap_snp_c2_1484	0.00067	1.5E-06	NS	NS	NS	NS
			8.9	*15.0*	*NS*	*NS*	*NS*	*NS*
Mco	1165	Mco_SNP1165	NS	NS	NS	NS	0.00407	NS
			NS	*NS*	*NS*	*NS*	*5.9*	*NS*
Mco	1166	Mco_SNP1166	0.00919	NS	NS	NS	NS	NS
			6.2	*NS*	*NS*	*NS*	*NS*	*NS*
Mco	1185	Mco_SNP1185	NS	0.00221	NS	NS	NS	0.00233
			NS	*5.0*	*NS*	*NS*	*NS*	*5.6*
Mco	1202	Mco_SNP1202	0.00248	0.00022	NS	NS	NS	NS
			4.9	*7.2*	*NS*	*NS*	*NS*	*NS*
Mco	1360	Mco_SNP1360	NS	0.00221	NS	NS	NS	0.00233
			NS	*5.0*	*NS*	*NS*	*NS*	*5.6*

ChoK	3011	solcap_snp_c2_44304	7.9E-05	NS	0.0011	0.00661	NS	NS
			11.1	*NS*	*8.4*	*6.5*	*NS*	*NS*
StHSL1	3637	solcap_snp_c1_8020	2.5E-05	7.3E-05	NS	1.9E-12	NS	0.00837
			12.4	*11.3*	*NS*	*27.2*	*NS*	*7.0*
StHSL1	3754	solcap_snp_c1_8019	1.6E-08	1.6E-08	NS	5.3E-19	NS	NS
			20.8	*20.8*	*NS*	*39.8*	*NS*	*NS*
StHSL1	3809	solcap_snp_c1_8018	2.3E-06	2.5E-06	NS	6.4E-07	0.00061	NS
			13.3	*13.2*	*NS*	*14.5*	*7.8*	*NS*
StHSL1	3996	StHSL1_SNP3996	2.4E-05	0.00044	NS	6.1E-11	NS	NS
			12.5	*9.5*	*NS*	*24.4*	*NS*	*NS*
StHSL1	4002	StHSL1_SNP4002	3.2E-05	8.5E-05	NS	3.1E-12	NS	0.00837
			12.2	*11.2*	*NS*	*26.9*	*NS*	*7.0*
Vps35	11477	Vps35_SNP11477	NS	NS	0.00494	NS	NS	NS
			NS	*NS*	*6.9*	*NS*	*NS*	*NS*
Vps35	11638	solcap_snp_c2_7902	0.00367	NS	0.00398	NS	NS	NS
			8.2	*NS*	*8.1*	*NS*	*NS*	*NS*
ATSIK	1117	solcap_snp_c2_25541	1.9E-08	1.3E-09	NS	1.4E-20	NS	NS
			22.0	*24.5*	*NS*	*44.7*	*NS*	*NS*
ATSIK	1139	solcap_snp_c2_25542	0.0002	0.00014	NS	2.4E-11	NS	NS
			10.9	*11.3*	*NS*	*26.6*	*NS*	*NS*
ATSIK	1150	ATSIK_SNP1150	1.6E-05	1.2E-05	NS	1E-06	0.00658	NS
			12.1	*12.5*	*NS*	*14.9*	*5.7*	*NS*
ATSIK	1157	ATSIK_SNP1157	1.5E-08	1.5E-09	NS	1.3E-20	NS	NS
			22.0	*24.2*	*NS*	*44.3*	*NS*	*NS*
ATSIK	1204	ATSIK_SNP1204	1.4E-08	9.9E-10	NS	8.6E-21	NS	NS
			22.0	*24.6*	*NS*	*44.6*	*NS*	*NS*
ATSIK	1262	ATSIK_SNP1262	1.4E-08	9.9E-10	NS	8.6E-21	NS	NS
			22.0	*24.6*	*NS*	*44.6*	*NS*	*NS*
ATSIK	1274	ATSIK_SNP1274	1.4E-08	9.9E-10	NS	8.6E-21	NS	NS
			22.0	*24.6*	*NS*	*44.6*	*NS*	*NS*
ATSIK	1294	ATSIK_SNP1294	7.2E-09	8.4E-10	NS	9.6E-21	NS	NS
			22.8	*24.8*	*NS*	*44.7*	*NS*	*NS*
ATSIK	1307	ATSIK_SNP1307	0.00011	0.00046	NS	7.1E-11	NS	NS
			11.6	*10.0*	*NS*	*25.5*	*NS*	*NS*
ATSIK	1308	ATSIK_SNP1308	1.4E-08	9.8E-10	NS	8.8E-21	NS	NS
			22.1	*24.6*	*NS*	*44.6*	*NS*	*NS*
ATSIK	1312	ATSIK_SNP1312	2.3E-08	1.5E-09	NS	7.2E-20	NS	NS
			21.6	*24.2*	*NS*	*43.1*	*NS*	*NS*
ATSIK	1332	ATSIK_SNP1332	5.8E-08	9.8E-09	NS	4.1E-20	NS	NS
			20.7	*22.4*	*NS*	*43.5*	*NS*	*NS*
ATSIK	1333	ATSIK_SNP1333	3.6E-08	9.8E-09	NS	1.1E-19	NS	NS
			21.2	*22.4*	*NS*	*42.9*	*NS*	*NS*
ATSIK	1353	ATSIK_SNP1353	1.4E-08	9.9E-10	NS	8.6E-21	NS	NS
			22.0	*24.6*	*NS*	*44.6*	*NS*	*NS*
ATSIK	1368	ATSIK_SNP1368	3.7E-09	7.3E-10	NS	1.2E-21	NS	NS
			23.7	*25.2*	*NS*	*46.4*	*NS*	*NS*
ATSIK	1376	ATSIK_SNP1376	8.4E-09	1E-09	NS	3.9E-21	NS	NS
			22.8	*24.8*	*NS*	*45.5*	*NS*	*NS*
ATSIK	1391	ATSIK_INDEL1391	1.30E-08	1.25E-09	NS	1.80E-20	NS	NS
			22.5	*24.7*	*NS*	*44.7*	*NS*	*NS*
Ala	954	Ala_SNP954	8E-09	1.3E-08	NS	2.2E-20	NS	NS
			21.3	*20.8*	*NS*	*41.8*	*NS*	*NS*
Ala	1017	Ala_SNP1017	0.00012	0.00053	NS	3.1E-10	NS	NS

			13.0	*11.4*	*NS*	*25.4*	*NS*	*NS*
Ala	1092	Ala_SNP1092	1E-08	1.2E-08	NS	1E-19	NS	NS
			21.0	*20.9*	*NS*	*40.7*	*NS*	*NS*
Ala	1256	Ala_SNP1256	2.2E-05	0.00035	NS	2.7E-10	NS	NS
			12.5	*9.7*	*NS*	*23.0*	*NS*	*NS*
Ala	1325	Ala_SNP1325	0.00012	0.00024	NS	2.5E-10	NS	0.00515
			10.9	*10.2*	*NS*	*23.3*	*NS*	*7.7*
Ala	1521	Ala_SNP1521	0.00011	0.00017	NS	9.1E-11	NS	0.00335
			11.1	*10.6*	*NS*	*24.3*	*NS*	*8.3*
Ala	1539	Ala_SNP1539	8.3E-05	0.00086	NS	4.7E-10	NS	0.00616
			11.3	*8.8*	*NS*	*22.8*	*NS*	*7.5*
Ghf	395	Ghf_SNP395	1.1E-07	2.9E-07	NS	7E-18	NS	NS
			19.9	*19.0*	*NS*	*39.7*	*NS*	*NS*
Ghf	397	Ghf_SNP397	0.00045	0.00184	NS	2.1E-09	NS	NS
			9.7	*8.2*	*NS*	*21.9*	*NS*	*NS*
Ghf	404	Ghf_SNP404	1.2E-06	1.3E-06	NS	4.6E-17	NS	NS
			17.4	*17.4*	*NS*	*38.2*	*NS*	*NS*
Ghf	415	solcap_snp_c2_25495	4.1E-06	1.1E-06	NS	8.9E-17	NS	NS
			15.9	*17.3*	*NS*	*37.2*	*NS*	*NS*
Ghf	448	solcap_snp_c2_25494	0.00373	0.00691	NS	1.7E-08	NS	NS
			7.4	*6.7*	*NS*	*20.0*	*NS*	*NS*
Ghf	484	Ghf_SNP484	NS	NS	NS	0.0003	NS	NS
			NS	*NS*	*NS*	*8.8*	*NS*	*NS*
Ghf	541	Ghf_SNP541	2.5E-05	7.6E-05	NS	4.1E-06	NS	NS
			11.4	*10.3*	*NS*	*13.2*	*NS*	*NS*
Ghf	553	solcap_snp_c2_25493	1.9E-07	1.2E-07	NS	1.7E-17	NS	NS
			18.9	*19.3*	*NS*	*38.2*	*NS*	*NS*
Ghf	637	Ghf_SNP637	0.00144	0.00862	NS	1.7E-08	NS	NS
			8.3	*6.3*	*NS*	*19.7*	*NS*	*NS*
Ghf	639	Ghf_SNP639	2.3E-07	6.7E-07	NS	8.4E-17	NS	NS
			18.5	*17.5*	*NS*	*36.7*	*NS*	*NS*
Ghf	648	Ghf_SNP648	0.00024	0.00043	NS	1.4E-09	NS	NS
			10.3	*9.7*	*NS*	*22.0*	*NS*	*NS*
Ghf	653	Ghf_SNP653	4.8E-05	0.00041	NS	3.6E-09	NS	NS
			12.0	*9.7*	*NS*	*21.2*	*NS*	*NS*
Ghf	659	Ghf_SNP659	4.6E-08	8.9E-07	NS	5.4E-17	0.00145	NS
			25.8	*22.8*	*NS*	*42.8*	*14.4*	*NS*
Ghf	666	Ghf_SNP666	8.4E-08	2.2E-07	NS	5.6E-17	NS	NS
			19.7	*18.8*	*NS*	*37.4*	*NS*	*NS*
Ghf	680	Ghf_SNP680	3.8E-05	0.00047	NS	1.8E-05	0.00317	NS
			12.4	*9.7*	*NS*	*13.1*	*7.6*	*NS*
Ghf	703	Ghf_SNP703	1.2E-07	5E-08	NS	1.1E-17	NS	NS
			19.3	*20.1*	*NS*	*38.4*	*NS*	*NS*
Ghf	718	Ghf_SNP718	2.3E-05	7.9E-05	NS	4.9E-06	0.00894	NS
			11.3	*10.1*	*NS*	*12.8*	*5.2*	*NS*
Ghf	737	Ghf_SNP737	9.6E-08	5.2E-08	NS	6E-19	NS	NS
			19.4	*19.9*	*NS*	*40.2*	*NS*	*NS*
Ghf	765	Ghf_SNP765	9.6E-08	5.2E-08	NS	6E-19	NS	NS
			19.4	*19.9*	*NS*	*40.2*	*NS*	*NS*
Ghf	786	Ghf_SNP786	2.1E-05	5.4E-05	NS	2.1E-06	NS	NS
			11.4	*10.5*	*NS*	*13.6*	*NS*	*NS*
Ghf	799	Ghf_SNP799	1E-07	6.1E-08	NS	7.7E-19	NS	NS
			19.4	*19.9*	*NS*	*40.2*	*NS*	*NS*
Ghf	854	Ghf_SNP854	8.9E-08	2.6E-07	NS	5.8E-18	NS	NS

			19.6	*18.6*	*NS*	*39.0*	*NS*	*NS*
CAST	129	CAST_SNP129	NS	0.00276	NS	2.2E-06	0.00813	NS
			NS	*9.6*	*NS*	*17.9*	*8.2*	*NS*
CAST	159	CAST_SNP159	NS	0.00734	NS	6.3E-05	NS	0.00116
			NS	*7.4*	*NS*	*12.7*	*NS*	*10.7*
CAST	175	CAST_SNP175	NS	NS	NS	0.00616	0.00431	0.00274
			NS	*NS*	*NS*	*6.6*	*7.0*	*8.4*
CAST	192	CAST_SNP192	NS	NS	NS	0.00241	NS	0.00876
			NS	*NS*	*NS*	*8.7*	*NS*	*8.1*
CAST	295	CAST_SNP295	5.1E-09	8E-10	0.00715	3.7E-09	NS	NS
			18.9	*20.6*	*5.3*	*19.2*	*NS*	*NS*
CAST	312	CAST_SNP312	NS	0.00194	NS	0.00084	0.00568	NS
			NS	*9.0*	*NS*	*9.9*	*7.8*	*NS*
CAST	335	CAST_SNP335	0.00082	0.00084	NS	1E-07	NS	NS
			10.0	*9.9*	*NS*	*19.1*	*NS*	*NS*
CAST	370	CAST_SNP370	0.00308	0.0001	NS	1.3E-07	NS	NS
			8.6	*12.3*	*NS*	*19.0*	*NS*	*NS*
CAST	459	solcap_snp_c2_49531	NS	NS	NS	0.00633	9.8E-05	NS
			NS	*NS*	*NS*	*8.1*	*13.1*	*NS*
CAST	504	solcap_snp_c2_49530	1.5E-06	7.8E-08	NS	4.4E-08	0.00071	NS
			13.9	*16.7*	*NS*	*17.2*	*7.8*	*NS*
AspAT	2817	AspAT_SNP2817	7.56E-05	6.21E-04	NS	1.45E-11	NS	6.09E-03
			11.5	*9.3*	*NS*	*26.0*	*NS*	*7.6*
AspAT	2831	AspAT_SNP2831	2.08E-10	5.93E-11	NS	1.49E-23	*NS*	NS
			24.7	*25.8*	*NS*	*46.7*	*NS*	*NS*
AspAT	2872	AspAT_SNP2872	1.86E-05	2.41E-04	NS	5.43E-12	*NS*	NS
			12.7	*10.1*	*NS*	*26.5*	*NS*	*NS*
AspAT	2883	AspAT_SNP2883	1.53E-05	3.28E-04	NS	9.73E-12	*NS*	NS
			12.7	*9.6*	*NS*	*25.6*	*NS*	*NS*
AspAT	2906	AspAT_SNP2906	2.09E-06	2.07E-06	NS	6.70E-07	9.24E-05	NS
			13.2	*13.3*	*NS*	*14.3*	*9.7*	*NS*
AspAT	2911	AspAT_SNP2911	1.99E-05	3.73E-04	NS	8.37E-12	NS	9.19E-03
			12.6	*9.6*	*NS*	*25.9*	*NS*	*6.9*
AspAT	2921	AspAT_SNP2921	3.25E-12	1.63E-10	7.48E-04	4.60E-21	NS	NS
			28.0	*24.7*	*10.0*	*42.8*	*NS*	*NS*
AspAT	2940	AspAT_SNP2940	1.74E-05	2.42E-04	NS	3.67E-12	NS	NS
			12.7	*10.0*	*NS*	*26.5*	*NS*	*NS*
AspAT	2972	AspAT_SNP2972	7.06E-06	2.23E-06	NS	6.27E-07	1.19E-04	NS
			12.2	*13.3*	*NS*	*14.5*	*9.5*	*NS*
AspAT	2975	AspAT_SNP2975	5.69E-04	5.39E-05	NS	2.21E-04	1.14E-03	NS
			7.9	*10.2*	*NS*	*8.8*	*7.2*	*NS*
AspAT	2986	AspAT_SNP2986	6.76E-06	1.63E-05	NS	5.03E-06	1.18E-03	NS
			13.7	*12.8*	*NS*	*14.0*	*8.4*	*NS*
AspAT	3006	AspAT_INDEL3006	1.45E-10	2.47E-10	8.51E-03	1.00E-22	NS	NS
			24.9	*24.5*	*7.3*	*45.4*	*NS*	*NS*

[1]Position of the SNP or INDEL. Counting starts from the presumed start codon (ATG) of the locus.

Table S 2. Comparison between SNP dosage measured by the Illumina array and sequencing by Sanger sequencing (solcap_snp_c2_45180) or pyrosequencing (solcap_snp_c1_2350 and solcap_snp_c2_45743). Allele dosages with a grey background showed different dosage between methods.

		c2_45180[1]		c1_2350		c2_45743	
Cultivar name	**BI**	**illumina**	**Sanger**	**illumina**	**pyro**	**illumina**	**pyro**
Susceptible							
Panda	60,2	GAAA	GAAA	AAAC	AAAC	CCTT	CCTT
Olga	66,2	GGAA	GAAA	AAAC	AACC	CCTT	CCTT
Maxilla	73,5	GAAA	GAAA	AACC	AACC	CTTT	CTTT
Logo	73,3	GAAA	GGAA	AAAC	AAAC	CCCT	CCCT
L.Rosetta	67,8	GGAA	GAAA	AAAC	AAAC	CCTT	CCTT
Kuba	71,1	GAAA	AAAA	AAAC	AAAC	CCCT	CCCT
Kolibri	72,3	GGAA	GGAA	AAAC	AACC	CCTT	CCTT
Fitis	55,5	GGAA	GGAA	ACCC	ACCC	CCTT	CCTT
Calla	56,1	GAAA	AAAA	AAAC	AAAC	CCTT	CCTT
Aspirant	73,7	GAAA	AAAA	AAAC	AACC	CCTT	CCTT
Resistant							
Remarka	11,2	GGGA	GGGA	AACC	AACC	CTTT	CTTT
Rafaela	11	GGAA	GGAA	ACCC	ACCC	CTTT	CTTT
Quarta	11,6	GGGA	GGAA	AACC	AACC	CTTT	CTTT
Marabel	11,4	GGGG	GGGG	ACCC	ACCC	CTTT	CTTT
Lolita	7,2	GGAA	GAAA	ACCC	ACCC	TTTT	TTTT
Krone	7,7	GGGG	GGGG	ACCC	ACCC	CTTT	CTTT
Gala	8,9	GGAA	GAAA	ACCC	ACCC	CTTT	CTTT
Elfe	9,4	GGGG	GGGG	ACCC	ACCC	CTTT	CTTT
Carmona	6,6	GGAA	GGAA	AACC	AACC	CCTT	CCTT
Agila	10,4	GGGA	GGAA	AACC	AACC	TTTT	TTTT

[1]SolCAP SNP ID after solcap_snp_..

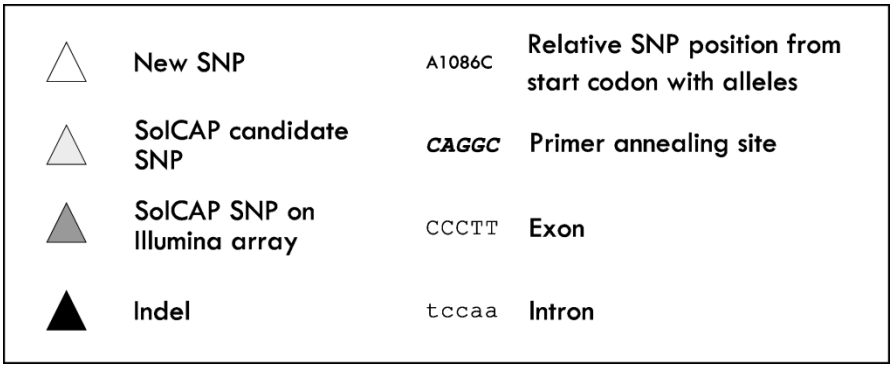


Figure S 2. Legend of figures containing sequence information of Sanger sequenced loci

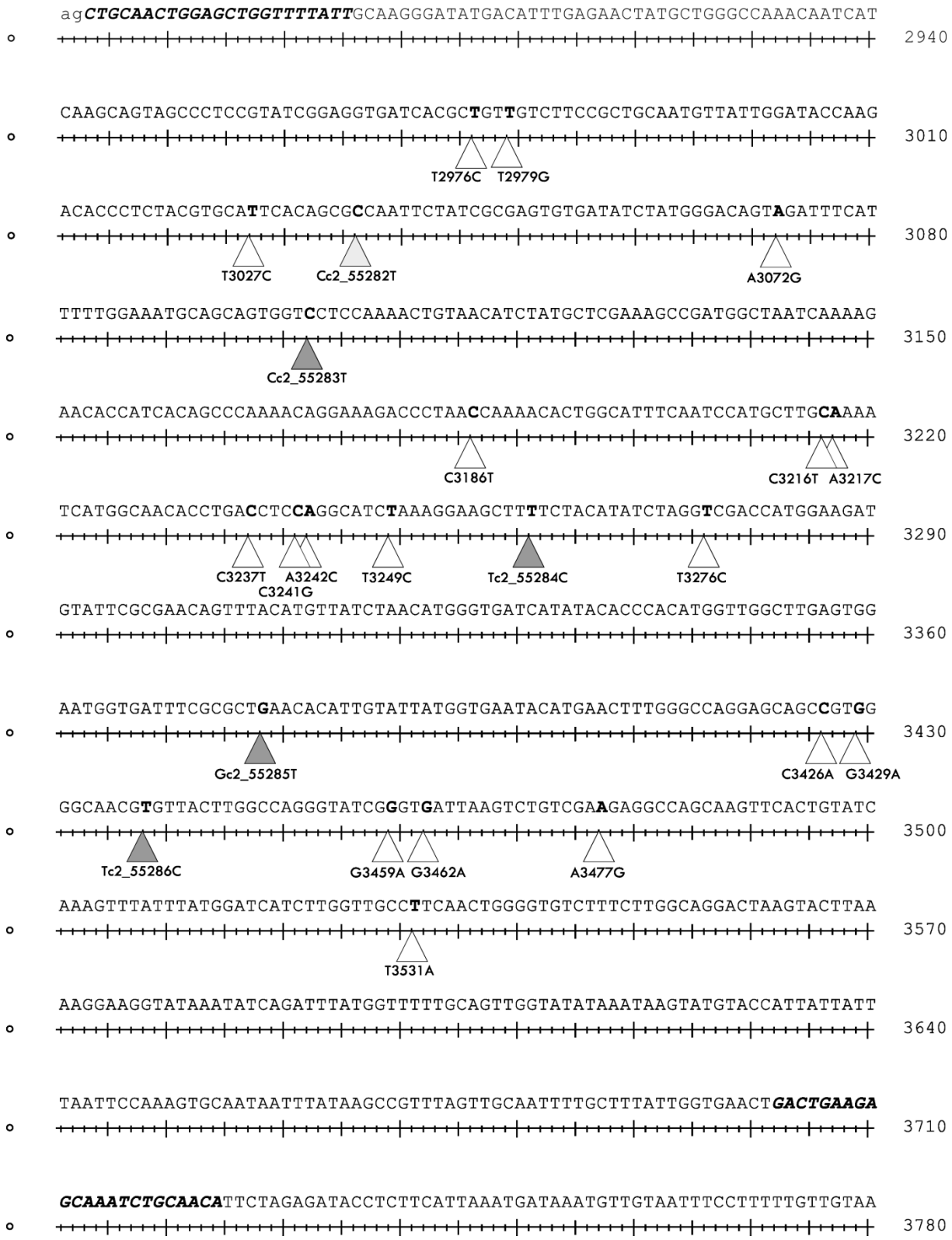


Figure S 3. Sequence of the amplicon sequenced with Sanger sequencing of the *Pest1* locus. The legend is displayed in Figure S 2.

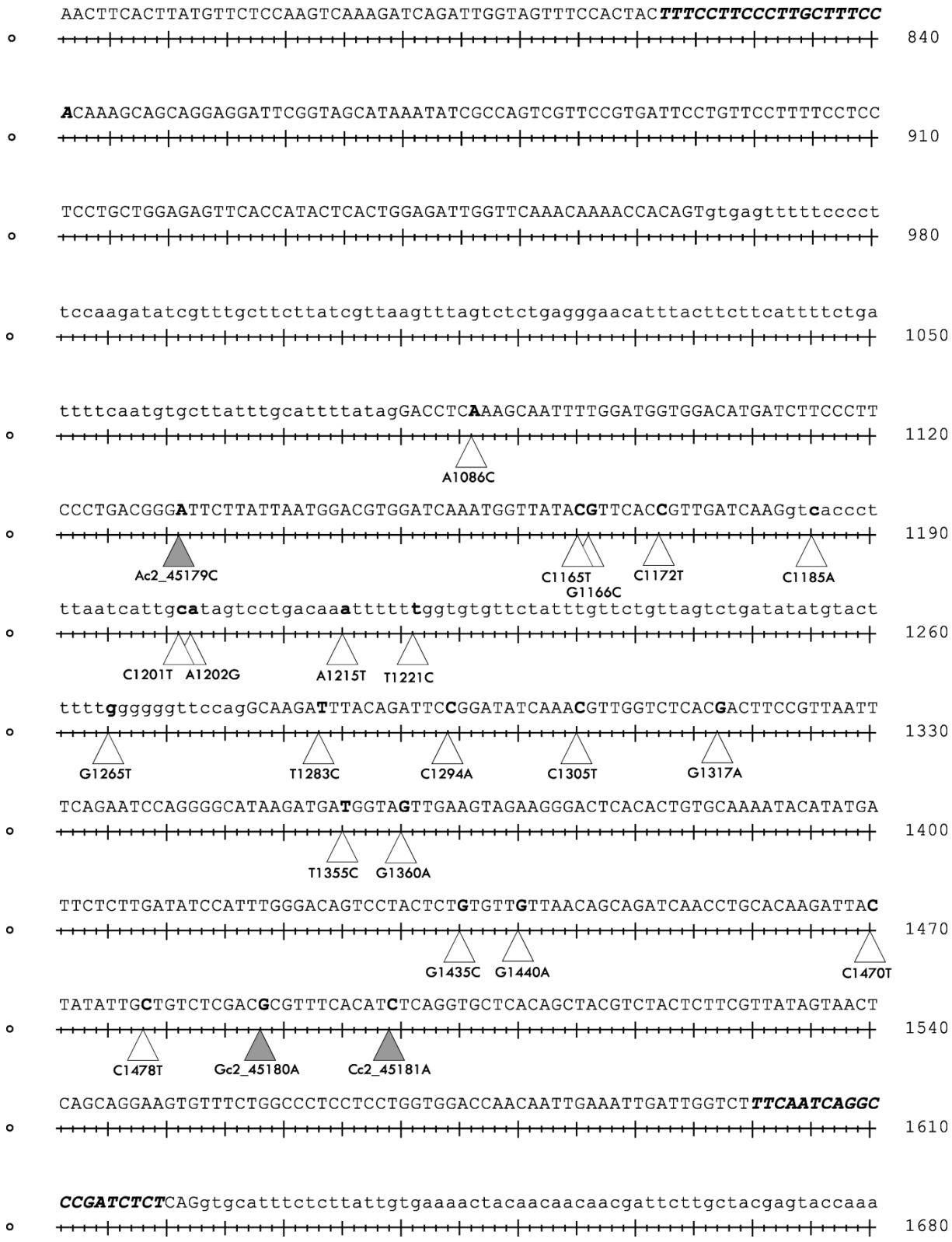


Figure S 4. Sequence of the amplicon sequenced with Sanger sequencing of the *Mco* locus. The legend is displayed in Figure S 2.

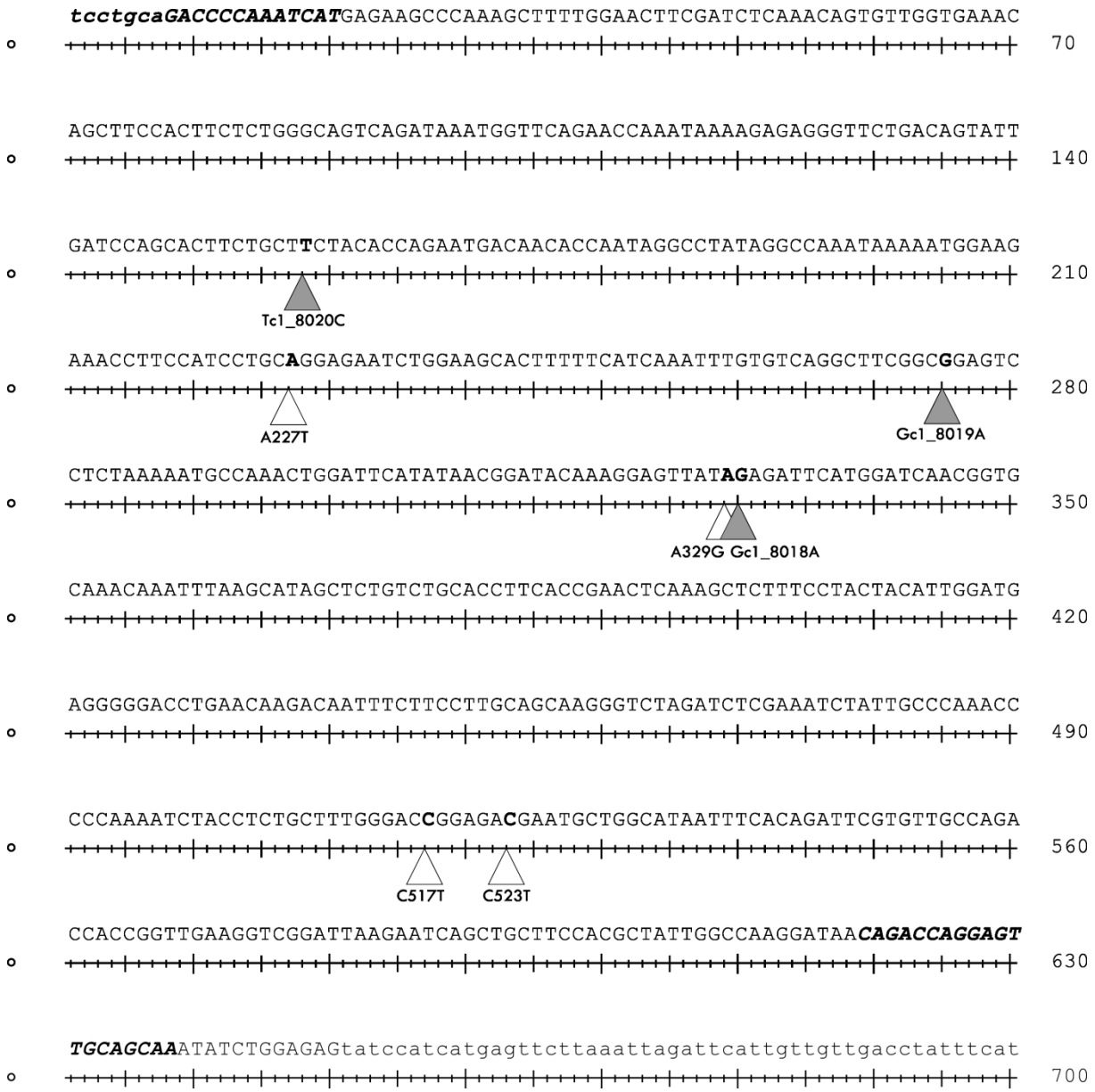


Figure S 5. Sequence of the amplicon sequenced with Sanger sequencing of the *StHSL1* locus. The legend is displayed in Figure S 2.

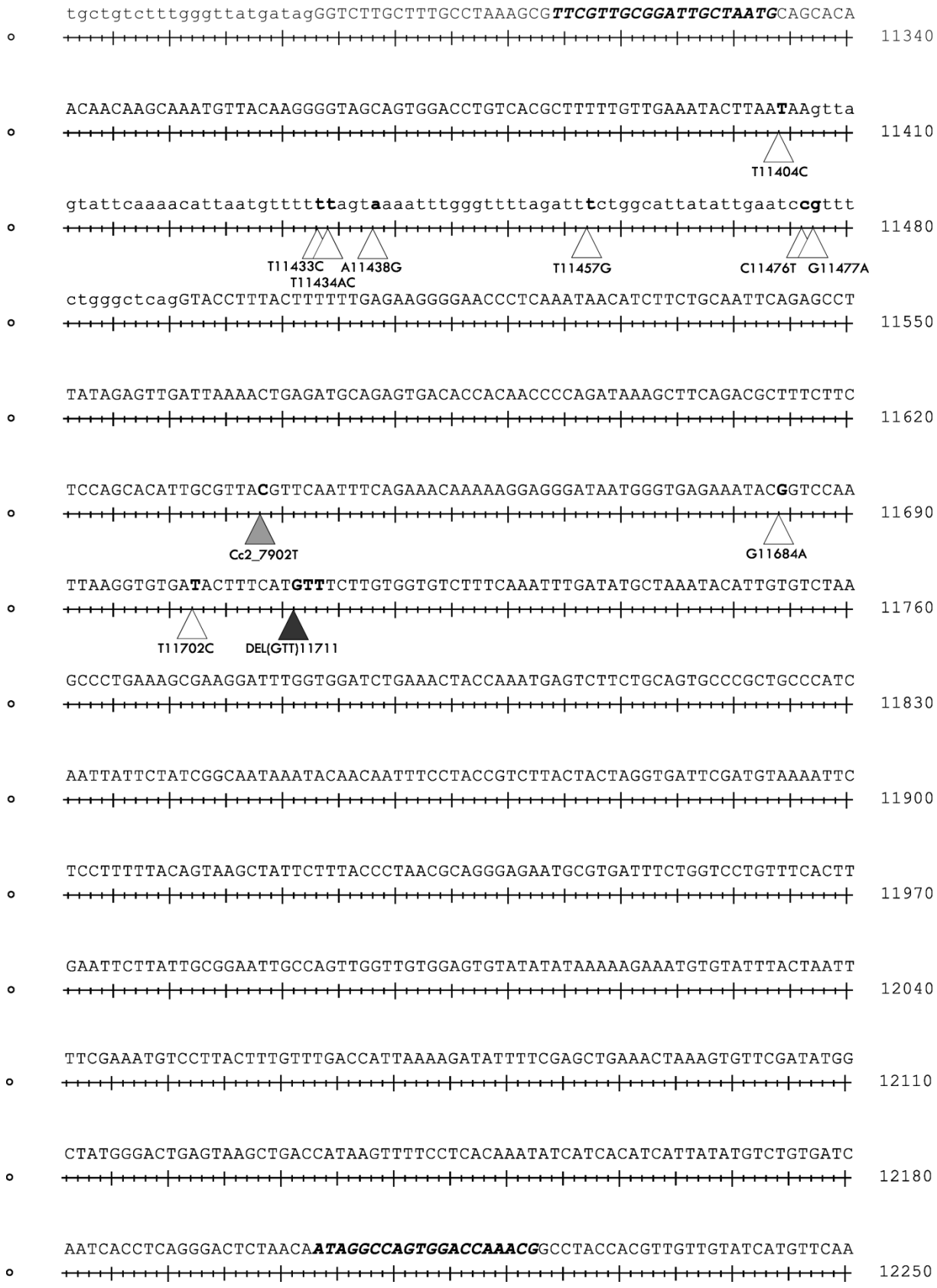


Figure S 6. Sequence of the amplicon sequenced with Sanger sequencing of the *Vps35* locus. The legend is displayed in Figure S 2.

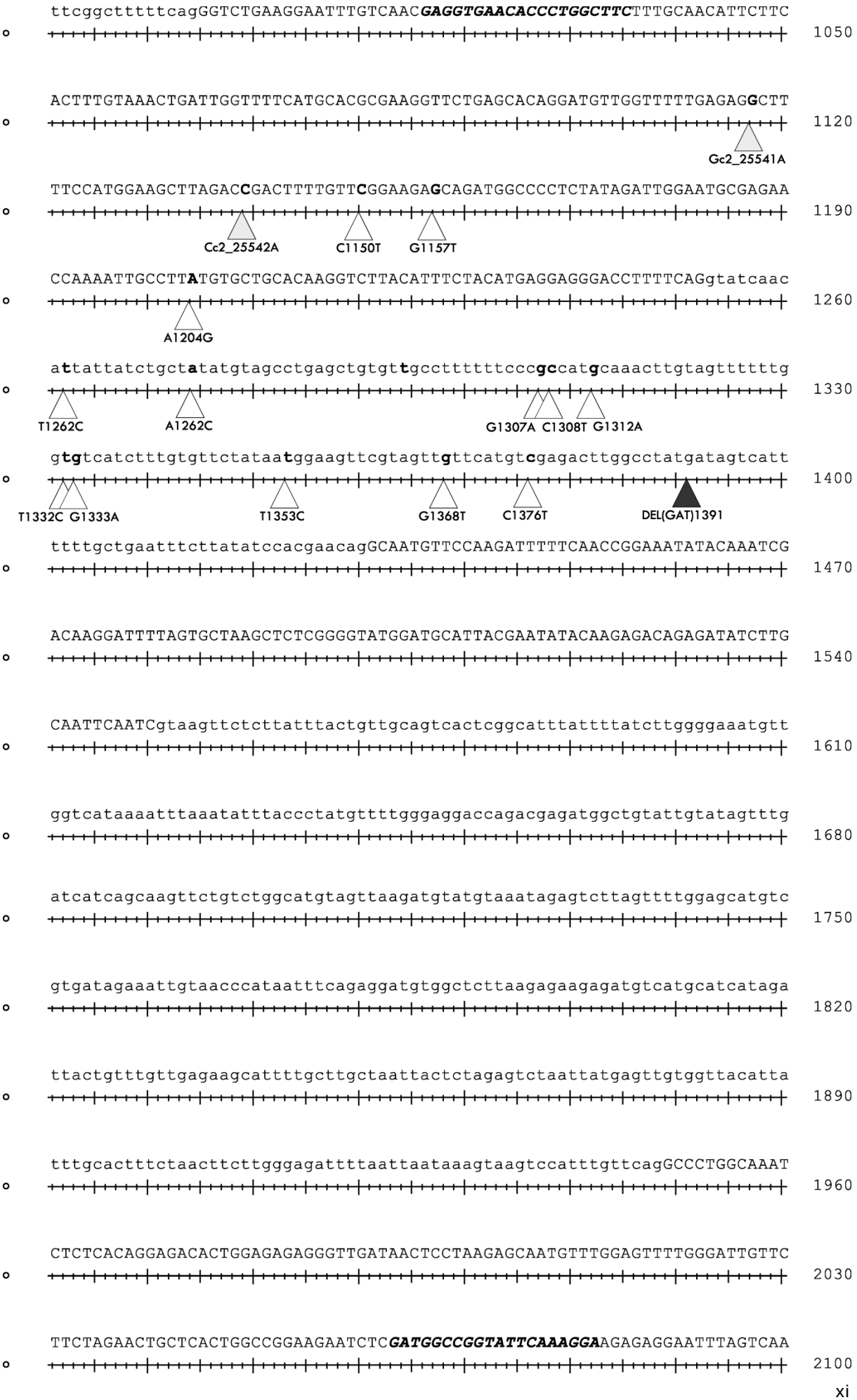

ttcggctttttcagGGTCTGAAGGAATTTGTCAAC***GAGGTGAACACCCTGGCTTC***TTTGCAACATTCTTC
1050
ACTTTGTAAACTGATTGGTTTTCATGCACGCGAAGGTTCTGAGCACAGGATGTTGGTTTTTGAGAG**G**CTT
1120
Gc2_25541A
TTCCATGGAAGCTTAGAC**C**GACTTTTGTT**C**GGAAGA**G**CAGATGGCCCCTCTATAGATTGGAATGCGAGAA
1190
Cc2_25542A
C1150T
G1157T
CCAAAATTGCCTT**A**TGTGCTGCACAAGGTCTTACATTTCTACATGAGGAGGGACCTTTTCAGgtatcaac
1260
A1204G
a**t**tattatctgct**a**tatgtagcctgagctgtgtt**t**gccttttttccc**gc**cat**g**caaacttgtagttttttg
1330
T1262C
A1262C
G1307A
C1308T
G1312A
g**tg**tcatctttgtgttctataa**t**ggaagttcgtagtt**g**ttcatgt**c**gagacttggcctatgatagtcatt
1400
T1332C
G1333A
T1353C
G1368T
C1376T
DEL(GAT)1391
ttttgctgaatttcttatatccacgaacagGCAATGTTCCAAGATTTTTCAACCGGAAATATACAAATCG
1470
ACAAGGATTTTAGTGCTAAGCTCTCGGGGTATGGATGCATTACGAATATACAAGAGACAGAGATATCTTG
1540
CAATTCAATCgtaagttctcttatttactgttgcagtcactcggcatttattttatcttggggaaatgtt
1610
ggtcataaaatttaaatatttaccctatgttttgggaggaccagacgagatggctgtattgtatagtttg
1680
atcatcagcaagttctgtctggcatgtagttaagatgtatgtaaatagagtcttagttttggagcatgtc
1750
gtgatagaaattgtaacccataatttcagaggatgtggctcttaagagaagagatgtcatgcatcataga
1820
ttactgtttgttgagaagcattttgcttgctaattactctagagtctaattatgagttgtggttacatta
1890
tttgcactttctaacttcttgggagattttaattaataaagtaagtccatttgttcagGCCCTGGCAAAT
1960
CTCTCACAGGAGACACTGGAGAGAGGGTTGATAACTCCTAAGAGCAATGTTTGGAGTTTTGGGATTGTTC
2030
TTCTAGAACTGCTCACTGGCCGGAAGAATCTC***GATGGCCGGTATTCAAAGGA***AGAGAGGAATTTAGTCAA
2100

Figure S 7 (previous page). Sequence of the amplicon sequenced with Sanger sequencing of the *ATSIK* locus. The legend is displayed in Figure S 2.

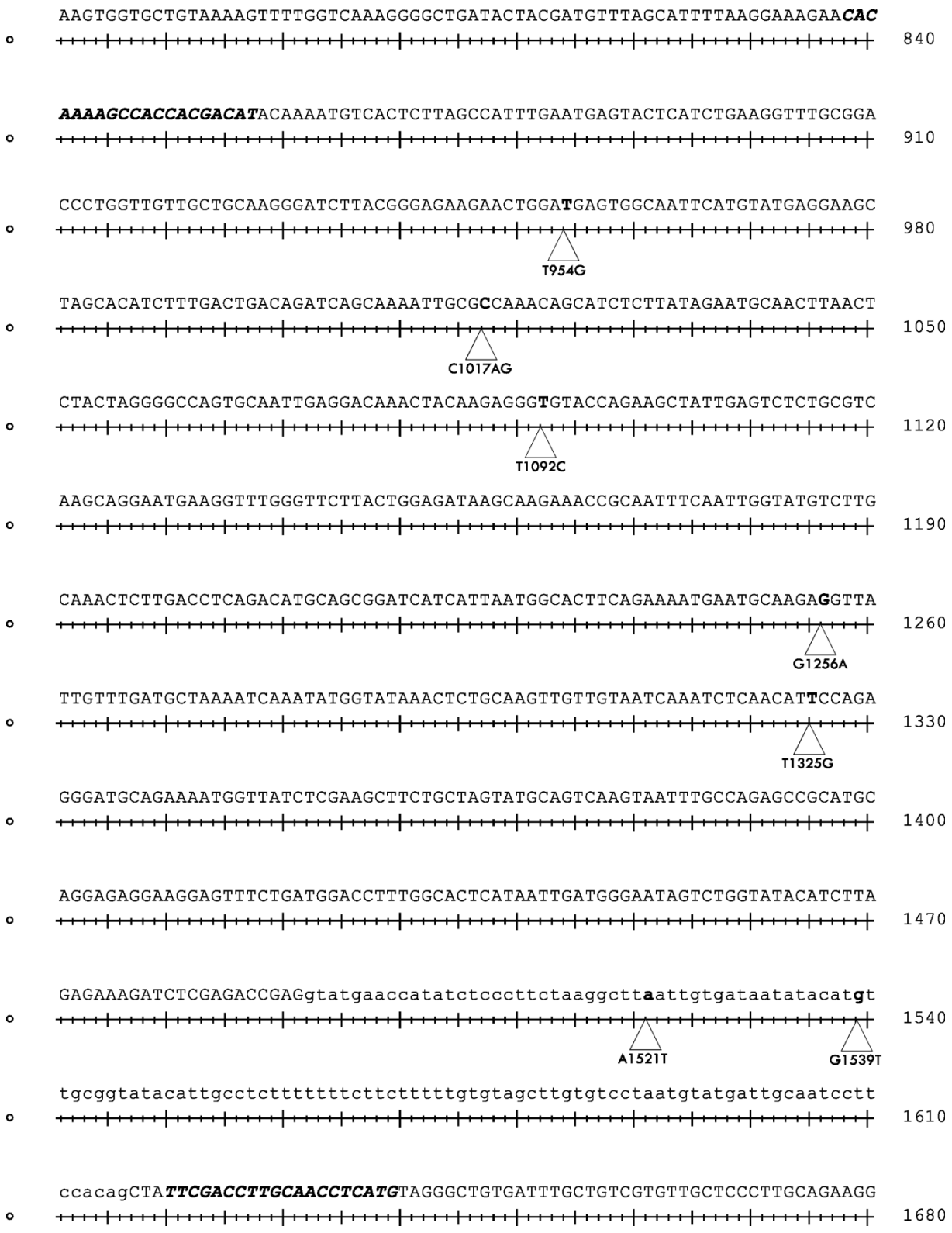


Figure S 8. Sequence of the amplicon sequenced with Sanger sequencing of the *Ala* locus. The legend is displayed in Figure S 2.

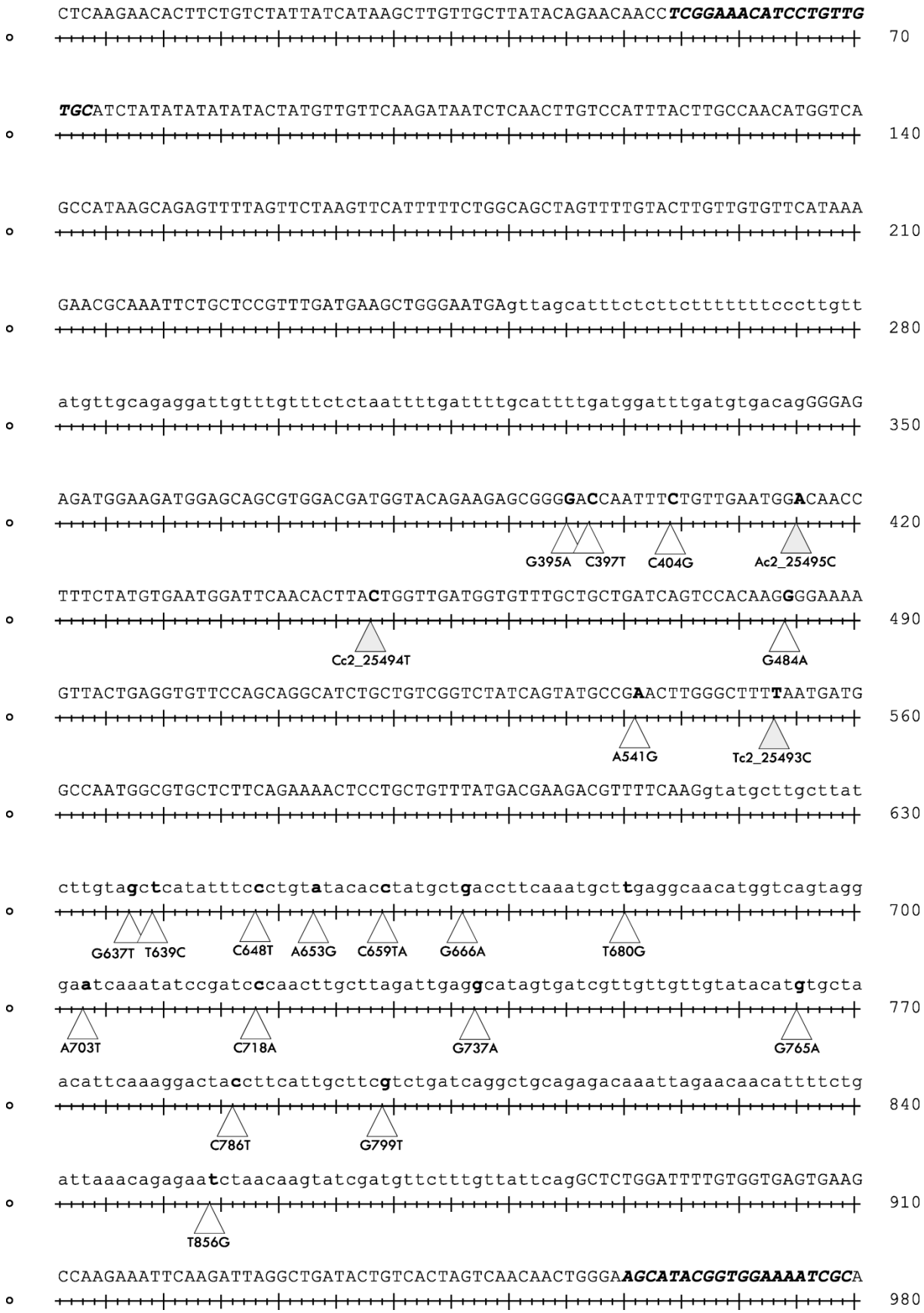


Figure S 9. Sequence of the amplicon sequenced with Sanger sequencing of the *Ghf* locus. The legend is displayed in Figure S 2.

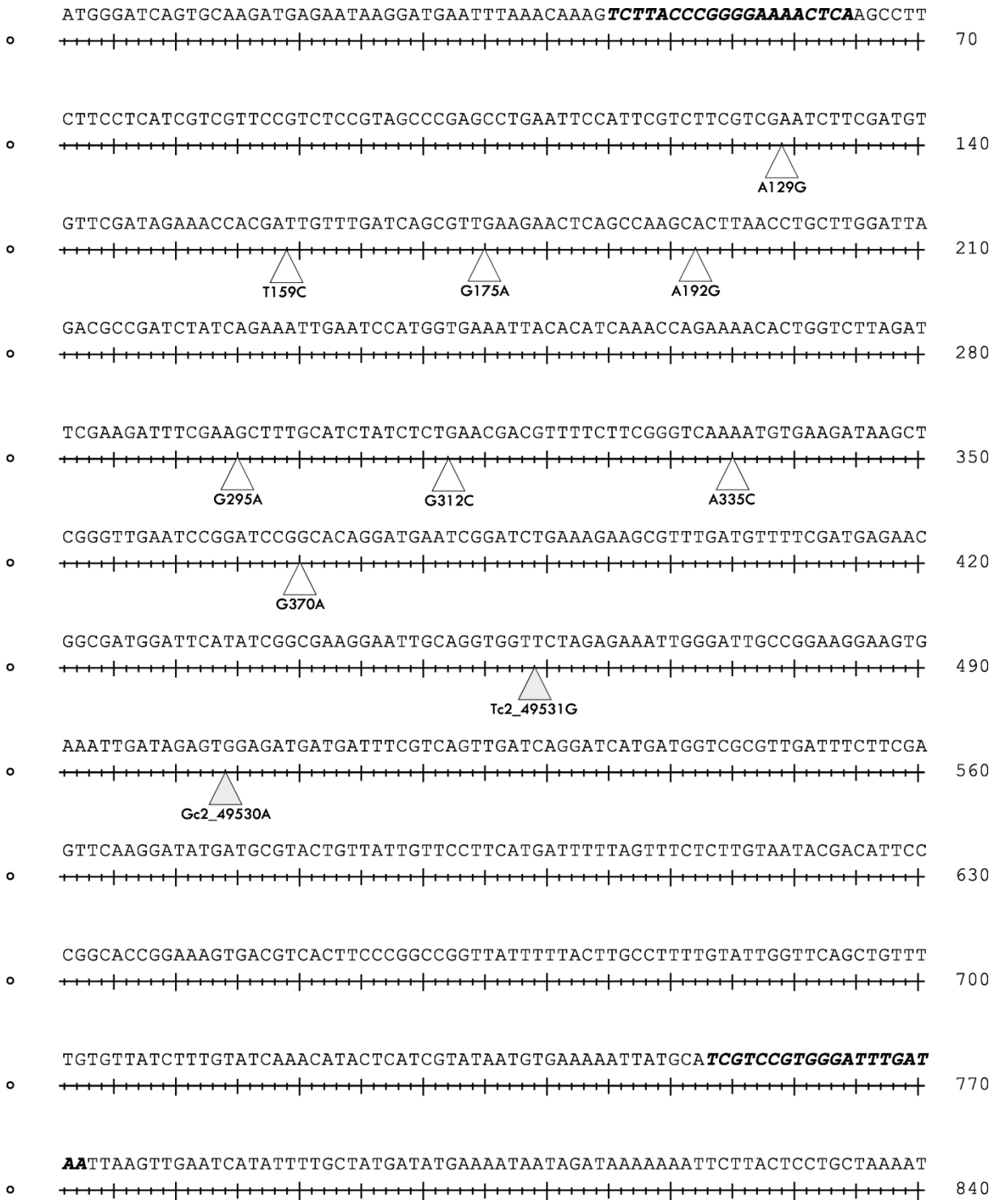


Figure S 10. Sequence of the amplicon sequenced with Sanger sequencing of the *CAST* locus. The legend is displayed in Figure S 2.

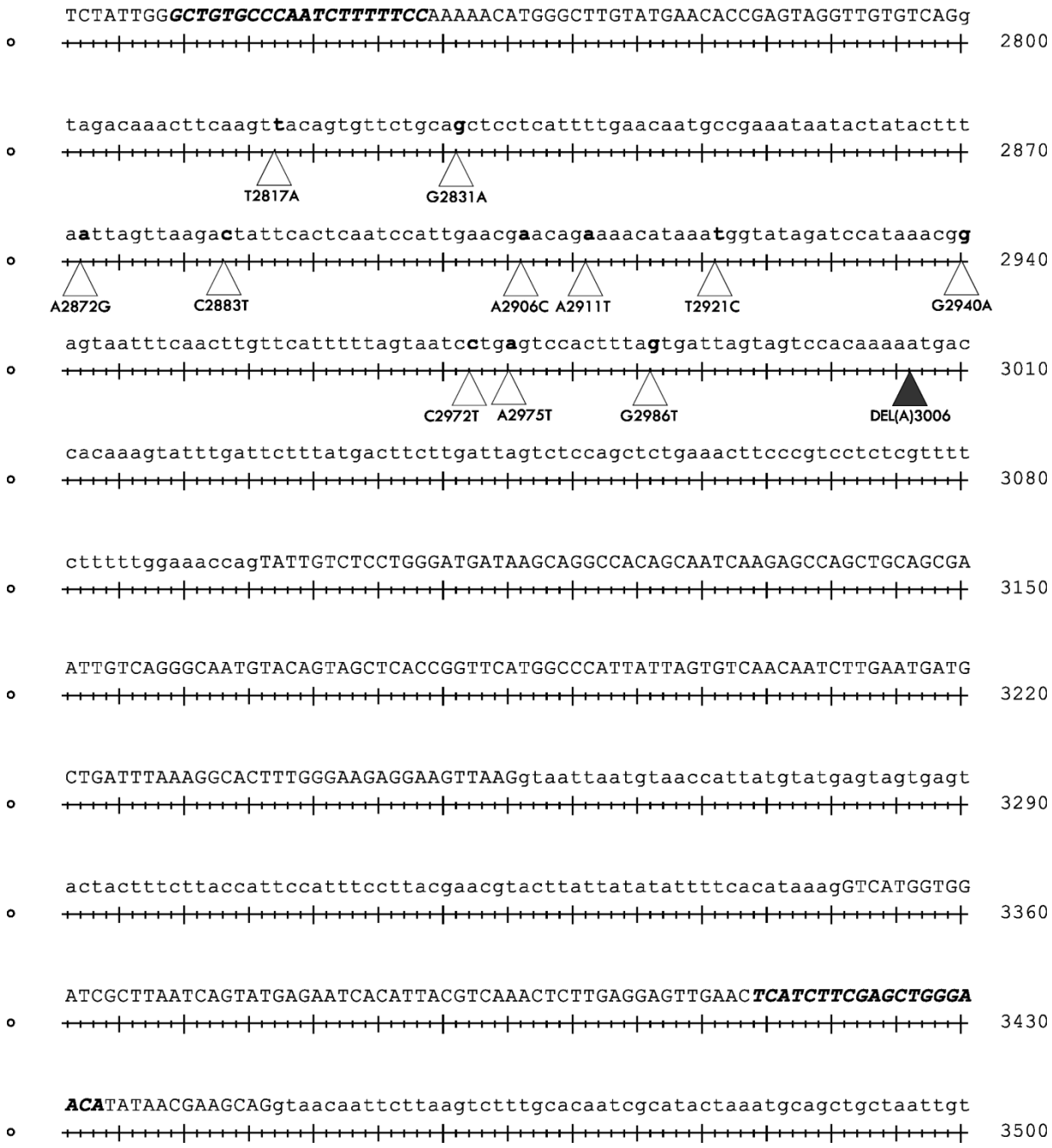


Figure S 11. Sequence of the amplicon sequenced with Sanger sequencing of the *AspAT* locus. The legend is displayed in Figure S 2.